Microphone Arrays : A Tutorial

Iain McCowan

April 2001

Abstract

This report presents a tutorial of fundamental array processing and beamforming theory relevant to microphone array speech processing. A microphone array consists of multiple microphones placed at different spatial locations. Built upon a knowledge of sound propagation principles, the multiple inputs can be manipulated to enhance or attenuate signals emanating from particular directions. In this way, microphone arrays provide a means of enhancing a desired signal in the presence of corrupting noise sources. Moreover, this enhancement is based purely on knowledge of the source location, and so microphone array techniques are applicable to a wide variety of noise types. Microphone arrays have great potential in practical applications of speech processing, due to their ability to provide both noise robustness and hands-free signal acquisition.

This report has been extracted from my PhD thesis, and can be referenced as :
I..A. McCowan. "Robust Speech Recognition using Microphone Arrays," PhD Thesis, Queensland University of Technology, Australia, 2001.

For a more in-depth discussion of key microphone processing techniques, the interested reader is refered to
M. Brandstein and D. Ward (Eds). "Microphone Arrays", Springer, 2001.

1 Array Processing Fundamentals

1.1 Introduction

Array processing involves the use of multiple sensors to receive or transmit a signal carried by propagating waves. Sensor arrays have application in a diversity of fields, such as sonar, radar, seismology, radio astronomy and tomography [1]. The focus of this article, is the use of microphone arrays to receive acoustic signals, or more specifically, speech signals. While the use of sensor arrays for speech processing is a relatively new area of research, the fundamental theory is well established as it is common to all sensor arrays, being based on the theory of wave propagation.

In general, sensor arrays can be considered as sampled versions of continuous apertures, and the principles governing their operation is best understood in this context. With this in mind, this section seeks to develop the principles of array processing by discussing the key areas of

- wave propagation,
- continuous apertures, and
- discrete sensor arrays.

While retaining a certain generality in its discussion of sensor arrays, the section is restricted in scope to the principles required to understand linear microphone arrays.

1.2 Wave Propagation

Sound waves propagate through fluids as longitudinal waves. The molecules in the fluid move back and forth in the direction of propagation, producing regions of compression and expansion. By using Newton's equations of motion to consider an infinitesimal volume of the fluid, an equation governing the wave's propagation can be developed. A generalised *wave equation* for acoustic waves is quite complex as it depends upon properties of the fluid, however, assuming an ideal fluid with zero viscosity, the wave equation can be derived as [2]

$$\nabla^2 x(t, \mathbf{r}) - \frac{1}{c^2} \frac{\delta^2}{\delta t^2} x(t, \mathbf{r}) = 0 \tag{1}$$

where $x(t, \mathbf{r})$ is a function representing the sound pressure at a point in time and space,

$$\mathbf{r} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{2}$$

and ∇^2 is the Laplacian operator. The speed of propagation, c, depends upon the pressure and density of the fluid, and is approximately $330 ms^{-1}$ in air. The wave equation of Equation 1 is known as the governing equation for a wide range of propagating waves, including electromagnetic waves.

The solution to the differential wave equation can be derived using the method of separation of variables. The solution is well known and for a monochromatic plane wave is given as [2]

$$x(t, \mathbf{r}) = A e^{j(\omega t - \mathbf{k} \cdot \mathbf{r})} \tag{3}$$

where A is the wave amplitude, $\omega = 2\pi f$ is the frequency in radians per second, and the *wavenumber vector* k indicates the speed and direction of wave propagation and is given by

$$\mathbf{k} = \frac{2\pi}{\lambda} \begin{bmatrix} \sin\theta \cos\phi & \sin\theta \sin\phi & \cos\theta \end{bmatrix} \tag{4}$$

where the wavelength λ is related to c by the simple relation $\lambda = c/f$. Alternately, the solution for a spherical wave can be derived as [2]

$$x(t, \mathbf{r}) = -\frac{A}{4\pi r} e^{j(\omega t - kr)} \tag{5}$$

where $r = |\mathbf{r}|$ is the radial distance from the source, and k is the scalar *wavenumber*, given by $2\pi/\lambda$. The spherical wave solution shows that the signal amplitude decays at a rate proportional to the distance from the source. This dependence of the amplitude on the distance has important implications for array processing algorithms when the source is in the near-field, as will be discussed in later sections. While sound waves are typically spherical in nature, they may be considered as plane waves at a sufficient distance from the source, and this approximation is often used to simplify mathematical analysis.

The plane wave solution in Equation 3 is expressed in terms of two variables, time and space. Due to the well defined propagation of the signal, these two variables are linked by a simple relation, and thus the solution can be expressed as function of a single variable. If we formulate the plane wave solution as

$$x(t, \mathbf{r}) = A e^{j\omega(t - \boldsymbol{\beta} \cdot \mathbf{r})} \tag{6}$$

where $\boldsymbol{\beta} = \frac{\mathbf{k}}{\omega}$, and we define a new variable u such that $u = t - \boldsymbol{\beta} \cdot \mathbf{r}$, then the solution can be expressed as

$$x(u) = A e^{j\omega u} \tag{7}$$

For spherical waves, with the substitution $u = t - r/c$, we have the similar expression

$$x(u) = -\frac{A}{4\pi r} e^{j\omega u} \tag{8}$$

Due to the linearity of the wave equation, the monochromatic solution can be expanded to the more general polychromatic case by considering the solution as a sum or integral of such complex exponentials. Fourier theory tells us that any function with a convergent Fourier integral can be expressed as a weighted superposition of complex exponentials. From this we can make the powerful conclusion that any signal with a valid Fourier transform, irrespective of its shape, satisfies the wave equation.

In this section, we have seen that propagating acoustic signals can be expressed as functions of a single variable, with time and space linked by a simple relation. In addition, the information in the signal is preserved as it propagates. These two conclusions imply that, for a band-limited signal, we can reconstruct the signal over all space and time by either

- *temporally* sampling the signal at a given *location in space*, or
- *spatially* sampling the signal at a given *instant of time*.

The latter implication is the basis for all aperture and sensor array signal processing techniques. Other implications from the above wave propagation analysis that are important for array processing applications are [3]

- The speed of propagation depends on the properties of the medium, and thus is constant for a given wave type and medium. For the specific case of acoustic waves in air, the speed of propagation is approximately $330ms^{-1}$.
- In general, waves propagate from their source as spherical waves, with the amplitude decaying at a rate proportional to the distance from the source.
- The superposition principle applies to propagating wave signals, allowing multiple waves to occur without interaction. To separate these signals, algorithms must be developed to distinguish the different signals based upon knowledge of their temporal and spatial characteristics.

The above discussion has retained the simplicity of assuming a homogeneous, lossless medium, and neglecting effects such as dispersion, diffraction, and changes in propagation speed. A thorough analysis of acoustic field theory can be found in Ziomek [2].

1.3 Continuous Apertures

The term *aperture* is used to refer to a spatial region that transmits or receives propagating waves. A transmitting aperture is referred to as an *active aperture*, while a receiving aperture is known as a *passive aperture*. For example, in optics, an aperture may be a hole in an opaque screen, and in electromagnetics it may be an electromagnetic antenna. In acoustics, an aperture is an electroacoustic transducer that converts acoustic signals into electrical signals (microphone), or vice-versa (loudspeaker).

1.3.1 Aperture Function

Consider a general receiving aperture of volume V where a signal $x(t, \mathbf{r})$ is received at time t and spatial location $\mathbf{r}$. Treating the infinitesimal volume dV at $\mathbf{r}$ as a linear filter having impulse response $a(t, \mathbf{r})$, the received signal is given by the convolution [2]

$$x_R(t, \mathbf{r}) = \int_{-\infty}^{\infty} x(\tau, \mathbf{r}) a(t - \tau, \mathbf{r}) d\tau \tag{9}$$

or, by taking the Fourier transform,

$$X_R(f, \mathbf{r}) = X(f, \mathbf{r}) A(f, \mathbf{r}) \tag{10}$$

The term $A(f, \mathbf{r})$ is known as the *aperture function* or the *sensitivity function*, and it defines the response as a function of spatial position along the aperture.

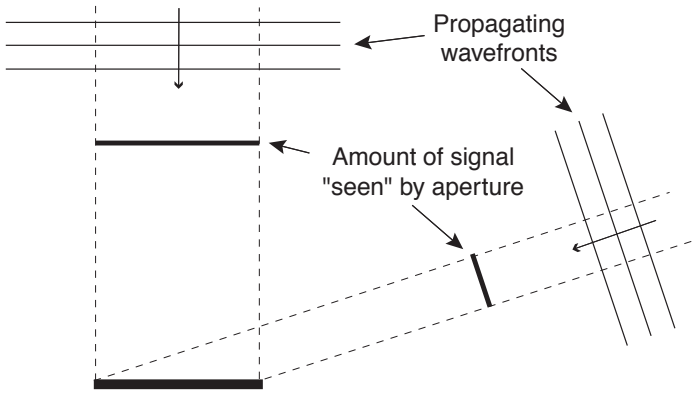


Figure 1: Signal received by a linear aperture

1.3.2 Directivity Pattern

The response of a receiving aperture is inherently directional in nature, because the amount of signal seen by the aperture varies with the direction of arrival. This principle is illustrated in Figure 1 (reproduced from Moore [4]) for the simple case of planar waves being received by a linear aperture.

The aperture response as a function of frequency and direction of arrival is known as the aperture *directivity pattern* or *beam pattern*. By manipulating the exact solution to the wave equation discussed in Section 1.2, the directivity pattern can be shown to be related to the aperture function by a Fourier transform relationship [2]. The far-field directivity pattern of a receiving aperture, having aperture function A_R, is given as

$$\begin{aligned} D_R(f, \boldsymbol{\alpha}) &= \mathcal{F}_{\mathbf{r}}\{A_R(f, \mathbf{r})\} \\ &= \int_{-\infty}^{\infty} A_R(f, \mathbf{r}) e^{j2\pi \boldsymbol{\alpha} \cdot \mathbf{r}} d\mathbf{r} \end{aligned} \tag{11}$$

where $\mathcal{F}_{\mathbf{r}}\{\cdot\}$ denotes the three dimensional Fourier transform,

$$\mathbf{r} = \begin{bmatrix} x_a \\ y_a \\ z_a \end{bmatrix} \tag{12}$$

is the spatial location of a point along the aperture, and

$$\begin{aligned} \boldsymbol{\alpha} &= f\beta \\ &= \frac{1}{\lambda} \begin{bmatrix} \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \end{bmatrix} \end{aligned} \tag{13}$$

is the direction vector of the wave, where the angles θ and ϕ are as shown in Figure 2. Note that the frequency dependence in the above equations is implicit in the wavelength term as $\lambda = c/f$.

