Chapter 8
Microphone Arrays: Fundamental Concepts

Jacek P. Dmochowski and Jacob Benesty

Abstract Microphone array beamforming is concerned with the extraction of a desired acoustic signal from noisy microphone measurements. The microphone array problem is a more difficult one than that of classical sensor array applications due to several reasons: the speech signal is naturally analog and wideband. Moreover, the acoustic channel exhibits strong multipath components and long reverberation times. However, the conventional metrics utilized to evaluate signal enhancement performance do not necessarily reflect these differences. In this chapter, we attempt to reformulate the objectives of microphone array processing in a unique manner, one which explicitly accounts for the broadband signal and the reverberant channel. A distinction is made between wideband and narrowband metrics. The relationships between broadband performance measures and the corresponding component narrowband measures are analyzed. The broadband metrics presented here provide measures which, when optimized, hopefully lead to beamformer designs tailored to the specific nature of the microphone array environment.

8.1 Introduction

Microphone arrays are becoming increasingly more common in the acquisition and de-noising of received acoustic signals. Additional microphones allow us to apply spatiotemporal filtering methods which are, at least in theory, significantly more powerful in their ability to rid the received signal of the unwanted additive noise than conventional temporal filtering techniques which simply emphasize certain temporal frequencies while de-emphasizing others.

Jacek P. Dmochowski
City College of New York, NY, USA, e-mail: jdmochowski@ccny.cuny.edu

Jacob Benesty
INRS-EMT, QC, Canada, e-mail: benesty@emt.inrs.ca

I. Cohen et al. (Eds.): Speech Processing in Modern Communication, STSP 3, pp. 199–223.
springerlink.com

It may be argued that the advantage of multiple microphones has not been fully realized in practice. In this chapter, we attempt to shed light on the fundamental problems and goals of microphone array beamforming by studying the metrics by which performance is measured.

The initial microphone array designs [1], [2] are formulated around rather lofty expectations. For example, the minimum variance distortionless response (MVDR) [3] beamformer has long been studied in the microphone array context. Notice that the MVDR beamformer attempts to perform both dereverberation *and* noise reduction simultaneously in each frequency bin. In a reverberant environment with unknown impulse responses, acoustic dereverberation is a challenging problem in of itself; constraining the frequency-domain solution to achieve perfect dereverberation while at the same time reducing additive noise is ambitious. It may be speculated that the coupling of dereverberation and noise reduction in adaptive beamformer designs leads to poor performance in practice [4].

This chapter attempts to clearly identify the challenges and define the metrics involved in the microphone array beamforming problem. Instead of attempting to develop new beamformer designs, we focus on clarifying the goals of microphone arrays, which will then hopefully lead to the development of more powerful beamformers that are tailored to the distinct nature of the environment.

8.2 Signal Model

Consider the conventional signal model in which an N-element microphone array captures a convolved source signal in some noise field. The received signals at the time instant t are expressed as [2], [5], [6]

$$\begin{aligned} y_n(t) &= g_n(t) * s(t) + v_n(t) \\ &= x_n(t) + v_n(t), \ n = 1, 2, \ldots, N, \end{aligned} \tag{8.1}$$

where $g_n(t)$ is the impulse response from the unknown source $s(t)$ to the nth microphone, $*$ stands for linear convolution, and $v_n(t)$ is the additive noise at microphone n. We assume that the signals $x_n(t)$ and $v_n(t)$ are uncorrelated and zero mean. By definition, $x_n(t)$ is coherent across the array. The noise signals $v_n(t)$ are typically only partially (if at all) coherent across the array. All previous signals are considered to be real and broadband.

Conventionally, beamforming formulations have attempted to recover $s(t)$ from the microphone measurements $y_n(t)$, $n = 1, \ldots, N$. This involves two processes: dereverberation and noise reduction. In this chapter, the desired signal is instead designated by the clean (but convolved) signal received at microphone 1, namely $x_1(t)$. The problem statement may be posed as follows: given N mixtures of two uncorrelated signals $x_n(t)$ and $v_n(t)$, our aim is to

preserve $x_1(t)$ while minimizing the contribution of the noise terms $v_n(t)$ in the array output. While the array processing does not attempt to perform any inversion of the acoustic channels g_n, (single-channel) dereverberation techniques may be applied to the beamformer output in a post-processing fashion. Such techniques are not considered in this chapter, however.

The main objective of this chapter is to properly define all relevant measures that aid us in recovering the desired signal $x_1(t)$, to analyze the signal components at the beamformer output, and to clarify the most important concepts in microphone arrays.

In the frequency domain, (8.1) can be rewritten as

$$\begin{aligned} Y_n(f) &= G_n(f)S(f) + V_n(f) \\ &= X_n(f) + V_n(f),\ n = 1, 2, \ldots, N, \end{aligned} \tag{8.2}$$

where $Y_n(f)$, $G_n(f)$, $S(f)$, $X_n(f) = G_n(f)S(f)$, and $V_n(f)$ are the frequency-domain representations of $y_n(t)$, $g_n(t)$, $s(t)$, $x_n(t)$, and $v_n(t)$, respectively, at temporal frequency f, and the time-domain signal

$$a(t) = \int_{-\infty}^{\infty} A(f)e^{j2\pi ft}df \tag{8.3}$$

is the inverse Fourier transform of $A(f)$.

The N microphone signals in the frequency domain are better summarized in a vector notation as

$$\begin{aligned} \mathbf{y}(f) &= \mathbf{g}(f)S(f) + \mathbf{v}(f) \\ &= \mathbf{x}(f) + \mathbf{v}(f) \\ &= \mathbf{d}(f)X_1(f) + \mathbf{v}(f), \end{aligned} \tag{8.4}$$

where

$$\begin{aligned} \mathbf{y}(f) &= \left[Y_1(f)\ Y_2(f) \cdots\ Y_N(f) \right]^T, \\ \mathbf{x}(f) &= \left[X_1(f)\ X_2(f) \cdots\ X_N(f) \right]^T, \\ &= S(f) \left[G_1(f)\ G_2(f) \cdots\ G_N(f) \right]^T \\ &= S(f)\mathbf{g}(f), \\ \mathbf{v}(f) &= \left[V_1(f)\ V_2(f) \cdots\ V_N(f) \right]^T, \\ \mathbf{d}(f) &= \left[1\ \tfrac{G_2(f)}{G_1(f)} \cdots\ \tfrac{G_N(f)}{G_1(f)} \right]^T \\ &= \frac{\mathbf{g}(f)}{G_1(f)}, \end{aligned}$$

and superscript ^{T} denotes transpose of a vector or a matrix. The vector $\mathbf{d}(f)$ is termed the *steering vector* or *direction vector* since it determines the

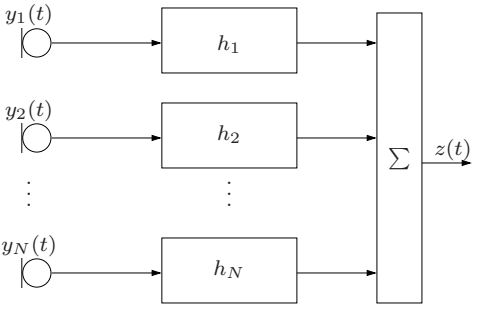

Fig. 8.1 Structure of a broadband beamformer, where h_n, $n = 1, 2, \ldots, N$, are finite impulse response (FIR) filters.

direction of the desired signal $X_1(f)$ [7], [8]. This definition is a generalization of the classical steering vector to a reverberant (multipath) environment. Indeed, the acoustic impulse responses ratios from a broadband source to the aperture convey information about the position of the source.

