47. Adaptive Beamforming and Postfiltering

S. Gannot, I. Cohen

In this chapter, we explore many of the basic concepts of array processing with an emphasis on adaptive beamforming for speech enhancement applications. We begin in Sect. 47.1 by formulating the problem of microphone array in a noisy and reverberant environment. In Sect. 47.2, we derive the frequency-domain linearly constrained minimum-variance (LCMV) beamformer, and its generalized sidelobe canceller (GSC) variant. The GSC components are explored in Sect. 47.3, and several commonly used special cases of these blocks are presented. As the GSC structure necessitates an estimation of the speech related acoustical transfer functions (ATFs), several alternative system identification methods are addressed in Sect. 47.4. Beamformers often suffer from sensitivity to signal mismatch. We analyze this phenomenon in Sect. 47.5 and explore several cures to this problem. Although the GSC beamformer yields a significant improvement in speech quality, when the noise field is spatially incoherent or diffuse, the noise reduction is insufficient and additional postfiltering is normally required. In Sect. 47.6, we present multi-microphone postfilters, based on either minimum mean-squared error (MMSE) or log-spectral amplitude estimate criteria. An interesting relation between the GSC and the Wiener filter is derived in this Section as well. In Sect. 47.7, we analyze the performance of the transfer-function GSC (TF-GSC), and in Sect. 47.8 demonstrate the advantage of multichannel postfiltering over single-channel postfiltering in nonstationary noise conditions.

Over the last four decades, array processing has become a well-established discipline, see e.g., [47.1–14]. In the mid 1980s, array processing and beamforming methods were adopted by the speech community to deal with data received by microphone arrays. Since then, beamforming techniques for microphone arrays have been used in many applications, such as speaker separation, speaker localization, speech dereverberation, acoustic echo cancellation, and speech enhancement.

Adaptive beamforming for speech signals requires particular consideration of problems that are specific to speech signals and to the acoustic environment. The speech signal is wide-band, highly nonstationary, and has a very wide dynamic range. An acoustic enclosure is usually modeled as a filter with very long impulse response due to multiple reflections from the room walls. In a typical office, the length of the filters may reach several thousand taps. Furthermore, the impulse response is often time varying due to speaker and objects movements.

The term beamforming refers to the design of a spatiotemporal filter. Broadband arrays comprise a set of filters, applied to each received microphone signal, followed by a summation operation. The main objective of the beamformer is to extract a desired signal, impinging on the array from a specific position, out of noisy measurements thereof. Usually, the interference signals occupy the same frequency band as the desired signal, rendering temporal-only filtering useless. The simplest structure is the delay-and-sum beamformer, which first compensates for the relative delay between distinct microphone signals and then sums the steered signal to form a single output. This beamformer, which is still widely used, can be very effective in mitigating noncoherent, i. e., spatially white, noise sources, provided that the number of microphones is relatively high. However, if the noise source is coherent, the noise reduction (NR) is strongly dependent on the direction of arrival of the noise signal. Consequently, the performance of the delay and sum beamformer in reverberant environments is often insufficient. *Jan* and *Flanagan* [47.15, 16] and *Rabinkin* et al. [47.17] extended the delay and sum concept by introducing the filter-and-sum beamformer. This structure, designed for multipath environments, namely reverberant enclosures, replaces the simpler delay compensator with a matched filter.

The array beam pattern can generally be designed to have a specified response. This can be done by properly setting the values of the multichannel filters' weights. However, the application of data-independent design methods is very limited in dynamic acoustical environments. Statistically optimal beamformers are designed based on the statistical properties of the desired and interference signals. In general, they aim at enhancing the desired signal, while rejecting the interference signal. Several criteria can be applied in the design of the beamformer, e.g., maximum signal-to-noise ratio (MSNR), minimum mean-squared error (MMSE), and linearly constrained minimum variance (LCMV). A summary of several design criteria can be found in [47.5, 7].

Beamforming methods use the signals' statistics (at least second-order statistics), which is usually not available and should be estimated from the data. Moreover, the acoustical environment is time varying, due to talker and objects movements, and abrupt changes in the noise characteristics (e.g., passing cars). Hence, adaptation mechanisms are required. An adaptive counterpart of each of the prespecified design criteria can be derived. Early contributions to the field of adaptive beamformers design can be attributed to *Sondhi* and *Elko* [47.18], to *Kaneda* and *Ohga* [47.19], and to *Van Compernolle* [47.20]. *Kellermann* [47.21] addressed the problem of joint echo cancellation and NR by incorporating echo cancellers into the beamformer design. *Nordholm* et al. [47.22, 23] used microphone arrays in a car environment, and designed a beamformer employing calibration signals to enhance the obtained performance. *Martin* [47.24] analyzed beamforming techniques for small microphone arrays. Many other applications of microphone arrays such as hearing aids, blind source separation (BSS), and dereverberation are addressed elsewhere in this handbook.

The minimization of the mean-squared error (MSE) in the context of array processing leads to the well-known multichannel Wiener filter [47.25]. *Doclo* and *Moonen* [47.26–28] proposed an efficient implementation of the Wiener filter based on the generalized singular-value decomposition (GSVD) of the microphone data matrix. This method yields an optimal estimation (in the MMSE sense) of the desired signal component of one of the microphone signals. The authors further proposed efficient schemes for recursive update of the GSVD. An optional, adaptive noise cancellation postfiltering stage is proposed as well. In that scheme, in addition to the optimal estimation of the desired speech signal, an optimal noise channel is also estimated. This estimated noise component can be used as a reference noise signal (similar to the one used in [47.25]), to further enhance the speech signal. *Spriet* et al. [47.29] proposed a subband implementation of the GSVD-based scheme, and *Rombouts* and

Moonen [47.30, 31] proposed to apply the efficient QR decomposition to the problem at hand.

In many adaptive array schemes the acoustical transfer-function (ATF) relating the speech source and the microphone should be known in advance, or at least estimated from the received data (note that in case of delay-only propagation, the acoustical transfer function reduces to a steering vector, consisting of phase-only components.) In contrast, the multichannel Wiener filter is uniquely based on estimates of the second-order statistics of the recorded noisy signal and the noise signal (estimated during noise-only segments), and does not make any a priori assumptions about the signal model. Unfortunately, as pointed by *Chen* et al. [47.32], the Wiener filter, which is optimal in the MMSE sense, cannot guarantee undistorted speech signal at its output. This drawback can however be mitigated by modifying the MMSE criterion to control the amount of imposed speech distortion. A method that employs this modification is presented in [47.33, 34]. It is also shown that the ATFs information (only a simple delay-only case is presented in the contributions) can be incorporated into the Wiener filter scheme (called the spatially preprocessed Wiener filter), resulting in an improved performance. The Wiener filter and its application to speech enhancement is addressed in a separate chapter of this handbook (6; 43).

In this chapter, we concentrate on a different adaptive structure based on the LCMV criterion. The LCMV beamformer, proposed by *Frost* [47.35], is aiming at minimizing the output power under linear constraints on the response of the array towards the desired speech signal. Frost proposed an adaptive scheme, which is based on a constrained least-mean-square (LMS)-type adaptation (for the LMS algorithm please refer to [47.25]). To avoid this constrained adaptation, *Griffiths* and *Jim* [47.36] proposed the GSC structure, which separates the output power minimization and the application of the constraint. The GSC structure is based on the assumption that the different sensors receive a delayed version of the desired signal, and therefore we refer to it as the delay generalized sidelobe canceller (D-GSC). The GSC structure was rederived in the frequency domain, and extended to deal with, the more-complicated general ATFs case by *Affes* and *Grenier* [47.37] and later by *Gannot* et al. [47.38]. This frequency-domain version, which takes into account the reverberant nature of the enclosure, was nicknamed the transfer-function generalized sidelobe canceller (TF-GSC). The GSC comprises three blocks: a fixed beamformer (FBF), which aligns the desired signal components, a blocking matrix (BM), which blocks the desired speech components resulting in reference noise signals, and a multichannel adaptive noise canceller (ANC), which eliminates noise components that leak through the sidelobes of the FBF.

Nordholm and *Leung* [47.39] analyze the limits of the obtainable NR of the GSC in an isotropic noise field. *Bitzer* et al. address the problem in [47.40, 41] and [47.42]. In [47.40], the authors derive an expression for the NR as a function of the noise field and evaluate the degradation as a function of the reverberation time (T_{60}). The special two-microphone case is treated in [47.41]. The additional NR due to the ANC branch of the GSC, implemented by a closed-form Wiener filter rather than the adaptive Widrow least-mean-square (LMS) procedure, is presented in [47.42]. The frequency-band nested subarrays structure is presented and its NR is theoretically analyzed by *Marro* et al. [47.43]. A more-complex dual GSC structure employing calibration signals was suggested and analyzed by *Nordholm* et al. [47.44]. *Huarng* and *Yeh* [47.45] addressed the distortion issue by evaluating the desired signal leakage into the reference noise branch of the GSC structure. However, the delay-only ATFs assumption is imposed and the expected degradation due to pointing errors alone is evaluated. The performance degradation due to constraining the Wiener filters to an finite impulse response (FIR) structure is demonstrated by *Nordholm* et al. in [47.46]. The resulting performance limits of the GSC structure strongly depend on the cross-correlation between the sensors' signals induced by the noise field, as shown in the above references and by *Cox* [47.47].

47.1 Problem Formulation

Consider an array of M sensors in a noisy and reverberant environment. The received signals generally include three components. The first is a desired speech signal, the second is some stationary interference signal, and the third is some nonstationary (transient) noise component. Our goal is to reconstruct the speech component from the received signals. Let $s(t)$ denote the desired source signal, let $a_m(t)$ represent the room impulse response (RIR) of the m-th sensor to the desired source, and let $n_m(t)$ denote the noise component at the m-th sensor.

The observed signal at the m-th sensor ($m = 1, \ldots, M$) is given by

$$\begin{aligned} z_m(t) &= a_m(t) * s(t) + n_m(t) \\ &= a_m(t) * s(t) + n_m^{\mathrm{s}}(t) + n_m^{\mathrm{t}}(t)\,, \end{aligned} \tag{47.1}$$

where $n_m^{\mathrm{s}}(t)$ and $n_m^{\mathrm{t}}(t)$ represent the stationary and nonstationary noise components at the m-th sensor, respectively, and $*$ denotes convolution. We assume that both noise components may comprise coherent (directional) noise component and diffused noise component.

The observed signals are divided in time into overlapping frames by the application of a window function and analyzed using the short-time Fourier transform (STFT). Assuming time-invariant transfer functions, we have in the time–frequency domain

$$\begin{aligned} Z_m(k,\ell) &\approx A_m(k)S(k,\ell) + N_m(k,\ell) \\ &\approx A_m(k)S(k,\ell) + N_m^{\mathrm{s}}(k,\ell) + N_m^{\mathrm{t}}(k,\ell)\,, \end{aligned} \tag{47.2}$$

where ℓ is the frame index and $k = 1, 2, \ldots, K$ represents the frequency bin index. (The equality in (47.2) is only justified for segments which are longer than the RIR length. Since RIRs tend to be very long, the conditions allowing for this representation to hold cannot be exactly met. We assume, however, that the STFT relation is a reasonable approximation.) $Z_m(k,\ell)$, $S(k,\ell)$, $N_m(k,\ell)$, $N_m^{\mathrm{s}}(k,\ell)$, and $N_m^{\mathrm{t}}(k,\ell)$ are the STFT of the respective signals. $A_m(k)$ is the ATF relating the speech source with the m-th sensor. The vector formulation of the equation set (47.2) is

$$\begin{aligned} \boldsymbol{Z}(k,\ell) &= \boldsymbol{A}(k)S(k,\ell) + \boldsymbol{N}(k,\ell) \\ &= \boldsymbol{A}(k)S(k,\ell) + \boldsymbol{N}_{\mathrm{s}}(k,\ell) + \boldsymbol{N}_{\mathrm{t}}(k,\ell)\,, \end{aligned} \tag{47.3}$$

where

$$\begin{aligned} \boldsymbol{Z}(k,\ell) &= \big(Z_1(k,\ell)\ \ Z_2(k,\ell)\ \ \cdots\ \ Z_M(k,\ell)\big)^{\mathrm{T}}\,, \\ \boldsymbol{A}(k) &= \big(A_1(k)\ \ A_2(k)\ \ \cdots\ \ A_M(k)\big)^{\mathrm{T}}\,, \\ \boldsymbol{N}(k,\ell) &= \big(N_1(k,\ell)\ \ N_2(k,\ell)\ \ \cdots\ \ N_M(k,\ell)\big)^{\mathrm{T}}\,, \\ \boldsymbol{N}_{\mathrm{s}}(k,\ell) &= \big(N_1^{\mathrm{s}}(k,\ell)\ \ N_2^{\mathrm{s}}(k,\ell)\ \ \cdots\ \ N_M^{\mathrm{s}}(k,\ell)\big)^{\mathrm{T}}\,, \\ \boldsymbol{N}_{\mathrm{t}}(k,\ell) &= \big(N_1^{\mathrm{t}}(k,\ell)\ \ N_2^{\mathrm{t}}(k,\ell)\ \ \cdots\ \ N_M^{\mathrm{t}}(k,\ell)\big)^{\mathrm{T}}. \end{aligned}$$

47.2 Adaptive Beamforming

Frost [47.35] proposed a beamformer that relies on the assumption that the ATFs between the desired source and the array of sensors can be uniquely determined by gain and delay values. In this section, we follow Frost's approach in the STFT domain and derive a beamforming algorithm for the arbitrary ATF case. We first obtain a closed form of the LCMV beamformer, and subsequently derive an adaptive solution. The outcome is a constrained LMS-type algorithm. We proceed, following the seminal work of *Griffiths* and *Jim* [47.36], with the formulation of an unconstrained adaptive solution namely, the transfer-function generalized sidelobe canceller (TF-GSC). We initially assume that the ATFs are known. Later, in Sect. 47.4, we present several alternatives for estimating the ATFs.

47.2.1 Frequency-Domain Frost Algorithm

Optimal Solution

Let $W_m^*(k,\ell)\,; m = 1, \ldots, M$ denote a set of M filters, and define

$$\boldsymbol{W}^{\mathrm{H}}(k,\ell) = \big(W_1^*(k,\ell)\ \ W_2^*(k,\ell)\ \ \cdots\ \ W_M^*(k,\ell)\big)\,,$$

where the superscript H denotes conjugation transpose. A filter-and-sum beamformer, depicted in Fig. 47.1, is realized by filtering each sensor signal by $W_m^*(k,\ell)$ and summing the outputs,

$$\begin{aligned} Y(k,\ell) &= \boldsymbol{W}^{\mathrm{H}}(k,\ell)\boldsymbol{Z}(k,\ell) \\ &= \boldsymbol{W}^{\mathrm{H}}(k,\ell)\boldsymbol{A}(k)S(k,\ell) + \boldsymbol{W}^{\mathrm{H}}(k,\ell)\boldsymbol{N}_{\mathrm{s}}(k,\ell) \\ &\quad + \boldsymbol{W}^{\mathrm{H}}(k,\ell)\boldsymbol{N}_{\mathrm{t}}(k,\ell) \\ &\triangleq Y_{\mathrm{s}}(k,\ell) + Y_{n,\mathrm{s}}(k,\ell) + Y_{n,\mathrm{t}}(k,\ell)\,, \end{aligned} \tag{47.4}$$

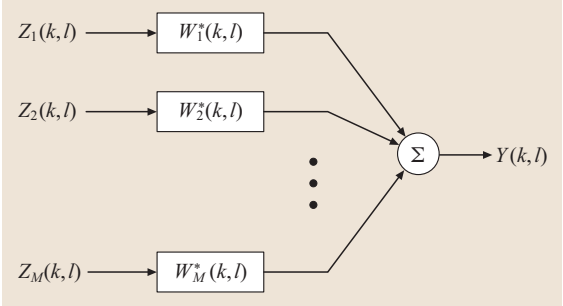

Fig. 47.1 Filter-and-sum beamformer

where $Y_{\mathrm{s}}(k,\ell)$ is the signal component and $Y_{n,\mathrm{s}}(k,\ell)$ and $Y_{n,\mathrm{t}}(k,\ell)$ are the stationary and nonstationary noise components, respectively. The output power of the beamformer is given by

$$\begin{aligned} &E\{Y(k,\ell)Y^*(k,\ell)\} \\ &= E\{\boldsymbol{W}^{\mathrm{H}}(k,\ell)\boldsymbol{Z}(k,\ell)\boldsymbol{Z}^{\mathrm{H}}(k,\ell)\boldsymbol{W}(k,\ell)\} \\ &= \boldsymbol{W}^{\mathrm{H}}(k,\ell)\boldsymbol{\phi}_{\boldsymbol{ZZ}}(k,\ell)\boldsymbol{W}(k,\ell)\,, \end{aligned}$$

where $\boldsymbol{\phi}_{\boldsymbol{ZZ}}(k,\ell) \triangleq E\{\boldsymbol{Z}(k,\ell)\boldsymbol{Z}^{\mathrm{H}}(k,\ell)\}$ is the power spectral density (PSD) matrix of the received signals. We want to minimize the output power subject to the following constraint on $Y_{\mathrm{s}}(k,\ell)$:

$$\begin{aligned} Y_{\mathrm{s}}(k,\ell) &= \boldsymbol{W}^{\mathrm{H}}(k,\ell)\boldsymbol{A}(k)S(k,\ell) \\ &= \mathcal{F}^*(k,\ell)S(k,\ell)\,, \end{aligned}$$

where $\mathcal{F}(k,\ell)$ is some prespecified filter, usually a simple delay. Without loss of generality we assume hereinafter that $\mathcal{F}(k,\ell)=1$. Hence, the minimization problem can be stated as

$$\begin{aligned} &\min_{\boldsymbol{W}} \left\{\boldsymbol{W}^{\mathrm{H}}(k,\ell)\boldsymbol{\phi}_{\boldsymbol{ZZ}}(k,\ell)\boldsymbol{W}(k,\ell)\right\} \\ &\text{subject to} \quad \boldsymbol{W}^{\mathrm{H}}(k,\ell)\boldsymbol{A}(k)=1\,. \end{aligned} \tag{47.5}$$

The minimization problem (47.5) is demonstrated in Fig. 47.2. The point where the equipower contours are tangent to the constraint plane is the optimum vector of beamforming filters. The perpendicular $\boldsymbol{F}(k)$ from the origin to the constraint plane will be calculated in the next section.

To solve (47.5) we first define the complex Lagrangian,

$$\begin{aligned} \mathcal{L}(\boldsymbol{W}) = &\boldsymbol{W}^{\mathrm{H}}(k,\ell)\boldsymbol{\Phi}_{\boldsymbol{ZZ}}(k,\ell)\boldsymbol{W}(k,\ell) \\ &+\lambda[\boldsymbol{W}^{\mathrm{H}}(k,\ell)\boldsymbol{A}(k)-1] \\ &+\lambda^*[\boldsymbol{A}^{\mathrm{H}}(k,\ell)\boldsymbol{W}(k)-1]\,, \end{aligned} \tag{47.6}$$

where λ is a Lagrange multiplier. Setting the derivative with respect to $\boldsymbol{W}^*$ to 0 [47.48] yields

$$\nabla_{\boldsymbol{W}^*}\mathcal{L}(W)\boldsymbol{\Phi}_{\boldsymbol{ZZ}}(k,\ell)\boldsymbol{W}(k,\ell)+\lambda\boldsymbol{A}(k)=0.$$

Now, recalling the constraint in (47.5), we obtain the LCMV optimal filter

$$\boldsymbol{W}^{\mathrm{LCMV}}(k,\ell) = \frac{\boldsymbol{\Phi}_{\boldsymbol{ZZ}}^{-1}(k,\ell)\boldsymbol{A}(k)}{\boldsymbol{A}^{\mathrm{H}}(k)\boldsymbol{\Phi}_{\boldsymbol{ZZ}}^{-1}(k,\ell)\boldsymbol{A}(k)}\,. \tag{47.7}$$

This closed-form solution is difficult to implement, and is not suitable for time-varying environments. Therefore we often have to resort to an adaptive solution, which is derived in the sequel.

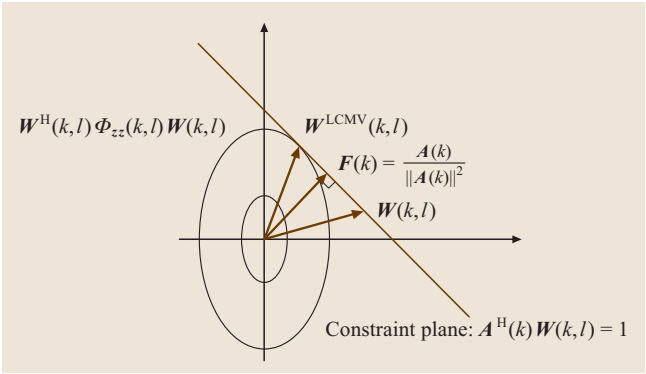

Fig. 47.2 Constrained minimization

It is interesting to show the equivalence between the LCMV solution (47.7) and the MSNR beamformer [47.7], which is obtained from

$$\max_{\boldsymbol{W}} \frac{|\boldsymbol{W}^{\mathrm{H}}(k,\ell)\boldsymbol{A}(k)|^2}{\boldsymbol{W}^{\mathrm{H}}(k,\ell)\boldsymbol{\Phi}_{\boldsymbol{NN}}(k,\ell)\boldsymbol{W}(k,\ell)}\,. \tag{47.8}$$

The well-known solution to (47.8) is the (colored-noise) matched filter

$$\boldsymbol{W}(k,\ell) \propto \boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1}(k,\ell)\boldsymbol{A}(k)\,.$$

If the array response is constrained to fulfil $\boldsymbol{W}^{\mathrm{H}}(k,\ell)\boldsymbol{A}(k)=1$, i. e., no distortion in the desired direction, we have

$$\boldsymbol{W}^{\mathrm{MSNR}}(k,\ell) = \frac{\boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1}(k,\ell)\boldsymbol{A}(k)}{\boldsymbol{A}^{\mathrm{H}}(k)\boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1}(k,\ell)\boldsymbol{A}(k)}\,. \tag{47.9}$$

Using (47.3) it can be verified that

$$\begin{aligned} \boldsymbol{\Phi}_{\boldsymbol{ZZ}} &= \phi_{\mathrm{ss}}(k,\ell)\boldsymbol{A}(k)\boldsymbol{A}^{\mathrm{H}}(k) \\ &\quad + \boldsymbol{\Phi}_{N_{\mathrm{s}}N_{\mathrm{s}}}(k,\ell) + \boldsymbol{\Phi}_{N_{\mathrm{t}}N_{\mathrm{t}}}(k,\ell) \\ &= \phi_{\mathrm{ss}}(k,\ell)\boldsymbol{A}(k)\boldsymbol{A}^{\mathrm{H}}(k) + \boldsymbol{\Phi}_{\boldsymbol{NN}}(k,\ell)\,, \end{aligned} \tag{47.10}$$

where $\boldsymbol{\Phi}_{\boldsymbol{NN}}(k,\ell) \triangleq \boldsymbol{\Phi}_{N_{\mathrm{s}}N_{\mathrm{s}}}(k,\ell)+\boldsymbol{\Phi}_{N_{\mathrm{t}}N_{\mathrm{t}}}(k,\ell)$, the overall noise PSD matrix. Using the matrix inversion lemma, it is shown in Appendix 47.B that

$$\boldsymbol{W}^{\mathrm{LCMV}}(k,\ell) = \frac{\boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1}(k,\ell)\boldsymbol{A}(k)}{\boldsymbol{A}^{\mathrm{H}}(k)\boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1}(k,\ell)\boldsymbol{A}(k)}\,. \tag{47.11}$$

This solution is identical to the solution of the MSNR beamformer.

While both methods are shown to be equal, provided that the ATFs $\boldsymbol{A}(k)$ are known, their behavior in the case of unknown ATFs is different. Analysis of these differences is given by *Cox* [47.47].

Note also that, due to the nonstationary noise component, the term $\boldsymbol{\Phi}_{NN}(k,\ell)$ depends on the frame index. This time dependence is one of the major factors leading to performance degradation in beamforming. We address this problem by introducing the multichannel postfilter in Sect. 47.6.