1.3.3 Linear Apertures

In order to investigate some properties of the aperture directivity pattern, it is useful to simplify the above equation by considering a linear aperture of length L along the x-axis, as shown in Figure 3. In

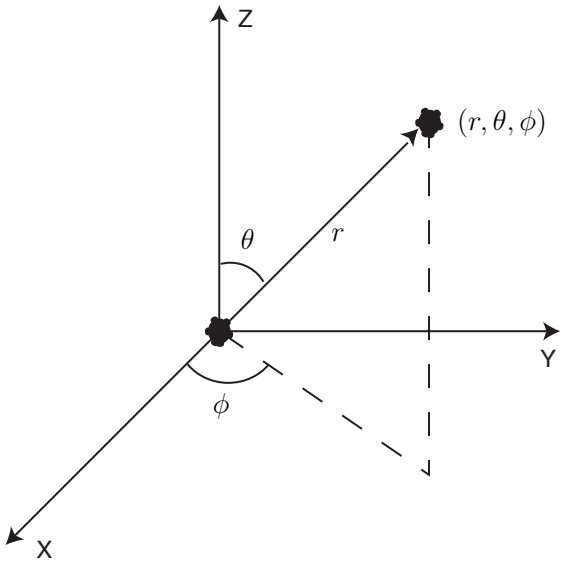


Figure 2: Spherical coordinate system

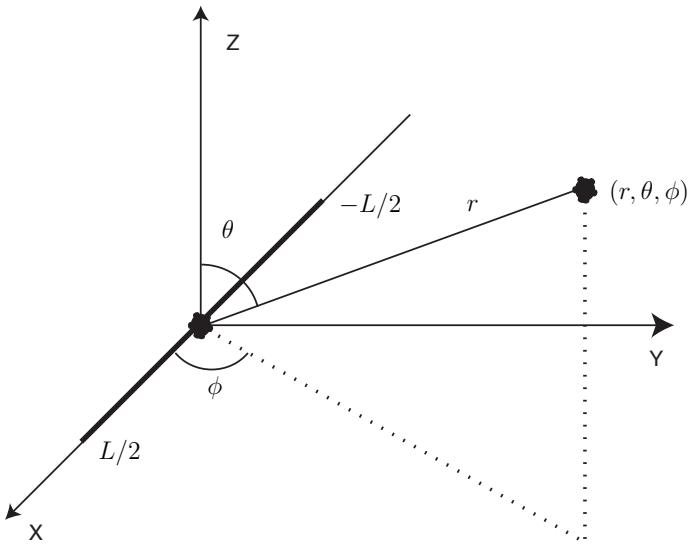


Figure 3: Continuous linear aperture

this case

$$\mathbf{r} = \begin{bmatrix} \mathbf{x_a} \\ 0 \\ 0 \end{bmatrix} \tag{14}$$

and the directivity pattern simplifies to

$$D_R(f, \alpha_x) = \int_{-L/2}^{L/2} A_R(f, x_a) e^{j2\pi\alpha_x x_a} dx_a \tag{15}$$

where

$$\alpha_x = \frac{\sin\theta\cos\phi}{\lambda} \tag{16}$$

if we write the equation as a function of angles θ and ϕ we obtain

$$D_R(f, \theta, \phi) = \int_{-L/2}^{L/2} A_R(f, x_a) e^{j\frac{2\pi}{\lambda}\sin\theta\cos\phi x_a} dx_a \tag{17}$$

The above expressions have been developed for plane waves and thus are only valid for the case of *far-field* sources. For a linear aperture, a wave source may be considered to come from the far-field of the aperture if [5]

$$|r| > \frac{2L^2}{\lambda} \tag{18}$$

For now the far-field assumption serves to simplify the discussion of aperture properties. The details of the more precise case of near-field sources will be considered later when discussing discrete linear sensor arrays.

Consider the case of a linear aperture with uniform, frequency-independent aperture function. The aperture function may be written as

$$A_R(x_a) = \text{rect}(x_a/L) \tag{19}$$

where

$$\text{rect}(x/L) \widehat{=} \begin{cases} 1 & |x| \leq L/2 \\ 0 & |x| > L/2 \end{cases} \tag{20}$$

The resulting directivity pattern is given by

$$D_R(f, \alpha_x) = \mathcal{F}\{\text{rect}(x_a/L)\} \tag{21}$$

which has the well known solution

$$D_R(f, \alpha_x) = L\,\text{sinc}(\alpha_x L) \tag{22}$$

where

$$\text{sinc}(x) \widehat{=} \frac{\sin(x)}{x} \tag{23}$$

Plots of the uniform aperture function and corresponding directivity pattern are shown in Figure 4. From the plot we see that zeros in the directivity pattern are located at $\alpha_x = m\lambda/L$, where m is an integer. The area of the directivity pattern in the range $-\lambda/L \leq \alpha_x \leq \lambda/L$ is referred to as the *main lobe* and its extent is termed the *beam width*. Thus we see that the beam width of a linear aperture is given by $2\lambda/L$, or in terms of frequency $2c/fL$. We note the important behaviour that the beam width is inversely proportional to the product fL, and so for a fixed aperture length, the beam width will decrease with increasing frequency.

It is often useful to consider the *normalised directivity pattern* of an aperture, as this serves to highlight the relative differences in array response over varying angles of arrival. As the *sinc* function is bounded

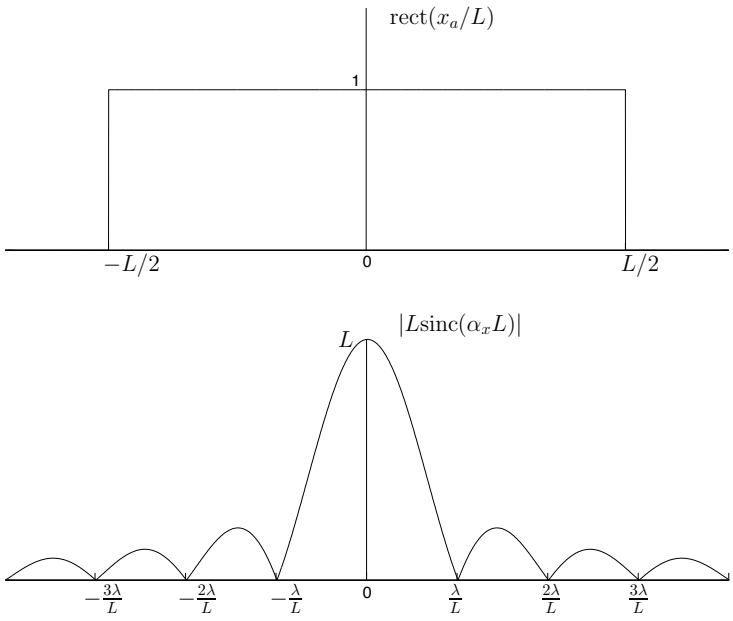


Figure 4: Uniform aperture function and directivity pattern

by $-1 \leq \text{sinc}(x) \leq 1$, the maximum possible value of the directivity pattern is $D_{\max} = L$, and the normalised directivity pattern is given as

$$D_N(f, \alpha_x) = \frac{D_R(f, \alpha_x)}{D_{\max}} = \text{sinc}(\alpha_x L) \tag{24}$$

or in terms of the angles θ and ϕ

$$D_N(f, \theta, \phi) = \text{sinc}(\frac{L}{\lambda} \sin\theta \cos\phi) \tag{25}$$

A common tool for examining the properties of the aperture response is a polar plot of the horizontal directivity pattern over angle ϕ, given by

$$D_N(f, \frac{\pi}{2}, \phi) = \text{sinc}(\frac{L}{\lambda} \cos\phi) \tag{26}$$

Polar plots of the horizontal directivity pattern are shown in Figure 5 for different values of L/λ, demonstrating the beam width's dependence on this ratio as discussed previously.

Although the directivity pattern given by Equation 22 can theoretically be evaluated for any value of α_x, because $\alpha_x = \sin\theta\cos\phi$, it is practically bounded by $-1 \leq \alpha_x \leq 1$. This interval is referred to as the *visible region* of the aperture. To examine the physical significance of key values of α_x we consider the horizontal directivity pattern, for which $\theta = \frac{\pi}{2}$. First, we see that $\alpha_x = 0$ implies that $\phi = \frac{\pi}{2}$ or $\phi = \frac{3\pi}{2}$, corresponding to a source that is situated perpendicular to the aperture axis, referred to as a *broadside* source. Conversely, $\alpha_x = \pm 1$ implies that $\phi = 0$ or $\phi = \pi$, corresponding to a source on the same axis as the aperture, termed an *endfire* source.

1.4 Discrete Sensor Arrays

A sensor array can be considered to be a sampled version of a continuous aperture, where the aperture is only excited at a finite number of discrete points. As each element can itself be considered as a continuous aperture, the overall response of the array can be determined as the superposition of each individual sensor response. This superposition of sensor responses results in an array response that approximates the equivalent (sampled) continuous aperture.

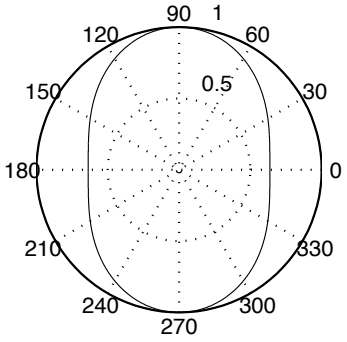
1.4.1 Linear Sensor Array

We consider the particular case of a linear array having an odd number of elements, as shown in Figure 6. In the general case where each element has a different complex frequency response $e_n(f, x)$, using the superposition principle we can express the complex frequency response of the array as

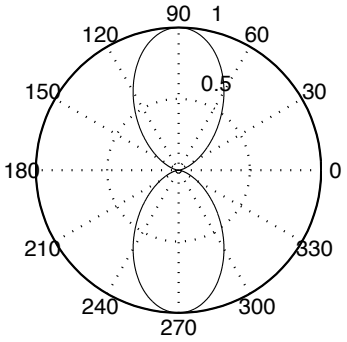
$$A(f, x_a) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f) e_n(f, x_a - x_n) \tag{27}$$

where $w_n(f)$ is the complex weight for element n, $e_n(f, x)$ is its complex frequency response or *element function*, and x_n is its spatial position on the x-axis. If we substitute this discrete aperture function into Equation 15 we obtain the far-field directivity pattern as

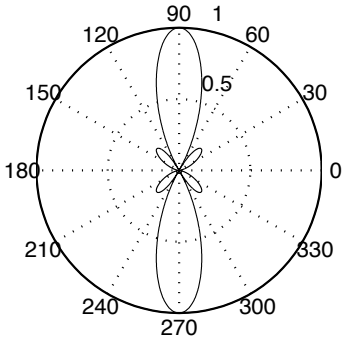
$$D(f, \alpha_x) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f) E_n(f, \alpha_x) e^{j2\pi\alpha_x x_n} \tag{28}$$



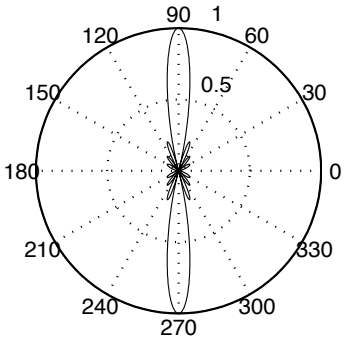
(a) $L/\lambda = 0.5$



(b) $L/\lambda = 1$



(c) $L/\lambda = 2$



(d) $L/\lambda = 4$

Figure 5: Polar plot of directivity pattern

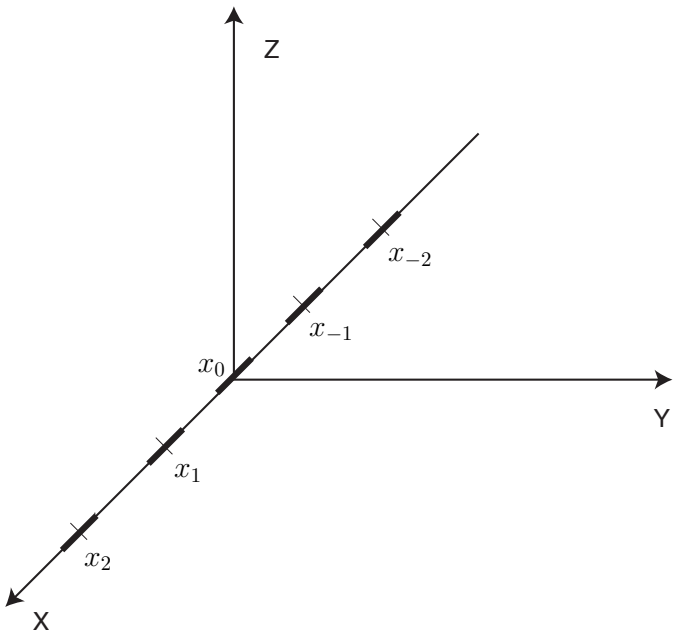


Figure 6: Discrete sensor array

where $E_n(f, \alpha_x)$ is the directivity pattern of element n.