8.3 Array Model

Usually, the array processing or beamforming is performed by applying a temporal filter to each microphone signal and summing the filtered signals (see Fig. 8.1). In the frequency domain, this is equivalent to adding a complex weight to the output of each sensor and summing across the aperture:

$$\begin{aligned} Z(f) &= \mathbf{h}^H(f)\mathbf{y}(f) \\ &= \mathbf{h}^H(f)\left[\mathbf{d}(f)X_1(f) + \mathbf{v}(f)\right] \\ &= X_{1,\mathrm{f}}(f) + V_{\mathrm{rn}}(f), \end{aligned} \tag{8.5}$$

where $Z(f)$ is the beamformer output signal,

$$\mathbf{h}(f) = \left[H_1(f)\ H_2(f)\ \cdots\ H_N(f)\right]^T$$

is the beamforming weight vector which is suitable for performing spatial filtering at frequency f, superscript ^{H} denotes transpose conjugation of a vector or a matrix, $X_{1,\mathrm{f}}(f) = \mathbf{h}^H(f)\mathbf{d}(f)X_1(f)$ is the filtered desired signal, and $V_{\mathrm{rn}}(f) = \mathbf{h}^H(f)\mathbf{v}(f)$ is the residual noise.

In the time domain, the beamformer output signal is

$$\begin{aligned} z(t) &= \int_{-\infty}^{\infty} Z(f) e^{j2\pi ft} df \\ &= \int_{-\infty}^{\infty} X_{1,\mathrm{f}}(f) e^{j2\pi ft} df + \int_{-\infty}^{\infty} V_{\mathrm{rn}}(f) e^{j2\pi ft} df \\ &= x_{1,\mathrm{f}}(t) + v_{\mathrm{rn}}(t). \end{aligned} \tag{8.6}$$

8.4 Signal-to-Noise Ratio

One of the most important measures in all aspects of speech enhancement is the *signal-to-noise ratio (SNR)*. The SNR is a second-order measure which quantifies the level of noise present relative to the level of the desired signal.

Since the processing of the array signals may be done either in the time- or frequency-domain, the SNR becomes domain-dependent. To that end, one must differentiate between the narrowband (i.e., at a single frequency) SNR and the broadband SNR (i.e., occurring across the entire frequency range). In any acoustic application, the broadband SNR is the more appropriate metric; however, since the array signals are often decomposed into narrowband bins and processed locally, the narrowband SNRs may be taken into account during the algorithmic processing.

We begin by defining the *broadband input SNR* as the ratio of the power of the time-domain desired signal over the power of the time-domain noise at the reference microphone, i.e.,

$$\begin{aligned} \mathrm{iSNR} &= \frac{E\left[x_1^2(t)\right]}{E\left[v_1^2(t)\right]} \\ &= \frac{\int_{-\infty}^{\infty} \phi_{x_1}(f) df}{\int_{-\infty}^{\infty} \phi_{v_1}(f) df}, \end{aligned} \tag{8.7}$$

where the component *narrowband input SNR* is written as

$$\mathrm{iSNR}(f) = \frac{\phi_{x_1}(f)}{\phi_{v_1}(f)}, \tag{8.8}$$

where

$$\phi_a(f) = E\left[|A(f)|^2\right] \tag{8.9}$$

denotes the power spectral density (PSD) of the wide sense stationary (WSS) process $a(t)$ at temporal frequency f.

To quantify the level of noise remaining in the beamformer output signal, $z(t)$, we define the *broadband output SNR* as the ratio of the power of the filtered desired signal over the power of the residual noise, i.e.,

$$\begin{aligned}\text{oSNR}(\mathbf{h}) &= \frac{E\left[x_{1,\text{f}}^2(t)\right]}{E\left[v_{\text{rn}}^2(t)\right]} \\ &= \frac{\int_{-\infty}^{\infty}\phi_{x_1}(f)\left|\mathbf{h}^H(f)\mathbf{d}(f)\right|^2 df}{\int_{-\infty}^{\infty}\mathbf{h}^H(f)\mathbf{\Phi}_v(f)\mathbf{h}(f)df},\end{aligned} \tag{8.10}$$

where $\mathbf{\Phi}_v(f) = E\left[\mathbf{v}(f)\mathbf{v}^H(f)\right]$ is the PSD matrix of the noise signals at the array.

The *narrowband output SNR* is given by

$$\text{oSNR}\left[\mathbf{h}(f)\right] = \frac{\phi_{x_1}(f)\left|\mathbf{h}^H(f)\mathbf{d}(f)\right|^2}{\mathbf{h}^H(f)\mathbf{\Phi}_v(f)\mathbf{h}(f)}. \tag{8.11}$$

In the particular case where we only have one microphone (no spatial processing), we get

$$\text{oSNR}\left[\mathbf{h}(f)\right] = \text{iSNR}\left(f\right). \tag{8.12}$$

Notice that the broadband input and output SNRs cannot be expressed as an integral of their narrowband counterparts:

$$\begin{aligned}\text{iSNR} &\neq \int_{-\infty}^{\infty}\text{iSNR}(f)df, \\ \text{oSNR}\left(\mathbf{h}\right) &\neq \int_{-\infty}^{\infty}\text{oSNR}\left[\mathbf{h}(f)\right]df.\end{aligned} \tag{8.13}$$

It is also important to understand that for all cases, the SNR has some limitations as a measure of beamforming "goodness." The measure considers signal power without taking into account distortion in the desired signal. As a result, additional measures need to be defined, as shown in upcoming sections.

8.5 Array Gain

The role of the beamformer is to produce a signal whose SNR is higher than that which was received. To that end, the *array gain* is defined as the ratio of the output SNR (after beamforming) over the input SNR (at the reference microphone) [1]. This leads to the following definitions:

- the *broadband array gain*,

$$\mathcal{A}(\mathbf{h}) = \frac{\text{oSNR}(\mathbf{h})}{\text{iSNR}} \tag{8.14}$$
$$= \frac{\int_{-\infty}^{\infty} \phi_{x_1}(f) \left|\mathbf{h}^H(f)\mathbf{d}(f)\right|^2 df}{\int_{-\infty}^{\infty} \phi_{x_1}(f) df} \frac{\int_{-\infty}^{\infty} \phi_{v_1}(f) df}{\int_{-\infty}^{\infty} \mathbf{h}^H(f)\mathbf{\Phi}_v(f)\mathbf{h}(f) df},$$

- and the *narrowband array gain*,

$$\mathcal{A}\left[\mathbf{h}(f)\right] = \frac{\text{oSNR}\left[\mathbf{h}(f)\right]}{\text{iSNR}(f)}$$
$$= \frac{\left|\mathbf{h}^T(f)\mathbf{d}(f)\right|^2}{\mathbf{h}^T(f)\mathbf{\Gamma}_v(f)\mathbf{h}(f)}, \tag{8.15}$$

where

$$\mathbf{\Gamma}_v(f) = \phi_{v_1}^{-1}(f)\mathbf{\Phi}_v(f) \tag{8.16}$$

is the spatial pseudo-coherence matrix of the noise. By inspection,

$$\mathcal{A}(\mathbf{h}) \neq \int_{-\infty}^{\infty} \mathcal{A}\left[\mathbf{h}(f)\right] df. \tag{8.17}$$

Assume that the noise is temporally and spatially white with variance σ_v^2 at all microphones; in this case, the pseudo-coherence matrix simplifies to

$$\mathbf{\Gamma}_v(f) = \mathbf{I}_N, \tag{8.18}$$

where $\mathbf{I}_N$ is the N-by-N identity matrix. As a result, the narrowband array gain simplifies to

$$\mathcal{A}\left[\mathbf{h}(f)\right] = \frac{\left|\mathbf{h}^H(f)\mathbf{d}(f)\right|^2}{\mathbf{h}^H(f)\mathbf{h}(f)}. \tag{8.19}$$

Using the Cauchy-Schwartz inequality, it is easy to obtain

$$\mathcal{A}\left[\mathbf{h}(f)\right] \leq \|\mathbf{d}(f)\|_2^2, \ \forall \ \mathbf{h}(f). \tag{8.20}$$

We deduce from (8.20) that the narrowband array gain never exceeds the square of the 2-norm of the steering vector $\mathbf{d}(f)$. For example, if the elements of $\mathbf{d}(f)$ are given by anechoic plane wave propagation

$$\mathbf{d}(f) = \left[\, 1 \ e^{-j2\pi f \tau_{12}} \ \cdots \ e^{-j2\pi f \tau_{1N}} \,\right], \tag{8.21}$$

where τ_{1n} is the relative delay between the reference microphone and microphone n, then it follows that

$$\begin{aligned}\mathcal{A}\left[\mathbf{h}(f)\right] &\leq \|\mathbf{d}(f)\|_2^2 \\ &\leq N,\end{aligned} \tag{8.22}$$

and the array gain is upper-bounded by the number of microphones. It is important to observe that the time-domain array gain is generally different from the narrowband array gain given at each frequency.