Adaptive Solution

Consider the following steepest-descent adaptive algorithm:

$$\begin{aligned}&\boldsymbol{W}(k,\ell+1)\\&=\boldsymbol{W}(k,\ell)-\mu\nabla_{\boldsymbol{W}^*}\mathcal{L}(k,\ell)\\&=\boldsymbol{W}(k,\ell)-\mu\left[\boldsymbol{\Phi}_{\boldsymbol{ZZ}}(k,\ell)\boldsymbol{W}(k,\ell)+\lambda\boldsymbol{A}(k)\right]\,.\end{aligned}$$

Imposing the look-direction constraint on $\boldsymbol{W}(\ell+1,k)$ yields

$$\begin{aligned}1&=\boldsymbol{A}^{\mathrm{H}}(k)\boldsymbol{W}(k,\ell+1)\\&=\boldsymbol{A}^{\mathrm{H}}(k)\boldsymbol{W}(k,\ell)-\mu\boldsymbol{A}^{\mathrm{H}}(k)\boldsymbol{\Phi}_{\boldsymbol{ZZ}}(k,\ell)\boldsymbol{W}(k,\ell)\\&\quad-\mu\boldsymbol{A}^{\mathrm{H}}(k)\boldsymbol{A}(k)\lambda\,.\end{aligned}$$

Solving for the Lagrange multiplier and applying further rearrangement of terms yields:

$$\begin{aligned}&\boldsymbol{W}(k,\ell+1)\\&=P(k)\boldsymbol{W}(k,\ell)-\mu P(k)\boldsymbol{\Phi}_{\boldsymbol{ZZ}}(k,\ell)\boldsymbol{W}(k,\ell)+\boldsymbol{F}(k)\,,\end{aligned}\tag{47.12}$$

where

$$P(k)\triangleq\mathbf{I}-\frac{\boldsymbol{A}(k)\boldsymbol{A}^{\mathrm{H}}(k)}{\|\boldsymbol{A}(k)\|^2}\tag{47.13}$$

and

$$\boldsymbol{F}(k)\triangleq\frac{\boldsymbol{A}(k)}{\|\boldsymbol{A}(k)\|^2}.\tag{47.14}$$

Further simplification can be obtained by replacing $\boldsymbol{\Phi}_{\boldsymbol{ZZ}}(k,\ell)$ by its instantaneous estimator, $\boldsymbol{Z}(k,\ell)\boldsymbol{Z}^{\mathrm{H}}(k,\ell)$, and recalling (47.4). We finally obtain,

$$\begin{aligned}&\boldsymbol{W}(k,\ell+1)\\&=P(k)[\boldsymbol{W}(k,\ell)-\mu\boldsymbol{Z}(k,\ell)Y^*(k,\ell)]+\boldsymbol{F}(k)\,.\end{aligned}$$

The entire algorithm is summarized in Table 47.1.

Table 47.1 Frequency-domain Frost algorithm

$\boldsymbol{W}(\ell=0,k)=\boldsymbol{F}(k)$
$\boldsymbol{W}(t+1,k)=P(k)\left[\boldsymbol{W}(k,\ell)-\mu\boldsymbol{Z}(k,\ell)Y^*(k,\ell)\right]+\boldsymbol{F}(k)$
$\ell=0,1,\ldots$
[$P(k)$ and $\boldsymbol{F}(k)$ are defined by (47.13) and (47.14)].

47.2.2 Frequency-Domain Generalized Sidelobe Canceller

In [47.36], *Griffiths* and *Jim* considered the case where each ATF is a delay element (with gain). They obtained an unconstrained adaptive enhancement algorithm, using the same linear constraint imposed by *Frost* [47.35]. The unconstrained algorithm is more tractable, reliable, and computationally more efficient in comparison with its constrained counterpart. In the adaptive solution Section, we obtained an adaptive algorithm for the case where each ATF is represented by an arbitrary linear time-invariant system. We now repeat the arguments of Griffiths and Jim for the arbitrary ATFs case, and derive an unconstrained adaptive enhancement algorithm. A detailed description can be found in [47.38].

Derivation

Consider the null space of $\boldsymbol{A}(k)$, defined by

$$\mathcal{N}(k)\triangleq\{\boldsymbol{W}\,|\,\boldsymbol{A}^{\mathrm{H}}(k)\boldsymbol{W}=0\}\,.$$

The constraint hyperplane,

$$\Lambda(k)\triangleq\{\boldsymbol{W}\,|\,\boldsymbol{A}^{\mathrm{H}}(k)\boldsymbol{W}=1\}$$

is parallel to $\mathcal{N}(k)$. In addition, let

$$\mathcal{R}(k)\triangleq\{\kappa\boldsymbol{A}(k)\,|\,\text{for any real }\kappa\}$$

be the column space. By the fundamental theorem of linear algebra [47.49], $\mathcal{R}(k)\perp\mathcal{N}(k)$. In particular, $\boldsymbol{F}(k)$ is perpendicular to $\mathcal{N}(k)$, since $\boldsymbol{F}(k)=\frac{1}{\|\boldsymbol{A}(k)\|^2}\boldsymbol{A}(k)\in\mathcal{R}(k)$. Furthermore,

$$\boldsymbol{A}^{\mathrm{H}}(k)\boldsymbol{F}(k)=\boldsymbol{A}^{\mathrm{H}}(k)\boldsymbol{A}(k)(\boldsymbol{A}^{\mathrm{H}}(k)\boldsymbol{A}(k))^{-1}=1\,.$$

Thus, $\boldsymbol{F}(k)\in\Lambda(k)$ and $\boldsymbol{F}(k)\perp\Lambda(k)$. Hence, $\boldsymbol{F}(k)$ is the perpendicular from the origin to the constraint hyperplane, $\Lambda(k)$. The matrix $P(k)$, defined in (47.13), is the projection matrix to the null space of $\boldsymbol{A}(k)$, $\mathcal{N}(k)$.

A vector in linear space can be uniquely split into a sum of two vectors in mutually orthogonal subspaces [47.49]. Hence,

$$\boldsymbol{W}(k,\ell)=\boldsymbol{W}_0(k,\ell)-\boldsymbol{V}(k,\ell)\,,\tag{47.15}$$

where $\boldsymbol{W}_0(k,\ell)\in\mathcal{R}(k)$ and $-\boldsymbol{V}(k,\ell)\in\mathcal{N}(k)$. By the definition of $\mathcal{N}(k)$,

$$\boldsymbol{V}(k,\ell)=\mathcal{H}(k)\boldsymbol{G}(k,\ell)\,,\tag{47.16}$$

where $\mathcal{H}(k)$ is a matrix such that its columns span the null space of $\boldsymbol{A}(k)$, i. e.,

$$\boldsymbol{A}^{\mathrm{H}}(k)\mathcal{H}(k)=0\,,\qquad\text{rank}\,\{\mathcal{H}(k)\}\le M-1\,,\tag{47.17}$$

where $\mathcal{H}(k)$ is usually called a BM (blocking matrix). The outputs of the BM will be denoted, for reasons that will be clear in the sequel, noise reference signals $\boldsymbol{U}(k,\ell)$, defined as

$$\boldsymbol{U}(k,\ell) = \mathcal{H}^{\mathrm{H}}(k)\boldsymbol{Z}(k,\ell)\,, \tag{47.18}$$

where

$$\boldsymbol{U}(k,\ell) = \big(U_2(k,\ell)\;\; U_3(k,\ell)\;\; \ldots\;\; U_M(k,\ell)\big)^{\mathrm{T}}\,.$$

The vector $\boldsymbol{G}(k,\ell)$ is a $\mathrm{rank}\{\mathcal{H}(k)\}\times 1$ vector of adjustable filters. We assume hereinafter that $\mathrm{rank}\{\mathcal{H}(k)\} = M-1$. Hence, the set of filters is defined as

$$\boldsymbol{G}(k,\ell) = \big(G_2(k,\ell)\;\; G_3(k,\ell)\;\; \ldots\;\; G_M(k,\ell)\big)^{\mathrm{T}}\,. \tag{47.19}$$

By the geometrical interpretation of Frost's algorithm,

$$\boldsymbol{W}_0(k,\ell) = \boldsymbol{F}(k) = \frac{\boldsymbol{A}(k)}{\|\boldsymbol{A}(k)\|^2} \tag{47.20}$$

(Recall that $\boldsymbol{F}(k)$ is the perpendicular from the origin to the constraint hyperplane, $\Lambda(k)$.) Now, using (47.4), (47.15), and (47.16) we obtain

$$Y(k,\ell) = Y_{\mathrm{FBF}}(k,\ell) - Y_{\mathrm{ANC}}(k,\ell)\,, \tag{47.21}$$

where

$$\begin{aligned} Y_{\mathrm{FBF}}(k,\ell) &= \boldsymbol{W}_0^{\mathrm{H}}(k,\ell)\boldsymbol{Z}(k,\ell) \\ Y_{\mathrm{ANC}}(k,\ell) &= \boldsymbol{G}^{\mathrm{H}}(k,\ell)\mathcal{H}^{\mathrm{H}}(k)\boldsymbol{Z}(k,\ell)\,. \end{aligned} \tag{47.22}$$

The output of the constrained beamformer is a difference of two terms, both operating on the input signal $\boldsymbol{Z}(k,\ell)$. The first term, $Y_{\mathrm{FBF}}(k,\ell)$, utilizes only fixed components (which depend on the ATFs), so it can be viewed as a FBF. The FBF coherently sums the desired speech components, while in general it destructively sums the noise components. Hence, it is expected that the signal-to-noise ratio (SNR) at the FBF output will be higher than the input SNR. However, this result cannot be guaranteed. We will elaborate on this issue while discussing the performance analysis in Sect. 47.7.

We now examine the second term $Y_{\mathrm{ANC}}(k,\ell)$. Note that

$$\begin{aligned} \boldsymbol{U}(k,\ell) &= \mathcal{H}^{\mathrm{H}}(k)\boldsymbol{Z}(k,\ell) \\ &= \mathcal{H}^{\mathrm{H}}(k)[\boldsymbol{A}(k)S(k,\ell) \\ &\quad + \boldsymbol{N}_{\mathrm{s}}(k,\ell) + \boldsymbol{N}_{\mathrm{t}}(k,\ell)] \\ &= \mathcal{H}^{\mathrm{H}}(k)[\boldsymbol{N}_{\mathrm{s}}(k,\ell) + \boldsymbol{N}_{\mathrm{t}}(k,\ell)]\,. \end{aligned} \tag{47.23}$$

The last transition is due to (47.17). It is worth mentioning that, when a perfect BM is applied, $\boldsymbol{U}(k,\ell)$ indeed contains only noise components. In general, however, $\mathcal{H}^{\mathrm{H}}(k)\boldsymbol{A}(k,\ell) \neq 0$, hence desired speech components may leak into the noise reference signals. If the speech component is indeed completely eliminated (blocked) by $\mathcal{H}(k)$, $Y_{\mathrm{ANC}}(k,\ell)$ becomes a pure noise term. The residual noise term in $Y_{\mathrm{FBF}}(k,\ell)$ can then be reduced by properly adjusting the filters $\boldsymbol{G}(k,\ell)$, using the minimum output power criterion. This minimization problem is in fact the classical multichannel noise cancellation problem. An adaptive LMS solution to the problem was proposed by *Widrow* [47.25].

To summarize, the beamformer is comprised of three parts. An FBF $\boldsymbol{W}_0$, which aligns the desired signal components, a BM $\mathcal{H}(k)$, which blocks the desired speech components resulting in the reference noise signals $\boldsymbol{U}(k,\ell)$, and a multichannel ANC $\boldsymbol{G}(k,\ell)$, which eliminates the stationary noise that leaks through the sidelobes of the FBF.

Noise Canceller Adaptation

The reference noise signals are emphasized by the ANC and subtracted from the output of the FBF, yielding

$$Y(k,\ell) = \big[\boldsymbol{W}_0^{\mathrm{H}}(k,\ell) - \boldsymbol{G}^{\mathrm{H}}(k,\ell)\mathcal{H}^{\mathrm{H}}(k)\big]\boldsymbol{Z}(k,\ell)\,. \tag{47.24}$$

Let three hypotheses $H_{0\mathrm{s}}$, $H_{0\mathrm{t}}$, and H_1 indicate, respectively, the absence of transients, the presence of an interfering transient, and the presence of a desired source transient at the beamformer output. The optimal solution for the filters $\boldsymbol{G}(k,\ell)$ is obtained by minimizing the power of the beamformer output during the stationary noise frames (i. e., when $H_{0\mathrm{s}}$ is true) [47.2]. We note, however, that no adaptation should be carried out during abrupt changes in the characteristics of the noise signal (e.g., a passing car). When the noise source position is constant and the noise statistics is slowly varying, the ANC filters can track the changes.

Let $\boldsymbol{\Phi}_{\boldsymbol{N}_{\mathrm{s}}\boldsymbol{N}_{\mathrm{s}}}(k,\ell) = E\{\boldsymbol{N}_{\mathrm{s}}(k,\ell)\boldsymbol{N}_{\mathrm{s}}^{\mathrm{H}}(k,\ell)\}$ denote the PSD matrix of the input stationary noise. Then, the power of the stationary noise at the beamformer output is minimized by solving the unconstrained optimization problem:

$$\begin{aligned} \min_{\boldsymbol{G}}\{&[\boldsymbol{W}_0(k,\ell) - \mathcal{H}(k,\ell)\boldsymbol{G}(k,\ell)]^{\mathrm{H}} \\ &\times \boldsymbol{\Phi}_{\boldsymbol{N}_{\mathrm{s}}\boldsymbol{N}_{\mathrm{s}}}(k,\ell)[\boldsymbol{W}_0(k,\ell) - \mathcal{H}(k,\ell)\boldsymbol{G}(k,\ell)]\}\,. \end{aligned} \tag{47.25}$$

A multichannel Wiener solution is given by (see also [47.42, 46])

$$\begin{aligned} \boldsymbol{G}(k,\ell) = &\big[\mathcal{H}^{\mathrm{H}}(k,\ell)\boldsymbol{\Phi}_{\boldsymbol{N}_{\mathrm{s}}\boldsymbol{N}_{\mathrm{s}}}(k,\ell)\mathcal{H}(k)\big]^{-1} \\ &\times \mathcal{H}^{\mathrm{H}}(k,\ell)\boldsymbol{\Phi}_{\boldsymbol{N}_{\mathrm{s}}\boldsymbol{N}_{\mathrm{s}}}(k,\ell)\boldsymbol{W}_0(k,\ell)\,. \end{aligned} \tag{47.26}$$

In practice, this optimization problem is solved by using the normalized LMS algorithm [47.2]:

$$\boldsymbol{G}(k,\ell+1) = \begin{cases} \boldsymbol{G}(k,\ell) + \frac{\mu_g}{P_{\text{est}}(k,\ell)}\boldsymbol{U}(k,\ell)Y^*(k,\ell) & H_{0s} \text{ true}, \\ \boldsymbol{G}(k,\ell)\,, & \text{otherwise}, \end{cases} \tag{47.27}$$

where

$$P_{\text{est}}(k,\ell) = \alpha_p P_{\text{est}}(k,\ell-1) + (1-\alpha_p)\|\boldsymbol{U}(k,\ell)\|^2 \tag{47.28}$$

represents the power of the noise reference signals, μ_g is a step size that regulates the convergence rate, and α_p is a smoothing parameter in the PSD estimation process.

To allow for the use of the STFT, we further assume that the ANC filters $\boldsymbol{g}_m$ have a time-varying finite impulse response (FIR) structure:

$$\boldsymbol{g}_m^{\mathrm{T}}(t) = \left(g_{m,-K_L}(t) \;\; \ldots \;\; g_{m,K_R}(t)\right). \tag{47.29}$$

Note, that the impulse responses are taken to be non-causal, to allow for relative delays between the FBF and the ANC branches.

In order to fulfill the FIR structure constraint in (47.29), the filters update is now given by

$$\tilde{\boldsymbol{G}}(\ell+1,k) = \boldsymbol{G}(k,\ell) + \mu\frac{\boldsymbol{U}(k,\ell)Y^*(k,\ell)}{P_{\text{est}}(k,\ell)}\,,$$
$$\boldsymbol{G}(\ell+1,k) \xleftarrow{\text{FIR}} \tilde{\boldsymbol{G}}(\ell+1,k)\,. \tag{47.30}$$

The operator $\xleftarrow{\text{FIR}}$ includes the following three stages, applied per filter: transformation of $\tilde{G}_m(\ell+1,k)$ to the time domain, truncation of the resulting impulse response to the interval $[-K_L, K_R]$ (i. e., imposing the FIR constraint), and transformation back to the frequency domain. The various filtering operations involved in the algorithm (multiplications in the transform domain) are realized using the overlap-and-save method [47.50, 51].

The resulting algorithm is merely an extension of the original Griffiths and Jim algorithm for the arbitrary ATF case. Figure 47.3 depicts a block diagram of the algorithm. The steps involved in the computation are summarized in Table 47.2. The matched beamformer $\boldsymbol{W}_0(k)$ and the BM $\mathcal{H}(k)$ are assumed to be known at this stage.

47.2.3 Time-Domain Generalized Sidelobe Canceller

The most commonly used GSC structure is the classical time-domain counterpart of the algorithm, proposed

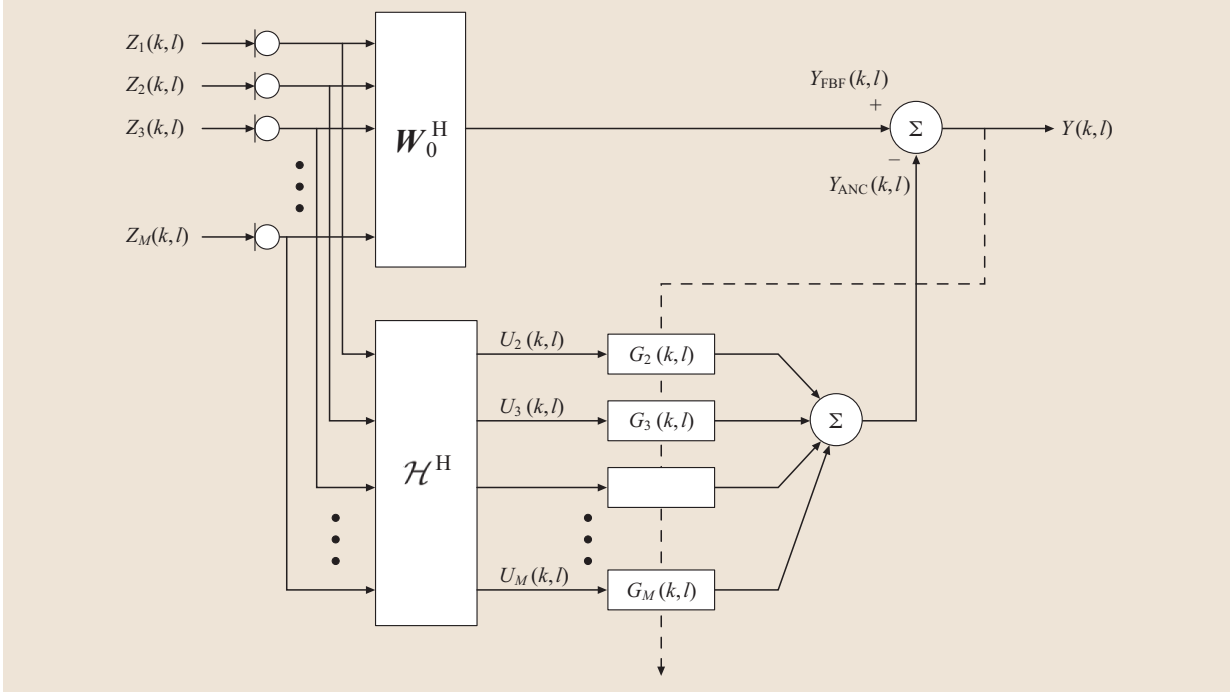

Fig. 47.3 Linearly constrained adaptive beamformer

Table 47.2 Frequency-domain GSC algorithm

1) Fixed beamformer:
$Y_{\mathrm{FBF}}(k,\ell) = \boldsymbol{W}_0^{\mathrm{H}}(k)\boldsymbol{Z}(k,\ell)$
2) Noise reference signals:
$\boldsymbol{U}(k,\ell) = \mathcal{H}^{\mathrm{H}}(k)\boldsymbol{Z}(k,\ell)$
3) Output signal:
$Y(k,\ell) = Y_{\mathrm{FBF}}(k,\ell) - \boldsymbol{G}^{\mathrm{H}}(k,\ell)\boldsymbol{U}(k,\ell)$
4) Filters update
$\tilde{\boldsymbol{G}}(t+1,k) = \boldsymbol{G}(k,\ell) + \mu_{\mathrm{g}} \frac{\boldsymbol{U}(k,\ell)Y^*(k,\ell)}{P_{\mathrm{est}}(k,\ell)}$
$\boldsymbol{G}(\ell+1,k) \overset{\mathrm{FIR}}{\longleftarrow} \tilde{\boldsymbol{G}}(\ell+1,k)$,
where
$P_{\mathrm{est}}(k,\ell) = \alpha_{\mathrm{p}} P_{\mathrm{est}}(\ell-1,k) + (1-\alpha_{\mathrm{p}}) \sum_m \lvert Z_m(k,\ell)\rvert^2$
5) Keep only nonaliased samples

by *Griffiths* and *Jim* [47.36]. For completeness of the exposition, we present now the time-domain algorithm.

Assuming that the array is steered towards the desired speech signal (refer to steering-related issues in Sect. 47.4), the FBF is given by

$$y_{\mathrm{FBF}}(t) = \sum_{m=1}^{M} z_m(t) ,$$

which is the simple delay-and-sum beamformer. Under the same delay-only steered array assumptions, it is evident that

$$u_m(t) = z_m(t) - z_1(t);\ m = 2, \ldots , M ,$$

are noise-only signals and that the desired speech component is cancelled out.

The filters $\boldsymbol{g}_m$ are updated in the time domain. The error signal (which is also the output of the enhancement algorithm) is given by,

$$y(t) = y_{\mathrm{FBF}}(t) - \sum_{m=2}^{M} \sum_{i=-K_L}^{K_R} g_{m,i}(t) u_m(t-i) . \tag{47.31}$$

Define, for $m = 2, \ldots , M$:

$$\boldsymbol{u}_m^{\mathrm{T}}(t) = \big(u_m(t+K_L) \ \cdots \ u_m(t) \ \cdots \ u_m(t-K_R)\big) .$$

Then, the adaptive normalized multichannel LMS solution is given by

$$\boldsymbol{g}_m(t+1) = \boldsymbol{g}_m(t) + \frac{\mu}{p_{\mathrm{est}}(t)} \boldsymbol{u}_m(t) y(t);\ m = 2, \ldots , M ,$$

where

$$p_{\mathrm{est}}(t) = \sum_{m=1}^{M} \lVert \boldsymbol{u}_m(t) \rVert^2 . \tag{47.32}$$

47.3 Fixed Beamformer and Blocking Matrix

In the previous section, we derived the generalized sidelobe canceller and showed that it includes a fixed beamformer, given in (47.20), a blocking matrix, given in (47.17), and a multichannel ANC, given in (47.26). Note that knowledge of the ATFs $\boldsymbol{A}(k)$ (assumed to be slowly time variant) suffices to determine both the FBF and BM. In this section we present three methods for determination of the fixed beamformer and the blocking matrix.

47.3.1 Using Acoustical Transfer Functions

A typical RIR is depicted in Fig. 47.4. It can be seen that the impulse response can get very long (several thousand taps), which makes the estimation task quite cumbersome. This impulse response was generated by the image method [47.52] proposed by Allen and Berkley (The authors thank E. A. P. Habets from TU Eindhoven, The Netherlands, for providing an efficient implementation

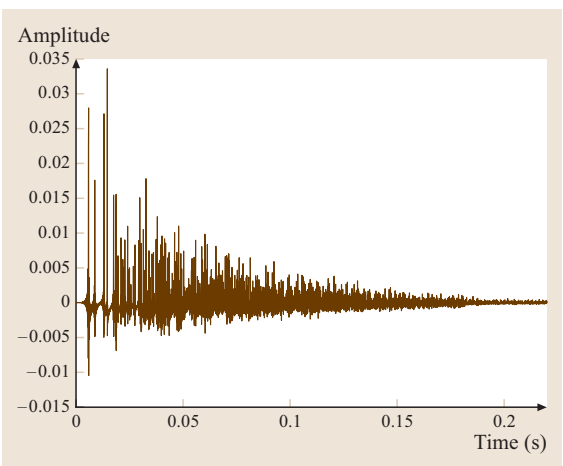

Fig. 47.4 A typical room impulse response with a reverberation time of 0.4 s

of the image method.). By this method, the RIR is generated by simulating multiple reflections of the sound source from the room walls. Two distinct segments of the impulse response can be observed. The first consists of the direct propagation path with a few early, distinguishable reflections. The second segment consists of overlapping random arrivals, with an exponentially decaying envelope. This segment is usually referred to as the tail of the RIR. Using this model we can estimate the various blocks of the GSC.

Assuming that $\boldsymbol{A}(k)$ is known, we have by (47.22), (47.20), and (47.3)

$$Y_{\mathrm{FBF}}(k,\ell) = S(k,\ell) + \frac{\boldsymbol{A}^{\mathrm{H}}(k)}{\|\boldsymbol{A}(k)\|^2}[\boldsymbol{N}_{\mathrm{s}}(k,\ell)+\boldsymbol{N}_{\mathrm{t}}(k,\ell)]\,. \quad (47.33)$$

The first term on the right-hand side is the signal term, and the second is the noise term. The FBF in this case is hence a matched filter-and-sum beamformer (see also [47.16, 17]).