In the case where all the elements have identical frequency response (that is $E_n(f, \alpha_x) = E(f, \alpha_x)$, $\forall n$), the aperture function can be simplified to

$$A(f, x_a) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f)\delta(x_a - x_n) \tag{29}$$

with the corresponding directivity pattern

$$D(f, \alpha_x) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f)e^{j2\pi\alpha_x x_n} \tag{30}$$

Equation 30 is the far-field directivity pattern for a linear array of N identical sensors, with arbitrary inter-element spacing. For the case where all elements are equally spaced by d metres, the directivity pattern becomes

$$D(f, \alpha_x) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f)e^{j2\pi\alpha_x nd} \tag{31}$$

Considering only the horizontal directivity pattern, we have

$$D(f, \phi) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f)e^{j\frac{2\pi}{\lambda}nd\cos\phi} \tag{32}$$

or, making the frequency dependence explicit

$$D(f, \phi) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f)e^{j\frac{2\pi f}{c}nd\cos\phi} \tag{33}$$

Equation 33 gives us the directivity pattern for a linear, equally spaced array of identical sensors. From the equation we see that the directivity pattern depends upon

- the number of array elements N
- the inter-element spacing d, and
- the frequency f.

Recall that a discrete sensor array approximates a continuous aperture. The *effective length* of a sensor array is the length of the continuous aperture which it samples, and is given by $L = Nd$. The actual *physical length* of the array, as given by the distance between the first and last sensors, is however $d(N-1)$. Several interesting characteristics of a linear, equally spaced sensor array can be observed by plotting the directivity pattern for the following scenarios

1. varying number of array elements N (L and f fixed).
2. varying effective array length $L = Nd$ (N and f fixed).
3. varying frequency f (N and L fixed).

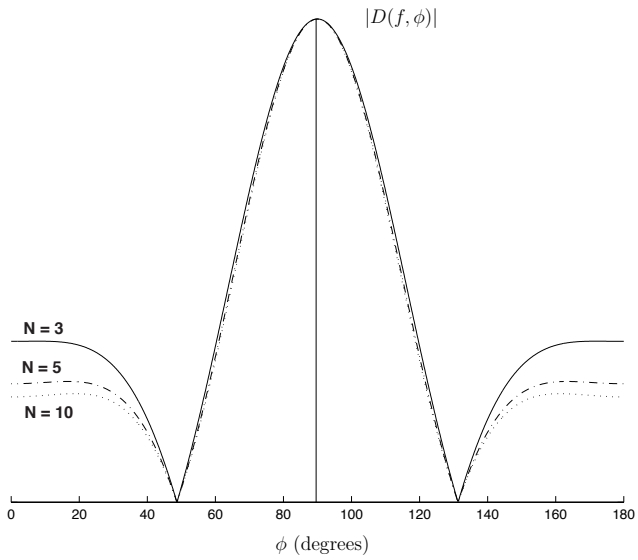


Figure 7: Directivity pattern for varying number of sensors (f=1 kHz,L=0.5 m)

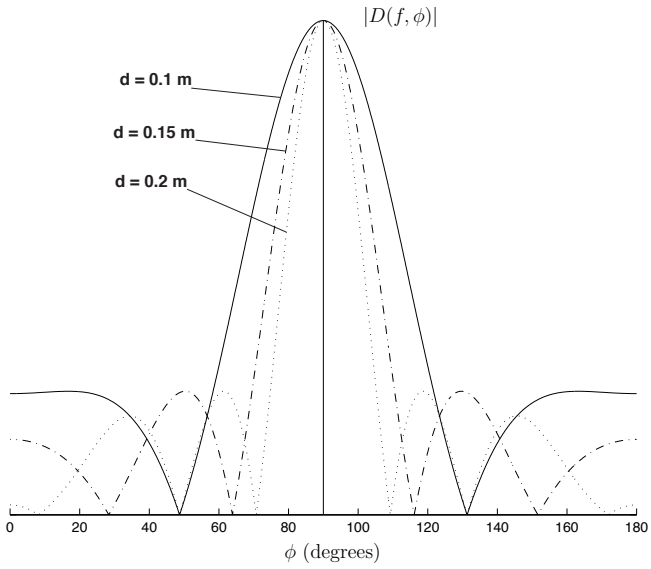


Figure 8: Directivity pattern for varying effective array length (f=1 kHz, N=5)

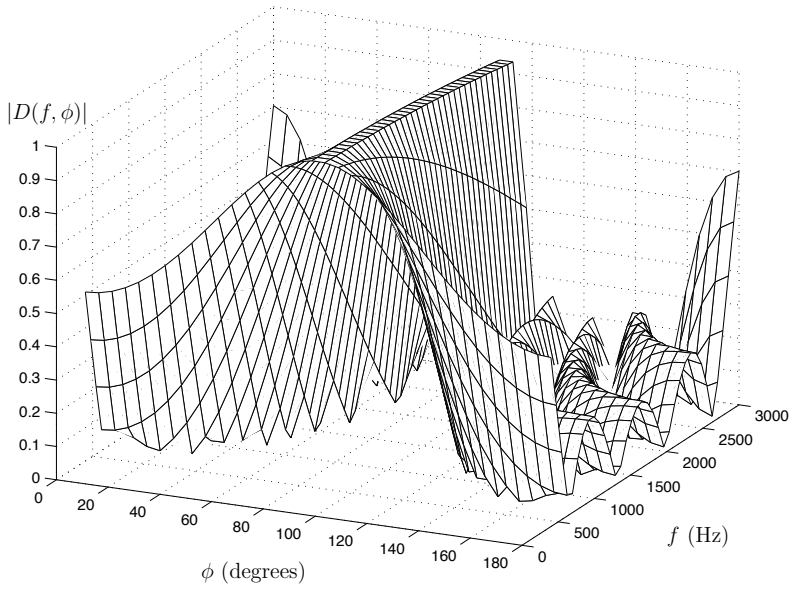


Figure 9: Directivity pattern for $400Hz \leq f \leq 3000Hz$ (N=5, d=.1 m)

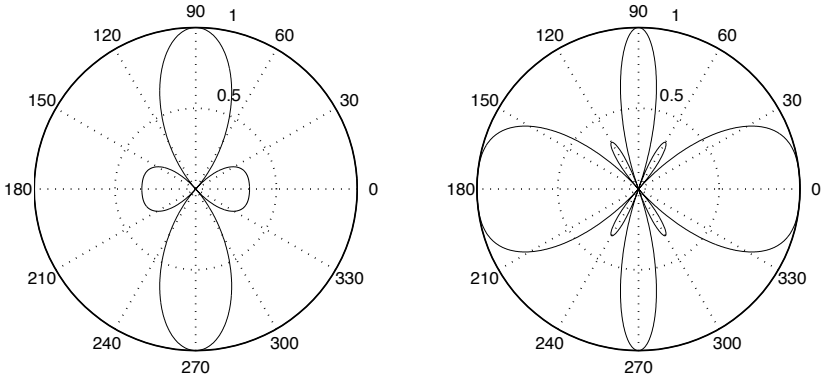
Figure 7 plots the directivity pattern for the first of these scenarios. We observe that the sidelobe level decreases with increasing spatial sampling frequency - that is, the more sensors we use, the lower the sidelobe level. The directivity pattern for the second scenario is shown in Figure 8. The plot shows that the beam width decreases as the effective array length (and thus the spacing) increases. In fact, the beam width is inversely proportional to the product fL, as seen in Figure 4. Given that $L = Nd$ and that N is fixed in this case, we see that to vary the beam width we must vary fd. It is more common however to require a constant beam width, in which case we must ensure that fd remains relatively constant. We thus see that, for a given frequency, two important characteristics of the array directivity pattern, namely the beam width and the sidelobe level, are directly determined by the inter-element spacing and the number of sensors respectively.

For a given array configuration, we note that the beam width will vary as a function of frequency : as the frequency increases, the beam width will decrease. This effect is shown in Figure 9, which plots the horizontal directivity pattern for the third scenario, where the frequency is varied over the range $400Hz \leq f \leq 3000Hz$.

1.4.2 Spatial Aliasing

A familiar principle in temporal sampling is that of the Nyquist frequency, which is the minimum sampling frequency required to avoid aliasing (the appearance of grating lobes) in the sampled signal [6]. In essence, sensor arrays implement spatial sampling, and an analogous requirement exists to avoid grating lobes in the directivity pattern. The temporal sampling theorem states that a signal must be sampled at a rate f_s (of period T_s) such that

$$f_s = \frac{1}{T_s} \geq 2f_{\max} \tag{34}$$



(a) $d = \frac{\lambda_{\min}}{2}$ (b) $d = \lambda_{\min}$

Figure 10: Example of spatial aliasing

where $f_{\max}$ is the maximum frequency component in the signal's frequency spectrum. Similarly, for spatial sampling we have the requirement that

$$f_{x_s} = \frac{1}{d} \geq 2 f_{x_{\max}} \tag{35}$$

where f_{x_s} is the spatial sampling frequency in samples per metre and $f_{x_{\max}}$ is the highest spatial frequency component in the angular spectrum of the signal. The spatial sampling frequency along the x-axis is given by

$$f_{x_s} = \frac{\sin\theta \cos\phi}{\lambda} \tag{36}$$

The maximum value of this ratio naturally occurs when the numerator is maximum and the denominator minimum. This leads to the relation

$$f_{x_{\max}} = \frac{1}{\lambda_{\min}} \tag{37}$$

and consequently the requirement that

$$d < \frac{\lambda_{\min}}{2} \tag{38}$$

where $\lambda_{\min}$ is the minimum wavelength in the signal of interest. Equation 38 is known as the *spatial sampling theorem*, and must be adhered to in order to prevent the occurrence of *spatial aliasing* in the directivity pattern of a sensor array. Figure 10 illustrates the effect of spatial aliasing on the polar plot of the horizontal directivity pattern.