8.6 Noise Rejection and Desired Signal Cancellation

The array gain fails to capture the presence of desired signal distortion introduced by the beamforming process. Thus, this section introduces two submeasures which treat signal distortion and noise reduction individually.

The *noise-reduction factor* [9], [10] or *noise-rejection factor* [11] quantifies the amount of noise being rejected by the beamformer. This quantity is defined as the ratio of the power of the noise at the reference microphone over the power of the noise remaining at the beamformer output. We provide the following definitions:

- the *broadband noise-rejection factor*,

$$\xi_{\text{nr}}(\mathbf{h}) = \frac{\int_{-\infty}^{\infty} \phi_{v_1}(f)df}{\int_{-\infty}^{\infty} \mathbf{h}^H(f)\mathbf{\Phi}_v(f)\mathbf{h}(f)df}, \tag{8.23}$$

- and the *narrowband noise-rejection factor*,

$$\begin{aligned}\xi_{\text{nr}}\left[\mathbf{h}(f)\right] &= \frac{\phi_{v_1}(f)}{\mathbf{h}^H(f)\mathbf{\Phi}_v(f)\mathbf{h}(f)} \\ &= \frac{1}{\mathbf{h}^H(f)\mathbf{\Gamma}_v(f)\mathbf{h}(f)}.\end{aligned} \tag{8.24}$$

The broadband noise-rejection factor is expected to be lower bounded by 1; otherwise, the beamformer amplifies the noise received at the microphones. The higher the value of the noise-rejection factor, the more the noise is rejected.

In practice, most beamforming algorithms distort the desired signal. In order to quantify the level of this distortion, we define the *desired-signal-reduction factor* [5] or *desired-signal-cancellation factor* [11] as the ratio of the variance of the desired signal at the reference microphone over the variance of the filtered desired signal at the beamformer output. It is easy to deduce the following mathematical definitions:

- the *broadband desired-signal-cancellation factor*,

$$\xi_{\text{dsc}}(\mathbf{h}) = \frac{\int_{-\infty}^{\infty} \phi_{x_1}(f) df}{\int_{-\infty}^{\infty} \phi_{x_1}(f) \left|\mathbf{h}^H(f)\mathbf{d}(f)\right|^2 df} \tag{8.25}$$

- and the *narrowband desired-signal-cancellation factor*,

$$\xi_{\text{dsc}}\left[\mathbf{h}(f)\right] = \frac{1}{\left|\mathbf{h}^H(f)\mathbf{d}(f)\right|^2}. \tag{8.26}$$

Once again, note that

$$\begin{aligned} \xi_{\text{nr}}(\mathbf{h}) &\neq \int_{-\infty}^{\infty} \xi_{\text{nr}}\left[\mathbf{h}(f)\right] df, \\ \xi_{\text{dsc}}(\mathbf{h}) &\neq \int_{-\infty}^{\infty} \xi_{\text{dsc}}\left[\mathbf{h}(f)\right] df. \end{aligned} \tag{8.27}$$

Another key observation is that the design of broadband beamformers that do not cancel the broadband desired signal requires the constraint

$$\mathbf{h}^H(f)\mathbf{d}(f) = 1, \forall f. \tag{8.28}$$

Thus, the desired-signal-cancellation factor is equal to 1 if there is no cancellation and expected to be greater than 1 when cancellation happens.

Lastly, by making the appropriate substitutions, one can derive the following relationships between the array gain, noise-rejection factor, and desired-signal-cancellation factor:

$$\begin{aligned} \mathcal{A}(\mathbf{h}) &= \frac{\xi_{\text{nr}}(\mathbf{h})}{\xi_{\text{dsc}}(\mathbf{h})}, \\ \mathcal{A}\left[\mathbf{h}(f)\right] &= \frac{\xi_{\text{nr}}\left[\mathbf{h}(f)\right]}{\xi_{\text{dsc}}\left[\mathbf{h}(f)\right]}. \end{aligned} \tag{8.29}$$

8.7 Beampattern

The *beampattern* is a convenient way to represent the response of the beamformer to the signal $x_1(t)$ as a function of the steering vector $\mathbf{d}(f)$ (or equivalently, the location of the source), assuming the absence of any noise or interference. This steering vector spans the ratios of acoustic impulse responses from any point in space to the array of sensors. Formally, the beampattern is defined as the ratio of the variance of the beamformer output when the source impinges with a steering vector $\mathbf{d}(f)$ to the variance of the desired signal $x_1(t)$. From this definition, we deduce

- the *broadband beampattern*,

$$\mathcal{B}(\mathbf{d}) = \frac{\int_{-\infty}^{\infty} \phi_{x_1}(f) \left|\mathbf{h}^H(f)\mathbf{d}(f)\right|^2 df}{\int_{-\infty}^{\infty} \phi_{x_1}(f) df}, \tag{8.30}$$

- and the *narrowband beampattern*,

$$\mathcal{B}\left[\mathbf{d}(f)\right] = \left|\mathbf{h}^H(f)\mathbf{d}(f)\right|^2. \tag{8.31}$$

It is interesting to point out that the broadband beampattern is a linear combination of narrowband beampatterns:

$$\mathcal{B}(\mathbf{d}) = \frac{\int_{-\infty}^{\infty} \phi_{x_1}(f)\mathcal{B}\left[\mathbf{d}(f)\right] df}{\int_{-\infty}^{\infty} \phi_{x_1}(f) df}, \tag{8.32}$$

as the denominator is simply a scaling factor. The contribution of each narrowband beampattern to the overall broadband beampattern is proportional to the power of the desired signal at that frequency.

It is also interesting to observe the following relations:

$$\mathcal{B}(\mathbf{d}) = \frac{1}{\xi_{\text{dsc}}(\mathbf{h})},$$
$$\mathcal{B}\left[\mathbf{d}(f)\right] = \frac{1}{\xi_{\text{dsc}}\left[\mathbf{h}(f)\right]}.$$

When the weights of the beamformer are chosen in such a way that there is no cancellation, the value of the beampattern is 1 in the direction of the source.