Considering the blocking matrix, there are many alternatives for blocking the desired speech signal in the reference channels. One alternative is calculation of

$$U_m(k,\ell) = A_m(k)Z_{m-1}(k,\ell) - A_{m-1}(k)Z_m(k,\ell)$$

for $m = 2,\ldots,M$. Any other combination of the microphone signals is applicable.

47.3.2 Using Delay-Only Filters

The simplest and yet the most widely used model for the ATF is a delay-only model. Arbitrarily defining the first microphone as the reference microphone we have

$$\boldsymbol{A}(k) = \left(1 \;\; \mathrm{e}^{-\mathrm{i}\frac{2\pi}{K}\tau_2} \;\; \mathrm{e}^{-\mathrm{i}\frac{2\pi}{K}\tau_3} \;\; \ldots \;\; \mathrm{e}^{-\mathrm{i}\frac{2\pi}{K}\tau_M}\right),$$

where $\tau_2,\ldots,\tau_M$ are the relative delays between each microphone and the reference microphone.

In the delay-only case, the FBF simplifies to the delay-and-sum beamformer, given by

$$\boldsymbol{W}_0(k) = \left(1 \;\; \mathrm{e}^{\mathrm{i}\frac{2\pi}{K}\tau_2} \;\; \mathrm{e}^{\mathrm{i}\frac{2\pi}{K}\tau_3} \;\; \ldots \;\; \mathrm{e}^{\mathrm{i}\frac{2\pi}{K}\tau_M}\right).$$

To avoid noncausal delays, a fixed amount of delay can be introduced.

It can easily be verified that the matrix

$$\mathcal{H}(k) = \begin{pmatrix} -\mathrm{e}^{\mathrm{i}\frac{2\pi}{K}\tau_2} & -\mathrm{e}^{\mathrm{i}\frac{2\pi}{K}\tau_2} & \ldots & -\mathrm{e}^{\mathrm{i}\frac{2\pi}{K}\tau_2} \\ 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ & \ldots & \ddots & \\ 0 & 0 & \ldots & 1 \end{pmatrix} \quad (47.34)$$

is a proper BM under the assumption of delay-only impulse response. This BM, originally proposed Griffiths and *Jim* [47.36], performs delay compensation and subtraction. It can be regarded as steering $M-1$ null beams towards the desired speech signal.

47.3.3 Using Relative Transfer Functions

We have shown, on the one hand, that the RIR can be very long and hence difficult to estimate. On the other hand, the use of delay-only model suffers from severe undermodeling problems. A good compromise is the use of the relative transfer function (RTF) between sensors. Define the RTFs as the ratio

$$\tilde{\boldsymbol{A}}^{\mathrm{T}}(k) \triangleq \left(1 \;\; \tfrac{A_2(k)}{A_1(k)} \;\; \tfrac{A_3(k)}{A_1(k)} \;\; \ldots \;\; \tfrac{A_M(k)}{A_1(k)}\right) = \frac{\boldsymbol{A}^{\mathrm{T}}(k)}{A_1(k)}\,. \quad (47.35)$$

Note that the ATF may have zeros outside the unit circle, as it is not necessarily a minimum-phase system. Thus to ensure stability of the RTFs we allow for noncausal systems. Therefore, we model the impulse response of the m-th ratio as

$$\tilde{\boldsymbol{a}}_m^{\mathrm{T}}(t) = \left(\tilde{a}_{m,-q_L}(t) \;\; \ldots \;\; \tilde{a}_{m,q_R}(t)\right). \quad (47.36)$$

It was experimentally shown that RTFs are usually much shorter than the corresponding ATFs [47.38], hence the FIR assumption may be justified.

Replacing in (47.20) the actual ATFs by the RTFs, the FBF becomes

$$\boldsymbol{W}_0(k) = \frac{\tilde{\boldsymbol{A}}(k)}{\|\tilde{\boldsymbol{A}}(k)\|^2}\,. \quad (47.37)$$

By (47.22) and (47.3) we then have

$$Y_{\mathrm{FBF}}(k,\ell) = A_1(k)S(k,\ell) + \frac{\tilde{\boldsymbol{A}}^{\mathrm{H}}(k)}{\|\tilde{\boldsymbol{A}}(k)\|^2}[\boldsymbol{N}_{\mathrm{s}}(k,\ell)+\boldsymbol{N}_{\mathrm{t}}(k,\ell)]\,. \quad (47.38)$$

Thus, when $\boldsymbol{W}_0(k,\ell)$ is given by (47.37), the signal term of $Y_{\mathrm{FBF}}(k,\ell)$ is the desired signal distorted only by the first ATF, $A_1(k)$. Note, however, that all the sensor outputs are summed coherently.

It can be easily verified that the use of the following BM suffices for completely eliminating the desired speech signal, provided that the RTFs are correctly modeled and estimated:

$$\mathcal{H}(k)=\begin{pmatrix} -\tilde{A}_2^*(k) & -\tilde{A}_3^*(k) & \dots & -\tilde{A}_M^*(k) \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & \dots & \ddots & \\ 0 & 0 & \dots & 1 \end{pmatrix} . \quad (47.39)$$

For this choice of the BM, the reference signals are given by,

$$U_m(k,\ell) = Z_m(k,\ell) - \tilde{A}_m(k) Z_1(k,\ell); \quad m = 2, 3, \dots, M . \quad (47.40)$$

47.4 Identification of the Acoustical Transfer Function

The specific choice of the ATFs model governs the applicable estimation method.

47.4.1 Signal Subspace Method

Affes and *Grenier* [47.37] prove that the identification of source-to-array impulse responses is possible by subspace tracking. Assume that the approximation in (47.3) is valid, i. e., the multichannel correlation matrix of the received signals is given by

$$\boldsymbol{\Phi}_{\boldsymbol{ZZ}}(k,\ell) = \phi_{\mathrm{ss}}(k,\ell)\boldsymbol{A}(k)\boldsymbol{A}^{\mathrm{H}}(k) + \boldsymbol{\Phi}_{\boldsymbol{NN}}(k,\ell) .$$

Assume also that the noise signal is spatially white, i. e., $\boldsymbol{\Phi}_{\boldsymbol{NN}}(k,\ell) = \sigma_n^2(k,\ell)\mathbf{I}$ (an extension for spatially nonwhite noise is described in [47.53]). The eigenvalues of the received signals correlation matrix are given by

$$\begin{aligned} &\lambda_l = \sigma_n^2(k,\ell) \quad l = 1, \dots, M-1 \\ &\lambda_M > \sigma_n^2(k,\ell) \ \text{ otherwise} . \end{aligned} \quad (47.41)$$

From the matrix structure we conclude that the most dominant eigenvector of $\boldsymbol{\Phi}_{\boldsymbol{ZZ}}(k,\ell)$ is given by $\boldsymbol{A}(k)$ (up to a scale factor). Hence, using the eigenvalue decomposition of $\boldsymbol{\Phi}_{\boldsymbol{ZZ}}(k,\ell)$, we can estimate the desired signal ATFs. In the case of spatially nonwhite noise signals the generalized eigenvalue decomposition (GEVD), using the known noise correlation matrix $\boldsymbol{\Phi}_{\boldsymbol{NN}}(k,\ell)$, can be applied instead [47.53].

Yang [47.54] proved that finding the most dominant eigenvectors is equivalent to minimizing a quadratic cost function. For the following derivation it is more convenient to deal with normalized terms. Let

$$\begin{aligned} \boldsymbol{Z}(k,\ell) &= \boldsymbol{A}(k)S(k,\ell) + \boldsymbol{N}(k,\ell) \\ &= \frac{\boldsymbol{A}(k)}{\|\boldsymbol{A}(k)\|}\|\boldsymbol{A}(k)\|S(k,\ell) + \boldsymbol{N}(k,\ell) \\ &\triangleq \bar{\boldsymbol{A}}(k)\bar{S}(k,\ell) + \boldsymbol{N}(k,\ell) , \end{aligned} \quad (47.42)$$

where $\bar{\boldsymbol{A}}(k) \triangleq \boldsymbol{A}(k)/\|\boldsymbol{A}(k)\|$ is the normalized ATFs vector, namely $\bar{\boldsymbol{A}}^{\mathrm{H}}(k)\bar{\boldsymbol{A}}(k) = 1$, and $\bar{S}(k,\ell) \triangleq \|\boldsymbol{A}(k)\| S(k,\ell)$ is the normalized desired speech signal.

When only the single most dominant eigenvector is required (as in the discussed case), *Affes* and *Grenier* [47.37] showed that Yang's criterion simplifies to the following minimization,

$$E\{\|[\mathbf{I} - \bar{\boldsymbol{A}}(k)\bar{\boldsymbol{A}}^{\mathrm{H}}(k)]\boldsymbol{Z}(k,\ell)\|^2\} . \quad (47.43)$$

Similar to the approximation used in the derivation of LMS algorithm, the received signal correlation matrix is approximated by its instantaneous value $\boldsymbol{\Phi}_{\boldsymbol{ZZ}}(k,\ell) \approx \boldsymbol{Z}(k,\ell)\boldsymbol{Z}^{\mathrm{H}}(k,\ell)$. Furthermore, approximating $\hat{\bar{\boldsymbol{A}}}^{\mathrm{H}}(k,\ell)\hat{\bar{\boldsymbol{A}}}(k,\ell) \approx 1$, where $\hat{\bar{\boldsymbol{A}}}(k,\ell)$ is an estimate of $\bar{\boldsymbol{A}}(k)$ at the current time instant, the following sequential procedure can be derived:

$$\begin{aligned} &\hat{\bar{\boldsymbol{A}}}(k,\ell+1) \\ &= \hat{\bar{\boldsymbol{A}}}(k,\ell) + \mu_{\mathrm{a}}(k,\ell)[\boldsymbol{Z}(k,\ell) \\ &\quad - \hat{\bar{\boldsymbol{A}}}(k,\ell)\hat{\bar{\boldsymbol{A}}}^{\mathrm{H}}(k,\ell)\boldsymbol{Z}(k,\ell)][\hat{\bar{\boldsymbol{A}}}^{\mathrm{H}}(k,\ell)\boldsymbol{Z}(k,\ell)]^* . \end{aligned} \quad (47.44)$$

Define $\bar{Y}_{\mathrm{FBF}}(k,\ell) \triangleq \bar{\boldsymbol{A}}^{\mathrm{H}}(k,\ell)\boldsymbol{Z}(k,\ell)$, and observe that the desired signal component at $\bar{Y}_{\mathrm{FBF}}(k,\ell)$ is $\bar{S}(k,\ell)$. Then,

$$\begin{aligned} &\hat{\bar{\boldsymbol{A}}}(k,\ell+1) \\ &= \hat{\bar{\boldsymbol{A}}}(k,\ell) + \mu_{\mathrm{a}}(k,\ell)[\boldsymbol{Z}(k,\ell) \\ &\quad - \hat{\bar{\boldsymbol{A}}}(k,\ell)\bar{Y}_{\mathrm{FBF}}(k,\ell)]\bar{Y}^*_{\mathrm{FBF}}(k,\ell) . \end{aligned} \quad (47.45)$$

Now, it is easy to verify that

$$\boldsymbol{Z}(k,\ell) - \hat{\bar{\boldsymbol{A}}}(k,\ell)\bar{Y}_{\mathrm{FBF}}(k,\ell)$$

are noise-only signals, provided that the estimate $\hat{\bar{\boldsymbol{A}}}(k,\ell)$ converges to the true normalized ATFs vector $\bar{\boldsymbol{A}}(k)$.

Hence, it can serve as a BM output, namely $\boldsymbol{U}(k,\ell)$. (There is a slight difference between the proposed BM and the conventional BM, as the number of the components in $\boldsymbol{U}(k,\ell)$ in the current scheme is M rather than $M-1$ as with the latter.) Collecting all terms we finally have

$$\hat{\hat{\boldsymbol{A}}}(k,\ell+1) = \hat{\hat{\boldsymbol{A}}}(k,\ell)+\mu_{\mathrm{a}}(k,\ell)\boldsymbol{U}(k,\ell)\bar{Y}^{*}_{\mathrm{FBF}}(k,\ell)\,. \tag{47.46}$$

Note that this procedure yields an estimate of the normalized ATFs rather than the ATFs themselves. *Affes* and *Grenier* [47.37] argue that $\|\boldsymbol{A}(k)\|$ is invariant to small talker movements and could be estimated in advance.

As a final remark, we would like to point out that the estimation procedure in (47.46) has many similarities to the method of *Hoshuyama* et al. [47.55] for robust design of the BM, and with the decorrelation criterion presented by *Weinstein* et al. [47.56] and further adapted to the GSC structure by *Gannot* [47.57]. We will elaborate on this issue in Sect. 47.5 when discussing the robust beamformers.

47.4.2 Time Difference of Arrival

When a delay-only steering is applied, an estimation of the time difference of arrival between the microphones suffices to model the entire impulse response. It should be noted however, that this procedure usually undermodels the RIR and is not sufficient for the problem at hand. Many algorithms were proposed for estimation the time difference of arrival (TDOA) [47.58, 59]. A survey of state-of-the-art methods for TDOA estimation can be found in [47.60]. This topic is beyond the scope of this chapter.

47.4.3 Relative Transfer Function Estimation

We present two methods for RTF estimation. The first is based on speech nonstationarity [47.38] , and the second employs the speech presence probability [47.61, 62].

Using Signal Nonstationarity

In this section, we review the system identification technique proposed by *Shalvi* and *Weinstein* [47.63] and later used by *Gannot* et al. [47.38] in the context of microphone arrays. This method relies on the assumptions that the background noise signal is stationary, that the desired signal $s(t)$ is nonstationary, and that the support of the relative impulse response between the sensors is finite and slowly time varying. (Note that the relative impulse response between the sensors is generally of infinite length, since it represents the ratio of ATFs. However, in real environments, the energy of the relative impulse response often decays much faster than the corresponding ATF [47.38]. Therefore, the finite-support assumption is practically not very restrictive.)

Rearranging terms in (47.40) we have

$$Z_m(k,\ell) = \tilde{A}_m(k)Z_1(k,\ell)+U_m(k,\ell)\,. \tag{47.47}$$

We assume that the RTFs are slowly changing in time compared to the time variations of the desired signal. We further assume that the statistics of the noise signal is slowly changing compared to the statistics of the desired signal. Consider some analysis interval during which the ATFs are assumed to be time invariant and the noise signal is assumed to be stationary. We divide that analysis interval into frames. Consider the i-th frame. By (47.47) we have

$$\boldsymbol{\Phi}^{(i)}_{z_m z_1}(k) = \tilde{A}_m(k)\boldsymbol{\Phi}^{(i)}_{z_1 z_1}(k)+\boldsymbol{\Phi}_{u_m z_1}(k)\,, \quad i=1,\ldots,I\,, \tag{47.48}$$

where I is the number of frames, $\boldsymbol{\Phi}^{(i)}_{z_i z_j}(k)$ is the cross-PSD between z_i and z_j at frequency bin k during the i-th frame, and $\boldsymbol{\Phi}_{u_m z_1}(k)$ is the cross-PSD between u_m and z_1 at frequency bin k, which is independent of the frame index due to the noise stationarity. Now, equations (47.2) and (47.40) imply that, when the signal is present,

$$U_m(k,\ell) = N^{\mathrm{s}}_m(k,\ell)-\tilde{A}_m(k)N^{\mathrm{s}}_1(k,\ell) \tag{47.49}$$

$$Z_1(k,\ell) = A_1(k)S(k,\ell)+N^{\mathrm{s}}_1(k,\ell)\,. \tag{47.50}$$

Let $\hat{\boldsymbol{\Phi}}^{(i)}_{z_1 z_1}(k)$, $\hat{\boldsymbol{\Phi}}^{(i)}_{z_m z_1}(k)$, and $\hat{\boldsymbol{\Phi}}^{(i)}_{u_m z_1}(k)$ be estimates of $\boldsymbol{\Phi}^{(i)}_{z_1 z_1}(k)$, $\boldsymbol{\Phi}^{(i)}_{z_m z_1}(k)$, and $\boldsymbol{\Phi}_{u_m z_1}(k)$, respectively. The estimates are obtained by replacing expectations with averages. Note that (47.48) also holds for the estimated values. Let $\varepsilon^{(i)}_m(k) = \hat{\boldsymbol{\Phi}}^{(i)}_{u_m z_1}(k)-\boldsymbol{\Phi}_{u_m z_1}(k)$ denote the estimation error of the cross-PSD between z_1 and u_m in the i-th frame. We then obtain,

$$\hat{\boldsymbol{\Phi}}^{(i)}_{z_m z_1}(k) = \tilde{A}_m(k)\hat{\boldsymbol{\Phi}}^{(i)}_{z_1 z_1}(k)+\boldsymbol{\Phi}_{u_m z_1}(k) +\varepsilon^{(i)}_m(k), \qquad i=1,\ldots,I\,. \tag{47.51}$$

If the noise reference signals $U_m(k,\ell)$, $m=2,\ldots,M$ were uncorrelated with $Z_1(k,\ell)$, then the standard system identification estimate, $\tilde{A}_m(k)=\hat{\boldsymbol{\Phi}}_{z_m z_1}(k)/\hat{\boldsymbol{\Phi}}_{z_1 z_1}(k)$, could be used to obtain an unbiased estimate of $\tilde{A}_m(k)$. Unfortunately, by (47.49) and (47.50), $U_m(k,\ell)$ and $Z_1(k,\ell)$ are in general correlated. Hence in [47.63] it is

proposed to obtain an unbiased estimate of $\tilde{A}_m(k)$ by applying the least-squares (LS) procedure to the following set of overdetermined equations

$$\begin{aligned}\boldsymbol{x} &\triangleq \begin{pmatrix} \hat{\boldsymbol{\Phi}}^{(1)}_{z_m z_1}(k) \\ \hat{\boldsymbol{\Phi}}^{(2)}_{z_m z_1}(k) \\ \vdots \\ \hat{\boldsymbol{\Phi}}^{(K)}_{z_m z_1}(k) \end{pmatrix} \\ &= \begin{pmatrix} \hat{\boldsymbol{\Phi}}^{(1)}_{z_1 z_1}(k) & 1 \\ \hat{\boldsymbol{\Phi}}^{(2)}_{z_1 z_1}(k) & 1 \\ \vdots & \\ \hat{\boldsymbol{\Phi}}^{(K)}_{z_1 z_1}(k) & 1 \end{pmatrix} \begin{pmatrix} \tilde{A}_m(k) \\ \boldsymbol{\Phi}_{u_m z_1}(k) \end{pmatrix} + \begin{pmatrix} \varepsilon^{(1)}_m(k) \\ \varepsilon^{(2)}_m(k) \\ \vdots \\ \varepsilon^{(K)}_m(k) \end{pmatrix} \\ &\triangleq \boldsymbol{G}\,\boldsymbol{\theta} + \boldsymbol{\epsilon}\,, \end{aligned} \tag{47.52}$$

where a separate set of equations is used for each $m = 2, \ldots, M$. The weighted least-squares (WLS) estimate of $\boldsymbol{\theta}$ is obtained by

$$\begin{aligned}\begin{pmatrix} \hat{\tilde{A}}_m(k) \\ \hat{\boldsymbol{\Phi}}_{u_m z_1}(k) \end{pmatrix} &= \hat{\boldsymbol{\theta}} \\ &= \arg\min_{\boldsymbol{\theta}} (\boldsymbol{x} - \boldsymbol{G}\,\boldsymbol{\theta})^{\mathrm{H}} \boldsymbol{W} (\boldsymbol{x} - \boldsymbol{G}\,\boldsymbol{\theta}) \\ &= (\boldsymbol{G}^{\mathrm{H}} \boldsymbol{W} \boldsymbol{G})^{-1} \boldsymbol{G}^{\mathrm{H}} \boldsymbol{W} \boldsymbol{x}\,, \end{aligned} \tag{47.53}$$

where $\boldsymbol{W}$ is a positive Hermitian weighting matrix, and $\boldsymbol{G}^{\mathrm{H}}\boldsymbol{W}\boldsymbol{G}$ is required to be invertible.

Shalvi and Weinstein suggested two alternative weighting matrices. One alternative is given by

$$W_{ij} = \begin{cases} T_i\,, & i = j \\ 0\,, & i \neq j \end{cases}\,, \tag{47.54}$$

where T_i is the length of subinterval i, so that longer intervals have higher weights. In this case, (47.53) reduces to

$$\hat{\tilde{A}}(k) = \frac{\langle \hat{\phi}_{z_m z_1}(k)\, \hat{\phi}_{z_1 z_1}(k) \rangle - \langle \hat{\phi}_{z_m z_1}(k) \rangle \langle \hat{\phi}_{z_1 z_1}(k) \rangle}{\langle \hat{\phi}^2_{z_1 z_1}(k) \rangle - \langle \hat{\phi}_{z_1 z_1}(k) \rangle^2} \tag{47.55}$$

with the average operation defined by

$$\langle \varphi(k) \rangle \triangleq \frac{\sum_{i=1}^{I} T_i \varphi^{(i)}(k)}{\sum_{i=1}^{I} T_i}\,. \tag{47.56}$$

Another alternative for $\boldsymbol{W}$, that minimizes the covariance of $\hat{\boldsymbol{\theta}}$, is given by

$$W_{ij} = \frac{B}{\hat{\phi}_{u_m u_m}(k)} \begin{cases} T_i / \hat{\phi}^{(i)}_{z_1 z_1}(k)\,, & i = j \\ 0\,, & i \neq j \end{cases}\,, \tag{47.57}$$

where B is related to the bandwidth of the window used for the cross-PSD estimation [47.63]. With this choice of the weighting function, (47.53) yields

$$\begin{aligned}\hat{\tilde{A}}(k) = &\frac{\langle 1/\hat{\phi}_{z_1 z_1}(k) \rangle \langle \hat{\phi}_{z_m z_1}(k) \rangle}{\langle \hat{\phi}_{z_1 z_1}(k) \rangle \langle 1/\hat{\phi}_{z_1 z_1}(k) \rangle - 1} \\ &- \frac{\langle \hat{\phi}_{z_m z_1}(k)\,/\,\hat{\phi}_{z_1 z_1}(k) \rangle}{\langle \hat{\phi}_{z_1 z_1}(k) \rangle \langle 1/\hat{\phi}_{z_1 z_1}(k) \rangle - 1} \end{aligned} \tag{47.58}$$

and the variance of $\hat{\tilde{A}}(k)$ is given by

$$\mathrm{var}\{\hat{\tilde{A}}(k)\} = \frac{1}{BT} \frac{\phi_{u_m u_m}(k) \langle 1/\phi_{z_1 z_1}(k) \rangle}{\langle \phi_{z_1 z_1}(k) \rangle \langle 1/\phi_{z_1 z_1}(k) \rangle - 1}\,, \tag{47.59}$$

where $T \triangleq \sum_{i=1}^{I} T_i$ is the total observation interval. Special attention should be given to choosing the frame length. On the one hand, it should be longer than the correlation length of $z_m(t)$, which must be longer than the length of the filter $a_m(t)$. On the other hand, it should be short enough for the filter time invariance and the noise quasistationarity assumptions to hold.

A major limitation of the WLS optimization in (47.53) is that both the identification of $\tilde{\boldsymbol{A}}(k)$ and the estimation of the cross-PSD $\phi_{u_m z_1}(k)$ are carried out using the same weight matrix $\boldsymbol{W}$. That is, each subinterval i is given the same weight, whether we are trying to find an estimate for $\tilde{\boldsymbol{A}}(k)$ or for $\phi_{u_m z_1}(k)$. However, subintervals with higher SNR values are of greater importance when estimating $\tilde{A}(k)$, whereas the opposite is true when estimating $\phi_{u_m z_1}(k)$. Consequently, the optimization criterion in (47.53) consists of two conflicting requirements: one is minimizing the error variance of $\tilde{\boldsymbol{A}}(k)$, which pulls the weight up to higher values on higher SNR subintervals. The other requirement is minimizing the error variance of $\phi_{u_m z_1}(k)$, which rather implies *smaller* weights on higher SNR subintervals. For instance, suppose we obtain observations on a relatively long low-SNR interval of length T_0, and on a relatively short high-SNR interval of length T_1 ($T_1 \ll T_0$). Then, the variance of $\tilde{\boldsymbol{A}}(k)$ in (47.59) is inversely proportional to the relative length of the high-SNR interval, $T_1/(T_0 + T_1)$. That is, including in the observation interval additional segments that do not contain speech (i. e., increasing T_0) increases the variance of $\tilde{\boldsymbol{A}}(k)$. This unnatural

consequence is a result of the desire to minimize the variance of $\phi_{u_m z_1}(k)$ by using larger weights on the segments that do not contain speech, while increasing the weights on such subintervals degrades the estimate for $\tilde{A}(k)$.