1.4.3 Array Gain and Directivity Factor

A key measure for sensor arrays is the *array gain*, which is defined as the improvement in signal-to-noise ratio between a reference sensor and the array output. The array gain can be expressed as

$$G_a = \frac{G_d}{G_n} \tag{39}$$

where G_d is the gain to the desired signal and G_n is the average gain to all noise sources. The gain to the desired signal corresponds to the power of the directivity pattern in the direction of arrival, while the noise gain naturally changes depending on the nature of the noise field.

A *diffuse* noise field is one in which noise of equal energy propagates in all directions at all times (see Section 2.2. In the case of a diffuse noise field, the array gain is also known as the *factor of directivity* and is given by

$$G_a(f, \theta_0, \phi_0) = \frac{|D(f, \theta_0, \phi_0)|^2}{\frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} |D(f, \theta, \phi)|^2 \sin\theta \, d\theta \, d\phi} \tag{40}$$

where the desired source is located in the direction (θ_0, ϕ_0).

1.5 Behaviour for Near-field Sources

To this point we have only considered the case of *far-field* sources. Recall that, for a linear aperture, a wave source may be considered to come from the far-field of the aperture if

$$|r| > \frac{2L^2}{\lambda} \tag{41}$$

Under this assumption, the wavefronts arriving at the aperture can be considered as plane waves, that is, the curvature of the wavefront can be neglected. For many practical applications of sensor arrays, particularly within the context of speech recognition, the above criterion is not satisfied and the signal source is said to be located within the *near-field* of the array. The derivation of equivalent near-field expressions for the general continuous and discrete directivity patterns is quite involved, but for the purpose of this discussion it is sufficient to consider the horizontal directivity pattern for a linear sensor array. Indeed, a simple derivation of a near-field expression is possible in this case.

Consider the arrival of planar wavefronts on different elements in a sensor array, as shown in Figure 11. From the diagram we see that the actual distance traveled by the wave between adjacent sensors is given by

$$d' = d \cos\phi \tag{42}$$

More generally, the distance traveled by the wave between the reference sensor $n = 0$ and the n^{th} sensor is given by

$$d' = nd \cos\phi \tag{43}$$

Figure 12 illustrates the arrival of spherical wavefronts on different elements in a sensor array. From the diagram we see that the actual distance traveled by the wave between the two sensors is given by

$$d' = d_1(r, \phi) - d_0(r, \phi) \tag{44}$$

and in general

$$d' = d_n(r, \phi) - d_0(r, \phi) \tag{45}$$

where $d_n(r, \phi)$ is the distance from the source to the n^{th} sensor as a function of the spherical coordinates of the source (in the horizontal plane) with respect to the reference sensor. Using trigonometric relations, it can be shown that this distance is given by [7]

$$d_n(r, \phi) = \left[r^2 + 2r(x_n - x_0)\cos\phi + (x_n - x_0)^2\right]^{\frac{1}{2}} \tag{46}$$

which, in the case of an equally spaced array, reduces to

$$d_n(r, \phi) = \left[r^2 + 2rnd\cos\phi + (nd)^2\right]^{\frac{1}{2}} \tag{47}$$

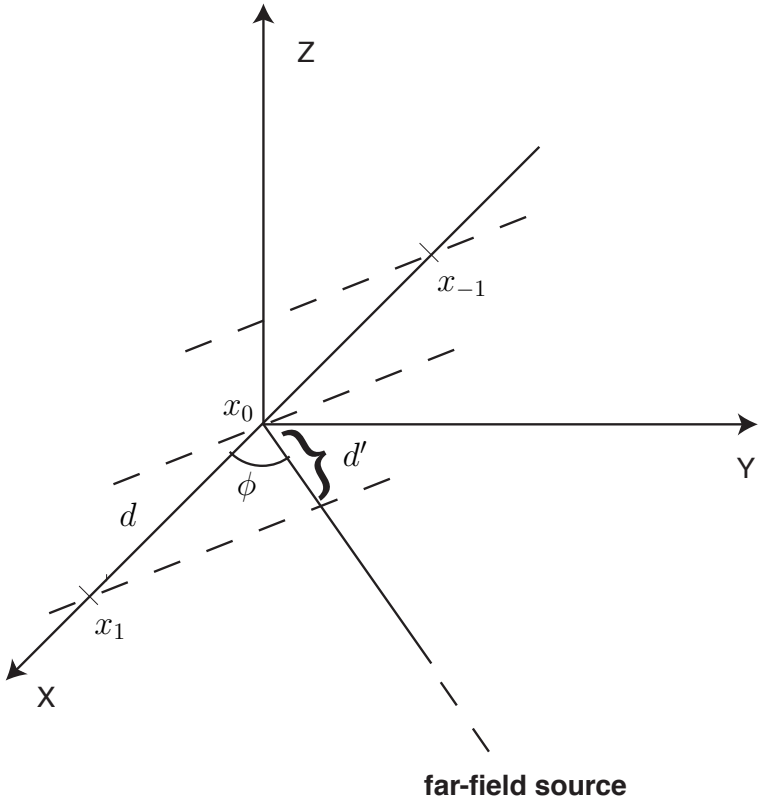


Figure 11: Arrival of wavefronts from a far-field source

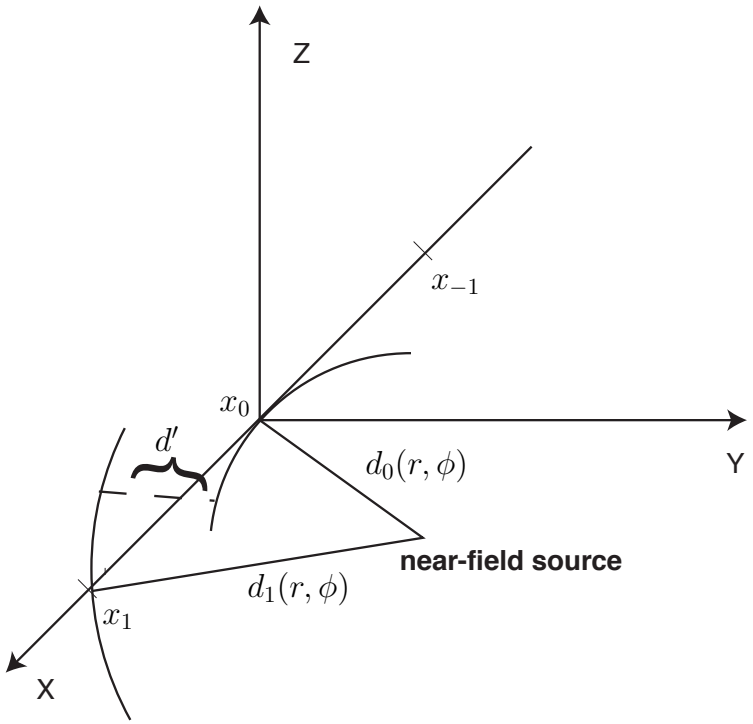


Figure 12: Arrival of wavefronts from a near-field source

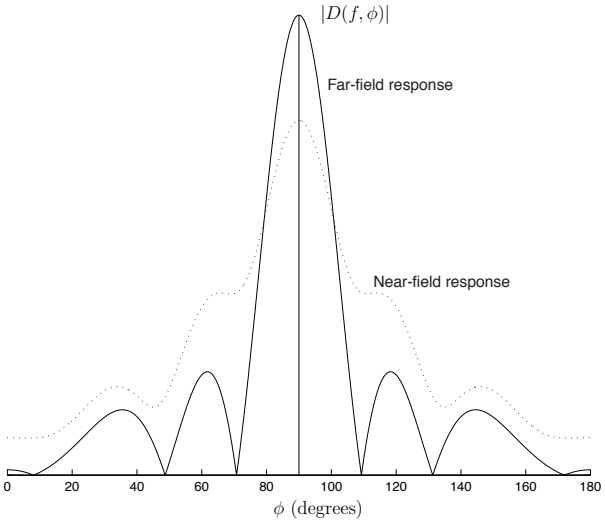


Figure 13: Directivity pattern for far-field and near-field (r=1 m) source (f=1 kHz, N=10, d=.1 m)

If we recall the far-field horizontal directivity pattern of a linear sensor array

$$D(f,\phi) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f) e^{j\frac{2\pi}{\lambda} nd\cos\phi} \tag{48}$$

we note that the exponential contains the term $nd\cos\phi$. We have seen that this corresponds to the distance traveled by the propagating wave between the reference sensor and the n^{th} sensor. Substituting in the equivalent expression for the near-field case we obtain

$$D'(f,\phi) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f) e^{j\frac{2\pi}{\lambda}(d_n(r,\phi)-d_0(r,\phi))} \tag{49}$$

In addition, we recall that for spherical acoustic waves, the amplitude decays at a rate proportional to the distance traveled. For far-field sources the amplitude differences between sensors can be considered to be negligible, however, these amplitude differences may be significant for near-field sources. Incorporating the amplitude dependency into the expression and normalising to give unity amplitude on the reference sensor we obtain the following expression for the horizontal directivity pattern for near-field sources

$$D_{nf}(f,\phi) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} \frac{d_0(r,\phi)}{d_n(r,\phi)} w_n(f) e^{j\frac{2\pi}{\lambda}(d_n(r,\phi)-d_0(r,\phi))} \tag{50}$$

Figure 13 plots the horizontal directivity pattern for both a far-field source and a near-field source for the same sensor array for r=1 m, illustrating the dependence of the pattern on the distance to the source.

If a sensor array is desired to operate in the near-field, the near-field directivity pattern can be made to match the corresponding far-field directivity pattern by compensating the frequency dependent sensor

weights $w_n(f)$. If we replace the far-field weights by the near-field compensated weights

$$w'_n(f) = \frac{d_n(r,\phi)}{d_0(r,\phi)} e^{j\frac{2\pi}{\lambda}(d_0(r,\phi)-d_n(r,\phi)+nd\cos\phi)} w_n(f) \tag{51}$$

then the near-field directivity pattern will match the far-field directivity pattern obtained using the original weights $w_n(f)$. This procedure is referred to as *near-field compensation* and allows us to approximate a desired far-field directivity pattern at a given point (r, ϕ) in the near-field.

1.6 Beamforming

We now consider the term $w_n(f)$ in the far-field horizontal directivity pattern of a linear sensor array

$$D(f,\alpha_x) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f) e^{j2\pi\alpha_x nd} \tag{52}$$

Up to this point of the discussion, we have assumed equally weighted sensors in calculating the directivity patterns, that is

$$w_n(f) = \frac{1}{N} \tag{53}$$

In general, the complex weighting can be expressed in terms of its magnitude and phase components as

$$w_n(f) = a_n(f) e^{j\varphi_n(f)} \tag{54}$$

where $a_n(f)$ and $\varphi_n(f)$ are real, frequency dependent amplitude and phase weights respectively. By modifying the amplitude weights, $a_n(f)$, we can modify the shape of the directivity pattern. Similarly, by modifying the phase weights, $\varphi_n(f)$, we can control the angular location of the response's main lobe. *Beamforming techniques* are algorithms for determining the complex sensor weights $w_n(f)$ in order to implement a desired *shaping* and *steering* of the array directivity pattern.