8.7.1 Anechoic Plane Wave Model

Consider the case of a far-field source impinging on the array in an anechoic environment. In that case, the transfer function from the source to each sensor is given by a phase-shift (neglecting any attenuation of the signal which is uniform across the array for a far-field source):

$$\mathbf{g}_{\text{a}}(f) = \left[\, e^{-j2\pi f\tau_1} \; e^{-j2\pi f\tau_2} \; \cdots \; e^{-j2\pi f\tau_N} \,\right]^T, \tag{8.33}$$

where τ_n is the propagation time from the source location to sensor n. The steering vector follows as

$$\begin{aligned}\mathbf{d}_{\text{a}}(f,\boldsymbol{\zeta}) &= \left[\, 1 \; e^{-j2\pi f(\tau_2-\tau_1)} \; \cdots \; e^{-j2\pi f(\tau_N-\tau_1)} \,\right]^T \\ &= \left[\, 1 \; e^{-j2\pi f\tau_{12}} \; \cdots \; e^{-j2\pi f\tau_{1N}} \,\right]^T, \end{aligned} \tag{8.34}$$

where $\tau_{1n} = \tau_n - \tau_1$. Moreover, the steering vector is now parameterized by the direction-of-arrival (DOA):

$$\boldsymbol{\zeta} = \begin{bmatrix} \sin\phi\cos\theta & \sin\phi\sin\theta & \cos\phi \end{bmatrix}^T, \tag{8.35}$$

where ϕ and θ are the incoming wave's elevation and azimuth angles, respectively. The relationship between the relative delays and the plane wave's DOA follows from the solution to the wave equation [1]:

$$\tau_{1n} = \frac{1}{c}\boldsymbol{\zeta}^T \left(\mathbf{r}_n - \mathbf{r}_1\right), \; n = 1, 2, \ldots, N, \tag{8.36}$$

where $\mathbf{r}_n$ is the location of the nth microphone. For a uniform linear array (ULA), $\mathbf{r}_n = \begin{bmatrix} nd & 0 & 0 \end{bmatrix}^T$ where d is the spacing between adjacent microphones; thus, one obtains

$$\tau_{1n} = \frac{(n-1)d}{c}\sin\phi\cos\theta, \; n = 1, 2, \ldots, N, \tag{8.37}$$

where c is the speed of sound propagation.

A conventional delay-and-sum beamformer (DSB) steered to DOA $\boldsymbol{\zeta}_o$ selects its weights according to

$$\mathbf{h}_{\text{dsb}}(f) = \frac{1}{N}\mathbf{d}_{\text{a}}(f, \boldsymbol{\zeta}_o). \tag{8.38}$$

This weighting time-aligns the signal component arriving from DOA $\boldsymbol{\zeta}_o$. As a result, the desired signal is coherently summed, while all other DOAs are incoherently added. The resulting narrowband beampattern of a DSB in an anechoic environment is given by

$$\begin{aligned} \mathcal{B}\left[\mathbf{h}_{\text{dsb}}(f)\right] &= \left| \frac{1}{N}\mathbf{d}_{\text{a}}^H(f, \boldsymbol{\zeta}_o)\mathbf{d}_{\text{a}}(f, \boldsymbol{\zeta}) \right|^2 \\ &= \frac{1}{N^2}\left| \sum_{n=0}^{N-1} e^{j2\pi f \frac{nd}{c}(\cos\theta_o - \cos\theta)} \right|^2 \\ &= \frac{1}{N^2}\left| \frac{1 - e^{j2\pi f N \frac{d}{c}(\cos\theta_o - \cos\theta)}}{1 - e^{j2\pi f \frac{d}{c}(\cos\theta_o - \cos\theta)}} \right|^2, \end{aligned} \tag{8.39}$$

where it has been assumed that $\phi = \phi_o = \frac{\pi}{2}$ (i.e., the source and sensors lie on a plane) for simplicity.

When processing a narrow frequency range centered around frequency f, $\mathcal{B}\left[\mathbf{h}_{\text{dsb}}(f)\right]$ depicts the spatial filtering capabilities of the resulting narrowband beamformer. For a wideband characterization, the broadband beampattern of the DSB is given by

$$\mathcal{B}\left(\mathbf{h}_{\text{dsb}}\right)=\frac{1}{N^2}\frac{\int_{-\infty}^{\infty}\phi_{x_1}(f)\left|\frac{1-e^{j2\pi fN\frac{d}{c}(\cos\theta_o-\cos\theta)}}{1-e^{j2\pi f\frac{d}{c}(\cos\theta_o-\cos\theta)}}\right|^2 df}{\int_{-\infty}^{\infty}\phi_{x_1}(f)df}. \tag{8.40}$$

8.8 Directivity

Acoustic settings frequently have a myriad of noise sources present. In order to model this situation, a spherically isotropic or "diffuse" noise field is one in which the noise power is constant and equal at all spatial frequencies (i.e., directions) [1], [12]. When designing beamformers, one would like to be able to quantify the ability of the beamformer to attenuate such a noise field. To that end, the *directivity* factor is classically defined as the array gain of a (narrowband) beamformer in an isotropic noise field. In the narrowband case, this is equivalent to the ratio of the beampattern in the direction of the source over the resulting residual noise power. Thus, we define

- the *broadband directivity factor*,

$$\mathcal{D}=\frac{\mathcal{B}(\mathbf{d})}{\int_{-\infty}^{\infty}\mathbf{h}^H(f)\mathbf{\Gamma}_{\text{diff}}(f)\mathbf{h}(f)df} \tag{8.41}$$

- and the *narrowband directivity factor*,

$$\mathcal{D}(f)=\frac{\mathcal{B}\left[\mathbf{d}(f)\right]}{\mathbf{h}^H(f)\mathbf{\Gamma}_{\text{diff}}(f)\mathbf{h}(f)}, \tag{8.42}$$

where

$$\begin{aligned}\left[\mathbf{\Gamma}_{\text{diff}}(f)\right]_{nm}&=\frac{\sin 2\pi f(m-n)dc^{-1}}{2\pi f(m-n)dc^{-1}}\\&=\text{sinc}\left[2\pi f(m-n)dc^{-1}\right]\end{aligned} \tag{8.43}$$

is the coherence matrix of a diffuse noise field [13].

The classical directivity index [11], [12] is simply

$$\mathcal{DI}(f)=10\log_{10}\mathcal{D}(f). \tag{8.44}$$

8.8.1 Superdirective Beamforming

As the name suggests, a superdirective beamformer is one which is designed to optimize the beamformer's directivity; to that end, consider a beamformer which maximizes the directivity while constraining the beampattern in the

direction of the source to unity. This leads to the following optimization problem for the beamforming weights $\mathbf{h}(f)$:

$$\mathbf{h}_{\text{sdb}}(f) = \arg\min_{\mathbf{h}(f)} \mathbf{h}^H(f)\mathbf{\Gamma}_{\text{diff}}(f)\mathbf{h}(f) \;\text{ subject to } \mathbf{h}^H(f)\mathbf{d}(f) = 1. \tag{8.45}$$

The solution to (8.45) is written as

$$\mathbf{h}_{\text{sdb}}(f) = \frac{\mathbf{\Gamma}_{\text{diff}}^{-1}(f)\mathbf{d}(f)}{\mathbf{d}^H(f)\mathbf{\Gamma}_{\text{diff}}^{-1}(f)\mathbf{d}(f)}. \tag{8.46}$$

Like the DSB, the superdirective beamformer of (8.46) is a fixed beamformer – that is, it is not data-dependent, as it *assumes* a particular coherence structure for the noise field and requires knowledge of the source DOA.

8.9 White Noise Gain

Notice also that the optimization of (8.45) may be performed for any noise field – it is not limited to a diffuse noise field. In the case of a spatially white noise field, the pseudo-coherence matrix is

$$\mathbf{\Gamma}_v(f) = \mathbf{I}_N, \tag{8.47}$$

and the beamformer that maximizes the so-called *white noise gain* (WNG) is found by substituting (8.47) into (8.46):

$$\begin{aligned}\mathbf{h}_{\text{wng}}(f) &= \frac{\mathbf{d}(f)}{\mathbf{d}^H(f)\mathbf{d}(f)} \\ &= \frac{1}{N}\mathbf{d}(f),\end{aligned} \tag{8.48}$$

which is indeed the DSB.