Another major limitation of RTF identification using nonstationarity is that the interfering signals are required to be stationary during the entire observation interval. The observation interval should include a certain number of subintervals that contain the desired signal, such that $\phi_{z_1 z_1}(k)$ is sufficiently nonstationary for all k. Unfortunately, if the desired signal is speech, the presence of the desired signal in the observed signals may be sparse in some frequency bands. This entails a very long observation interval, thus constraining the interfering signals to be stationary over long intervals. Furthermore, the RTF $\tilde{\boldsymbol{A}}(k)$ is assumed to be constant during the observation interval. Hence, very long observation intervals also restrict the capability of the system identification technique to track varying $\tilde{\boldsymbol{A}}(k)$ (e.g., tracking moving talkers in reverberant environments).

Using Speech Presence Probability

In this section, we present a system identification approach that is adapted to speech signals. Specifically, the presence of the desired speech signal in the time–frequency domain is uncertain, and the speech presence probability is utilized to separate the tasks of system identification and cross-PSD estimation. An estimate for $\tilde{\boldsymbol{A}}(k)$ is derived based on subintervals that contain speech, while subintervals that do not contain speech are of more significance when estimating the components of $\phi_{u_m z_1}(k)$.

Let the observed signals be divided in time into overlapping frames by the application of a window function and analyzed using the STFT. Under the same considerations leading to the estimation procedure based on speech nonstationarity (and based on the assumption that the RTFs can be modelled by short filters), (47.48) is still valid. Now using (47.48)-(47.50) and the fact that the desired signal $s(t)$ is uncorrelated with the interfering signals $n_m^{\mathrm{s}}(t)$; $m = 1, 2, \ldots, M$, we have

$$\begin{aligned}\phi_{z_m z_1}(k,\ell) &= \tilde{A}_m(k)|A_1|^2(k)\phi_{\mathrm{s\,s}}(k,\ell) \\ &\quad + \phi_{n_m^{\mathrm{s}} n_1^{\mathrm{s}}}(k,\ell)\,. \end{aligned} \tag{47.60}$$

Writing this equation in terms of the PSD estimates, we obtain for $m = 2, 3, \ldots, M$

$$\begin{aligned}\hat{\phi}_{z_m z_1}(k,\ell) &= \tilde{A}_m(k)|\hat{A}_1|^2(k)\hat{\phi}_{\mathrm{s\,s}}(k,\ell) \\ &\quad + \hat{\phi}_{n_m^{\mathrm{s}} n_1^{\mathrm{s}}}(k,\ell) + \varepsilon_m(k,\ell) \\ &= \tilde{A}_m(k)\hat{\phi}_{\tilde{\mathrm{s}}\,\tilde{\mathrm{s}}}(k,\ell) + \hat{\phi}_{n_m^{\mathrm{s}} n_1^{\mathrm{s}}}(k,\ell) + \varepsilon_m(k,\ell)\,, \end{aligned} \tag{47.61}$$

where $\varepsilon_m(k,\ell)$ denotes an estimation error and $\hat{\phi}_{\tilde{\mathrm{s}}\,\tilde{\mathrm{s}}}(k,\ell) = |\hat{A}_1|^2(k)\hat{\phi}_{\mathrm{s\,s}}(k,\ell)$ represents the PSD of the speech signal component in microphone 1. This gives us L equations, which may be written in a matrix form as

$$\begin{aligned}\hat{\boldsymbol{\psi}}_m(k) &\triangleq \begin{pmatrix} \hat{\phi}_{z_m z_1}(k,1) - \hat{\phi}_{n_m^{\mathrm{s}} n_1^{\mathrm{s}}}(k,1) \\ \hat{\phi}_{z_m z_1}(k,2) - \hat{\phi}_{n_m^{\mathrm{s}} n_1^{\mathrm{s}}}(k,2) \\ \vdots \\ \hat{\phi}_{z_m z_1}(k,L) - \hat{\phi}_{n_m^{\mathrm{s}} n_1^{\mathrm{s}}}(k,L) \end{pmatrix} \\ &= \begin{pmatrix} \hat{\phi}_{\tilde{\mathrm{s}}\,\tilde{\mathrm{s}}}(k,1) \\ \hat{\phi}_{\tilde{\mathrm{s}}\,\tilde{\mathrm{s}}}(k,2) \\ \vdots \\ \hat{\phi}_{\tilde{\mathrm{s}}\,\tilde{\mathrm{s}}}(k,L) \end{pmatrix} \tilde{A}_m(k) + \begin{pmatrix} \varepsilon_m(k,1) \\ \varepsilon_m(k,2) \\ \vdots \\ \varepsilon_m(k,L) \end{pmatrix} \\ &\triangleq \hat{\boldsymbol{\phi}}_{\tilde{\mathrm{s}}\,\tilde{\mathrm{s}}}(k)\,\tilde{A}_m(k) + \boldsymbol{\varepsilon}_m(k)\,. \end{aligned} \tag{47.62}$$

Since the RTF $\tilde{\boldsymbol{A}}_m(k)$ represents the coupling between the primary and reference sensors with respect to the *desired* source signal, the optimization criterion for the identification of $\tilde{A}_m(k)$ has to take into account only short-time frames which contain desired signal components. Specifically, let $I(k,\ell)$ denote an indicator function for the signal presence [$I(k,\ell) = 1$ if $\phi_{\tilde{\mathrm{s}}\,\tilde{\mathrm{s}}}(k,\ell) \neq 0$, i. e., during H_1, and $I(k,\ell) = 0$ otherwise], and let $\mathbf{I}(k)$ represent a diagonal matrix with the elements $[I(k,1), I(k,2), \ldots, I(k,L)]$ on its diagonal. Then the WLS estimate of $\tilde{A}(k)$ is obtained by

$$\begin{aligned}\hat{\tilde{A}}_m &= \arg\min_{\tilde{A}_m} \left\{ [\mathbf{I}\boldsymbol{\varepsilon}_m]^{\mathrm{H}} \mathbf{W} [\mathbf{I}\boldsymbol{\varepsilon}_m] \right\} \\ &= \arg\min_{\tilde{A}_m} \Big\{ [\hat{\boldsymbol{\psi}}_m - \hat{\boldsymbol{\phi}}_{\tilde{\mathrm{s}}\tilde{\mathrm{s}}} \tilde{A}_m]^{\mathrm{H}} \\ &\quad \mathbf{I}\mathbf{W}\mathbf{I} [\hat{\boldsymbol{\psi}}_m - \hat{\boldsymbol{\phi}}_{\tilde{\mathrm{s}}\tilde{\mathrm{s}}} \tilde{A}_m] \Big\} \\ &= [\hat{\boldsymbol{\phi}}_{\tilde{\mathrm{s}}\tilde{\mathrm{s}}}^{\mathrm{T}} \mathbf{I}\mathbf{W}\mathbf{I} \hat{\boldsymbol{\phi}}_{\tilde{\mathrm{s}}\tilde{\mathrm{s}}}]^{-1} \hat{\boldsymbol{\phi}}_{\tilde{\mathrm{s}}\tilde{\mathrm{s}}}^{\mathrm{T}} \mathbf{I}\mathbf{W}\mathbf{I} \hat{\boldsymbol{\psi}}_m\,, \end{aligned} \tag{47.63}$$

where the argument k has been omitted for notational simplicity. Recognizing the product $\mathbf{I}\,\mathbf{W}\,\mathbf{I}$ as the equivalent weight matrix, the variance of $\hat{\tilde{A}}$ is given by ([47.64] p. 405)

$$\begin{aligned}\mathrm{var}\{\hat{\tilde{A}}_m\} &= \left(\boldsymbol{\phi}_{\tilde{\mathrm{s}}\,\tilde{\mathrm{s}}}^{\mathrm{T}} \mathbf{I}\mathbf{W}\mathbf{I} \boldsymbol{\phi}_{\tilde{\mathrm{s}}\,\tilde{\mathrm{s}}}\right)^{-1} \boldsymbol{\phi}_{\tilde{\mathrm{s}}\,\tilde{\mathrm{s}}}^{\mathrm{T}} \mathbf{I}\mathbf{W}\mathbf{I}\,\mathrm{cov}(\boldsymbol{\varepsilon}_m) \\ &\quad \times \mathbf{I}\mathbf{W}\mathbf{I} \boldsymbol{\phi}_{\tilde{\mathrm{s}}\,\tilde{\mathrm{s}}} \left(\boldsymbol{\phi}_{\tilde{\mathrm{s}}\,\tilde{\mathrm{s}}}^{\mathrm{T}} \mathbf{I}\mathbf{W}\mathbf{I} \boldsymbol{\phi}_{\tilde{\mathrm{s}}\,\tilde{\mathrm{s}}}\right)^{-1}\,, \end{aligned} \tag{47.64}$$

where $\mathrm{cov}(\boldsymbol{\varepsilon}_m)$ is the covariance matrix of $\boldsymbol{\varepsilon}_m$. The matrix $\mathbf{W}$ that minimizes the variance of $\hat{\tilde{A}}$ therefore satisfies ([47.64] prop. 8.2.4)

$$\mathbf{I}\mathbf{W}\mathbf{I} = \mathbf{I}\,[\mathrm{cov}(\boldsymbol{\varepsilon}_m)]^{-1}\,\mathbf{I}\,. \tag{47.65}$$

This choice of $\mathbf{W}$ yields an asymptotically unbiased estimator

$$\hat{A} = \left\{\hat{\boldsymbol{\phi}}_{\tilde{s}\tilde{s}}{}^{\mathrm{T}}\mathbf{I}[\mathrm{cov}(\boldsymbol{\varepsilon}_m)]^{-1}\mathbf{I}\hat{\boldsymbol{\phi}}_{\tilde{s}\tilde{s}}\right\}^{-1} \times \hat{\boldsymbol{\phi}}_{\tilde{s}\tilde{s}}{}^{\mathrm{T}}\mathbf{I}[\mathrm{cov}(\boldsymbol{\varepsilon}_m)]^{-1}\mathbf{I}\hat{\boldsymbol{\psi}}_m \,, \tag{47.66}$$

which is known as the *minimum variance* or Gauss–Markov estimator. Substituting (47.65) into (47.64), we obtain the variance of the resulting estimator

$$\mathrm{var}\{\hat{A}\} = \left\{\boldsymbol{\phi}_{\tilde{s}\tilde{s}}{}^{\mathrm{T}}\mathbf{I}[\mathrm{cov}(\boldsymbol{\varepsilon}_m)]^{-1}\mathbf{I}\,\boldsymbol{\phi}_{\tilde{s}\tilde{s}}\right\}^{-1} . \tag{47.67}$$

The elements of $\mathrm{cov}(\boldsymbol{\varepsilon}_m)$ are asymptotically given by (see [47.61])

$$\mathrm{cov}[\varepsilon_m(k,\ell), \varepsilon_m(k,\ell')] = \begin{cases} \phi_{\tilde{s}\tilde{s}}(k,\ell)\,\phi_{u_m u_m}(k,\ell) \,, & \text{if } \ell = \ell' \,, \\ 0 \,, & \text{otherwise.} \end{cases} \tag{47.68}$$

Under the assumption that hypothesis $H_{0\mathrm{t}}$ is false, i. e., the noise is stationary, $\phi_{u_m u_m}(k,\ell)$ is independent of the frame index ℓ (in practice, it suffices that the statistics of the interfering signals is slowly changing compared with the statistics of the desired signal). Denoting by $\langle\cdot\rangle_\ell$ an average operation over the frame index ℓ

$$\langle\varphi(k,\ell)\rangle_\ell \triangleq \frac{1}{L}\sum_{\ell=1}^{L}\varphi(k,\ell) \,, \tag{47.69}$$

and substituting (47.68) into (47.66) and (47.67) we obtain

$$\hat{\tilde{A}}_m(k) = \frac{\langle I(k,\ell)[\hat{\phi}_{z_m z_1}(k,\ell) - \hat{\phi}_{n_m^{\mathrm{s}} n_1^{\mathrm{s}}}(k,\ell)]\rangle_\ell}{\langle I(k,\ell)\,\hat{\phi}_{\tilde{s}\tilde{s}}(k,\ell)\rangle_\ell} \,, \tag{47.70}$$

$$\mathrm{var}\{\hat{\tilde{A}}_m(k)\}\frac{\phi_{u_m u_m}(k)}{L\,\langle I(k,\ell)\,\phi_{\tilde{s}\tilde{s}}(k,\ell)\rangle_\ell} \,. \tag{47.71}$$

Note that, for a given frequency-bin index k, only frames that contain speech $[I(k,\ell) \neq 0]$ influence the values of $\hat{\tilde{A}}_m(k)$ and $\mathrm{var}\{\hat{\tilde{A}}_m(k)\}$. In contrast to the nonstationarity method, including in the observation interval additional segments that do not contain speech does not increase the variance of $\hat{\tilde{A}}_m(k)$ for any k. However, the proposed identification approach requires an estimate for $I(k,\ell)$, i. e. identifying which time–frequency bins (k,ℓ) contain the desired signal. In practice, the speech presence probability $p(k,\ell)$ can be estimated from the beamformer output (see Sect. 47.6), and an estimate for the indicator function is obtained by

$$\hat{I}(k,\ell) = \begin{cases} 1 \,, & \text{if } p(k,\ell) \geq p_0 \,, \\ 0 \,, & \text{otherwise,} \end{cases} \tag{47.72}$$

where p_0 $(0 \leq p_0 < 1)$ is a predetermined threshold. Note, also that the output of the beamformer consists of the filtered version of the desired speech $\tilde{s}(t)$, when the RTFs is used in the FBF branch. The parameter p_0 controls the trade-off between the detection and false alarm probabilities, which are defined by $P_{\mathrm{D}} \triangleq \mathcal{P}\{p(k,\ell) \geq p_0 \,|\, I(k,\ell) = 1\}$ and $P_{\mathrm{FA}} \triangleq \mathcal{P}\{p(k,\ell) \geq p_0 \,|\, I(k,\ell) = 0\}$. A smaller value of p_0 increases the detection probability and allows for more short-time frames to be involved in the estimation of $\tilde{\boldsymbol{A}}(k)$. However, a smaller value of p_0 also increases the false alarm probability, which may cause a mismodification of $\hat{\tilde{\boldsymbol{A}}}(k)$ due to frames that do not contain desired speech components.

The identification algorithm based on speech presence probability requires estimates for $\phi_{z_m z_1}(k,\ell)$, $\phi_{\tilde{s}\tilde{s}}(k,\ell)$, and $\phi_{n_m^{\mathrm{s}} n_1^{\mathrm{s}}}(k,\ell)$. An estimate for $\phi_{z_m z_1}(k,\ell)$ is obtained by applying a first-order recursive smoothing to the cross-periodogram of the observed signals, $Z_m(k,\ell)\,Z_1^*(k,\ell)$. Specifically,

$$\hat{\phi}_{z_m z_1}(k,\ell) = \alpha_{\mathrm{s}}\,\hat{\phi}_{z_m z_1}(k,\ell-1) + (1-\alpha_{\mathrm{s}})Z_m(k,\ell)\,Z_1^*(k,\ell) \,, \tag{47.73}$$

where the smoothing parameter α_{s} $(0 \leq \alpha_{\mathrm{s}} < 1)$ determines the equivalent number of cross-periodograms that are averaged, $N_\ell \approx (1+\alpha_{\mathrm{s}})/(1-\alpha_{\mathrm{s}})$. Typically, speech periodograms are recursively smoothed with an equivalent rectangular window of $T_{\mathrm{s}} = 0.2$ s long, which represents a good compromise between smoothing the noise and tracking the speech spectral variations [47.65]. Therefore, for a sampling rate of 8 kHz, an STFT window length of 256 samples and a frame update step of 128 samples, we use $\alpha_{\mathrm{s}} = (T_{\mathrm{s}} \cdot 8000/128 - 1)/(T_{\mathrm{s}} \cdot 8000/128 + 1) \approx 0.85$.

To obtain an estimate for the PSD of the desired signal, we can use the output of the multichannel postfilter discussed in Sect. 47.6, during speech presence, i. e., H_1 is true.

$$\hat{\phi}_{\tilde{s}\tilde{s}}(k,\ell) = \alpha_{\mathrm{s}}\,\hat{\phi}_{\tilde{s}\tilde{s}}(k,\ell-1) + (1-\alpha_{\mathrm{s}})I(k,\ell)\,|Y(k,\ell)|^2 \,. \tag{47.74}$$

The cross-PSD of the interfering signals, $n_m^{\mathrm{s}}(t)$ and $n_1^{\mathrm{s}}(t)$, is estimated by using the minima-controlled

recursive averaging (MCRA) approach [47.66, 67]. Specifically, past spectral cross-power values of the noisy observed signals are recursively averaged with a time-varying frequency-dependent smoothing parameter during periods for which H_{0s} is true

$$\hat{\phi}_{n_m^s n_1^s}(k,\ell) = \tilde{\alpha}_u(k,\ell)\,\hat{\phi}_{n_m^s n_1^s}(k,\ell-1) + \beta\,[1-\tilde{\alpha}_u(k,\ell)]\,Z_m(k,\ell)\,Z_1^*(k,\ell)\,, \tag{47.75}$$

where $\tilde{\alpha}_u(k,\ell)$ is the smoothing parameter ($0 < \tilde{\alpha}_u(k,\ell) \le 1$), and β ($\beta \ge 1$) is a factor that compensates the bias when the desired signal is absent [47.67]. The smoothing parameter is determined by the signal presence probability, $p(k,\ell)$, and a constant α_u ($0 < \alpha_u < 1$) that represents its minimal value

$$\tilde{\alpha}_u(k,\ell) = \alpha_u + (1-\alpha_u)p(k,\ell)\,. \tag{47.76}$$

The value of $\tilde{\alpha}_u$ is close to 1 when the desired signal is present to prevent the noise cross-PSD estimate from increasing as a result of signal components. It decreases linearly with the probability of signal presence to allow a faster update of the noise estimate. The value of α_u compromises between the tracking rate (response rate to abrupt changes in the noise statistics) and the variance of the noise estimate. Typically, in the case of high levels of nonstationary noise, a good compromise is obtained by $\alpha_u = 0.85$ [47.67]. Substituting these spectral estimates into (47.70) we obtain an estimate for $\tilde{A}_m(k)$.

47.5 Robustness and Distortion Weighting

Beamformers often suffer from sensitivity to signal mismatch. The GSC in particular suffers from two basic problems. First, nonideal FBF can lead to noncoherent filter-and-sum operation. *Doclo* and *Moonen* [47.27] and *Nordholm* et al. [47.68] use spatial and frequency-domain constraints to improve the robustness of beamformers. The second problem, which is the concern of this survey, is the leakage phenomenon, caused by imperfect BM. If the desired speech leaks into the noise reference signals $\boldsymbol{U}(k,\ell)$ the noise canceller filters will subtract speech components from the FBF output, causing self-cancellation of the desired speech, and hence a severe distortion. Note that, even when the ANC filters are adapted during noise-only periods, the self-cancellation is unavoidable. The goal of this section is to present several concepts for increasing the robustness of the GSC structure and reducing its sensitivity to signal mismatch.

Cox et al. [47.7] presented a thorough analysis of array sensitivity. The array output SNR is evidently given by

$$\mathrm{SNR}_{out}(k,\ell) = \frac{\phi_{ss}(k,\ell)\boldsymbol{W}^H(k)\boldsymbol{A}(k)\boldsymbol{A}^H(k)\boldsymbol{W}(k)}{\boldsymbol{W}^H(k)\boldsymbol{\Phi}_{\boldsymbol{NN}}(k,\ell)\boldsymbol{W}(k)}\,.$$

Now assume that the signal's ATFs are different from the ATFs used for designing the LCMV beamformer, i. e., $\breve{\boldsymbol{A}}(k) = \boldsymbol{A} + \boldsymbol{\epsilon}(k)$. Assume also that the spatial correlation matrix of the perturbation $\boldsymbol{\epsilon}(k)$ is given by $E\{\boldsymbol{\epsilon}(k)\boldsymbol{\epsilon}^H(k)\} = \sigma_\epsilon^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix of dimensions $M \times M$. Namely, we assume that the perturbation components are uncorrelated. Hence the expected output SNR is given by

$$E\{\mathrm{SNR}_{out}(k,\ell)\} = \frac{\phi_{ss}(k,\ell)\boldsymbol{W}^H(k)\big(\boldsymbol{A}(k)\boldsymbol{A}^H(k) + \sigma_\epsilon^2\mathbf{I}\big)\boldsymbol{W}(k)}{\boldsymbol{W}^H(k)\boldsymbol{\Phi}_{\boldsymbol{NN}}(k,\ell)\boldsymbol{W}(k)}\,. \tag{47.77}$$

Define the sensitivity of the array to the ATFs perturbation as $J(k,\ell)$

$$J(k,\ell) \triangleq \frac{\frac{\partial}{\partial\sigma_\epsilon^2}E\{\mathrm{SNR}_{out}(k,\ell)\}}{E\{\mathrm{SNR}_{out}(k,\ell)\}|_{\sigma_\epsilon^2=0}} = \frac{\boldsymbol{W}^H(k)\boldsymbol{W}(k)}{\boldsymbol{W}^H(k)\boldsymbol{A}(k)\boldsymbol{A}^H(k)\boldsymbol{W}(k)}\,. \tag{47.78}$$

The resulting expression is the reciprocal of the white-noise gain of the array. Using the array constraint $\boldsymbol{W}^H(k)\boldsymbol{A}(k) = 1$ we finally obtain the following expression for the sensitivity of the array to the ATFs perturbation,

$$J(k,\ell) = \boldsymbol{W}^H(k)\boldsymbol{W}(k)\,. \tag{47.79}$$

Specifically, the array sensitivity is equal to the norm of the beamformer weights. Hence, reducing the sensitivity of the array is equivalent to constraining the norm of the array filter coefficients. Due to (47.15) the array filters can be decomposed into two orthogonal filters, $\boldsymbol{W}(k) = \boldsymbol{W}_0(k) - \boldsymbol{V}(k) = \boldsymbol{W}_0(k) - \mathcal{H}(k)\boldsymbol{G}(k,\ell)$. It is therefore sufficient to constrain the adaptive filter norm, namely $\boldsymbol{G}^H(k,\ell)\boldsymbol{G}(k,\ell) = \|\boldsymbol{G}(k,\ell)\|^2 \le \Omega(k,\ell)$, where $\Omega(k,\ell)$ is a prespecified norm. The GSC structure is

modified to fulfil the norm constraint, as follows:

$$\tilde{\boldsymbol{G}}'(k) = \boldsymbol{G}(k,\ell) + \mu \frac{\boldsymbol{U}(k,\ell)Y^*(k,\ell)}{P_{\text{est}}(k,\ell)}\,, \tag{47.80}$$

$$\tilde{\boldsymbol{G}}(\ell+1,k) = \begin{cases} \tilde{\boldsymbol{G}}'(k) & \|\tilde{\boldsymbol{G}}'(k)\|^2 \le \Omega(k,\ell+1) \\ \frac{\sqrt{\Omega(k,\ell+1)}}{\|\tilde{\boldsymbol{G}}'(k)\|}|\tilde{\boldsymbol{G}}'(k) & \text{otherwise.} \end{cases} \tag{47.81}$$

Finally, the conventional FIR constraint is imposed on the norm-constrained filters

$$\boldsymbol{G}(\ell+1,k) \xleftarrow{\text{FIR}} \tilde{\boldsymbol{G}}(\ell+1,k)\,.$$

Based on this concept, *Hoshuyama* et al. [47.55, 69] proposed several methods for addressing the robustness issue, concentrating on the self-cancellation phenomenon, caused by the leakage of the desired speech signal to the BM outputs $\boldsymbol{U}(k,\ell)$. This phenomenon is emphasized in reverberant environments, in the case where the BM only compensates for the relative delay [as in (47.34)]. In general there are two ways to mitigate this leakage problem. First, an improved spatial filtering can be incorporated into the design of the BM. *Claesson* and *Nordholm* [47.22] proposed to apply spatial high-pass filter to cancel out all signals within a specified frequency and angular range. *Huarng* and *Yeh* [47.45] analyzed the leakage phenomenon and applied a derivative constraint on the array response, yielding wider tolerance to pointing errors.

A second cure for the leakage problems involves applying constraints on the ANC filters. *Hoshuyama* et al. [47.55] proposed several structures combining modifications for both the BM and the ANC blocks. The conventional delay-compensation BM is replaced by an adaptive BM based on signal cancellers. Two constraining strategies may be applied to the involved filters. The first strategy uses norm-constraint, and the second uses the leaky LMS adaptation scheme. *Haykin* [47.1] proved that both strategies are equivalent. The modified BM outputs, for $m = 1, \ldots, M$, are given by

$$U_m(k,\ell) = Z_m(k,\ell) - H_m^*(k,\ell)Y_{\text{FBF}}(k,\ell)\,, \tag{47.82}$$

where $H_m(k,\ell)$ are updated as to minimize the power of $U_m(k,\ell)$, by cancelling all desired speech components. Whenever, the SNR in $Y_{\text{FBF}}(k,\ell)$ is sufficiently high, the blocking ability of the structure is improved.