To illustrate the concept of beam steering, we consider the case where the sensor amplitude weights $a_n(f)$ are set to unity, resulting in the directivity pattern

$$D(f,\phi) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} e^{j(2\pi\alpha_x nd+\varphi_n(f))} \tag{55}$$

If we use the phase weights

$$\varphi_n(f) = -2\pi\alpha'_x nd \tag{56}$$

where

$$\alpha'_x = \frac{\sin\theta'\cos\phi'}{\lambda} \tag{57}$$

then the directivity pattern becomes

$$D'(f,\alpha_x) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} e^{j\frac{2\pi}{\lambda}nd\left(\alpha_x-\alpha'_x\right)} \tag{58}$$

which can be expressed as

$$D'(f,\alpha_x) = D(f,\alpha_x-\alpha'_x) \tag{59}$$

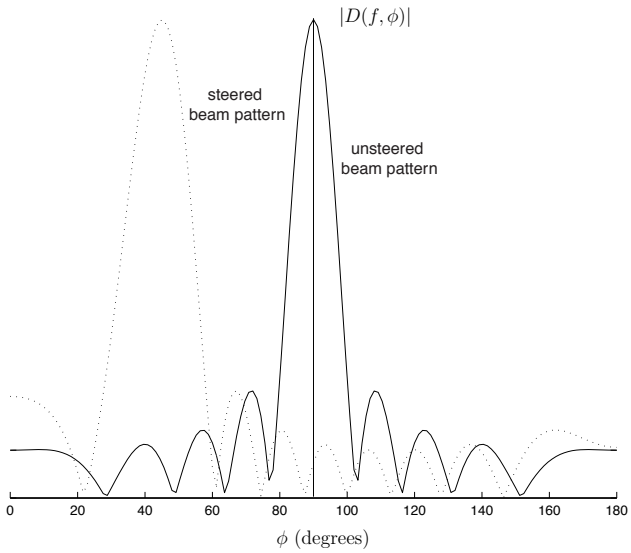


Figure 14: Unsteered and steered directivity patterns (ϕ'=45 degrees, f=1 kHz, N=10, d=.15 m)

The effect of such a phase weight on the beam pattern is thus to steer the main lobe of the beam pattern to the direction cosine $\alpha_x = \alpha'_x$, and thus to the directions $\theta = \theta'$ and $\phi = \phi'$. While the beam pattern remains unchanged apart from the shift along the α_x axis, when plotted as a function of angle, the beam shape will change as α_x is actually a function of $\sin\theta$ and $\cos\phi$. The horizontal directivity pattern is shown in in Figure 14, where the beam pattern has been shifted to $\phi' = 45^o$.

Fourier transform theory tells us that a negative phase shift in the frequency domain corresponds to a time delay in the time domain [6], and so beam steering can effectively be implemented by applying time delays to the sensor inputs. Considering only the horizontal plane, we see that the delay for the n^{th} sensor is given by

$$\begin{aligned} \tau_n &= \frac{\varphi_n}{2\pi f} \\ &= \frac{2\pi f n d \cos\phi'}{2\pi f c} \\ &= \frac{nd\cos\phi'}{c} \end{aligned} \tag{60}$$

which is seen to be equivalent to the time the plane wave takes to travel between the reference sensor and the n^{th} sensor. This is the principle of the simplest of all beamforming techniques, known as *delay-sum beamforming*, where the time domain sensor inputs are first delayed by τ_n seconds, and then summed to give a single array output. While we have seen here that the mathematics of discrete sensor arrays assures a main lobe of increased gain in the direction of the desired signal, the signal enhancement and noise reduction provided by the delay-sum beamformer can intuitively be attributed to the constructive (in phase) interference of the desired propagating wave and the destructive (out of phase) interference of waves from all other directions. Other more complicated beamforming techniques will be discussed in detail in the following section.

2 Microphone Array Beamforming Techniques

2.1 Introduction

The previous section presented the fundamental theory of sensor arrays, and introduced the concept of beamforming algorithms. This chapter continues the discussion by presenting the theory of a number of key microphone array beamforming techniques.

Beamforming techniques can be broadly classified as being either data-independent, or data-dependent. *Data-independent*, or *fixed*, beamformers are so named because their parameters are fixed during operation. Conversely, *data-dependent*, or *adaptive*, beamforming techniques continuously update their parameters based on the received signals.

As different beamforming techniques are appropriate for different noise conditions, the chapter begins by defining the noise fields encountered in microphone array applications. Following this, the principles of a number of key beamforming techniques are described in detail. The chapter concludes with a summary of the beamforming techniques, indicating their advantages, disadvantages and applicability in different noise conditions.

2.2 Noise Fields

There are three main categories of noise fields to be defined for microphone array applications. These categories are characterised by the degree of correlation between noise signals at different spatial locations. A commonly used measure of the correlation is the *coherence*, which is defined as [8]

$$\Gamma_{ij}(f) \widehat{=} \frac{\Phi_{ij}(f)}{\sqrt{\Phi_{ii}(f)\Phi_{jj}(f)}} \tag{61}$$

where Φ_{ij} is the cross-spectral density between signals i and j. The coherence is essentially a normalised cross-spectral measure, as the magnitude squared coherence can be seen to be bounded by $0 \leq |\Gamma_{ij}(f)|^2 \leq 1$.

A more comprehensive analysis of noise fields can be found in Templeton and Saunders [9].

2.2.1 Coherent Noise Fields

A *coherent noise field* is one in which noise signals propagate to the microphones directly from their sources without undergoing any form of reflection, dispersion or dissipation due to the acoustic environment. In a coherent noise field, the noise signals on different microphones in an array are strongly correlated, and hence $|\Gamma_{ij}(f)|^2 \approx 1$. In practice, coherent noise fields occur in open air environments where there are no major obstacles to sound propagation and where wind or thermal turbulence effects are minimal.

2.2.2 Incoherent Noise Fields

In an *incoherent noise field*, the noise measured at any given spatial location is uncorrelated with the noise measured at all other locations, that is $|\Gamma_{ij}(f) \approx 0|^2$. Such an ideal incoherent noise field is difficult to achieve and is seldom encountered in practical situations. In the case of microphone arrays however, electrical noise in the microphones is generally randomly distributed and can be considered to be a source of incoherent noise. Incoherent noise is also said to be *spatially white*.

2.2.3 Diffuse Noise Fields

In a *diffuse noise field*, noise of equal energy propagates in all directions simultaneously. Thus sensors in a diffuse noise field will receive noise signals that are lowly correlated, but have approximately the same

energy. Many practical noise environments can be characterised by a diffuse noise field, such as office or car noise. The coherence between the noise at any two points in a diffuse noise field is a function of the distance between the sensors, and can be modeled as [10]

$$\Gamma_{ij}(f) = \text{sinc}\left(\frac{2\pi f d_{ij}}{c}\right) \tag{62}$$

where d_{ij} is the distance between sensors i and j, and the sinc function has been defined in Equation 23. It can be seen that the coherence approaches unity for closely spaced sensors and decreases sharply with increasing distance.

2.3 Classical Beamforming

2.3.1 Delay-sum Beamforming

The simplest of all microphone array beamforming techniques is delay-sum beamforming, as discussed in Section 1.6. We recall that, by applying phase weights to the input channels, we can steer the main lobe of the directivity pattern to a desired direction. Considering the horizontal directivity pattern, if we use the phase weights

$$\varphi_n = \frac{-2\pi(n-1)d\cos\phi' f}{c} \tag{63}$$

we obtain the directivity pattern

$$D(f,\phi) = \sum_{n=1}^{N} e^{j\frac{2\pi f(n-1)d\left(\cos\phi - \cos\phi'\right)}{c}} \tag{64}$$

and the directivity pattern's main lobe will be moved to the direction $\phi = \phi'$, as illustrated in Figure 15 for $\phi' = 45^o$. Note that in this chapter we have made a simple modification to the formulae from Chapter 1 in order to change the microphone index range from $-\frac{N-1}{2} \leq n \leq \frac{N-1}{2}$ to the more convenient $1 \leq n \leq N$.

The negative phase shift in the frequency domain can effectively be implemented by applying time delays to the sensor inputs, where the delay for the n^{th} sensor is given by

$$\tau_n = \frac{(n-1)d\cos\phi'}{c} \tag{65}$$

which is the time the plane wave takes to travel between the reference sensor and the n^{th} sensor.

Delay-sum beamforming is so-named because the time domain sensor inputs are first delayed by τ_n seconds, and then summed to give a single array output. Usually, each channel is given an equal amplitude weighting in the summation, so that the directivity pattern demonstrates unity gain in the desired direction. This leads to the complex channel weights

$$w_n(f) = \frac{1}{N} e^{j\frac{-2\pi f}{c}(n-1)d\cos\phi'} \tag{66}$$

Expressing the array output as the sum of the weighted channels we obtain

$$y(f) = \frac{1}{N}\sum_{n=1}^{N} x_n(f) e^{j\frac{-2\pi f}{c}(n-1)d\cos\phi'} \tag{67}$$

Equivalently, in the time domain we have

$$y(t) = \frac{1}{N}\sum_{n=1}^{N} x_n(t-\tau_n) \tag{68}$$

where τ_n is defined in Equation 65.

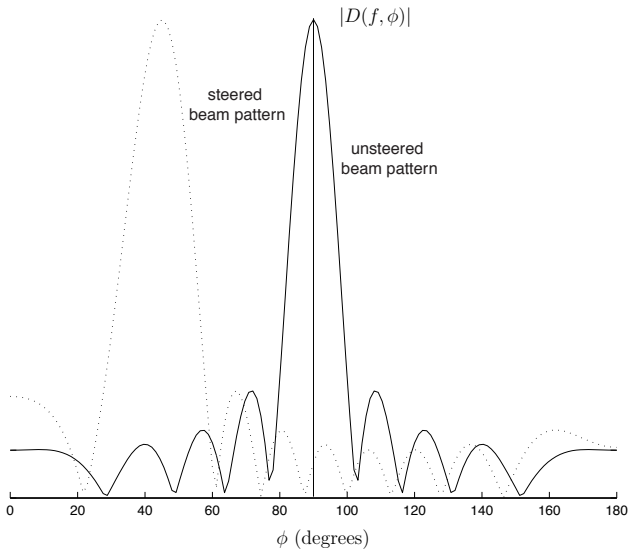


Figure 15: Unsteered and steered directivity patterns (ϕ'=45 degrees, f=1 kHz, N=10, d=.15 m)

2.3.2 Filter-sum Beamforming

The delay-sum beamformer belongs to a more general class known as *filter sum beamformers*, in which both the amplitude and phase weights are frequency dependent. In practice, most beamformers are a class of filter-sum beamformer. The output of a filter-sum beamformer is given as

$$y(f) = \sum_{n=1}^{N} w_n(f) x_n(f) \tag{69}$$

It is often convenient to use matrix algebra to simplify the notation when describing microphone array techniques. The above equation can be rewritten using matrix notation as

$$y(f) = \mathbf{w}(f)^T \mathbf{x}(f) \tag{70}$$

where the weight vector $\mathbf{w}(f)$ and data vector $\mathbf{x}(f)$ are defined as

$$\mathbf{w}(f) = \left[\begin{array}{ccccccc} w_1(f) & \cdots & w_n(f) & \cdots & w_N(f) \end{array} \right]^T \tag{71}$$

and

$$\mathbf{x}(f) = \left[\begin{array}{ccccccc} x_1(f) & \cdots & x_n(f) & \cdots & x_N(f) \end{array} \right]^T \tag{72}$$

where $(\cdot)^T$ denotes matrix transpose. A block diagram showing the structure of a general filter-sum beamformer is given in Figure 16.