The narrowband WNG is formally defined as the array gain with a spatially white noise field:

$$\begin{aligned}\mathcal{W}\left[\mathbf{h}(f)\right] &= \frac{\left|\mathbf{h}(f)\mathbf{d}(f)\right|^2}{\mathbf{h}^H(f)\mathbf{h}(f)} \\ &= \frac{\mathcal{B}\left[\mathbf{d}(f)\right]}{\mathbf{h}^H(f)\mathbf{h}(f)}.\end{aligned} \tag{8.49}$$

Analogously, we define the broadband WNG as

$$\mathcal{W}(\mathbf{h}) = \frac{\mathcal{B}(\mathbf{d})}{\int_{-\infty}^{\infty}\mathbf{h}^H(f)\mathbf{h}(f)df}. \tag{8.50}$$

8.10 Spatial Aliasing

The phenomenon of aliasing is classically viewed as an artifact of performing spectral analysis on a sampled signal. Sampling introduces a periodicity into the Fourier transform; if the bandwidth of the signal exceeds half of the sampling frequency, the spectral replicas overlap, leading to a distortion in the observed spectrum.

Spatial aliasing is analogous to its temporal counterpart: in order to reconstruct a spatial sinusoid from a set of uniformly-spaced discrete spatial samples, the spatial sampling period must be less than half of the sinusoid's wavelength. This principle has long been applied to microphone arrays in the following sense: the spacing between adjacent microphone elements should be less than half of the wavelength corresponding to the highest temporal frequency of interest. Since microphone arrays are concerned with the naturally wideband speech (i.e. the highest frequency of interest is in the order of 4 kHz), the resulting arrays are quite small in size.

Notice that the spatial sampling theorem is formulated with respect to a temporally narrowband signal. On the other hand, microphone arrays sample a temporally broadband signal; one may view the wideband nature of sound as diversity. In this section, it is shown that this diversity allows us to increase the microphone array spacing beyond that allowed by the Nyquist theorem without suffering any aliasing artifacts.

Denote the value of the sound field by the four-dimensional function $s(\mathbf{x},t) = s(x,y,z,t)$, where $\mathbf{x} = \left[\, x \;\; y \;\; z \,\right]^T$ denotes the observation point in Cartesian co-ordinates. One may express this function as a multidimensional inverse Fourier transform:

$$s(\mathbf{x},t) = \frac{1}{(2\pi)^4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S(\mathbf{k},\omega) e^{j\left(\omega t - \mathbf{k}^T\mathbf{x}\right)} d\mathbf{k} d\omega, \tag{8.51}$$

where $\omega = 2\pi f$ is the angular temporal frequency,

$$\mathbf{k} = \left[\, k_x \;\; k_y \;\; k_z \,\right]^T$$

is the angular spatial frequency vector, and $S(\mathbf{k},\omega)$ are the coefficients of the basis functions $e^{j\left(\omega t - \mathbf{k}^T\mathbf{x}\right)}$, which are termed *monochromatic plane waves*, as the value of each basis function at a fixed time instant is constant along any plane of the form $\mathbf{k}^T\mathbf{x} = K$, where K is some constant. Thus, any sound field may be represented as a linear combination of propagating narrowband plane waves. It is also interesting to note that according to the physical constraints posed by the wave equation to propagating waves, the following relationship exists among spatial and temporal frequencies [1]:

$$\mathbf{k} = \frac{\omega}{c}\boldsymbol{\zeta}. \tag{8.52}$$

Thus, the spatial frequency vector points in the direction of propagation $\boldsymbol{\zeta}$, while its magnitude is linearly related to the temporal frequency.

The values of the weighting coefficients follow from the multidimensional Fourier transform:

$$S(\mathbf{k},\omega) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} s(\mathbf{x},t)e^{-j\left(\omega t-\mathbf{k}^T\mathbf{x}\right)}d\mathbf{x}dt. \tag{8.53}$$

Notice that each coefficient may be written as

$$\begin{aligned} S(\mathbf{k},\omega) &= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} s(\mathbf{x},t)e^{-j\omega t}dt\right]e^{j\mathbf{k}^T\mathbf{x}}d\mathbf{x} \\ &= \int_{-\infty}^{\infty} S_{\mathbf{x}}(\omega)e^{j\mathbf{k}^T\mathbf{x}}d\mathbf{x}, \end{aligned} \tag{8.54}$$

where

$$S_{\mathbf{x}}(\omega) = \int_{-\infty}^{\infty} s(\mathbf{x},t)e^{-j\omega t}dt$$

is the temporal Fourier transform of the signal observed at position $\mathbf{x}$. Thus, multidimensional spectrum of a space-time field is equal to the spatial Fourier transform of the temporal Fourier coefficients across space. From the duality of the Fourier transform, one may also write

$$S_{\mathbf{x}}(\omega) = \frac{1}{(2\pi)^3}\int_{-\infty}^{\infty} S(\mathbf{k},\omega)e^{-j\mathbf{k}^T\mathbf{x}}d\mathbf{k},$$

which expresses the Fourier transform of the signal observed at an observation point $\mathbf{x}$ as a Fourier integral through a three-dimensional slice of the space-time frequency.

In microphone array applications, the goal is to estimate the temporal signal propagating from a certain DOA; this task may be related to the estimation of the space-time Fourier coefficients $S(\mathbf{k},\omega)$. To see this, consider forming a wideband signal by integrating the space-time Fourier transform across temporal frequency, while only retaining the portion that propagates from the desired DOA $\boldsymbol{\zeta}_o$:

$$s_{\boldsymbol{\zeta}_\mathrm{o}}(t) \triangleq \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\mathbf{k}_\mathrm{o},\omega)e^{j\omega t}d\omega, \tag{8.55}$$

where $\mathbf{k}_\mathrm{o} = \frac{\omega}{c}\boldsymbol{\zeta}_\mathrm{o}$ is the spatial frequency vector whose direction is that of the desired plane wave.

By substituting (8.54) into (8.55), one obtains the following expression for the resulting broadband beam:

$$
\begin{aligned}
s_{\boldsymbol{\zeta}_\mathrm{o}}(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} S_{\mathbf{x}}(\omega)e^{j\mathbf{k}_\mathrm{o}{}^{T}\mathbf{x}}d\mathbf{x}\right]e^{j\omega t}d\omega \\
&= \int_{-\infty}^{\infty}\left[\frac{1}{2\pi}\int_{-\infty}^{\infty} S_{\mathbf{x}}(\omega)e^{j\omega\left(t+\frac{1}{c}\boldsymbol{\zeta}_\mathrm{o}{}^{T}\mathbf{x}\right)}d\omega\right]d\mathbf{x} \\
&= \int_{-\infty}^{\infty} s\left(\mathbf{x}, t+\frac{1}{c}\boldsymbol{\zeta}_\mathrm{o}{}^{T}\mathbf{x}\right)d\mathbf{x}. \qquad (8.56)
\end{aligned}
$$

It is evident from (8.56) that the broadband beam is formed by integrating the time-delayed (or advanced) space-time field across space. This operation is indeed the limiting case of the DSB [1] as the number of microphones tends to infinity.

When the space-time field $s(\mathbf{x},t)$ consists of a plane wave propagating from DOA $\boldsymbol{\zeta}$, one may write

$$
s(\mathbf{x},t) = s\left(t-\frac{1}{c}\boldsymbol{\zeta}^{T}\mathbf{x}\right). \qquad (8.57)
$$

This simplifies the expression for the resulting broadband beam,

$$
\begin{aligned}
s_{\boldsymbol{\zeta}_\mathrm{o}}(t) &= \int_{-\infty}^{\infty} s\left(\mathbf{x}, t+\frac{1}{c}\boldsymbol{\zeta}_\mathrm{o}{}^{T}\mathbf{x}\right)d\mathbf{x} \\
&= \int_{-\infty}^{\infty} s\left[t+\frac{1}{c}\left(\boldsymbol{\zeta}_\mathrm{o}-\boldsymbol{\zeta}\right)^{T}\mathbf{x}\right]d\mathbf{x}. \qquad (8.58)
\end{aligned}
$$

Now that we have an expression for the broadband beam, we can analyze the effect of estimating this signal using a discrete spatial aperture. The general expression for the discrete-space broadband beam follows from (8.56) as

$$
s^{\mathrm{d}}_{\boldsymbol{\zeta}_\mathrm{o}}(t) = \sum_{n=-\infty}^{\infty} s\left(\mathbf{x}_n, t+\frac{1}{c}\boldsymbol{\zeta}_\mathrm{o}^{T}\mathbf{x}_n\right), \qquad (8.59)
$$

where $\mathbf{x}_n$ is the nth spatial sample. When the space-time field consists of a single plane wave, the discrete beam simplifies to

$$
s^{\mathrm{d}}_{\boldsymbol{\zeta}_\mathrm{o}}(t) = \sum_{n=-\infty}^{\infty} s\left[t+\frac{1}{c}\left(\boldsymbol{\zeta}_\mathrm{o}-\boldsymbol{\zeta}\right)^{T}\mathbf{x}_n\right]. \qquad (8.60)
$$

We now delve into analyzing the effect of the spatial sampling frequency on the resulting broadband beams.