Gannot [47.57] showed that this equation, in conjunction with the expression for the beamformer output

$$Y(k,\ell) = Y_{\text{FBF}}(k,\ell) - \boldsymbol{G}^{\text{H}}(k,\ell)\boldsymbol{U}(k,\ell)\,,$$

is closely related to the decorrelation criterion proposed by *Weinstein* et al. [47.56]. A different decorrelation based structure was later proposed by *Fancourt* and *Parra* [47.70].

Two alternative schemes are proposed for adapting the filters $H_m(k,\ell)$. (Originally, *Hoshuyama* et al. [47.55] stated their formulation in the time domain using the original GSC structure. Here we state the frequency-domain counterpart of the proposed algorithm. The first to propose frequency domain implementation of Hoshuyama's concepts were *Herbordt* and *Kellermann* [47.71, 72].) The first scheme is the leaky LMS,

$$H_m(k,\ell+1) = (1-\delta)H_m(k,\ell) + \frac{\mu_{\text{h}}}{|Y_{\text{FBF}}(k,\ell)|^2}U_m(k,\ell)Y_{\text{FBF}}^*(k,\ell) \tag{47.83}$$

for $m = 1, \ldots, M$. The regular FIR constraint, omitted for the clarity of the exposition, is then applied. The second scheme constrains the filter coefficients to a predefined mask, yielding for $m = 1, \ldots, M$:

$$H'_m(k,\ell+1) = H_m(k,\ell) + \frac{\mu_{\text{h}}}{|Y_{\text{FBF}}(k,\ell)|^2}U_m(k,\ell)Y_{\text{FBF}}^*(k,\ell) \tag{47.84}$$

and

$$H(\ell+1,k) = \begin{cases} \phi_{\text{low}}(k,\ell+1) & H'_m(k,\ell+1) \ge \phi_{\text{low}}(k,\ell+1)\,, \\ \phi_{\text{up}}(k,\ell+1) & H'_m(k,\ell+1) \le \phi_{\text{up}}(k,\ell+1)\,, \\ H'_m(k,\ell+1) & \text{otherwise.} \end{cases} \tag{47.85}$$

The ANC filter is either adapted by the leaky LMS algorithm or the norm-constrained adaptation mechanism proposed by *Cox* (see (47.81)). As a concluding remark summarizing Hoshuyama's methods, we draw the reader attention to the resemblance of the proposed adaptation of the BM filters and the subspace tracking procedure presented by *Affes* and *Grenier* depicted in (47.46).

Spriet et al. [47.33] adopted a different approach to mitigating the leakage problem, by modifying the adaptation criterion for the ANC filters, $\boldsymbol{G}(k,\ell)$. The minimization criterion in (47.25) is altered to deal with the leakage problem. Let, $Y_{\text{FBF}}^{\text{s}}(k,\ell)$ be the speech component at the FBF output. Let $\boldsymbol{U}^{\text{s}}(k,\ell)$ and $\boldsymbol{U}^{\text{n}}(k,\ell)$ be the speech and noise (without distinction between stationary and transient noise signals) components in the

reference signals, respectively. Then, the filters $\boldsymbol{G}(k,\ell)$ minimize the following expression

$$E\{\|Y^{\mathrm{s}}_{\mathrm{FBF}}(k,\ell)-\boldsymbol{G}^{\mathrm{H}}(k,\ell)(\boldsymbol{U}^{\mathrm{s}}(k,\ell)+\boldsymbol{U}^{\mathrm{n}}(k,\ell))\|^2\}\,.$$

Since the speech and noise signals are uncorrelated, the above expression can be restated as

$$E\{\|\boldsymbol{G}^{\mathrm{H}}(k,\ell)\boldsymbol{U}^{\mathrm{n}}(k,\ell)\|^2\} + E\{\|Y^{\mathrm{s}}_{\mathrm{FBF}}(k,\ell)-\boldsymbol{G}^{\mathrm{H}}(k,\ell)\boldsymbol{U}^{\mathrm{s}}(k,\ell)\|^2\}\,. \quad (47.86)$$

Note, that the first term is related to the noise signal and the second term to the speech distortion. Hence, the Wiener filter design criterion can be easily generalized [47.26] to allow for a trade-off between speech distortion and NR, by incorporating a weighting factor $\mu\in[0,\infty)$. The resulting criterion is then given by

$$\mu E\{\|\boldsymbol{G}^{\mathrm{H}}(k,\ell)\boldsymbol{U}^{\mathrm{n}}(k,\ell)\|^2\} + E\{\|Y^{\mathrm{s}}_{\mathrm{FBF}}(k,\ell)-\boldsymbol{G}^{\mathrm{H}}(k,\ell)\boldsymbol{U}^{\mathrm{s}}(k,\ell)\|^2\}\,. \quad (47.87)$$

It is easily verified that the corresponding minimizer is

$$\boldsymbol{G}(k,\ell)=\left(\frac{1}{\mu}\boldsymbol{\Phi}_{\boldsymbol{U}^{\mathrm{s}}\boldsymbol{U}^{\mathrm{s}}}+\boldsymbol{\Phi}_{\boldsymbol{U}^{\mathrm{n}}\boldsymbol{U}^{\mathrm{n}}}\right)^{-1}\boldsymbol{\phi}_{\boldsymbol{U}Y^{\mathrm{s}}_{\mathrm{FBF}}}\,, \quad (47.88)$$

where

$$\boldsymbol{\Phi}_{\boldsymbol{U}^{\mathrm{s}}\boldsymbol{U}^{\mathrm{s}}}=E\{\boldsymbol{U}^{\mathrm{s}}(k,\ell)(\boldsymbol{U}^{\mathrm{s}}(k,\ell))^{\mathrm{H}}\}\,,$$
$$\boldsymbol{\Phi}_{\boldsymbol{U}^{\mathrm{n}}\boldsymbol{U}^{\mathrm{n}}}=E\{\boldsymbol{U}^{\mathrm{n}}(k,\ell)(\boldsymbol{U}^{\mathrm{n}}(k,\ell))^{\mathrm{H}}\}\,,$$
$$\boldsymbol{\phi}_{\boldsymbol{U}Y^{\mathrm{s}}_{\mathrm{FBF}}}=E\{\boldsymbol{U}^{\mathrm{s}}(k,\ell)(Y^{\mathrm{s}}_{\mathrm{FBF}}(k,\ell))^{*}\}\,.$$

Using the reference signals definitions we have

$$\boldsymbol{\Phi}_{\boldsymbol{U}^{\mathrm{s}}\boldsymbol{U}^{\mathrm{s}}}=\mathcal{H}^{\mathrm{H}}(k,\ell)\boldsymbol{\Phi}_{\boldsymbol{SS}}(k,\ell)\mathcal{H}(k,\ell)\,,$$
$$\boldsymbol{\Phi}_{\boldsymbol{U}^{\mathrm{n}}\boldsymbol{U}^{\mathrm{n}}}=\mathcal{H}^{\mathrm{H}}(k,\ell)\boldsymbol{\Phi}_{\boldsymbol{NN}}(k,\ell)\mathcal{H}(k,\ell)\,,$$
$$\boldsymbol{\phi}_{\boldsymbol{U}^{\mathrm{s}}Y^{\mathrm{s}}_{\mathrm{FBF}}}=\mathcal{H}^{\mathrm{H}}(k,\ell)\boldsymbol{\Phi}_{\boldsymbol{SS}}(k,\ell)\boldsymbol{W}_0(k,\ell)\,. \quad (47.89)$$

Since $\boldsymbol{\Phi}_{\boldsymbol{SS}}(k,\ell)$ is not available it can be evaluated using

$$\boldsymbol{\Phi}_{\boldsymbol{SS}}(k,\ell)=\boldsymbol{\Phi}_{\boldsymbol{ZZ}}(k,\ell)-\boldsymbol{\Phi}_{\boldsymbol{NN}}(k,\ell)$$

and $\boldsymbol{\Phi}_{\boldsymbol{NN}}(k,\ell)$ is estimated while the speech signal is absent. *Doclo* and *Moonen* [47.73] prove that the output SNR after NR with the above speech distortion weighted multichannel Wiener filter (SDW-MWF) is always larger than or equal to the input SNR, for any filter length, and for any value of the trade-off parameter μ between NR and speech distortion.

This solution for the ANC filters constitutes the speech distortion regularized generalized sidelobe canceller (SDR-GSC) structure. *Spriet* et al. [47.33] further proposed to incorporate a single-channel postfilter, which compensates for the distortion imposed by the structure in case of speech leakage into the reference signals. Further discussion of this structure is beyond the scope of this survey. In [47.74], the authors propose a stochastic gradient-based implementation of their criterion. The robustness of both the multichannel Wiener filter and the GSC structures are analyzed by *Spriet* et al. [47.75] in the context of hearing-aid application.

Improving the robustness of the BM is an ongoing research topic. An interesting direction was taken by *Low* et al. [47.76]. The authors propose to incorporate concepts adopted from the BSS discipline to improve the separation of the speech and noise signals and hence reducing the amount of leakage of the desired signal into the reference noise signals.

47.6 Multichannel Postfiltering

Postfiltering methods for multimicrophone speech enhancement algorithms have recently attracted an increased interest. It is well known that beamforming methods yield a significant improvement in speech quality [47.9]. However, when the noise field is spatially incoherent or diffuse, the NR is insufficient [47.77] and additional postfiltering is normally required [47.78]. Furthermore, as nonstationary noise cannot, in general, be distinguished from speech signals, a significant performance degradation is expected in nonstationary noise environment.

Most multimicrophone speech enhancement methods consist of a multichannel part (either delay and sum beamformer or GSC [47.36]) followed by a postfilter, which is based on Wiener filtering (sometimes in conjunction with spectral subtraction). Numerous articles have been published on the subject, e.g., [47.79–87] to mention just a few.

In general, the postfilters can be divided into two groups. The first is a single-channel postfilter which acts as a single-microphone speech enhancement algorithm on the beamformer output. Multichannel postfilters, on the other hand, explicitly use the spatial information, extracted by the GSC structure, to gain better distinction between the speech signal and the transient noise.

47.6.1 MMSE Postfiltering

Simmer et al. [47.78] address the general problem of the single-channel postfilter. They first derive the multichannel Wiener filter for estimating the speech signal, $S(k,\ell)$ from the microphone signals $\boldsymbol{Z}(k,\ell)$ given in (47.3). Then, they show that the Wiener filter can be factorized into a multiplication of the LCMV (Frost) beamformer given in (47.7) and a single-channel Wiener filter that depends on the output speech and noise signals.

To show this, we will start with the multichannel Wiener filter, which is given by

$$\boldsymbol{W}^{\text{Wiener}}(k,\ell) = \boldsymbol{\Phi}_{\boldsymbol{ZZ}}^{-1}(k,\ell)\boldsymbol{\Phi}_{\boldsymbol{Z}S}(k,\ell)\,, \tag{47.90}$$

where $\boldsymbol{\Phi}_{\boldsymbol{ZZ}}$ is given in (47.10) and

$$\begin{aligned}\boldsymbol{\Phi}_{\boldsymbol{Z}S}(k,\ell) &= E\{\boldsymbol{Z}(k,\ell)S^*(k,\ell)\}\\ &= \boldsymbol{A}(k)\phi_{\text{ss}}(k,\ell)\,.\end{aligned} \tag{47.91}$$

Hence,

$$\begin{aligned}&\boldsymbol{W}^{\text{Wiener}}(k,\ell)\\ &= \left(\phi_{\text{ss}}(k,\ell)\boldsymbol{A}(k)\boldsymbol{A}^{\text{H}}(k) + \boldsymbol{\Phi}_{\boldsymbol{NN}}(k,\ell)\right)^{-1}\\ &\quad\times\phi_{\text{ss}}(k,\ell)\boldsymbol{A}(k)\,.\end{aligned} \tag{47.92}$$

Omitting, for the clarity of the exposition, the explicit time- and frequency-domain dependence and using the matrix inversion lemma yields

$$\begin{aligned}&\boldsymbol{W}^{\text{Wiener}}(k,\ell)\\ &= \left(\boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1} - \frac{\phi_{\text{ss}}\boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1}\boldsymbol{A}\boldsymbol{A}^{\text{H}}\boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1}}{1+\phi_{\text{ss}}\boldsymbol{A}^{\text{H}}\boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1}\boldsymbol{A}}\right)\phi_{\text{ss}}\boldsymbol{A}\\ &= \left(1 - \frac{\phi_{\text{ss}}\boldsymbol{A}\boldsymbol{A}^{\text{H}}\boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1}}{1+\phi_{\text{ss}}\boldsymbol{A}^{\text{H}}\boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1}\boldsymbol{A}}\right)\boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1}\phi_{\text{ss}}\boldsymbol{A}\\ &= \left(\frac{\phi_{\text{ss}}}{1+\phi_{\text{ss}}\boldsymbol{A}^{\text{H}}\boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1}\boldsymbol{A}}\right)\boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1}\boldsymbol{A}\\ &= \left(\frac{\phi_{\text{ss}}}{\phi_{\text{ss}}+(\boldsymbol{A}^{\text{H}}\boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1}\boldsymbol{A})^{-1}}\right)\frac{\boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1}\boldsymbol{A}}{\boldsymbol{A}^{\text{H}}\boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1}\boldsymbol{A}}\,.\end{aligned} \tag{47.93}$$

(The derivation here is a slight modification of the results introduced in [47.78].) The reader can easily identify the second multiplier as the MSNR beamformer (47.8) which was shown to be equivalent to the LCMV beamformer (47.11).

We turn now to analyzing the first term in the multiplicative expression. The PSD of desired signal component at the output of the LCMV beamformer is given by

$$\begin{aligned}&\phi_{Y_{\text{s}}Y_{\text{s}}}(k,\ell)\\ &= \phi_{\text{ss}}(\boldsymbol{W}^{\text{LCMV}})^{\text{H}}\boldsymbol{A}\boldsymbol{A}^{\text{H}}\boldsymbol{W}^{\text{LCMV}}\\ &= \phi_{\text{ss}}\left(\frac{\boldsymbol{A}^{\text{H}}\boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1}\boldsymbol{A}}{\boldsymbol{A}^{\text{H}}\boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1}\boldsymbol{A}}\right)^2 = \phi_{\text{ss}}\,.\end{aligned} \tag{47.94}$$

As expected, the LCMV is a distortionless beamformer. The noise component at the beamformer output is given by,

$$\begin{aligned}&\phi_{Y_nY_n}(k,\ell)\\ &= (\boldsymbol{W}^{\text{LCMV}})^{\text{H}}\boldsymbol{\Phi}_{\boldsymbol{NN}}\boldsymbol{W}^{\text{LCMV}} = \frac{\boldsymbol{A}^{\text{H}}\boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1}\boldsymbol{A}}{\left(\boldsymbol{A}^{\text{H}}\boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1}\boldsymbol{A}\right)^2}\\ &= \frac{1}{\boldsymbol{A}^{\text{H}}\boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1}\boldsymbol{A}}\,.\end{aligned} \tag{47.95}$$

Using (47.94) and (47.95), the first term in the multichannel Wiener filter can be rewritten as

$$\begin{aligned}&\frac{\phi_{\text{ss}}}{\phi_{\text{ss}}+\left(\boldsymbol{A}^{\text{H}}\boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1}\boldsymbol{A}\right)^{-1}}\\ &= \frac{\phi_{Y_{\text{s}}Y_{\text{s}}}(k,\ell)}{\phi_{Y_nY_n}(k,\ell)+\phi_{Y_{\text{s}}Y_{\text{s}}}(k,\ell)}\,,\end{aligned} \tag{47.96}$$

which is evidently recognized as the single-channel Wiener filter applied to the LCMV beamformer output. It was therefore proven that the multichannel Wiener filter can be factorized into a product of the LCMV beamformer and a single-channel Wiener postfilter applied to the beamformer output,

$$\begin{aligned}&\boldsymbol{W}^{\text{Wiener}}(k,\ell)\\ &= \underbrace{\frac{\phi_{Y_{\text{s}}Y_{\text{s}}}(k,\ell)}{\phi_{Y_nY_n}(k,\ell)+\phi_{Y_{\text{s}}Y_{\text{s}}}(k,\ell)}}_{\text{Wiener postfilter}}\\ &\quad\times\underbrace{\frac{\boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1}(k,\ell)\boldsymbol{A}(k)}{\boldsymbol{A}^{\text{H}}(k)\boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1}(k,\ell)\boldsymbol{A}(k)}}_{\text{LCMV beamformer}}\,.\end{aligned} \tag{47.97}$$

Several algorithms have been proposed for designing the postfilter; all differ in their treatment of the single-channel Wiener postfilter estimation. *Zelinski* [47.79] was probably the first to apply a postfilter to the output of a microphone array (delay and sum beamformer in his formulation). Zelinski proposed the following Wiener

filter estimation

$$\boldsymbol{W}^{\text{Zelinski}}(k,\ell) = \frac{\frac{2}{M(M-1)}\sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\text{Re}[Z_i(k,\ell)Z_j^*(k,\ell)]}{\frac{1}{M}\sum_{i=1}^{M}|Z_i(k,\ell)|^2}\,. \tag{47.98}$$

Later, postfiltering was incorporated into the Griffiths and Jim GSC beamformer by *Bitzer* et al. [47.86, 87]. The authors proposed to use two postfilters in succession. The first is applied to the FBF branch, and the second to the GSC output. In directional noise source and in the low-frequency band of a diffused noise field, correlation between the noise components at each sensor exists. While the first postfilter is useless in this case, the latter suppresses the noise.

Simmer and *Wasiljeff* [47.88] showed that Zelinski's postfilter has two disadvantages. First, only minor SNR improvement can be expected in frequencies, for which the coherence function between the received noise signals is high (i. e., coherent noise sources). Second, for frequencies with a low coherence function, the noise PSD is overestimated by a factor M (the number of microphones). They propose to mitigate the second disadvantage by slightly modifying the noise estimation, to obtain an estimate of the noise PSD at the output of the beamformer, rather than at its input.

In diffused noise field the noise coherence function tends to be high in lower frequencies, whereas in higher-frequency bands it tends to be low. The cutoff frequency depends on the distance between microphones (see further discussion in Sect. 47.7). This property is the cause for the first drawback of Zelinski's postfilter. It was therefore proposed by *Fischer* and *Kammeyer* [47.83] to split the beamformer into three nonoverlapping subarrays (with different intermicrophone distances) for which the noise coherence function is kept low. To avoid grating lobes, bandpass filters with corresponding cutoff frequencies, are applied to the beamformer output. *Marro* et al. [47.43] improved this concept and further modified the Wiener postfilter estimation. A comprehensive survey of these postfiltering methods can be found in [47.78]. *McCowan* and *Bourlard* [47.89, 90] develop a more-general expression of the postfilter estimation, based on an assumed knowledge of the complex coherence function of the noise field. This general expression can be used to construct a more-appropriate postfilter in a variety of different noise fields.

47.6.2 Log-Spectral Amplitude Postfiltering

A major drawback of single-channel postfiltering techniques is that highly nonstationary noise components are not addressed. The time variation of the interfering signals is assumed to be sufficiently slow, such that the postfilter can track and adapt to the changes in the noise statistics. Unfortunately, transient interferences are often much too brief and abrupt for the conventional tracking methods.

Transient Beam-to-Reference Ratio

Generally, the TF-GSC output comprises three components: a nonstationary desired source component, a pseudostationary noise component, and a transient interference. Our objective is to determine which category a given time–frequency bin belongs to, based on the beamformer output and the reference signals.

Recall the three hypotheses $H_{0\text{s}}$, $H_{0\text{t}}$, and H_1 that indicate, respectively, the absence of transients, the presence of an interfering transient, and the presence of a desired source transient at the beamformer output (the pseudostationary interference is present in any case). Then, if transients have not been detected at the beamformer output and the reference signals, we can accept the $H_{0\text{s}}$ hypothesis. If a transient is detected at the beamformer output but not at the reference signals, the transient is likely a source component and therefore we determine that H_1 is true. On the contrary, a transient that is detected at one of the reference signals but not at the beamformer output is likely an interfering component, which implies that $H_{0\text{t}}$ is true. If a transient is simultaneously detected at the beamformer output and at one of the reference signals, a further test is required, which involves the ratio between the transient power at beamformer output and the transient power at the reference signals. The discussion here is partly based on [47.77]. A real-time version of the method that incorporates adaptive estimation of the ATFs is introduced in [47.91].

Let $\mathcal{S}$ be a smoothing operator in the power-spectral domain,

$$\mathcal{S}Y(k,\ell) = \alpha_{\text{s}}\cdot\mathcal{S}Y(k,\ell-1) + (1-\alpha_{\text{s}})\sum_{i=-w}^{w} b_i|Y(k-i,\ell)|^2\,, \tag{47.99}$$

where α_{s} $(0 \le \alpha_{\text{s}} \le 1)$ is a forgetting factor for the smoothing in time, and b is a normalized window function $(\sum_{i=-w}^{w} b_i = 1)$ that determines the order of

smoothing in frequency. Let $\mathcal{M}$ denote an estimator for the PSD of the background pseudostationary noise, derived using the MCRA approach [47.66,67]. The decision rules for detecting transients at the TF-GSC output and reference signals are

$$\Lambda_Y(k,\ell) \triangleq \mathcal{S}Y(k,\ell)/\,\mathcal{M}Y(k,\ell) > \Lambda_0\,, \tag{47.100}$$

$$\Lambda_{\boldsymbol{U}}(k,\ell) \triangleq \max_{2\le i\le M}\left\{\frac{\mathcal{S}U_i(k,\ell)}{\mathcal{M}U_i(k,\ell)}\right\} > \Lambda_1\,, \tag{47.101}$$

respectively, where Λ_Y and Λ_U denote measures of the local nonstationarities, and Λ_0 and Λ_1 are the corresponding threshold values for detecting transients [47.92]. The transient beam-to-reference ratio (TBRR) is defined by the ratio between the transient power of the beamformer output and the transient power of the strongest reference signal

$$\Omega(k,\ell) = \frac{\mathcal{S}Y(k,\ell) - \mathcal{M}Y(k,\ell)}{\max\limits_{2\le i\le M}\{\mathcal{S}U_i(k,\ell) - \mathcal{M}U_i(k,\ell)\}}\,. \tag{47.102}$$

Transient signal components are relatively strong at the beamformer output, whereas transient noise components are relatively strong at one of the reference signals. Hence, we expect $\Omega(k,\ell)$ to be large for signal transients, and small for noise transients. Assuming there exist thresholds $\Omega_{\text{high}}(k)$ and $\Omega_{\text{low}}(k)$ such that

$$\Omega(k,\ell)|_{H_{0t}} \le \Omega_{\text{low}}(k) \le \Omega_{\text{high}}(k) \le \Omega(k,\ell)|_{H_1} \tag{47.103}$$

the decision rule for differentiating desired signal components from the transient interference components is

$$\begin{aligned} H_{0t} &: \gamma_{\text{s}}(k,\ell) \le 1 \text{ or } \Omega(k,\ell) \le \Omega_{\text{low}}(k)\,,\\ H_1 &: \gamma_{\text{s}}(k,\ell) \ge \gamma_0 \text{ and } \Omega(k,\ell) \ge \Omega_{\text{high}}(k)\,,\\ H_{\text{r}} &: \text{otherwise}, \end{aligned} \tag{47.104}$$

where

$$\gamma_{\text{s}}(k,\ell) \triangleq \frac{|Y(k,\ell)|^2}{\mathcal{M}Y(k,\ell)} \tag{47.105}$$

represents the a posteriori SNR at the beamformer output with respect to the pseudostationary noise, γ_0 denotes a constant satisfying $\mathcal{P}\,(\gamma_{\text{s}}(k,\ell) \ge \gamma_0 \mid H_{0\text{s}}) < \epsilon$ for a certain significance level ϵ, and H_{r} designates a *reject* option where the conditional error of making a decision between $H_{0\text{t}}$ and H_1 is high.

Figure 47.5 summarizes a block diagram for the hypothesis testing. The hypothesis testing is carried out in the time–frequency plane for each frame and frequency bin. $H_{0\text{s}}$ is accepted when transients have neither been detected at the beamformer output nor at the reference signals. If a transient is detected at the beamformer output but not at the reference signals, we accept H_1. On the other hand, if a transient is detected at one of the reference signals but not at the beamformer output, we accept $H_{0\text{t}}$. If a transient is detected simultaneously at the beamformer output and at one of the reference signals, we compute the TBRR $\Omega(k,\ell)$ and the a posteriori SNR at the beamformer output with respect to the pseudostationary noise $\gamma_{\text{s}}(k,\ell)$ and decide on the hypothesis according to (47.104).