2.3.3 Sub-array Beamforming

From the equation for the directivity pattern of a uniformly spaced sensor array, it is seen that the characteristics of the array response depend on the frequency of interest, the inter-element spacing (or effective length, as $L = Nd$), and the number of elements in the array. The dependency on the operating

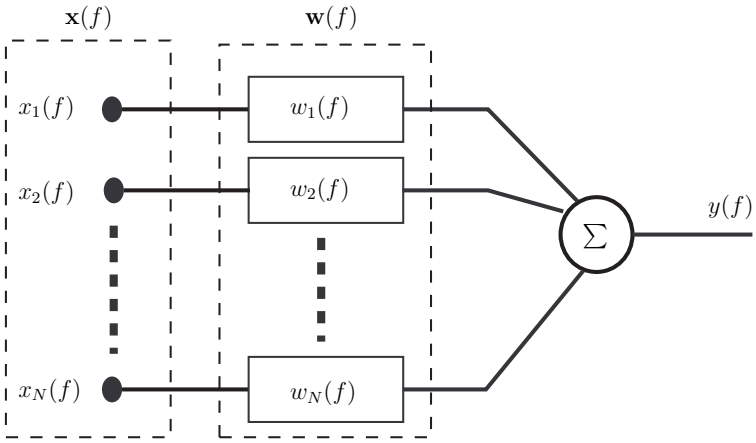


Figure 16: Filter-sum beamformer structure

frequency means that the response characteristics (beam-width, sidelobe level) will only remain constant for narrow-band signals, where the bandwidth is not a significant proportion of the centre frequency. Speech, however, is a broad-band signal, meaning that a single linear array design is inadequate if a frequency invariant beam-pattern is desired.

One simple method of covering broadband signals is to implement the array as a series of sub-arrays, which are themselves linear arrays with uniform spacing. These sub-arrays are designed to give desired response characteristics for a given frequency range. Due to the dependencies discussed in Section 1.4.1, as the frequency increases, a smaller array length is required to maintain constant beam-width. In addition, to ensure the sidelobe level remains the same across different frequency bands, the number of elements in each sub-array should remain the same. The sub-arrays are generally implemented in a nested fashion, such that any given sensor may be used in more than one sub-array. Each sub-array is restricted to a different frequency range by applying band-pass filters, and the overall broad-band array output is formed by recombining the outputs of the band-limited sub-arrays. An example of such a nested sub-array structure for delay-sum beamforming, designed to cover 4 different frequency bands, is shown in Figure 17. The sub-arrays employ 3, 5, 5 and 5 microphone respectively, but, due to the nested structure, the 4 sub-arrays can be implemented using a total of 9 microphones.

For a general sub-array broadband beamformer, the beamforming channel filters are band-pass filtered between the specified upper and lower frequencies for each sub-band. At the output of each channel filter we have

$$v_{s,i}(f) = w_{s,i}(f)x_i(f) \tag{73}$$

where $x_i(f)$ is the input to channel i of the array, and the subscript s represents the sub-array index. The output of sub-array s, is then given by the sum across channels as

$$y_s(f) = \sum_{i=1}^{N} v_{s,i}(f) \tag{74}$$

where there are N microphones in the array. The summation in each sub-array is shown up to N for simplicity of notation, although in practice only the channels belonging to each sub-array are used. The

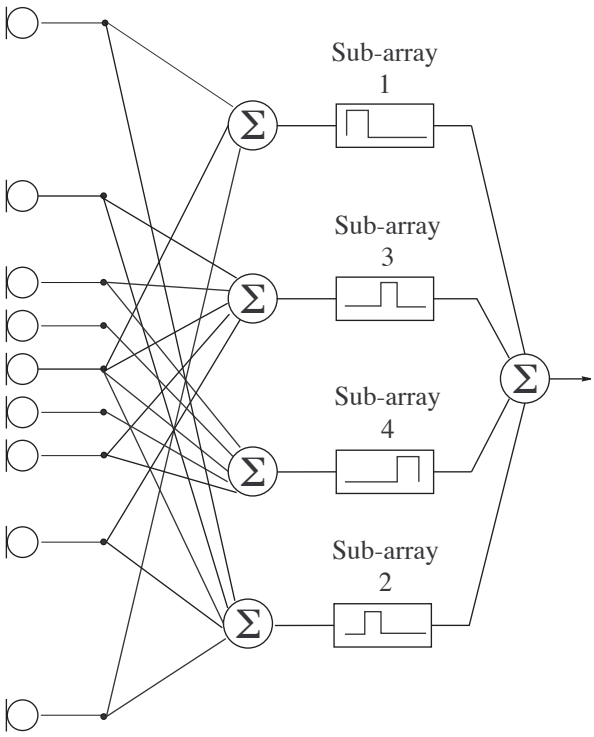


Figure 17: Sample nested sub-array structure

overall array output is then calculated as

$$y(f) = \sum_{s=1}^{S} y_s(f) \tag{75}$$

where there are S sub-arrays.

Any technique may be used to design the filters for a sub-array filter-sum beamformer. The most common technique is simply to use conventional delay-sum beamforming within each sub-array.

2.4 Superdirective Beamforming

Conventional linear arrays with sensors spaced at $d \approx \lambda/2$ have directivity that is approximately proportional to the number of sensors, N. It has been found that the directivity of linear endfire arrays theoretically approaches N^2 as the spacing approaches zero in a diffuse (spherically isotropic) noise field [11, 12]. Beamforming techniques that exploit this capability for closely spaced endfire arrays are termed *superdirective beamformers*. The channel filters of superdirective beamformers are typically formulated to maximise the array gain, or factor of directivity. In Section 1.4.3 the factor of directivity was defined as

$$G_a(f, \theta_0, \phi_0) = \frac{|D(f, \theta_0, \phi_0)|^2}{\frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} |D(f, \theta, \phi)|^2 \sin\theta \, d\theta \, d\phi} \tag{76}$$

Recall that the horizontal directivity pattern is given by

$$D(f, \phi) = \sum_{n=1}^{N} w_n(f) e^{j\frac{2\pi f}{c}(n-1)d\cos\phi} \tag{77}$$

Using the filter-sum weight vector, $\mathbf{w}(f)$, and defining the propagation vector as

$$\mathbf{d}(f) = \begin{bmatrix} 1 & \cdots & e^{-j\frac{2\pi f}{c}(n-1)d\cos\phi} & \cdots & e^{-j\frac{2\pi f}{c}(N-1)d\cos\phi} \end{bmatrix}^T \tag{78}$$

we can formulate the directivity pattern in matrix notation as

$$D(f, \phi) = \mathbf{w}(f)^H \mathbf{d}(f) \tag{79}$$

where $(\cdot)^H$ denotes matrix transpose conjugate. Expressing the factor of directivity in matrix notation, and noting that $\mathbf{w}$ is independent of direction, we obtain

$$G_a(f, \theta_0, \phi_0) = \frac{\left|\mathbf{w}(f)^H \mathbf{d}(f)\right|^2}{\mathbf{w}(f)^H \left(\frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} \mathbf{d}(f)\mathbf{d}(f)^H \sin\theta \, d\theta \, d\phi\right) \mathbf{w}(f)} \tag{80}$$

and if we define the matrix $\mathbf{\Gamma}$ as

$$\mathbf{\Gamma} = \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} \mathbf{d}(f)\mathbf{d}(f)^H \sin\theta \, d\theta \, d\phi \tag{81}$$

we can express the factor of directivity concisely as

$$G_a = \frac{\left|\mathbf{w}^H \mathbf{d}\right|^2}{\mathbf{w}^H \mathbf{\Gamma} \mathbf{w}} \tag{82}$$

where the frequency dependence has been omitted for notational simplicity.

From the discussion in Section 1.4.3 we recall that the factor of directivity is the array gain for a diffuse noise field. The diffuse noise field is characterised by the matrix $\mathbf{\Gamma}$, which represents the cross-spectral density of the noise between sensors. For a general noise field with noise cross-spectral matrix $\mathbf{Q}$, the array gain can be expressed as [13]

$$G_a = \frac{\left|\mathbf{w}^H\mathbf{d}\right|^2}{\mathbf{w}^H\mathbf{Q}\mathbf{w}} \tag{83}$$

Superdirective beamformers aim to calculate the weight vector $\mathbf{w}$ that maximises the array gain, that is :

$$\max_{\mathbf{w}} \frac{\left|\mathbf{w}^H\mathbf{d}\right|^2}{\mathbf{w}^H\mathbf{Q}\mathbf{w}} \tag{84}$$

Cox [13] gives the solution using the Lagrange method as

$$\mathbf{w} = \alpha\mathbf{Q}^{-1}\mathbf{d} \tag{85}$$

where α is an arbitrary complex constant. If we choose α to produce unity signal response with zero phase shift (that is, $\mathbf{w}^H\mathbf{d} = 1$), then the solution is [13] :

$$\mathbf{w} = \frac{\mathbf{Q}^{-1}\mathbf{d}}{\mathbf{d}^H\mathbf{Q}^{-1}\mathbf{d}} \tag{86}$$

In practice, the above solution can lead to undesirable gain of incoherent noise due to electrical sensor noise, channel mismatch and errors in microphone spacing. To prevent excessive amplification of the incoherent noise, a more robust solution can be found if a constraint is placed on the white noise gain :

$$\frac{\left|\mathbf{w}^H\mathbf{d}\right|^2}{\mathbf{w}^H\mathbf{w}} = \delta^2 \leq M \tag{87}$$

where the constraining value δ^2 must be chosen to be less than or equal to the maximum possible white noise gain M.

The above solution can also be expanded for the more general case where the optimisation is subjected to multiple linear constraints (including that of unity response for the desired signal), expressed as

$$\mathbf{C}^H\mathbf{w} = \mathbf{g} \tag{88}$$

The solution under such a set of linear constraints, as well as a constraint on the white noise gain, is given by Cox [13] as

$$\mathbf{w} = [\mathbf{Q} + \epsilon\mathbf{I}]^{-1}\mathbf{C}\left\{\mathbf{C}^H[\mathbf{Q} + \epsilon\mathbf{I}]^{-1}\mathbf{C}\right\}^{-1}\mathbf{g} \tag{89}$$

where ϵ is a Lagrange multiplier that is iteratively adjusted until the white noise gain constraint is satisfied. The white noise gain is the array gain for spatially white (incoherent) noise, that is, $\mathbf{Q} = \mathbf{I}$. A constraint on the white noise gain is necessary, as an unconstrained superdirective solution will in fact result in significant gain to any incoherent noise, particularly at low frequencies. Cox [13] states that the technique of adding a small amount to each diagonal matrix element prior to inversion is in fact the optimum means of solving this problem. A study of the relationship between the multiplier ϵ and the desired white noise gain δ^2, shows that the white noise gain increases monotonically with increasing ϵ. One possible means of obtaining the desired value of ϵ is thus an iterative technique employing a binary search algorithm between a specified minimum and maximum value for ϵ. The computational expense of the iterative procedure is not critical, as the beamformer filters depend only on the source location and array geometry, and thus must only be calculated once for a given configuration.