8.10.1 Monochromatic Signal

Consider first a narrowband plane wave:

$$s(\mathbf{x}, t) = e^{j\omega\left(t - \frac{1}{c}\boldsymbol{\zeta}^T\mathbf{x}\right)}. \tag{8.61}$$

Substituting (8.61) into (8.60) results in

$$s^{\mathrm{d}}_{\boldsymbol{\zeta}_\mathrm{o}}(t) = A\left(\boldsymbol{\zeta}_\mathrm{o}, \boldsymbol{\zeta}\right) e^{j\omega t}, \tag{8.62}$$

where

$$A\left(\boldsymbol{\zeta}_\mathrm{o}, \boldsymbol{\zeta}\right) = \sum_{n=-\infty}^{\infty} e^{j\frac{\omega}{c}\left(\boldsymbol{\zeta}_\mathrm{o} - \boldsymbol{\zeta}\right)^T \mathbf{x}_n}$$

and

$$s^{\mathrm{d}}_{\boldsymbol{\zeta}_\mathrm{o}}(0) = A\left(\boldsymbol{\zeta}_\mathrm{o}, \boldsymbol{\zeta}\right).$$

Thus, the beam is a complex weighted version of the desired signal $s(t) = e^{j\omega t}$. It is instructive to analyze the values of the complex amplitude $A\left(\boldsymbol{\zeta}_\mathrm{o}, \boldsymbol{\zeta}\right)$ in terms of the spatial sampling rate.

To ease the analysis, assume that the spatial sampling is in the x-direction only and with a sampling period of d:

$$\mathbf{x}_n = \begin{bmatrix} nd & 0 & 0 \end{bmatrix}^T. \tag{8.63}$$

As a result,

$$A\left(\theta_\mathrm{o}, \theta\right) = \sum_{n=-\infty}^{\infty} e^{j\omega\frac{nd}{c}\left(\sin\phi_\mathrm{o}\cos\theta_\mathrm{o} - \sin\phi\cos\theta\right)}. \tag{8.64}$$

Moreover, assume that the source lies on the x-y plane and that we are only concerned with the azimuthal component of the spatial spectrum; that is, $\phi_o = \phi = \frac{\pi}{2}$. In that case

$$A\left(\theta_\mathrm{o}, \theta\right) = \sum_{n=-\infty}^{\infty} e^{j\omega\frac{nd}{c}\left(\cos\theta_\mathrm{o} - \cos\theta\right)}. \tag{8.65}$$

Recall the following property of an infinite summation of complex exponentials:

$$\sum_{n=-\infty}^{\infty} e^{j\omega nT} = \frac{1}{T}\sum_{k=-\infty}^{\infty}\delta\left(\frac{\omega}{2\pi}-\frac{k}{T}\right)$$
$$= \frac{2\pi}{T}\sum_{k=-\infty}^{\infty}\delta\left(\omega-\frac{2\pi k}{T}\right), \tag{8.66}$$

where $\delta(\cdot)$ is the delta-Dirac function. Substituting (8.66) into (8.65),

$$A(\theta_{\mathrm{o}},\theta) = \frac{2\pi}{\frac{d}{c}(\cos\theta_{\mathrm{o}}-\cos\theta)}\sum_{k=-\infty}^{\infty}\delta\left[\omega-\frac{2\pi k}{\frac{d}{c}(\cos\theta_{\mathrm{o}}-\cos\theta)}\right]. \tag{8.67}$$

Consider now the conditions for the argument of the delta Dirac function in (8.67) to equal zero. This requires

$$\omega = \frac{2\pi k}{\frac{d}{c}(\cos\theta_{\mathrm{o}}-\cos\theta)}, \tag{8.68}$$

which is equivalent to

$$d(\cos\theta_{\mathrm{o}}-\cos\theta) = k\lambda, \tag{8.69}$$

where $\lambda = 2\pi\frac{c}{\omega}$ is the wavelength. For $k=0$, (8.69) holds if $\cos\theta_{\mathrm{o}} = \cos\theta$, meaning that

$$s_{\theta_{\mathrm{o}}}^{\mathrm{d}}(t) = A(\theta_{\mathrm{o}},\theta)\,e^{j\omega t}$$
$$= \infty \quad \text{for } \cos\theta_{\mathrm{o}} = \cos\theta, \tag{8.70}$$

which is the desired result; indeed, this is the true (i.e., non-aliased) spectral peak. Note that for $\cos\theta = \cos\theta_o$, the factor $A(\theta_{\mathrm{o}},\theta)$ is infinite since the analysis assumes that the number of microphones $N \to \infty$.

Given the result of (8.70), a rigorous definition of spatial aliasing in broadband applications may be proposed: aliasing occurs whenever

$$\exists\ \theta_{\mathrm{o}} \neq \theta \ \text{ such that } \ s_{\theta_{\mathrm{o}}}^{\mathrm{d}}(t) = \infty. \tag{8.71}$$

In other words, spatial aliasing occurs when the discrete-space broadband beam $s_{\boldsymbol{\zeta}_{\mathrm{o}}}^{\mathrm{d}}(t)$ tends to infinity even though the steered DOA θ_{o} does not match the true DOA θ.

The steered range for a ULA is $0 \leq \theta \leq \pi$ and the cosine function is one-to-one over this interval. Thus

$$\cos\theta_{\mathrm{o}} = \cos\theta \Rightarrow \theta_{\mathrm{o}} = \theta, \quad 0 \leq \theta_{\mathrm{o}}, \theta \leq \pi. \tag{8.72}$$

It is now straightforward to determine the aliasing conditions. Under the assumption of a narrowband signal, the beam $s_{\theta_{\mathrm{o}}}^{\mathrm{d}}(t)$ tends to infinity if there

exists an integer $k \in \mathbb{Z}$ such that

$$\omega = \frac{2\pi k}{\frac{d}{c}(\cos\theta_{\mathrm{o}} - \cos\theta)}, \tag{8.73}$$

or

$$\frac{d}{\lambda} = \frac{k}{\cos\theta_{\mathrm{o}} - \cos\theta}. \tag{8.74}$$

Take $k = 1$; over the range $0 \leq \theta_{\mathrm{l}} \leq 2\pi$,

$$|\cos\theta_{\mathrm{l}} - \cos\theta_0| \leq 2, \tag{8.75}$$

meaning that to prevent aliasing, one needs to ensure that

$$\frac{d}{\lambda} < \frac{1}{\cos\theta_{\mathrm{o}} - \cos\theta}, \tag{8.76}$$

or

$$d < \frac{\lambda}{2}, \tag{8.77}$$

which is indeed the classical narrowband aliasing criterion. Note that for $|k| > 1$, the condition (8.77) also prevents (8.73).