Log-Spectral Amplitude Estimation

We address now the problem of estimating the time-varying PSD of the TF-GSC output noise component, and present the multichannel postfiltering technique. Figure 47.6 describes a block diagram of the multichannel postfiltering. Following the hypothesis testing, an estimate $\hat{q}(k,\ell)$ for the a priori signal absence probability is produced. Subsequently, we derive an estimate $p(k,\ell) \triangleq \mathcal{P}(H_1 \mid Y, \boldsymbol{U})$ for the signal presence probability, and an estimate $\hat{\lambda}_{\text{d}}(k,\ell)$ for the noise PSD.

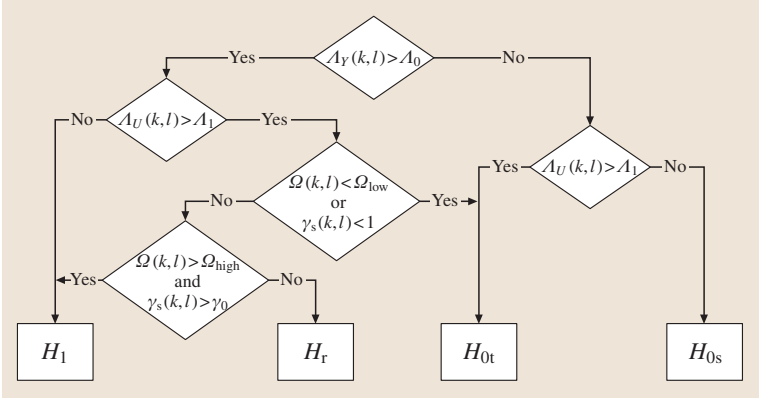

Fig. 47.5 Block diagram for hypothesis testing

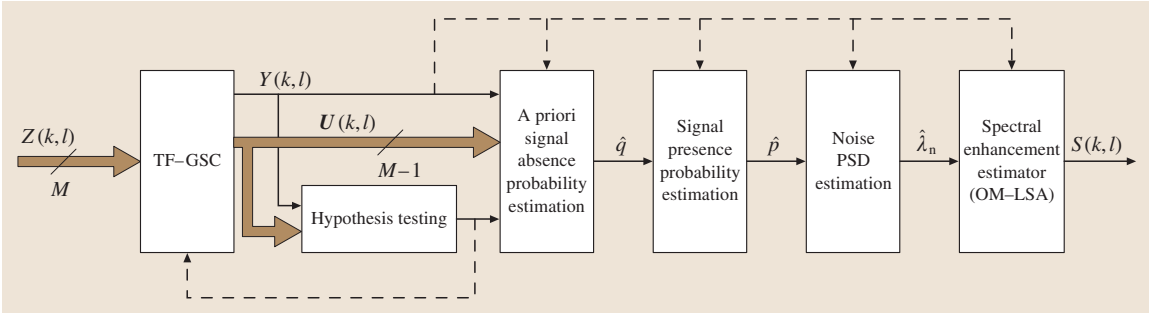

Fig. 47.6 Block diagram of multichannel postfiltering

Finally, spectral enhancement of the beamformer output is achieved by applying the optimally-modified log-spectral amplitude (OM-LSA) gain function [47.93], which minimizes the MSE of the log-spectral amplitude under signal presence uncertainty.

Based on a Gaussian statistical model [47.94], the signal presence probability is given by

$$p(k,\ell) = \left\{1+\frac{q(k,\ell)}{1-q(k,\ell)}[1+\xi(k,\ell)]\exp[-\upsilon(k,\ell)]\right\}^{-1}, \quad (47.106)$$

where $\xi(k,\ell) \triangleq \lambda_s(k,\ell)/\lambda_n(k,\ell)$ is the a priori SNR, $\lambda_s(k,\ell)$ is the desired signal PSD at the beamformer output $\lambda_n(k,\ell)$ is the noise PSD at the beamformer output, $\upsilon(k,\ell) \triangleq \gamma(k,\ell)\,\xi(k,\ell)/[1+\xi(k,\ell)]$, and $\gamma(k,\ell) \triangleq |Y(k,\ell)|^2/\lambda_n(k,\ell)$ is the a posteriori SNR. The a priori signal absence probability $\hat{q}(k,\ell)$ is set to 1 if the signal absence hypotheses (H_{0s} or H_{0t}) are accepted, and is set to 0 if the signal presence hypothesis (H_1) is accepted. In the case of the reject hypothesis H_r, a soft signal detection is accomplished by letting $\hat{q}(k,\ell)$ be inversely proportional to $\Omega(k,\ell)$ and $\gamma_s(k,\ell)$:

$$\hat{q}(k,\ell) = \max\left\{\frac{\gamma_0-\gamma_s(k,\ell)}{\gamma_0-1}, \frac{\Omega_{high}-\Omega(k,\ell)}{\Omega_{high}-\Omega_{low}}\right\}. \quad (47.107)$$

The a priori SNR is estimated by [47.93]

$$\hat{\xi}(k,\ell) = \alpha K_{H_1}^2(k,\ell-1)\gamma(k,\ell-1) + (1-\alpha)\max\{\gamma(k,\ell)-1, 0\}, \quad (47.108)$$

where α is a weighting factor that controls the trade-off between NR and signal distortion, and

$$K_{H_1}(k,\ell) \triangleq \frac{\xi(k,\ell)}{1+\xi(k,\ell)}\exp\left(\frac{1}{2}\int_{\upsilon(k,\ell)}^{\infty}\frac{e^{-t}}{t}\,dt\right) \quad (47.109)$$

is the spectral gain function of the log-spectral amplitude (LSA) estimator when signal is surely present [47.95]. An estimate for noise PSD is obtained by recursively averaging past spectral power values of the noisy measurement, using a time-varying frequency-dependent smoothing parameter. The recursive averaging is given by

$$\hat{\lambda}_n(k,\ell+1) = \tilde{\alpha}_n(k,\ell)\hat{\lambda}_n(k,\ell)+\beta[1-\tilde{\alpha}_n(k,\ell)]|Y(k,\ell)|^2, \quad (47.110)$$

where the smoothing parameter $\tilde{\alpha}_n(k,\ell)$ is determined by the signal presence probability $p(k,\ell)$,

$$\tilde{\alpha}_n(k,\ell) \triangleq \alpha_n+(1-\alpha_n)\,p(k,\ell), \quad (47.111)$$

and β is a factor that compensates the bias when signal is absent. The constant α_n ($0<\alpha_n<1$) represents the minimal smoothing parameter value. The smoothing parameter is close to 1 when signal is present, to prevent an increase in the noise estimate as a result of signal components. It decreases when the probability of signal presence decreases, to allow a fast update of the noise estimate.

Table 47.3 Values of the parameters used in the implementation of the log-spectral amplitude postfiltering for a sampling rate of 8 kHz

Normalized LMS:	$\alpha_p = 0.9$	$\mu_h = 0.05$
ATF identification:	$N = 10$	$R = 10$
Hypothesis testing:	$\alpha_s = 0.9$	$\gamma_0 = 4.6$
	$\Lambda_0 = 1.67$	$\Lambda_1 = 1.81$
	$\Omega_{low} = 1$	$\Omega_{high} = 3$
	$b = (0.25\ \ 0.5\ \ 0.25)$	
Noise PSD estimation:	$\alpha_n = 0.85$	$\beta = 1.47$
Spectral enhancement:	$\alpha = 0.92$	
	$K_{min} = -20\,\text{dB}$	

The estimate of the clean signal STFT is finally given by

$$\hat{S}(k,\ell) = K(k,\ell)Y(k,\ell)\,, \tag{47.112}$$

where

$$K(k,\ell) = \left\{K_{H_1}(k,\ell)\right\}^{p(k,\ell)} K_{\min}^{1-p(k,\ell)} \tag{47.113}$$

is the OM-LSA gain function and $K_{\min}$ denotes a lower bound constraint for the gain when signal is absent. The implementation of the integrated TF-GSC and multichannel postfiltering algorithm is summarized in Fig. 47.6.

Typical values of the respective parameters, for a sampling rate of 8 kHz, are given in Table 47.3. The STFT and its inverse are implemented with biorthogonal Hamming windows of 256 samples length (32 ms) and 64 samples frame update step (75% overlapping windows).

47.7 Performance Analysis

The use of actual signals (such as noisy speech recordings in room environment) demonstrates the ability of the TF-GSC algorithm to reduce the noise while maintaining the desired signal spectral content (Sect. 47.8). However, it is also beneficial to perform analytical evaluation of the expected performance, especially for determining the performance limits. While the D-GSC [47.36] is widely analyzed in the literature, the more-realistic arbitrary ATFs scenario, is only superficially treated. In more-complex environments such as reverberating room this assumption is not valid, and may result in severe degradation in the performance. Furthermore, most references address the NR obtained by the algorithm but do not present any measure of distortion imposed on the desired signal, even in the simple delay-only ATFs. In this section we will analyze both the NR of the TF-GSC structure (while using the RTFs rather than the ATFs themselves) and the distortion it imposes. For a thorough performance evaluation of the multichannel postfilter (for the two-channel case) please refer to [47.96].

47.7.1 The Power Spectral Density of the Beamformer Output

Using (47.24) and (47.37), the algorithm's output is given by

$$Y(k,\ell) = \frac{\hat{\tilde{\boldsymbol{A}}}^{\mathrm{H}}(k)}{\|\hat{\tilde{\boldsymbol{A}}}(k)\|^2}\boldsymbol{Z}(k,\ell) - \boldsymbol{G}^{\mathrm{H}}(k,\ell)\widehat{\mathcal{H}}^{\mathrm{H}}(k)\boldsymbol{Z}(k,\ell)\,,$$

where only estimates of the RTFs, $\hat{\tilde{\boldsymbol{A}}}(k)$ [and $\widehat{\mathcal{H}}(k)$], rather than their exact values, are assumed to be known. Using this expression, the PSD of the output signal is given by

$$\begin{aligned}\boldsymbol{\Phi}_{\mathrm{yy}}(k,\ell) &= E[Y(k,\ell)Y^*(k,\ell)] \\ &= E\left\{\left[\frac{1}{\|\hat{\tilde{\boldsymbol{A}}}(k)\|^2}\hat{\tilde{\boldsymbol{A}}}^{\mathrm{H}}(k)\boldsymbol{Z}(k,\ell)\right.\right. \\ &\quad \left.- \boldsymbol{G}^{\mathrm{H}}(k,\ell)\widehat{\mathcal{H}}^{\mathrm{H}}(k)\boldsymbol{Z}(k,\ell),\right] \\ &\quad \times\left[\frac{1}{\|\hat{\tilde{\boldsymbol{A}}}(k)\|^2}\hat{\tilde{\boldsymbol{A}}}^{\mathrm{H}}(k)\boldsymbol{Z}(k,\ell)\right. \\ &\quad \left.\left.- \boldsymbol{G}^{\mathrm{H}}(k,\ell)\widehat{\mathcal{H}}^{\mathrm{H}}(k)\boldsymbol{Z}(k,\ell),\right]^{\mathrm{H}}\right\}.\end{aligned} \tag{47.114}$$

Opening brackets and using the PSD definition $\boldsymbol{\Phi}_{\boldsymbol{ZZ}}(k,\ell) = E\{\boldsymbol{Z}(k,\ell)\boldsymbol{Z}^{\mathrm{H}}(k,\ell)\}$ yields,

$$\begin{aligned}\boldsymbol{\Phi}_{\mathrm{yy}}(k,\ell) &= \frac{1}{\|\hat{\tilde{\boldsymbol{A}}}(k)\|^4}\hat{\tilde{\boldsymbol{A}}}^{\mathrm{H}}(k)\boldsymbol{\Phi}_{\boldsymbol{ZZ}}(k,\ell)\hat{\tilde{\boldsymbol{A}}}(k) \\ &\quad - \frac{1}{\|\hat{\tilde{\boldsymbol{A}}}(k)\|^2}\boldsymbol{G}^{\mathrm{H}}(k,\ell)\widehat{\mathcal{H}}^{\mathrm{H}}(k) \\ &\quad \times\boldsymbol{\Phi}_{\boldsymbol{ZZ}}(k,\ell)\hat{\tilde{\boldsymbol{A}}}(k) \\ &\quad - \frac{1}{\|\hat{\tilde{\boldsymbol{A}}}(k)\|^2}\hat{\tilde{\boldsymbol{A}}}^{\mathrm{H}}(k)\boldsymbol{\Phi}_{\boldsymbol{ZZ}}(k,\ell)\widehat{\mathcal{H}}(k)\boldsymbol{G}(k) \\ &\quad + \boldsymbol{G}^{\mathrm{H}}(k,\ell)\widehat{\mathcal{H}}^{\mathrm{H}}(k)\boldsymbol{\Phi}_{\boldsymbol{ZZ}}(k,\ell)\widehat{\mathcal{H}}(k)\boldsymbol{G}(k)\,.\end{aligned} \tag{47.115}$$

The output PSD depends on the input signal $\boldsymbol{Z}(k,\ell)$ and the optimal multichannel Wiener filter, given by (47.26), calculated during frames for which hypothesis $H_{0\mathrm{s}}$ is true. Using the independence of the desired signal and the noise signal, the NR and the distortion imposed by the algorithm can be calculated separately by deriving

expressions for the output PSD in the following two situations:

$$\boldsymbol{\Phi}_{\boldsymbol{ZZ}}(k,\ell)=\begin{cases}\boldsymbol{\Phi}_{\boldsymbol{N}_{\mathrm{s}}\boldsymbol{N}_{\mathrm{s}}}(k,\ell) & H_{0\mathrm{s}}\ ,\\ \phi_{\mathrm{ss}}(k,\ell)\boldsymbol{A}(k)\boldsymbol{A}^{\mathrm{H}}(k) & H_1\ ,\end{cases}$$

yielding

$$\boldsymbol{\Phi}_{\mathrm{yy}}(k,\ell)=\begin{cases}\boldsymbol{\Phi}^{\mathrm{n}}_{\mathrm{yy}}(k,\ell) & H_{0\mathrm{s}} \quad \Rightarrow \text{noise reduction},\\ \boldsymbol{\Phi}^{\mathrm{s}}_{\mathrm{yy}}(k,\ell) & H_1 \quad \Rightarrow \text{distortion}\ .\end{cases}$$

Note that for simplicity we calculate the NR only during $H_{0\mathrm{s}}$, i. e., while only stationary noise signal is present. For a performance evaluation in the nonstationary case, please refer to [47.96].

Using (47.26) and (47.115), we obtain the output signal PSD:

$$\begin{aligned}\boldsymbol{\Phi}_{\mathrm{yy}}(k,\ell)=\frac{1}{\|\hat{\tilde{\boldsymbol{A}}}(k)\|^4}\times\big\{&\hat{\tilde{\boldsymbol{A}}}^{\mathrm{H}}(k)\boldsymbol{\Phi}_{\boldsymbol{ZZ}}(k,\ell)\hat{\tilde{\boldsymbol{A}}}(k)\\ &-\hat{\tilde{\boldsymbol{A}}}^{\mathrm{H}}(k)\boldsymbol{\Phi}_{\boldsymbol{N}_{\mathrm{s}}\boldsymbol{N}_{\mathrm{s}}}(k,\ell)\widehat{\mathcal{H}}(k)\\ &\quad\big[\widehat{\mathcal{H}}^{\mathrm{H}}(k)\boldsymbol{\Phi}_{\boldsymbol{N}_{\mathrm{s}}\boldsymbol{N}_{\mathrm{s}}}(k,\ell)\widehat{\mathcal{H}}(k)\big]^{-1}\\ &\quad\widehat{\mathcal{H}}^{\mathrm{H}}(k)\boldsymbol{\Phi}_{\boldsymbol{ZZ}}(k,\ell)\hat{\tilde{\boldsymbol{A}}}(k)\\ &-\hat{\tilde{\boldsymbol{A}}}^{\mathrm{H}}(k)\boldsymbol{\Phi}_{\boldsymbol{ZZ}}(k,\ell)\widehat{\mathcal{H}}(k)\\ &\quad\big[\widehat{\mathcal{H}}^{\mathrm{H}}(k)\boldsymbol{\Phi}_{\boldsymbol{N}_{\mathrm{s}}\boldsymbol{N}_{\mathrm{s}}}(k,\ell)\widehat{\mathcal{H}}(k)\big]^{-1}\\ &\quad\widehat{\mathcal{H}}^{\mathrm{H}}(k)\boldsymbol{\Phi}_{\boldsymbol{N}_{\mathrm{s}}\boldsymbol{N}_{\mathrm{s}}}(k,\ell)\hat{\tilde{\boldsymbol{A}}}(k)\\ &+\hat{\tilde{\boldsymbol{A}}}^{\mathrm{H}}(k)\boldsymbol{\Phi}_{\boldsymbol{N}_{\mathrm{s}}\boldsymbol{N}_{\mathrm{s}}}(k,\ell)\widehat{\mathcal{H}}(k)\\ &\quad\big[\widehat{\mathcal{H}}^{\mathrm{H}}(k)\boldsymbol{\Phi}_{\boldsymbol{N}_{\mathrm{s}}\boldsymbol{N}_{\mathrm{s}}}(k,\ell)\widehat{\mathcal{H}}(k)\big]^{-1}\\ &\quad\widehat{\mathcal{H}}^{\mathrm{H}}(k)\boldsymbol{\Phi}_{\boldsymbol{ZZ}}(k,\ell)\widehat{\mathcal{H}}(k)\\ &\times\big[\widehat{\mathcal{H}}^{\mathrm{H}}(k)\boldsymbol{\Phi}_{\boldsymbol{N}_{\mathrm{s}}\boldsymbol{N}_{\mathrm{s}}}(k,\ell)\widehat{\mathcal{H}}(k)\big]^{-1}\\ &\quad\widehat{\mathcal{H}}^{\mathrm{H}}(k)\boldsymbol{\Phi}_{\boldsymbol{N}_{\mathrm{s}}\boldsymbol{N}_{\mathrm{s}}}(k,\ell)\hat{\tilde{\boldsymbol{A}}}(k)\big\}\ .\end{aligned}\tag{47.116}$$

This complicated expression depends on various parameters: the input signal PSD, $\boldsymbol{\Phi}_{\boldsymbol{ZZ}}(k,\ell)$, the stationary noise PSD used for calculating the optimal filters, $\boldsymbol{\Phi}_{\boldsymbol{N}_{\mathrm{s}}\boldsymbol{N}_{\mathrm{s}}}(k,\ell)$, and the RTFs estimate $\hat{\tilde{\boldsymbol{A}}}(k)$ [which is also used for the BM $\widehat{\mathcal{H}}(k)$]. This expression will be used in Subsects. 47.7.2 and 47.7.3 for deriving general expressions for the distortion imposed by the algorithm and the obtainable NR, respectively. These general expressions will be evaluated for several interesting cases.

47.7.2 Signal Distortion

The distortion imposed by the algorithm can be calculated by the general expression given in (47.116) for a signal $\boldsymbol{Z}(k,\ell)=\boldsymbol{A}(k)s(k,\ell)$. Assume a perfect estimate of the RTFs $\tilde{\boldsymbol{A}}(k)$, is available, i. e., $\hat{\tilde{\boldsymbol{A}}}(k)=\tilde{\boldsymbol{A}}(k)=\frac{\boldsymbol{A}(k)}{A_1(k)}$. Thus, using the signal PSD expression $\boldsymbol{\Phi}_{\boldsymbol{ZZ}}(k,\ell)=\phi_{\mathrm{ss}}(k,\ell)\boldsymbol{A}(k)\boldsymbol{A}^{\mathrm{H}}(k)$ and the identity $\mathcal{H}^{\mathrm{H}}(k)\boldsymbol{A}(k)=0$, expression (47.116) reduces to

$$\begin{aligned}&\boldsymbol{\Phi}^{\mathrm{s}}_{\mathrm{yy}}(k,\ell)\\ &=\phi_{\mathrm{ss}}(k,\ell)\frac{|\mathcal{F}(k)|^2}{\|\tilde{\boldsymbol{A}}(k)\|^4}\tilde{\boldsymbol{A}}^{\mathrm{H}}(k)\boldsymbol{A}(k)\boldsymbol{A}^{\mathrm{H}}(k)\tilde{\boldsymbol{A}}(k)\\ &\quad\phi_{\mathrm{ss}}(k,\ell)|A_1(k)|^2\ .\end{aligned}\tag{47.117}$$

The filter $A_1(k)$ is the ATF relating the source signal and the first (arbitrarily chosen as the reference) sensor. This distortion cannot be eliminated by the algorithm. Note that this distortion is due to the use of the RTFs, rather than the ATFs themselves. Using direct estimate of the ATFs will avoid this distortion affect. Actually, it imposes on the output signal the same amount of distortion imposed on the arbitrary reference sensor. Hence, we define the total distortion caused by the algorithm by normalizing the output,

$$\mathrm{DIS}(k,\ell)=\frac{\boldsymbol{\Phi}^{\mathrm{s}}_{\mathrm{yy}}(k,\ell)}{|A_1(k)|^2\phi_{\mathrm{ss}}(k,\ell)}\ .\tag{47.118}$$

Hence, a value of $\mathrm{DIS}(k,\ell)=1$ indicates a distortionless output. This value is obtained whenever an exact knowledge of the RTFs is available.

The distortion level demonstrates only weak dependence on the noise field, both for the delay-only and complex ATF cases. For details please refer to [47.77].

47.7.3 Stationary Noise Reduction

We calculate now the amount of obtainable stationary NR. When the noise is nonstationary, the use of a postfilter becomes more important. A performance analysis of the multichannel postfilter (for the two-channel case) in nonstationary noise environment can be found in [47.96].

We will use again the general expression for the output signal given by (47.116), this time with a noise signal as the input signal, i. e., $\boldsymbol{Z}(k,\ell)=\boldsymbol{N}_{\mathrm{s}}(k,\ell)$ [the same noise signal used for calculating the optimal Wiener filter (47.26)]. The expression (47.116) now reduces to

$$\begin{aligned}
&\Phi_{yy}^{n}(k,\ell)\\
&= \frac{\hat{\tilde{A}}^{H}(k)\Phi_{N_sN_s}(k,\ell)\hat{\tilde{A}}(k)}{\|\hat{\tilde{A}}(k)\|^4} - \frac{\hat{\tilde{A}}^{H}(k)}{\|\hat{\tilde{A}}(k)\|^4}\\
&\quad\times\Phi_{N_sN_s}(k,\ell)\widehat{\mathcal{H}}(k)\left[\widehat{\mathcal{H}}^{H}(k)\Phi_{N_sN_s}(k,\ell)\right]^{-1}\\
&\quad\times\widehat{\mathcal{H}}^{H}(k)\Phi_{N_sN_s}(k,\ell)\hat{\tilde{A}}(k)\,. \qquad (47.119)
\end{aligned}$$

The first term of the equation can be identified as $\Phi_{FBF}^{n}(k,\ell)$, the FBF component of the output PSD when a noise-only signal is applied to the array,

$$\begin{aligned}
\Phi_{FBF}^{n}(k,\ell) &= E\left\{Y_{FBF}^{n}(k,\ell)\left[Y_{FBF}^{n}(k,\ell)\right]^{*}\right\}\\
&= \frac{\hat{\tilde{A}}^{H}(k)\Phi_{N_sN_s}(k,\ell)\hat{\tilde{A}}(k)}{\|\hat{\tilde{A}}(k)\|^4}\,. \qquad (47.120)
\end{aligned}$$

Another interesting figure of merit is the extra NR obtained by the noise cancelling branch (see also [47.40]),

$$\mathrm{NR}_{ANC}(k,\ell) = \frac{\Phi_{FBF}^{n}(k,\ell)}{\Phi_{yy}^{n}(k,\ell)}\,. \qquad (47.121)$$

Expressions (47.119), (47.120), and (47.121) can be used for calculating the NR obtained by the algorithm and to determine the major contributor for this NR. Assuming small errors regime, the error in estimating $\tilde{A}(k)$ has only minor influence on the NR. Therefore, we will assume, throughout the NR analysis, perfect knowledge of the RTFs, i. e., $\hat{\tilde{A}}(k) = \tilde{A}(k)$.

The resulting expressions for the noise cancellation depends on the noise PSD at the sensors. We calculate now the expected NR of the algorithm for three important noise fields: coherent (point source), diffused (spherically isotropic), and incoherent (noise signals generated at the sensors; e.g., amplifier noise, are assumed to be uncorrelated).