For speech processing applications, superdirective methods are useful for obtaining acceptable array performance at low frequencies for realistic array dimensions. The wavelength for acoustic waves at 500 Hz is approximately 0.66 m, and so sensor elements spaced closer than 0.33 m in an endfire configuration can be used in the low frequency range to improve performance.

2.5 Near-field Superdirective Beamforming

Low frequency performance is problematic for conventional beamforming techniques because large wavelengths give negligible phase differences between closely spaced sensors, leading to poor directive discrimination. Täger [14] states that delay-weight-sum beamformers can roughly cover the octave band $0.25 < d/\lambda < 0.5$ (where d is the inter-element spacing) before excessive loss of directivity occurs. A frequency of 100 Hz corresponds to a wavelength of 3.4 m for sound waves, and so to cater for this frequency range requires that 0.85m $< d <$ 1.7m. For a sub-array of 5 elements, this would give an array dimension of 3.4m $< L <$ 6.8m, which is impractical for many applications. For example, in the context of a multimedia workstation, it is desirable that the array dimension does not exceed the monitor width, which will be approximately 17 inches, or 40 cm. Thus methods providing good low frequency performance with realistic array dimensions are required.

One such method is a technique proposed by Täger [14, 15], called near-field superdirectivity. As its name implies, near-field superdirectivity is a modification of the standard superdirective technique presented in Section 2.4, in which the propagation vector $\mathbf{d}$ is replaced by one formulated for a near-field source.

We recall from Section 1.5 that the near-field directivity pattern can be expressed as

$$D_{nf}(f,\phi) = \sum_{n=1}^{N} \frac{d_1(r,\phi)}{d_n(r,\phi)} w_n(f) e^{j\frac{2\pi}{\lambda}(d_n(r,\phi)-d_1(r,\phi))} \tag{90}$$

If we define the time difference between sensor n and the reference sensor ($n = 1$) as

$$\tau_n = \frac{d_n(r,\phi) - d_1(r,\phi)}{c} \tag{91}$$

and the amplitude attenuation factor between sensor n and the reference sensor as

$$\alpha_n = \frac{d_1(r,\phi)}{d_n(r,\phi)} \tag{92}$$

then the near-field modified propagation vector can be expressed as

$$\mathbf{d}_{nf}(f) = \begin{bmatrix} \alpha_1 e^{-j2\pi f\tau_1} & \cdots & \alpha_n e^{-j2\pi f\tau_n} & \cdots & \alpha_N e^{-j2\pi f\tau_N} \end{bmatrix}^T \tag{93}$$

Near-field superdirectivity uses the above near-field propagation vector in the standard superdirective formulation, while maintaining the assumption of a (far-field) diffuse noise field in the noise cross-spectral matrix $\mathbf{\Gamma}$. In this way, as well as providing directional sensitivity, the technique gives a level of discrimination between the array's near- and far-fields. Expressed formally we have

$$\max_{\mathbf{w}} \frac{\left|\mathbf{w}^H \mathbf{d}_{nf}\right|^2}{\mathbf{w}^H \mathbf{\Gamma} \mathbf{w}} \tag{94}$$

where $\mathbf{\Gamma}$ was defined in Equation 81 using the far-field propagation vector. Thus, similar to standard superdirectivity, the solution under a set of linear constraints, $\mathbf{C}^H\mathbf{w} = \mathbf{g}$ (including $\mathbf{w}^H\mathbf{d}_{nf} = 1$), and a robustness constraint on the white noise gain is given as

$$\mathbf{w} = [\mathbf{\Gamma} + \epsilon\mathbf{I}]^{-1}\mathbf{C}\left\{\mathbf{C}^H[\mathbf{\Gamma} + \epsilon\mathbf{I}]^{-1}\mathbf{C}\right\}^{-1}\mathbf{g} \tag{95}$$

Near-field superdirectivity succeeds in achieving greater performance than standard techniques for near-field sources at low frequencies. This is due to the fact that it takes the amplitude differences into account as well as the phase differences. While the phase differences are negligible at low frequencies, the amplitude differences are significant, particularly when the sensors are placed in an endfire configuration as this maximises the difference in the distance from the source to each microphone. A simple illustration of the effect of the amplitude compensation is given in Täger [14].

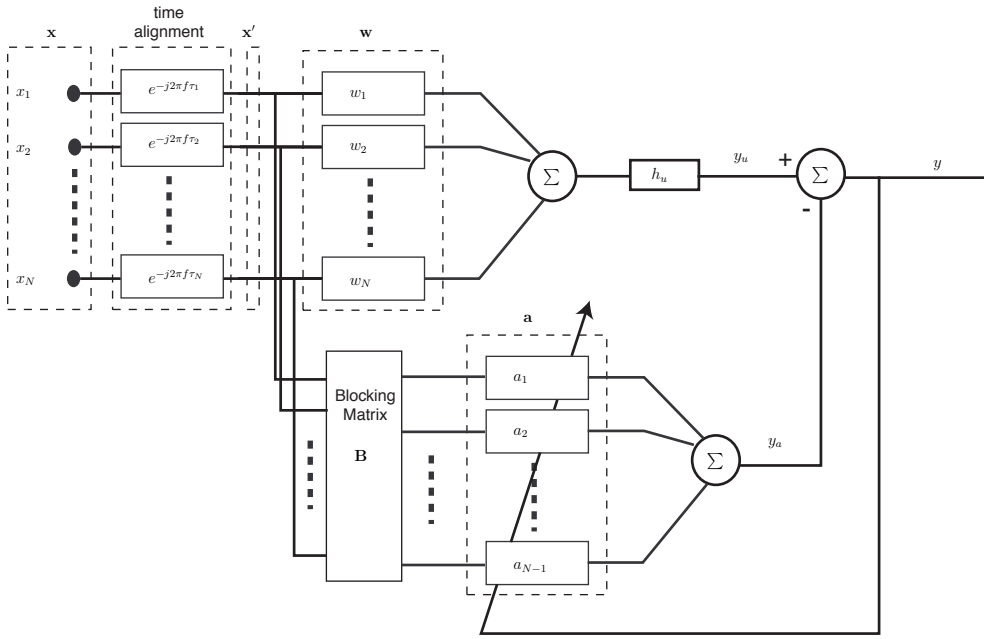


Figure 18: Generalised sidelobe canceler structure

2.6 Generalised Sidelobe Canceler (GSC)

Data-dependent beamforming techniques attempt to adaptively filter the incoming signals in order to pass the signal from the desired direction while rejecting noises coming from other directions. For their optimisation criterion, most adaptive techniques rely on the minimisation of the mean-square error between a reference signal that is highly correlated to the desired signal, and the output signal. Unfortunately, the normal least means square (LMS) algorithm can degrade the desired signal as it seeks purely to minimise the mean-square error - it places no conditions upon the distortion to the desired signal. The most famous adaptive beamforming technique that addresses this limitation is known as *Frost's algorithm* [16]. Frost's algorithm treats the filter estimation process as a problem in constrained least mean-square minimisation - the solution minimises the mean-square error while maintaining a specified transfer function for the desired signal. This constraint is normally designed to ensure that the response to the desired signal has constant gain and linear phase. Frost's algorithm belongs to a class of beamformers known as *linearly constrained minimum variance* (LCMV) beamformers.

Perhaps the most commonly used LCMV beamforming technique is the *generalised sidelobe canceler* (GSC) [17]. GSC is a beamforming structure that can be used to implement a variety of linearly constrained adaptive array processors, including Frost's algorithm. It separates the adaptive beamformer into two main processing paths. The first of these implements a standard fixed beamformer, with constraints on the desired signal. The second path is the adaptive portion, which provides a set of filters that adaptively minimise the power in the output. The desired signal is eliminated from this second path by a blocking matrix, ensuring that it is the noise power that is minimised. The block structure of the generalised sidelobe canceler is shown in Figure 18.

Examining the upper path, the inputs are first time aligned and then passed through a filter-sum beamformer to give the fixed beamformed signal y'_u as

$$y'_u(f) = \mathbf{w}_c(f)^T \mathbf{x}'(f) \tag{96}$$

where

$$\mathbf{w}(f) = \begin{bmatrix} w_1(f) & \cdots & w_n(f) & \cdots & w_N(f) \end{bmatrix}^T \tag{97}$$

are the fixed amplitude weights for each of the N channels, and

$$\mathbf{x}'(f) = \begin{bmatrix} x'_1(f) & \cdots & x'_n(f) & \cdots & x'_N(f) \end{bmatrix}^T \tag{98}$$

are the time aligned input signals.

The output of the fixed beamformer is then filtered by the constraint filter h_u which ensures a specified gain and phase response for the desired signal. The output of the upper path is thus given by

$$y_u(f) = h_u(f) y'_u(f) \tag{99}$$

The lower path of the structure is the adaptive portion. It consists of two major parts. The first of these is the blocking matrix, $\mathbf{B}$, whose purpose is to remove the desired signal from the lower path. As the desired signal is common to all the time-aligned channel inputs, blocking will occur if the rows of the blocking matrix sum to zero. If $\mathbf{x}''$ denotes the signals at the output of the blocking matrix, then

$$\mathbf{x}''(f) = \mathbf{B}\mathbf{x}'(f) \tag{100}$$

where each row of the blocking matrix sums to zero, and the rows are linearly independent. As $\mathbf{x}'$ can have at most $N-1$ linearly independent components, the number of rows in $\mathbf{B}$ must be $N-1$ or less. The standard Griffiths-Jim blocking matrix is [17]

$$\mathbf{B} = \begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \\ \vdots & \cdots & \ddots & \ddots & \cdots & \vdots \\ 0 & \cdots & 0 & 1 & -1 & 0 \\ 0 & \cdots & 0 & 0 & 1 & -1 \end{bmatrix} \tag{101}$$

Following application of the blocking matrix, $\mathbf{x}''$ is adaptively filtered and summed to give the lower path output y_a. If we denote the lower path adaptive filters as $\mathbf{a}$, then we have

$$y_a(f) = \mathbf{a}(f)^T \mathbf{x}''(f) \tag{102}$$

Due to the blocking matrix, the lower path output only contains noise signals. The overall system output is calculated as the difference of the upper and lower path outputs as

$$y(f) = y_u(f) - y_a(f) \tag{103}$$

Because the upper path contains the constrained desired signal estimate, and the lower path only contains noise and interference terms, finding the set of filter coefficients $\mathbf{a}$ which minimise the power in y is effectively equivalent to finding the linearly constrained minimum variance beamforming solution. As the signal is constrained in the upper path, the unconstrained LMS algorithm can be used to adapt the lower path filter coefficients

$$\mathbf{a}_{k+1}(f) = \mathbf{a}_k(f) + \mu y_k(f) \mathbf{x}''_k(f) \tag{104}$$

where μ is the step size and k is the frame number.