8.10.2 Broadband Signal

Consider now a wideband signal with arbitrary temporal frequency content $S(\omega)$:

$$s(t) = \int_{-\infty}^{\infty} S(\omega) e^{j\omega t} d\omega. \tag{8.78}$$

Assuming a one-dimensional sampling scheme and considering only spatial frequencies with $\phi_{\mathrm{o}} = \phi = \frac{\pi}{2}$, the continuous version of the broadband beam corresponding to this signal follows from (8.58) as

$$s_{\theta_{\mathrm{o}}}(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S(\omega) e^{j\omega\left[t + \frac{x}{c}(\cos\theta_{\mathrm{o}} - \cos\theta)\right]} d\omega dx. \tag{8.79}$$

By setting $x = nd$ and replacing the integral with an infinite summation, we obtain the discrete version of (8.79):

$$
\begin{aligned}
s_{\theta_\mathrm{o}}^{\mathrm{d}}(t) &= \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} S(\omega) e^{j\omega\left[t+\frac{nd}{c}(\cos\theta_\mathrm{l}-\cos\theta)\right]} d\omega \qquad (8.80)\\
&= \int_{-\infty}^{\infty} S(\omega) e^{j\omega t} \left[\sum_{n=-\infty}^{\infty} e^{j\omega\frac{nd}{c}(\cos\theta_\mathrm{o}-\cos\theta)}\right] d\omega\\
&= \int_{-\infty}^{\infty} S(\omega) e^{j\omega t} \frac{2\pi}{\frac{d}{c}(\cos\theta_\mathrm{o}-\cos\theta)} \sum_{k=-\infty}^{\infty} \delta\left[\omega - \frac{2\pi k}{\frac{d}{c}(\cos\theta_\mathrm{o}-\cos\theta)}\right] d\omega\\
&= \frac{2\pi}{\frac{d}{c}(\cos\theta_\mathrm{o}-\cos\theta)} \sum_{k=-\infty}^{\infty} S\left[\frac{2\pi k}{\frac{d}{c}(\cos\theta_\mathrm{o}-\cos\theta)}\right] e^{j\frac{2\pi k}{\frac{d}{c}(\cos\theta_\mathrm{o}-\cos\theta)}t}.
\end{aligned}
$$

Examining (8.80), it follows that the discrete-space broadband beam for an arbitrary wideband signal takes the form of a series of weighted complex exponentials. For any temporal signal which obeys

$$
\sum_{k=-\infty}^{\infty} S\left[\frac{2\pi k}{\frac{d}{c}(\cos\theta_\mathrm{o}-\cos\theta)}\right] e^{j\frac{2\pi k}{\frac{d}{c}(\cos\theta_\mathrm{o}-\cos\theta)}t} < \infty, \quad \cos\theta_\mathrm{o} \neq \cos\theta, \qquad (8.81)
$$

one can state that the beam exhibits an infinite peak only when the scaling factor

$$
\frac{2\pi}{\frac{d}{c}(\cos\theta_\mathrm{o}-\cos\theta)} = \infty, \qquad (8.82)
$$

which implies $\theta_\mathrm{o} = \theta$. Thus, for wideband signals with spectra of the form (8.81), under the definition of (8.71), spatial aliasing does not result, regardless of the spatial sampling period d.

The condition of (8.81) refers to signals which are band-limited and not dominated by a strong harmonic component. The presence of such harmonic components at integer multiples of $\frac{2\pi}{\frac{d}{c}(\cos\theta_\mathrm{o}-\cos\theta)}$ may drive the broadband beam to infinity at DOAs not matching the true DOA.

8.11 Mean-Squared Error

The ultimate aim of a beamformer is to reproduce the desired signal, free of any noise or interference, in the array output. To that end, the *mean-squared error (MSE)* is a key measure for designing optimal beamforming algorithms.

Let us first write the time-domain (broadband) error signal between the beamformer output signal and the desired signal, i.e.,

$$\begin{aligned} e(t) &= z(t) - x_1(t) \\ &= x_{1,\mathrm{f}}(t) - x_1(t) + v_{\mathrm{rn}}(t) \\ &= \int_{-\infty}^{\infty} \left[\mathbf{h}^H(f)\mathbf{d}(f) - 1\right] X_1(f) e^{j2\pi ft} df + \int_{-\infty}^{\infty} \mathbf{h}^H(f)\mathbf{v}(f) e^{j2\pi ft} df \\ &= e_{x_1}(t) + e_v(t), \end{aligned} \tag{8.83}$$

where

$$\begin{aligned} e_{x_1}(t) &= \int_{-\infty}^{\infty} \left[\mathbf{h}^H(f)\mathbf{d}(f) - 1\right] X_1(f) e^{j2\pi ft} df \\ &= \int_{-\infty}^{\infty} \mathcal{E}_{x_1}(f) e^{j2\pi ft} df, \end{aligned} \tag{8.84}$$

is the desired signal distortion, and

$$\begin{aligned} e_v(t) &= \int_{-\infty}^{\infty} \mathbf{h}^H(f)\mathbf{v}(f) e^{j2\pi ft} df \\ &= \int_{-\infty}^{\infty} \mathcal{E}_v(f) e^{j2\pi ft} df \end{aligned} \tag{8.85}$$

represents the broadband residual noise.

The variance of the time-domain error signal is the *broadband MSE*:

$$\begin{aligned} J(\mathbf{h}) &= E\left[e^2(t)\right] \\ &= E\left[e_{x_1}^2(t)\right] + E\left[e_v^2(t)\right] \\ &= \int_{-\infty}^{\infty} \phi_{x_1}(f) \left|\mathbf{h}^H(f)\mathbf{d}(f) - 1\right|^2 df + \int_{-\infty}^{\infty} \mathbf{h}^H(f)\mathbf{\Phi}_v(f)\mathbf{h}(f) df. \end{aligned} \tag{8.86}$$

For the particular filter $\mathbf{h}(f) = \mathbf{i},\ \forall\ f$, where

$$\mathbf{i} = \left[1\ 0 \ldots 0\right]^T, \tag{8.87}$$

we obtain

$$J(\mathbf{i}) = \int_{-\infty}^{\infty} \phi_{v_1}(f) df. \tag{8.88}$$

Therefore, we define the broadband normalized MSE (NMSE) as

$$\tilde{J}(\mathbf{h}) = \frac{J(\mathbf{h})}{J(\mathbf{i})}, \tag{8.89}$$

which can be rewritten as

$$\tilde{J}(\mathbf{h}) = \mathrm{iSNR} \cdot \upsilon_{\mathrm{dsd}}(\mathbf{h}) + \frac{1}{\xi_{\mathrm{nr}}(\mathbf{h})}, \tag{8.90}$$

where

$$\upsilon_{\text{dsd}}(\mathbf{h}) = \frac{\int_{-\infty}^{\infty} \phi_{x_1}(f) \left|\mathbf{h}^H(f)\mathbf{d}(f) - 1\right|^2 df}{\int_{-\infty}^{\infty} \phi_{x_1}(f) df}, \tag{8.91}$$

is the *broadband desired-signal-distortion index* [14].

From the broadband MSE we can deduce the *narrowband MSE*

$$\begin{aligned} J\left[\mathbf{h}(f)\right] &= E\left[\left|\mathcal{E}_{x_1}(f)\right|^2\right] + E\left[\left|\mathcal{E}_v(f)\right|^2\right] \\ &= \phi_{x_1}(f)\left|\mathbf{h}^H(f)\mathbf{d}(f) - 1\right|^2 + \mathbf{h}^H(f)\mathbf{\Phi}_v(f)\mathbf{h}(f). \end{aligned} \tag{8.92}$$

We can also deduce the narrowband NMSE:

$$\tilde{J}\left[\mathbf{h}(f)\right] = \text{iSNR}\left[\mathbf{h}(f)\right] \cdot \upsilon_{\text{dsd}}\left[\mathbf{h}(f)\right] + \frac{1}{\xi_{\text{nr}}\left[\mathbf{h}(f)\right]}, \tag{8.93}$$

where

$$\upsilon_{\text{dsd}}\left[\mathbf{h}(f)\right] = \left|\mathbf{h}^H(f)\mathbf{d}(f) - 1\right|^2, \tag{8.94}$$

is the *narrowband desired-signal-distortion index* [14].