Coherent Noise Field

Assume a single stationary point source noise signal with PSD $\Phi_{n_sn_s}(k,\ell)$ and assume that $b_m(t)$ are slowly time-varying ATFs relating the noise source and the m-th sensor. Define,

$$\boldsymbol{N}_s(k,\ell) = \boldsymbol{B}(k)N_s(k,\ell)\,,$$

where

$$\boldsymbol{B}^{T}(k) = (B_1(k)\;\; B_2(k)\;\; \cdots\;\; B_M(k))\,.$$

The PSD matrix of the noise component at the sensors' signals is given by,

$$\Phi_{N_sN_s}(k,\ell) = \Phi_{n_sn_s}(k,\ell)\boldsymbol{B}(k)\boldsymbol{B}^{H}(k) + \varepsilon\mathbf{I}\,,$$

where $\mathbf{I}$ is an $M\times M$ identity matrix, and $\varepsilon\to 0$. The last term is added for stability reasons (see Appendix 47.A). For $\boldsymbol{B}(k)\neq\boldsymbol{A}(k)$, the achievable NR is infinite, i. e.,

$$\Phi_{yy}^{n}(k,\ell) = 0 \quad \text{for} \quad \boldsymbol{B}(k)\neq\boldsymbol{A}(k)\,.$$

Thus, perfect noise cancellation is achieved. The derivation of this result is given in Appendix 47.A. Note that this is not a surprising result, since for $M\geq 2$ the Wiener filter can entirely eliminate the noise component. This result is valid for all ATFs $\boldsymbol{B}(k)$ provided that $\boldsymbol{B}(k)\neq\boldsymbol{A}(k)$, i. e., the noise and the signal do not originate from the same point. If $\boldsymbol{B}(k) = \boldsymbol{A}(k)$ the noise and desired signal are indistinguishable and no NR is expected. If $\boldsymbol{B}(k)\neq\boldsymbol{A}(k)$ the proposed algorithm can eliminate any point source noise signal as good as the D-GSC can eliminate a directional noise signal in the delay-only propagation case (see [47.40]).

It is also interesting to explore the contribution of the FBF block to the NR,

$$\begin{aligned}
&\Phi_{FBF}^{n}(k,\ell)\\
&= \frac{\tilde{A}^{H}(k)\Phi_{N_sN_s}(k,\ell)\tilde{A}(k)}{\|\tilde{A}(k)\|^4}\\
&= \frac{\tilde{A}^{H}(k)\Phi_{n_sn_s}(k,\ell)\boldsymbol{B}(k)\boldsymbol{B}^{H}(k)\tilde{A}(k)}{\|\tilde{A}(k)\|^4}\\
&= \frac{\Phi_{n_sn_s}(k,\ell)}{\|\tilde{A}(k)\|^4}\tilde{A}^{H}(k)\boldsymbol{B}(k)\left[\tilde{A}^{H}(k)\boldsymbol{B}(k)\right]^{H}\,.
\end{aligned}$$

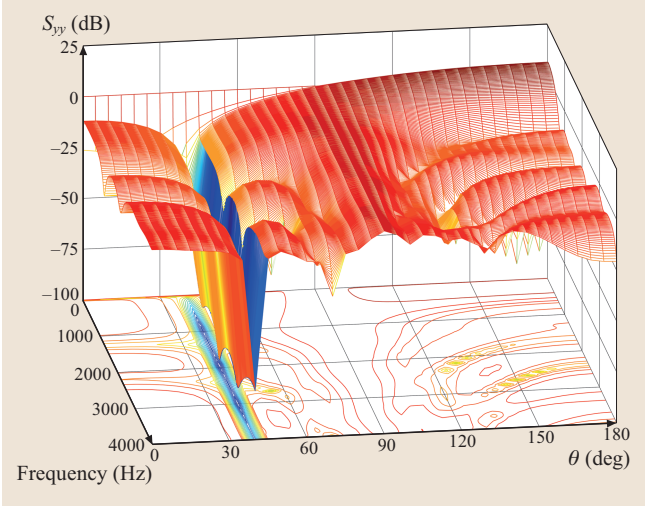

Fig. 47.7 Array output for directional noise field for linear array with $M = 5$ sensors for delay-only ATFs

Although the FBF combines the desired signal coherently, it does not necessarily improves the SNR. The infinite NR is due to the ANC branch.

The expected NR in the simple case of delay-only ATFs is shown in Fig. 47.7. We optimize a linear array of $M = 5$ microphones to cancel noise impinging on the array from the direction $\theta = 40°$. We present the output signal of the array as a function of the frequency and the array steering angle θ. It is clearly shown that the main lobe is maintained (i. e., low distortion), while a null is constructed at all frequencies at the noise angle. The main lobe is wider in the lower-frequency band. This result is in good agreement with the theory, since at $\omega = 0\,\text{rad/s}$ there is no phase difference between the signals at the sensors. Similar results were obtained by *Bitzer* et al. [47.40,41]. The general ATFs case is further explored in [47.77, 97].

Diffused Noise Field

In highly reverberant acoustical environment, such as a car enclosure, the noise field tends to be diffused (see for instance [47.42, 98]). A diffused noise source is assumed to be equidistributed on a sphere in the far field of the array. The cross-coherence function between signals received by two sensors (i, j) with distance d_{ij} is given by

$$\Gamma_{N_i N_j}(k) = \frac{\Phi_{N_i N_j}(k)}{\sqrt{\Phi_{N_i N_i}(k)\Phi_{N_j N_j}(k)}} = \frac{\sin(kd_{ij}/c)}{kd_{ij}/c} ,$$

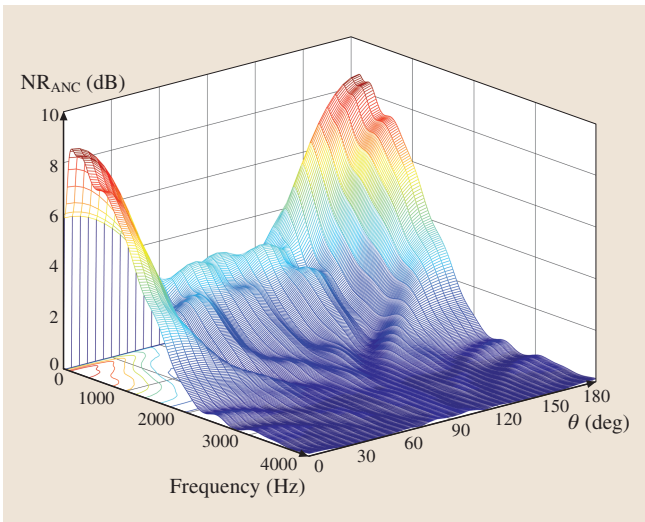

Fig. 47.8 Extra noise reduction of noise cancelling branch for diffused noise field using a linear array with $M = 5$ sensors and delay-only ATFs

where c is the speed of sound [47.98]. Thus, the coherence matrix is given by,

$$\Gamma(k) = \begin{pmatrix} 1 & \Gamma_{N_1 N_2}(k) & \cdots & \Gamma_{N_1 N_M}(k) \\ \Gamma_{N_2 N_1}(k) & 1 & \cdots & \Gamma_{N_2 N_M}(k) \\ \vdots & & \ddots & \vdots \\ \Gamma_{N_M N_1}(k) & & & 1 \end{pmatrix} .$$

The noise PSD at the sensors input is

$$\boldsymbol{\Phi}_{\boldsymbol{N}_\mathrm{s}\boldsymbol{N}_\mathrm{s}}(k, \ell) = \boldsymbol{\Phi}_{n_\mathrm{s} n_\mathrm{s}}(k, \ell)\Gamma(k) .$$

Using (47.120), the noise PSD at the FBF output is given by

$$\boldsymbol{\Phi}^\mathrm{n}_\mathrm{FBF}(k, \ell) = \frac{\phi_\mathrm{nn}(k, \ell)\tilde{\boldsymbol{A}}^\mathrm{H}(k)\Gamma(k)\tilde{\boldsymbol{A}}(k)}{\|\tilde{\boldsymbol{A}}(k)\|^4} .$$

The extra NR obtained by the ANC is given by

$$\begin{aligned} &\mathrm{NR}_\mathrm{ANC}(k, \ell) \\ &= \left\{1 - \left[\tilde{\boldsymbol{A}}^\mathrm{H}(k)\Gamma(k)\mathcal{H}(k)\left[\mathcal{H}^\mathrm{H}(k)\Gamma(k)\mathcal{H}(k)\right]^{-1}\right.\right. \\ &\quad \left.\left. \times \mathcal{H}^\mathrm{H}(k)\Gamma(k)\tilde{\boldsymbol{A}}(k)[\tilde{\boldsymbol{A}}^\mathrm{H}(k)\Gamma(k)\tilde{\boldsymbol{A}}(k)]^{-1}\right]\right\}^{-1} . \end{aligned} \tag{47.122}$$

This expression depends on the RTFs $\tilde{\boldsymbol{A}}(k)$, assumed to be error free, and on the coherence function $\Gamma(k)$.

The same $M = 5$ microphone array used for the directional noise case is now used for the diffused noise field. In Fig. 47.8 we show the extra NR obtained by the ANC for various steering angles and for the entire frequency band. It is clear that almost no NR is obtained in the high-frequency band and only relatively low NR in the low-frequency band. The obtained results are in accordance with the results in [47.39, 42].

Incoherent Noise Field

For incoherent noise field we assume that the noise at the sensors has no spatial correlation.

$$\boldsymbol{\Phi}_{\boldsymbol{N}_\mathrm{s}\boldsymbol{N}_\mathrm{s}}(k, \ell) = \boldsymbol{\Phi}_{n_\mathrm{s} n_\mathrm{s}}(k, \ell)\mathbf{I} ,$$

where $\mathbf{I}$ is an $M \times M$ identity matrix. Using (47.119) with perfect knowledge of the RTFs, i. e., $\hat{\tilde{\boldsymbol{A}}}(k) = \tilde{\boldsymbol{A}}(k)$, and with the prespecified $\boldsymbol{\Phi}_{\boldsymbol{N}_\mathrm{s}\boldsymbol{N}_\mathrm{s}}(k, \ell)$ we obtain,

$$\begin{aligned} &\boldsymbol{\Phi}^\mathrm{n}_\mathrm{yy}(k, \ell) \\ &= \frac{\boldsymbol{\Phi}_{n_\mathrm{s} n_\mathrm{s}}(k, \ell)}{\|\tilde{\boldsymbol{A}}(k)\|^4}\tilde{\boldsymbol{A}}^\mathrm{H}(k) \\ &\quad \times \left\{\mathbf{I} - \mathcal{H}(k)\left[\mathcal{H}^\mathrm{H}(k)\mathcal{H}(k)\right]^{-1}\mathcal{H}^\mathrm{H}(k)\right\}\tilde{\boldsymbol{A}}(k) . \end{aligned}$$

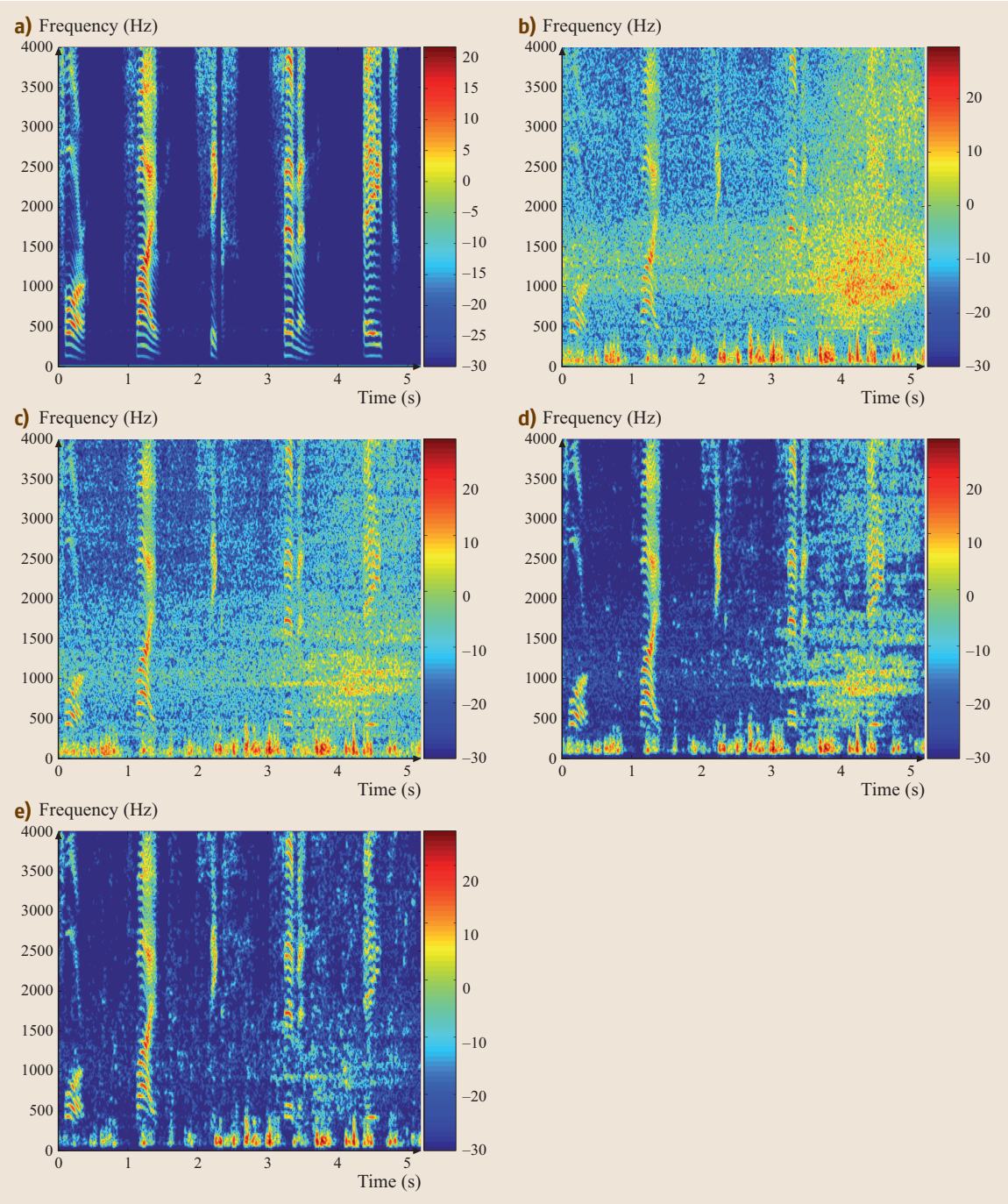

Fig. 47.9 (a) Original clean speech signal at microphone #1: "Five six seven eight nine". (b) Noisy signal at microphone #1. (c) TF-GSC output. (d) Single-channel postfiltering output. (e) Multichannel postfiltering.

It can be easily verified that,

$$\tilde{\boldsymbol{A}}^{\mathrm{H}}(k)\mathcal{H}(k) = \mathbf{0}_{1\times(M-1)} \,.$$

Furthermore, as $\mathcal{H}^{\mathrm{H}}(k)\mathcal{H}(k)$ is a positive matrix, its inverse always exists. Thus, the contribution of the noise cancelling branch is zero, and the NR is only attributed to the FBF. The noise power at the output is thus,

$$\begin{aligned}\boldsymbol{\Phi}^{\mathrm{n}}_{\mathrm{yy}}(k,\ell) = \boldsymbol{\Phi}^{\mathrm{n}}_{\mathrm{FBF}}(k,\ell) &= \boldsymbol{\Phi}_{n_{\mathrm{s}}n_{\mathrm{s}}}(k,\ell)\frac{\tilde{\boldsymbol{A}}^{\mathrm{H}}(k)\tilde{\boldsymbol{A}}(k)}{\|\tilde{\boldsymbol{A}}(k)\|^4} \\ &= \frac{\boldsymbol{\Phi}_{n_{\mathrm{s}}n_{\mathrm{s}}}(k,\ell)}{\|\tilde{\boldsymbol{A}}(k)\|^2} \,.\end{aligned}$$

Again, no NR is guaranteed by this structure, and the result depends on the RTFs involved.

In the case of delay-only ATFs, the FBF branch becomes a simple delay-and-sum beamformer, thus the expected NR is M, the number of sensors.

47.8 Experimental Results

In this section, we compare the performance of a system, consisting of the TF-GSC and multichannel postfilter, to a system consisting of a TF-GSC and a single-channel postfilter.

A linear array, consisting of four microphones with 5 cm spacing, is mounted in a car on the visor. Clean speech signals are recorded at a sampling rate of 8 kHz in the absence of background noise (standing car, silent environment). A car noise signal is recorded while the car speed is about 60 km/h, and the window next to the driver is slightly open (about 5 cm; the other windows are closed). The input microphone signals are generated by mixing the speech and noise signals at various SNR levels in the range $[-5, 10]$ dB.

Offline TF-GSC beamforming [47.38] is applied to the noisy multichannel signals, and its output is enhanced using the OM-LSA estimator [47.93]. The result is referred to as single-channel postfiltering output. Alternatively, the proposed TF-GSC and multichannel postfiltering is applied to the noisy signals. A subjective comparison between multichannel and single-channel postfiltering was conducted using speech sonograms and validated by informal listening tests. Typical examples of speech sonograms are presented in Fig. 47.9. For audio samples please refer to [47.99]. The noise PSD at the beamformer output varies substantially due to the residual interfering components of speech, wind blows, and passing cars. The TF-GSC output is characterized by a high level of noise. Single-channel postfiltering suppresses pseudostationary noise components, but is inefficient at attenuating the transient noise components. By contrast, the system which consists a TF-GSC and multichannel postfilter achieves superior noise attenuation, while preserving the desired source components. This is verified by subjective informal listening tests.

47.9 Summary

In this chapter, we concentrated on the GSC beamformer, and presented a comprehensive study of its components. We showed, that the GSC structure is closely related to other array optimization criteria, such as the Wiener filter. We described multimicrophone postfilters, based on either the MMSE or the log-spectral estimation criteria, which are designed for improving the amount of obtainable NR, with minimal degradation of speech quality. The robustness of the GSC structure to imperfect estimation of its components was analyzed. Various methods were proposed for increasing the robustness of the GSC, and especially, for avoiding leakage of the desired signal into the noise reference signals. Finally, the performance of the TF-GSC was theoretically analyzed and experimentally evaluated under nonstationary noise conditions.

47.A Appendix: Derivation of the Expected Noise Reduction for a Coherent Noise Field

For clarity of the exposition we will omit the time and frequency dependence in the derivation. Recall (47.119) and define

$$\mathcal{X}(k,\ell) \triangleq \mathcal{X} = \boldsymbol{\Phi}_{\boldsymbol{N}_s\boldsymbol{N}_s}\mathcal{H}\left(\mathcal{H}^{\mathrm{H}}\boldsymbol{\Phi}_{\boldsymbol{N}_s\boldsymbol{N}_s}\mathcal{H}\right)^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{\Phi}_{\boldsymbol{N}_s\boldsymbol{N}_s}\,. \tag{47.A1}$$

Denote,

$$\mathcal{X} \triangleq \mathcal{K}\times\mathcal{L}\times\mathcal{M}\,,$$

where, $\mathcal{K} = \boldsymbol{\Phi}_{\boldsymbol{N}_s\boldsymbol{N}_s}\mathcal{H}$, $\mathcal{L} = (\mathcal{H}^{\mathrm{H}}\boldsymbol{\Phi}_{\boldsymbol{N}_s\boldsymbol{N}_s}\mathcal{H})^{-1}$, and $\mathcal{M} = \mathcal{H}^{\mathrm{H}}\boldsymbol{\Phi}_{\boldsymbol{N}_s\boldsymbol{N}_s}$. Thus, $\mathcal{X}$ is a multiplication of three terms.

Starting from $\mathcal{L}$ and using the detailed noise structure,

$$\begin{aligned}\mathcal{L} &= \left[\mathcal{H}^{\mathrm{H}}\left(\phi_{\mathrm{nn}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}+\varepsilon\mathbf{I}\right)\mathcal{H}\right]^{-1}\\ &= \left[\phi_{\mathrm{nn}}(\mathcal{H}^{\mathrm{H}}\boldsymbol{B})(\mathcal{H}^{\mathrm{H}}\boldsymbol{B})^{\mathrm{H}}+\varepsilon\mathcal{H}^{\mathrm{H}}\mathcal{H}\right]^{-1}.\end{aligned}$$

If $\boldsymbol{B}=\boldsymbol{A}$, i. e., the noise source is located exactly at the desired signal position, then $\mathcal{H}^{\mathrm{H}}\boldsymbol{B}=0$, and the calculation of the inverse is straightforward, yielding $\mathcal{L}=(\varepsilon\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}$, $\mathcal{K}=\varepsilon\mathcal{H}$, and $\mathcal{M}=\varepsilon\mathcal{H}^{\mathrm{H}}$. Collecting all terms we obtain, $\mathcal{X}=\frac{\varepsilon^2}{\varepsilon}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\overset{\varepsilon\to0}{\longrightarrow}0$, i. e., the signal at the BM output is zero, as expected. The total noise part of the output is given by,

$$\begin{aligned}\boldsymbol{\Phi}^{\mathrm{n}}_{\mathrm{yy}} &= \boldsymbol{\Phi}^{\mathrm{n}}_{\mathrm{FBF}}\\ &= \frac{\hat{\boldsymbol{A}}^{\mathrm{H}}\boldsymbol{\Phi}_{\boldsymbol{N}_s\boldsymbol{N}_s}\hat{\boldsymbol{A}}}{\|\hat{\boldsymbol{A}}\|^4} = \frac{\hat{\boldsymbol{A}}^{\mathrm{H}}\left(\phi_{\mathrm{nn}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}+\varepsilon\mathbf{I}\right)\hat{\boldsymbol{A}}}{\|\hat{\boldsymbol{A}}\|^4}\\ &\overset{\varepsilon\to0}{\longrightarrow}\phi_{nn}|A_1|^2\,,\end{aligned}$$

where the last transition is due to $\boldsymbol{B}=\boldsymbol{A}$. This is exactly the no-distortion result obtained in Sect. 47.7.2, for the desired signal direction.

Table 47.4 Twelve terms used for calculating $\mathcal{X}=\mathcal{K}\mathcal{L}\mathcal{M}$

$\mathcal{K}_1\mathcal{L}_1\mathcal{M}_1$	$= \frac{1}{\varepsilon}\boldsymbol{\Phi}^2_{\mathrm{nn}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}$;	(I)
$\mathcal{K}_1\mathcal{L}_2\mathcal{M}_1$	$= -\frac{1}{\varepsilon}\boldsymbol{\Phi}^3_{\mathrm{nn}}\times\frac{\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}}{\phi_{\mathrm{nn}}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}}$	(II)
	$= -\frac{1}{\varepsilon}\boldsymbol{\Phi}^2_{\mathrm{nn}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}$;	
$\mathcal{K}_1\mathcal{L}_3\mathcal{M}_1$	$= \frac{\boldsymbol{\Phi}^3_{\mathrm{nn}}\boldsymbol{B}(\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B})^2\boldsymbol{B}^{\mathrm{H}}}{(\phi_{\mathrm{nn}}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B})^2}$	(III)
	$= \phi_{\mathrm{nn}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}$;	
$\mathcal{K}_2\mathcal{L}_1\mathcal{M}_1$	$= \phi_{\mathrm{nn}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}$;	(IV)
$\mathcal{K}_2\mathcal{L}_2\mathcal{M}_1$	$= -\boldsymbol{\Phi}^2_{\mathrm{nn}}\frac{\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}}{\phi_{\mathrm{nn}}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}}$	(V)
	$= -\phi_{\mathrm{nn}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}$;	
$\mathcal{K}_2\mathcal{L}_3\mathcal{M}_1$	$= \varepsilon\frac{\boldsymbol{\Phi}^2_{\mathrm{nn}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}}{(\phi_{\mathrm{nn}}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B})^2}$	(VI)
	$= \varepsilon\frac{\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}}{\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}}$;	
$\mathcal{K}_1\mathcal{L}_1\mathcal{M}_2$	$= \phi_{\mathrm{nn}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}$;	(VII)
$\mathcal{K}_1\mathcal{L}_2\mathcal{M}_2$	$= -\boldsymbol{\Phi}^2_{\mathrm{nn}}\frac{\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}}{\phi_{\mathrm{nn}}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}}$	(VIII)
	$= -\phi_{\mathrm{nn}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}$;	
$\mathcal{K}_1\mathcal{L}_3\mathcal{M}_2$	$= \varepsilon\frac{\boldsymbol{\Phi}^2_{\mathrm{nn}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}}{(\phi_{\mathrm{nn}}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B})^2}$	(IX)
	$= \varepsilon\frac{\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}}{\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}}$;	
$\mathcal{K}_2\mathcal{L}_1\mathcal{M}_2$	$= \varepsilon\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}$;	(X)
$\mathcal{K}_2\mathcal{L}_2\mathcal{M}_2$	$= -\varepsilon\frac{\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}}{\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}}$;	(XI)
$\mathcal{K}_2\mathcal{L}_3\mathcal{M}_2$	$= \varepsilon^2\frac{\phi_{\mathrm{nn}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}}{(\phi_{\mathrm{nn}}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B})^2}$.	(XII)

For the general case, $\boldsymbol{B} \neq \boldsymbol{A}$, we use the *matrix inversion lemma*, yielding:

$$\mathcal{L} = \frac{1}{\varepsilon}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1} - \frac{\frac{1}{\varepsilon^2}\phi_{\mathrm{nn}}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}}{1+\frac{1}{\varepsilon}\phi_{\mathrm{nn}}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}} .$$

Now, using the approximation $\frac{1}{1+\mu} \approx 1-\mu$, for $\mu \to 0$ (μ properly defined), yields,

$$\mathcal{L} = \frac{1}{\varepsilon}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1} - \frac{\frac{1}{\varepsilon}\phi_{\mathrm{nn}}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}}{\phi_{\mathrm{nn}}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}} + \frac{\phi_{\mathrm{nn}}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}}{\left(\phi_{\mathrm{nn}}\boldsymbol{B}^{\mathrm{H}}\mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathcal{H}^{\mathrm{H}}\boldsymbol{B}\right)^2} . \quad (47.A2)$$

Now, calculating $\mathcal{X}$,

$$\begin{aligned}\mathcal{X} &= \mathcal{K}\mathcal{L}\mathcal{M} = \boldsymbol{\Phi}_{NN}\mathcal{H}\mathcal{L}\mathcal{H}^{\mathrm{H}}\boldsymbol{\Phi}_{NN} \\ &= \left(\phi_{\mathrm{nn}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}+\varepsilon\mathbf{I}\right)\mathcal{H}\mathcal{L}\mathcal{H}^{\mathrm{H}}\left(\phi_{\mathrm{nn}}\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}+\varepsilon\mathbf{I}\right) \\ &\triangleq (\mathcal{K}_1+\mathcal{K}_2)\mathcal{H}(\mathcal{L}_1+\mathcal{L}_2+\mathcal{L}_3)\mathcal{H}^{\mathrm{H}}(\mathcal{M}_1+\mathcal{M}_2) , \end{aligned} \quad (47.A3)$$

with the obvious definitions of $\mathcal{K}_1$, $\mathcal{K}_2$, $\mathcal{L}_1$, $\mathcal{L}_2$, $\mathcal{L}_3$, $\mathcal{M}_1$, and $\mathcal{M}_2$. Opening the brackets we have twelve terms given in Table 47.4.