Note that the broadband MSE is a linear combination of the underlying narrowband MSEs:

$$J(\mathbf{h}) = \int_{-\infty}^{\infty} J\left[\mathbf{h}(f)\right] df. \tag{8.95}$$

8.11.1 Wiener Filter

Intuitively, we would like to derive a beamformer which minimizes the MSE at every frequency. This is the essence of the multichannel Wiener filter. The conventional minimum MSE (MMSE) minimizes the narrowband MSE:

$$\mathbf{h}_{\text{W}}(f) = \arg\max_{\mathbf{h}(f)} J\left[\mathbf{h}(f)\right]. \tag{8.96}$$

Taking the gradient of $J\left[\mathbf{h}(f)\right]$ with respect to $\mathbf{h}^H(f)$ results in

$$\nabla_{\mathbf{h}^H(f)} J\left[\mathbf{h}(f)\right] = \phi_{x_1}(f)\left[\mathbf{d}(f)\mathbf{d}^H(f)\mathbf{h}(f) - \mathbf{d}(f)\right] + \mathbf{\Phi}_v(f)\mathbf{h}(f). \tag{8.97}$$

Setting (8.97) to zero and solving for $\mathbf{h}(f)$ reveals the conventional (narrowband) MMSE solution:

$$\mathbf{h}_{\text{W}}(f) = \phi_{x_1}(f)\mathbf{\Phi}_y^{-1}(f)\mathbf{d}(f), \tag{8.98}$$

where

$$\begin{aligned}\mathbf{\Phi}_y(f) &= E\left[\mathbf{y}(f)\mathbf{y}^H(f)\right] \\ &= \phi_{x_1}(f)\mathbf{d}(f)\mathbf{d}^H(f) + \mathbf{\Phi}_v(f)\end{aligned} \tag{8.99}$$

is the PSD matrix of the array measurements. The filter $\mathbf{h}_{\mathrm{W}}(f)$ minimizes the difference between array output and desired signal at the single frequency f.

Since the broadband MSE is a linear combination of the narrowband MSEs, minimizing the MSE at every frequency guarantees the minimization of the broadband MSE. Thus, applying the narrowband solution $\mathbf{h}_{\mathrm{W}}(f)$ at every component frequency results in the broadband MMSE solution.

8.11.2 Minimum Variance Distortionless Response

The celebrated minimum variance distortionless response (MVDR) beamformer proposed by Capon [3], [15] is also easily derived from the narrowband MSE. Indeed, minimizing $E\left[\left|\mathcal{E}_v(f)\right|^2\right]$ with the constraint that $E\left[\left|\mathcal{E}_{x_1}(f)\right|^2\right] = 0$ [or $\mathbf{h}^H(f)\mathbf{d}(f) - 1 = 0$], we obtain the classical MVDR filter:

$$\mathbf{h}_{\mathrm{MVDR}}(f) = \frac{\mathbf{\Phi}_v^{-1}(f)\mathbf{d}(f)}{\mathbf{d}^H(f)\mathbf{\Phi}_v^{-1}(f)\mathbf{d}(f)}. \tag{8.100}$$

Using the fact that $\mathbf{\Phi}_x(f) = E\left[\mathbf{x}(f)\mathbf{x}^H(f)\right] = \phi_{x_1}(f)\mathbf{d}(f)\mathbf{d}^H(f)$, the explicit dependence of the above filter on the steering vector is eliminated to obtain the following forms [5]:

$$\begin{aligned}\mathbf{h}_{\mathrm{MVDR}}(f) &= \frac{\mathbf{\Phi}_v^{-1}(f)\mathbf{\Phi}_x(f)}{\mathrm{tr}\left[\mathbf{\Phi}_v^{-1}(f)\mathbf{\Phi}_x(f)\right]}\mathbf{i} \\ &= \frac{\mathbf{\Phi}_v^{-1}(f)\mathbf{\Phi}_y(f) - \mathbf{I}_N}{\mathrm{tr}\left[\mathbf{\Phi}_v^{-1}(f)\mathbf{\Phi}_y(f)\right] - N}\mathbf{i},\end{aligned} \tag{8.101}$$

where tr[·] denotes the trace of a square matrix.

The MVDR beamformer rejects the maximum level of noise allowable without distorting the desired signal at each frequency; however, the level broadband noise rejection is unclear. On the other hand, since the constraint is verified at all frequencies, the MVDR filter guarantees zero desired signal distortion at every frequency.

8.12 Conclusions

This chapter has reformulated the objectives of microphone arrays taking into account the wideband nature of the speech signal and the reverberant properties of acoustic environments. The SNR, array gain, noise-reduction factor, desired-signal-cancellation factor, beampattern, directivity factor, WNG, and MSE were defined in both narrowband and broadband contexts. Additionally, an analysis of spatial aliasing with broadband signals revealed that the spatial Nyquist criterion may be relaxed in microphone array applications. To this point in time, microphone array designs have been mainly focused on optimizing narrowband measures at each frequency bin. The broadband criteria presented in this chapter will hopefully serve as the metrics which future beamformer designs will focus on.

References

1. D. H. Johnson and D. E. Dudgeon, *Array Signal Processing–Concepts and Techniques*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
2. M. Brandstein and D. B. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*. Berlin, Germany: Springer-Verlag, 2001.
3. J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.
4. J. Benesty, J. Chen, Y. Huang, and J. Dmochowski, "On microphone-array beamforming from a MIMO acoustic signal processing perspective," *IEEE Trans. Audio, Speech, Language Processing*, vol. 15, pp. 1053–1065, Mar. 2007.
5. J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Berlin, Germany: Springer-Verlag, 2008.
6. S. Gannot and I. Cohen, "Adaptive beamforming and postfiltering," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds., Berlin, Germany: Springer-Verlag, 2008, Chapter 47, pp. 945–978.
7. B. D. Van Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE Acoust., Speech, Signal Process. Mag.*, vol. 5, pp. 4–24, Apr. 1988.
8. W. Herbordt and W. Kellermann, "Adaptive beamforming for audio signal acquisition," in *Adaptive Signal Processing: Applications to Real-World Problems*, J. Benesty and Y. Huang, Eds., Berlin, Germany: Springer-Verlag, 2003, Chapter 6, pp. 155–194.
9. J. Benesty, J. Chen, Y. Huang, and S. Doclo, "Study of the Wiener filter for noise reduction," in *Speech Enhancement*, J. Benesty, S. Makino, and J. Chen, Eds., Berlin, Germany: Springer-Verlag, 2005, Chapter 2, pp. 9–41.
10. J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, pp. 1218–1234, July 2006.
11. W. Herbordt, *Combination of Robust Adaptive Beamforming with Acoustic Echo Cancellation for Acoustic Human/Machine Interfaces*. PhD Thesis, Erlangen-Nuremberg University, Germany, 2004.
12. G. W. Elko and J. Meyer, "Microphone arrays," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds., Berlin, Germany: Springer-Verlag, 2008, Chapter 48, pp. 1021–1041.

13. A. Spriet, *Adaptive Filtering Techniques for Noise Reduction and Acoustic Feedback Cancellation in Hearing Aids*. PhD Thesis, Katholieke Universiteit Leuven, Belgium, 2004.
14. J. Benesty, J. Chen, Y. Huang, and I. Cohen, *Noise Reduction in Speech Processing*. Berlin, Germany: Springer-Verlag, 2009.
15. R. T. Lacoss, "Data adaptive spectral analysis methods," *Geophysics*, vol. 36, pp. 661–675, Aug. 1971.