Note, that terms I and II, terms IV and V, and terms VII and VIII eliminate each other, and that terms VI, IX, X, XI, and XII vanish as ε approaches to zero. Only term III is left, i. e., $\mathcal{X} = \phi_{\mathrm{nn}}BB^{\mathrm{H}}$. Substituting $\mathcal{X}$ into (47.119) we have

$$\boldsymbol{\Phi}_{\mathrm{yy}}^{\mathrm{n}} = \frac{\tilde{\boldsymbol{A}}^{\mathrm{H}}\boldsymbol{\Phi}_{N_{\mathrm{s}}N_{\mathrm{s}}}\tilde{\boldsymbol{A}}}{\|\tilde{\boldsymbol{A}}\|^4} - \frac{\tilde{\boldsymbol{A}}^{\mathrm{H}}\left(\phi_{\mathrm{nn}}BB^{\mathrm{H}}\right)\tilde{\boldsymbol{A}}}{\|\tilde{\boldsymbol{A}}\|^4} = 0 . \quad (47.A4)$$

47.B Appendix: Equivalence Between Maximum SNR and LCMV Beamformers

The LCMV beamformer is given by

$$W^{\mathrm{LCMV}} = \frac{\left(\phi_{\mathrm{ss}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}+\boldsymbol{\Phi}_{NN}\right)^{-1}\boldsymbol{A}}{\boldsymbol{A}^{\mathrm{H}}\left(\phi_{\mathrm{ss}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}+\boldsymbol{\Phi}_{NN}\right)^{-1}\boldsymbol{A}} . \quad (47.B1)$$

Using the matrix inversion lemma we have in the denominator

$$\begin{aligned} &\boldsymbol{A}^{\mathrm{H}}\left(\boldsymbol{\Phi}_{NN}+\phi_{\mathrm{ss}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\right)^{-1}\boldsymbol{A} \\ &= \boldsymbol{A}^{\mathrm{H}}\left(\boldsymbol{\Phi}_{NN}^{-1} - \frac{\phi_{\mathrm{ss}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Phi}_{NN}^{-1}}{1+\phi_{\mathrm{ss}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}}\right)\boldsymbol{A} \\ &= \boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A} - \boldsymbol{A}^{\mathrm{H}}\frac{\phi_{\mathrm{ss}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Phi}_{NN}^{-1}}{1+\phi_{\mathrm{ss}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}}\boldsymbol{A} \\ &= \boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A} - \phi_{\mathrm{ss}}\frac{\left(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}\right)^2}{1+\phi_{\mathrm{ss}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}} \\ &= \frac{\left(1+\phi_{\mathrm{ss}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}\right)\left(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}\right)}{1+\phi_{\mathrm{ss}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}} \\ &\quad - \frac{\phi_{\mathrm{ss}}\left(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}\right)^2}{1+\phi_{\mathrm{ss}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}} \\ &= \frac{\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}}{1+\phi_{\mathrm{ss}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}} \end{aligned} \quad (47.B2)$$

Similarly, we have in the numerator

$$\begin{aligned} &\left(\phi_{\mathrm{ss}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}+\boldsymbol{\Phi}_{NN}\right)^{-1}\boldsymbol{A} \\ &= \boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A} - \frac{\phi_{\mathrm{ss}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}}{1+\phi_{\mathrm{ss}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}} \\ &= \frac{\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}\left(1+\phi_{\mathrm{ss}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}\right)}{1+\phi_{\mathrm{ss}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}} \\ &\quad - \frac{\phi_{\mathrm{ss}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}}{1+\phi_{\mathrm{ss}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}} \\ &= \frac{\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}}{1+\phi_{\mathrm{ss}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}} . \end{aligned} \quad (47.B3)$$

Dividing (47.B3) by (47.B2) we have

$$W^{\mathrm{LCMV}} = \frac{\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}}{\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Phi}_{NN}^{-1}\boldsymbol{A}} \equiv W^{\mathrm{MSNR}} . \quad (47.B4)$$

References

47.1 S. Haykin: *Adaptive Filter Theory*, 4th edn. (Prentice Hall, Upper Saddle River 2002)
47.2 B. Widrow, S.D. Stearns: *Adaptive Signal Processing* (Prentice-Hall, Upper Saddle River 1985)
47.3 D. Monzingo, T. Miller: *Introduction to Adaptive Arrays* (Wiley, New York 1980)
47.4 H.L.V. Trees: *Detection, Estimation, and Modulation Theory*, Vol. IV (Wiley, New York 2002)
47.5 B.V. Veen, K. Buckley: Beamforming: A versatile approach to spatial filtering, IEEE Acoust. Speech Signal Process. Mag. **5**(2), 4–24 (1988)
47.6 H. Krim, M. Viberg: Two decades of array signal processing research: The parametric approach, IEEE Signal Proc. Mag. **13**, 67–94 (1996)
47.7 H. Cox, R. Zeskind, M. Owen: Robust Adaptive Beamforming, IEEE Trans. Acoust. Speech Signal Process. **35**(10), 1365–1376 (1987)
47.8 S.L. Gay, J. Benesty: *Acoustic Signal Processing for Telecommunication* (Kluwer Academic, Dordrecht 2001)
47.9 M.S. Brandstein, D.B. Ward: *Microphone Arrays: Signal Processing Techniques and Applications* (Springer, Berlin, Heidelberg 2001)
47.10 J. Benesty, Y. Huang (Eds.): *Adaptive Signal Processing: Applications to Real-World Problems* (Springer, Berlin, Heidelberg 2003)
47.11 E. Hänsler, G. Schmidt: *Acoustic Echo and Noise Control: A Pracical Approach* (Wiley, New York 2004)
47.12 J. Benesty, S. Makino, J. Chen: *Speech Enhancement* (Springer, Berlin, Heidelberg 2005)
47.13 W. Herbordt: *Sound Capture for Human/Machine Interfaces – Practical Aspects of Microphone Array Signal Processing* (Springer, Berlin, Heidelberg 2005)
47.14 S. Gannot, J. Benesty, J. Bitzer, I. Cohen, S. Doclo, R. Martin, S. Nordholm: Advances in multimicrophone speech processing, EURASIP J. Appl. Signal Process. **2006**(ID46357) (2006)
47.15 E. Jan, J. Flanagan: Microphone arrays for speech processing, International Symposium on Signals, Systems, and Electronics (ISSSE) (URSI, San Francisco 1995) pp. 373–376
47.16 E. Jan, J. Flanagan: Sound capture from spatial volumes: matched-filter processing of microphone arrays having randomly-distributed sensors, Proc. ICASSP (Atlanta, 1996) pp. 917–920
47.17 D. Rabinkin, R. Renomeron, J. Flanagan, D. Macomber: Optimal truncation time for matched filter array processing, Proc. ICASSP (Seattle, 1998) pp. 3269–3272
47.18 M. Sondhi, G. Elko: Adaptive optimization of microphone arrays under a nonlinear constraint, Proc. ICASSP, Vol. 11 (Tokyo, 1986) pp. 981–984
47.19 Y. Kaneda, J. Ohga: Adaptive microphone-array system for noise reduction, IEEE Trans. Acoust. Speech Signal Process. **34**(6), 1391–1400 (1986)
47.20 D.V. Compernolle: Switching adaptive filters for enhancing noisy and reverberant speech from microphone array recordings, Proc. ICASSP (Albuquerque, 1990) pp. 833–836
47.21 W. Kellermann: Acoustic echo cancellation for beamforming microphone arrays. In: *Microphone Arrays: Signal Processing Techniques and Applications*, ed. by M.S. Brandstein, D.B. Ward (Springer, Berlin, Heidelberg 2001) pp. 281–306
47.22 I. Claesson, S. Nordholm: A spatial filtering approach to robust adaptive beamforming, IEEE Trans. Antennas Propag. **40**(1), 1093–1096 (1992)
47.23 S. Nordholm, I. Claesson, N. Grbić: Optimal and adaptive microphone arrays for speech input in automobiles. In: *Microphone Arrays: Signal Processing Techniques and Applications*, ed. by M.S. Brandstein, D.B. Ward (Springer, Berlin, Heidelberg 2001) pp. 307–330
47.24 R. Martin: Small microphone arrays with postfilters for noise and acoustic echo reduction. In: *Microphone Arrays: Signal Processing Techniques and Applications*, ed. by M.S. Brandstein, D.B. Ward (Springer, Berlin, Heidelberg 2001) pp. 255–280
47.25 B. Widrow, J.R. Glover Jr., J.M. McCool, J. Kaunitz, C.S. Williams, R.H. Hearn, J.R. Zeidler, E. Dong Jr, R.C. Goodlin: Adaptive noise cancelling: Principals and applications, Proc. IEEE **63**(12), 1692–1716 (1975)
47.26 S. Doclo, M. Moonen: GSVD-based optimal filtering for single and multimicrophone speech enhancement, IEEE Trans. Speech Audio Process. **50**(9), 2230–2244 (2002)
47.27 S. Doclo, M. Moonen: Design of far-field and near-field broadband beamformers using eigenfilters, Signal Process. **83**(12), 2641–2673 (2003)
47.28 S. Doclo, M. Moonen: Multi-microphone noise reduction using recursive GSVD-based optimal filtering with ANC postprocessing stage, IEEE Trans. Speech Audio Process. **13**(1), 53–69 (2005)
47.29 A. Spriet, M. Moonen, J. Wouters: A multichannel subband GSVD approach for speech enhancement, Eur. Trans. Telecommun. **13**(2), 149–158 (2002), Special Issue on Acoustic Echo and Noise Control
47.30 G. Rombouts, M. Moonen: QRD-based unconstrained optimal filtering for acoustic noise reduction, Signal Process. **83**(9), 1889–1904 (2003)
47.31 G. Rombouts, M. Moonen: Fast QRD-lattice-based unconstrained optimal filtering for acoustic noise reduction, IEEE Trans. Speech Audio Process. **13**(6), 1130–1143 (2005)

47.32 J. Chen, J. Benesty, Y. Huang, S. Doclo: New insights into the noise reduction Wiener filter, IEEE Trans. Speech Audio Process. **14**(4), 1218–1234 (2006)

47.33 A. Spriet, M. Moonen, J. Wouters: Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction, Signal Process. **84**(12), 2367–2387 (2004)

47.34 S. Doclo, A. Spriet, J. Wouters, M. Moonen: Speech distortion weighted multichannel wiener filtering techniques for noise reduction. In: *Speech Enhancement*, Signals and Communication Technology, ed. by J. Benesty, S. Makino, J. Chen (Springer, Berlin, Heidelberg 2005) pp. 199–228

47.35 O.L. Frost: An algorithm for linearly constrained adaptive array processing, Proc. IEEE **60**(8), 926–935 (1972)

47.36 L.J. Griffiths, C.W. Jim: An alternative approach to linearly constrained adaptive beamforming, IEEE Trans. Antennas Propag. **30**(1), 27–34 (1982)

47.37 S. Affes, Y. Grenier: A source subspace tracking array of microphones for double talk situations, Proc. ICASSP (Munich, 1997) pp. 269–272

47.38 S. Gannot, D. Burshtein, E. Weinstein: Signal enhancement using beamforming and nonstationarity with applications to speech, IEEE Trans. Signal Proces. **49**(8), 1614–1626 (2001)

47.39 S. Nordholm, Y.H. Leung: Performance limits of the broadband generalized sidelobe cancelling structure in an isotropic noise field, J. Acoust. Soc. Am. **107**(2), 1057–1060 (2000)

47.40 J. Bitzer, K.U. Simmer, K.D. Kammeyer: Theoretical noise reduction limits of the generalized sidelobe canceller (GSC) for speech enhancement, Proc. ICASSP (Phoenix, 1999) pp. 2965–2968

47.41 J. Bitzer, K. Simmer, K. Kammeyer: Multichannel noise reduction – algorithms and theoretical limits, Proc. EUSIPCO, Vol. I (Rhodes, 1998) pp. 105–108

47.42 J. Bitzer, K.-D. Kammeyer, K. Simmer: An alternative implementation of the superdirective beamformer, Workshop on Application of Signal Processing to Audio and Acoustics (IEEE, New Paltz, NY 1999)

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

47.44 S. Nordholm, I. Claesson, M. Dahl: Adaptive microphone array employing calibration signals: an analytical evaluation, IEEE Trans. Speech Audio Process. **7**(3), 241–252 (1999)

47.45 K.-C. Huarng, C.-C. Yeh: Performance analysis of derivative constraint adaptive arrays with pointing errors, IEEE Trans. Acoust. Speech Signal Process. **38**(2), 209–219 (1990)

47.46 S. Nordholm, I. Claesson, P. Eriksson: The broadband wiener solution for griffiths-jim beamformers, IEEE Trans. Signal Process. **40**(2), 474–478 (1992)

47.47 H. Cox: Resolving power and sensitivity to mismatch of optimum array processors, J. Acoust. Soc. Am. **54**(3), 771–785 (1973)

47.48 D. Brandwood: A complex gradient operator and its application in adaptive array theory, Proc. IEEE **130**(1 Parts F and H), 11–16 (1983)

47.49 G. Strang: *Linear Algebra and its Application*, 2nd edn. (Academic, New York 1980)

47.50 R. Crochiere: A weighted overlap-add method for short-time Fourier analysis/synthesis, IEEE Trans. Acoust. Speech Signal Process. **28**(1), 99–102 (1980)

47.51 J. Shynk: Frequency-domain and multirate and adaptive filtering, IEEE Signal Process. Mag. **9**(1), 14–37 (1992)

47.52 J. Allen, D. Berkley: Image method for efficiently simulating small-room acoustics, J. Acoust. Soc. Am. **65**(4), 943–950 (1979)

47.53 S. Doclo, M. Moonen: Combined frequency-domain dereverberation and noise reduction technique for multi-microphone speech enhancement, Int. Workshop Acoust. Echo Noise Control (IWAENC) (Darmstadt 2001) pp. 31–34

47.54 B. Yang: Projection approximation subspace tracking, IEEE Trans. Signal Process. **43**(1), 95–107 (1995)

47.55 O. Hoshuyama, A. Sugiyama: Robust adaptive beamforming. In: *Microphone Arrays*, ed. by M. Brandstein, D. Ward (Springer, Berlin, Heidelberg 2001) pp. 87–109

47.56 E. Weinstein, M. Feder, A. Oppenheim: Multichannel signal separation by decorrelation, IEEE Trans. Speech Audio Process. **1**(4), 405–413 (1993)

47.57 S. Gannot: *Array Processing of Nonstationary Signals with Application to Speech* (Tel-Aviv University, Tel-Aviv 2000), available on line, http://www.biu.ac.il/ gannot

47.58 C. Knapp, G. Carter: The generalized correlation method for estimation of time delay, IEEE Trans. Acoust. Speech Signal Process. **24**(4), 320–327 (1976)

47.59 T. Dvorkind, S. Gannot: Time difference of arrival estimation of speech source in a noisy and reverberant environment, Signal Process. **85**(1), 177–204 (2005)

47.60 J. Chen, J. Benesty, Y. Huang: Time delay estimation in room acoustic environments: an overview, EURASIP J. Appl. Signal Process. **26**(503), 19 (2006)

47.61 I. Cohen: Relative transfer function identification using speech signals, IEEE Trans. Speech Audio Process. **12**(5), 451–459 (2004)

47.62 I. Cohen: Identification of speech source coupling between sensors in reverberant noisy environments, IEEE Signal Process. Lett. **11**(7), 613–616 (2004)

47.63 O. Shalvi, E. Weinstein: System identification using nonstationary signals, IEEE Trans. Signal Process. **44**(8), 2055–2063 (1996)

47.64 D.G. Manolakis, V.K. Ingle, S.M. Kogan: *Statistical and Adaptive Signal Processing: Spectral Estima-*

tion, Signal Modeling, Adaptive Filtering and Array Processing (McGraw-Hill, Singapore 2000)
47.65 R. Martin: Noise power spectral density estimation based on optimal smoothing and minimum statistics, IEEE Trans. Speech Audio Process. **9**(5), 504–512 (2001)
47.66 I. Cohen, B. Berdugo: Noise estimation by minima controlled recursive averaging for robust speech enhancement, IEEE Signal Process. Lett. **9**(1), 12–15 (2002)
47.67 I. Cohen: Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging, IEEE Trans. Speech Audio Process. **11**(5), 466–475 (2003)
47.68 S. Nordholm, H.Q. Dam, N. Grbić, S.Y. Low: Adaptive microphone array employing spatial quadratic soft constraints and spectral shaping. In: *Speech Enhancement*, Signals and Communication Technology, ed. by J. Benesty, S. Makino, J. Chen (Springer, Berlin, Heidelberg 2005) pp. 229–246
47.69 O. Hoshuyama, A. Sugiyama, A. Hirano: A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters, IEEE Trans. Signal Process. **47**(10), 2677–2684 (1999)
47.70 C. Fancourt, L. Parra: The generalized sidelobe decorrelator, Proc. IEEE Workshop on Applications of Signal Process. to Audio and Acoustics (New Paltz, NY 2001) pp. 167–170
47.71 W. Herbordt, W. Kellermann: Computationally efficient frequency-domain robust generalized sidelobe canceller, Int. Workshop Acoust. Echo Noise Control (IWAENC) (Darmstadt 2001) pp. 51–54
47.72 W. Herbordt, W. Kellermann: Adaptive beamforming for audio signal acquisition. In: *Adaptive Signal Processing: Applications to Real-World Problems*, Signals and Communication Technology, ed. by J. Benesty, Y. Huang (Springer, Berlin, Heidelberg 2003) pp. 155–194
47.73 S. Doclo, M. Moonen: On the output SNR of the speech-distortion weighted multichannel Wiener filter, IEEE Signal Process. Lett. **11**(12), 809–811 (2005)
47.74 A. Spriet, M. Moonen, J. Wouters: Stochastic gradient-based implementation of spatially preprocessed speech distortion weighted multichannel wiener filtering for noise reduction in hearing aids, IEEE Trans. Signal Process. **53**(3), 911–925 (2005)
47.75 A. Spriet, M. Moonen, J. Wouters: Robustness analysis of multichannel wiener filtering and generalized sidelobe cancellation for multimicrophone noise reduction in hearing aid applications, IEEE Trans. Speech Audio Process. **13**(4), 487–503 (2005)
47.76 S. Low, S. Nordholm, R. Togneri: Convolutive blind signal separation with post-processing, IEEE Trans. Speech Audio Process. **12**(5), 539–548 (2004)
47.77 S. Gannot, D. Burshtein, E. Weinstein: Analysis of the power spectral deviation of the general transfer function GSC, IEEE Trans. Signal Process. **52**(4), 1115–1121 (2004)
47.78 K.U. Simmer, J. Bitzer, C. Marro: Post-filtering techniques. In: *Microphone Arrays: Signal Processing Techniques and Applications*, ed. by M.S. Brandstein, D.B. Ward (Springer, Berlin, Heidelberg 2001) pp. 39–60
47.79 R. Zelinski: A Microphone array with adaptive post-filtering for noise reduction in reverberant rooms, IEEE ICASSP (New York, 1988) pp. 2578–2581
47.80 R. Zelinski: Noise reduction based on microphone array with LMS adaptive post-filtering, Electron. Lett. **26**(24), 2036–2581 (1990)
47.81 S. Fischer, K.U. Simmer: An adaptive microphone array for hands-free communication, Int. Workshop Acoust. Echo Noise Control (IWAENC) (Røros, 1995) pp. 44–47
47.82 S. Fischer, K.U. Simmer: Beamforming microphone arrays for speech acquisition in noisy environments, Speech Commun. **20**(3–4), 215–227 (1996)
47.83 S. Fischer, K.-D. Kammeyer: Broadband beamforming with adaptive postfiltering for speech acquisition in noisy environment, Proc. ICASSP, Vol. 1 (Munich, 1997) pp. 359–362
47.84 J. Meyer, K.U. Simmer: Multi-channel speech enhancement in a car environment using Wiener filtering and spectral subtraction, Proc. ICASSP (Munich, 1997) pp. 21–24
47.85 K.U. Simmer, S. Fischer, A. Wasiljeff: Suppression of coherent and incoherent noise using a microphone array, Annales des Télécommun. **49**(7–8), 439–446 (1994)
47.86 J. Bitzer, K. Simmer, K.-D. Kammeyer: Multimicrophone noise reduction by post-filter and superdirective beamformer, Int. Workshop Acoust. Echo Noise Control (IWAENC) (Pocono Manor, 1999) pp. 100–103
47.87 J. Bitzer, K.U. Simmer, K.-D. Kammeyer: Multimicrophone noise reduction techniques as front-end devices for speech recognition, Speech Commun. **34**, 3–12 (2001)
47.88 K.U. Simmer, A. Wasiljeff: Adptive microphone arrays for noise supression in the frequency domain, Proc. Second Cost 229 Workshop on Adaptive Algorithms in Communications (Bordeaux, 1992) pp. 185–194
47.89 I. McCowan, H. Bourlard: Microphone array post-filter for diffuse noise field, Proc. ICASSP, Vol. 1. (Orlando, 2002) pp. 905–908
47.90 I. McCowan, H. Bourlard: Microphone array post-filter based on noise field coherence, IEEE Trans. Speech Audio Process. **11**(6), 709–716 (2003)
47.91 I. Cohen, S. Gannot, B. Berdugo: An integrated real-time beamforming and postfiltering system for non-stationary noise environments, EURASIP J. Appl. Signal Process. **2003**(11), 1064–1073 (2003), special issue on Signal Processing for Acoustic Communication Systems

47.92 I. Cohen, B. Berdugo: Microphone array post-filtering for non-stationary noise suppression, Proc. ICASSP (Orlando, 2002) pp. 901–904
47.93 I. Cohen, B. Berdugo: Speech enhancement for non-stationary noise environments, Signal Process. **81**(11), 2403–2418 (2001)
47.94 Y. Ephraim, D. Malah: Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator, IEEE Trans. Acoust. Speech Signal Process. **ASSP-32**(6), 1109–1121 (1984)
47.95 Y. Ephraim, D. Malah: Speech enhancement using a minimum mean square error log-spectral amplitude estimator, IEEE Trans. Acoust. Speech Signal Process. **33**(2), 443–445 (1985)
47.96 I. Cohen: Analysis of two-channel generalized sidelobe canceller (GSC) with post-filtering, IEEE Trans. Speech Audio Process. **11**(6), 684–699 (2003)
47.97 S. Gannot, D. Burshtein, E. Weinstein: Theoretical analysis of the general transfer function GSC, Int. Workshop Acoust. Echo Noise Control (IWAENC) (Darmstadt 2001) pp. 27–30
47.98 N. Dal-Degan, C. Prati: Acoustic noise analysis and speech enhancement techniques for mobile radio application, Signal Process. **15**(4), 43–56 (1988)
47.99 S. Gannot, I. Cohen: Audio Sample Files http://www.biu.ac.il/ gannot (2005)