The GSC is a flexible structure due to the separation of the beamformer into a fixed and adaptive portion, and it is the most widely used adaptive beamformer. In practice, the GSC can cause a degree of distortion to the desired signal, due to a phenomenon known as signal leakage. Signal leakage occurs when the blocking matrix fails to remove all of the desired signal from the lower noise canceling path. This can be particularly problematic for broad-band signals, such as speech, as it is difficult to ensure perfect signal cancellation across a broad frequency range.

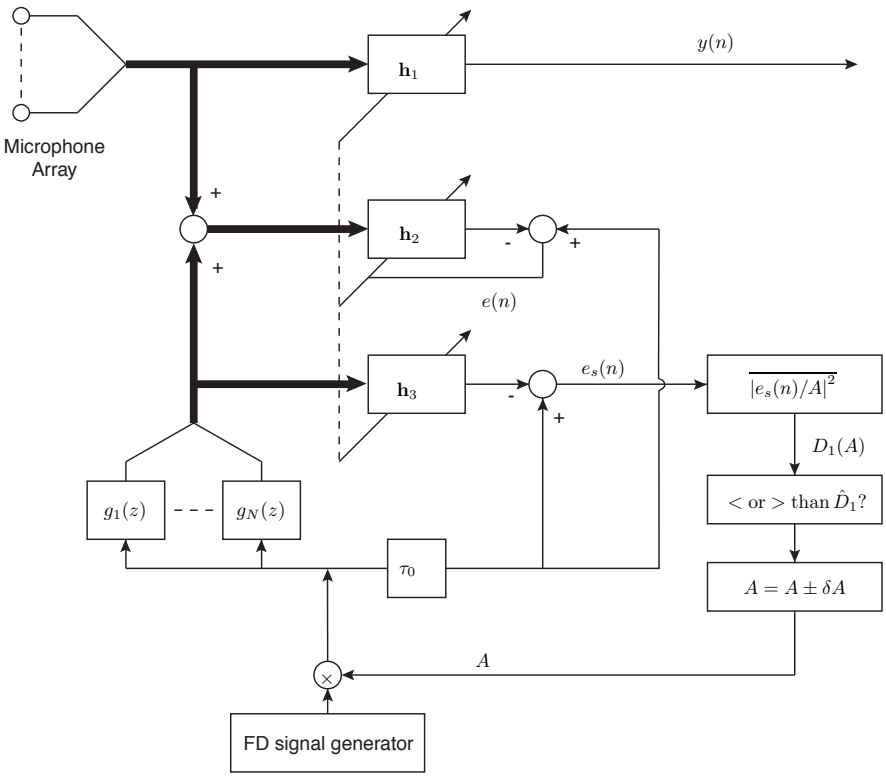


Figure 19: AMNOR system structure

2.7 AMNOR

While LCMV algorithms are theoretically powerful, they often encounter a number of problems in practice. Because of the hard constraint of one permissible value for the desired signal's transfer function, LCMV techniques can fail to sufficiently reduce the noise level due to the lack of freedom in the choice of filters. Evaluation of the human auditory system shows that a certain level of distortion in the desired signal can be tolerated and so in some situations it may be permissible, and even desirable, to allow some signal distortion in order to achieve better noise reduction.

A technique incorporating such a 'soft' constraint, named the AMNOR (*a*daptive *m*icrophone-array system for *no*ise *r*eduction) system was proposed by Kaneda [18]. Instead of allowing only one response for the desired signal, the system adopts a soft constraint that allows a class of responses whose degradation is less than some pre-determined permissible level.

Figure 19 shows the structure of the AMNOR system. The system is essentially composed of three filter blocks, $\mathbf{h}_1$, $\mathbf{h}_2$ and $\mathbf{h}_3$.

The filter block $\mathbf{h}_1$ contains the beamforming filters that are applied to the multi-channel input to give the system output, $y(n)$. The impulse response of the acoustic path between the source and array element i is modeled by the filter $g_i(z)$. The frequency response of the beamformer to the desired signal is therefore

$$F(z) = \sum_{i=1}^{N} h_{1,i}(z) g_i(z) \tag{105}$$

The second set of filters $\mathbf{h}_2$ are adaptively updated to satisfy the criterion of minimum output noise power for a given level of degradation to the desired signal. The adaptation only occurs during noise-only periods, during which time a *fictitious desired* signal is introduced into the system. This fictitious desired signal is a white noise signal with unity power that is magnified by a variable amplitude factor A. The fictitious desired signal is filtered by the acoustic path impulse responses $g_i(z)$ in order to simulate the presence of a known desired signal during noise-only periods.

It can be shown [18] that the mean square error in the output is related to the degradation to the desired signal $\mathbf{D}_1$, the output noise power $\mathbf{D}_2$, and the amplitude of the fictitious desired signal A, according to

$$\overline{|e(n)|^2} = A^2 \cdot \mathbf{D}_1 + \mathbf{D}_2 \tag{106}$$

In addition, it can be shown that $\mathbf{D}_1$ and $\mathbf{D}_2$ are monotonically decreasing and increasing functions of A respectively. This has the powerful implication that the level of signal degradation and output noise power can be adjusted by varying a single parameter - namely the amplitude of the fictitious desired signal, A.

The third set of filters $\mathbf{h}_3$ are used to estimate the response degradation $\mathbf{D}_1$ in order to adapt the amplitude of the fictitious desired signal to achieve the desired levels of degradation and output noise power.

Full details of the algorithm are given in Kaneda [18], and further work is presented in Kaneda [19] and Kataoka *et al* [20]. The AMNOR technique has the limitations of requiring accurate speech/silence detection and knowledge of the impulse responses of the acoustic paths between the source and each microphone. Due to the fixed filters during speech periods, the technique implicitly assumes slowly-varying noise characteristics. In practice, the acoustic paths are modeled using simple time delays, as for delay-sum beamforming.

2.8 Post-filtering

In practice, the basic filter-sum beamformer seldom exhibits the level of improvement that the theory promises and further enhancement is desirable. One method of improving the system performance is to add a post-filter to the output of the beamformer.

Zelinski [21] proposed a Wiener post-filter formulated using the cross-spectral densities between channels in a microphone array. Incorporating a post-filter with a beamformer allows use of knowledge obtained in spatial filtering to also allow effective frequency filtering of the signal. In using both spatial and frequency domain enhancement, the use of information about the signal is maximised, where this knowledge is solely the direction of arrival of the signal.

The use of such a post-filter with a filter-sum microphone array was thoroughly investigated by Marro [22, 23] who demonstrated the mathematical interaction of the post-filter and the beamformer, and determined an optimal array structure for their combination. A diagram illustrating the system is presented in Figure 20.

At the output of the channel filters we have the time-aligned channel inputs

$$v_i(f) = w_i(f) x_i(f) \tag{107}$$

These signals contain an aligned version of the desired signal plus a noise component

$$v_i(f) = s(f) + n_i(f) \tag{108}$$

where s is the desired signal and n_i is the noise on microphone i.

The general Wiener filter expression for a microphone array is given as [22]

$$h_{opt}(f) = \frac{\Phi_{ss}(f)}{\Phi_{ss}(f) + \Phi_{\bar{n}\bar{n}}(f)} \tag{109}$$

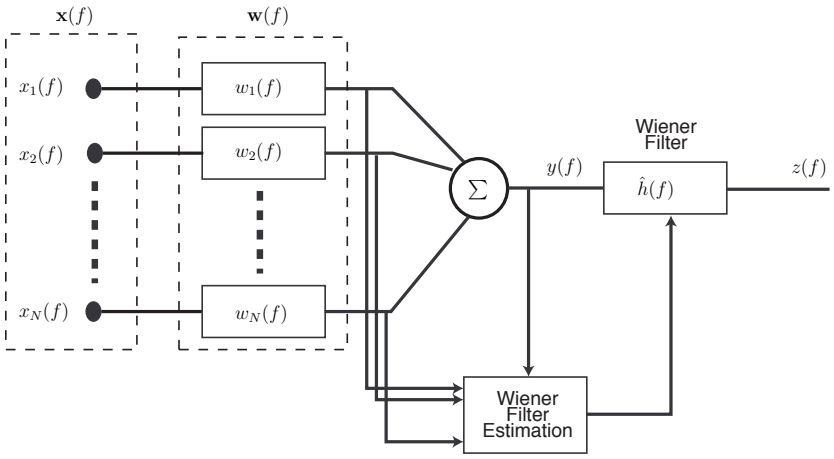


Figure 20: Filter-sum beamformer with post-filter

where $\Phi_{ss}(f)$ and $\Phi_{\bar{n}\bar{n}}(f)$ are respectively the auto-spectral density of the desired signal $s(f)$ and the noise at the output of the beamformer $\bar{n}(f)$.

A common problem with Wiener filters is the estimation of the signal and noise auto-spectral densities. The multi-channel approach provides an interesting solution to this problem. Under the assumptions that

1. The signal arriving at each microphone can be modeled by the sum of the desired signal and noise, according to Equation 108.
2. The noises $n_i(n)$ and desired signal $s(n)$ are uncorrelated.
3. The power spectral density of the noise is the same on each microphone $\Phi_{n_i n_i}(f) = \Phi_{nn}(f), i = 1, \ldots, N$.
4. The noises are uncorrelated between different microphones $\Phi_{n_i n_j}(f) = 0, i \neq j$.
5. The input signals $v_i(n)$ are restored into perfect phase alignment with $s(n)$.

we have

$$\Phi_{v_i v_i}(f) = \Phi_{ss}(f) + \Phi_{nn}(f) \tag{110}$$

and

$$\Phi_{v_i v_j}(f) = \Phi_{ss}(f) \tag{111}$$

and by averaging these spectral densities, we can estimate the Wiener filter equation as [22]

$$\hat{h}(f) = \frac{\sum_{i=1}^{N} |w_i(f)|^2}{\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_i(f) w_j^*(f)} \frac{\Re\{\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\Phi}_{v_i v_j}(f)\}}{\sum_{i=1}^{N} \hat{\Phi}_{v_i v_i}(f)} \tag{112}$$

The real operator $\Re\{\cdot\}$ is used because $\Phi_{ss}(f)$ is necessarily real. An incoherent noise field is the ideal condition for such a post-filter, however a diffuse noise field also provides a reasonable approximation of the above assumptions for the noise signals on different sensors. For this reason, the post-filter is best suited to incoherent or diffuse noise. The overall system output is given by

$$z(f) = \hat{h}(f) y(f) \tag{113}$$

where $y(f)$ is the beamformer output.

In Marro [22], equations are developed for the post-filter transfer function in terms of beamformer characteristics such as the noise reduction factor, signal to noise ratio and array gain for the following adverse input conditions :

- the presence of diffuse noise;
- the presence of a coherent noise source;
- a minor fault in the pointing direction of the array; and
- the presence of noise that is correlated with the desired signal.

By investigating the dependence of the post-filter upon these properties, it is shown that a post-filter enhances the beamformer output in the following ways :

- The post-filter cancels any incoherent noise.
- The post-filter further enhances the beamformer's rejection of coherent correlated or uncorrelated noise sources not emanating from the steered direction.
- The post-filter displays robustness to minor errors in the pointing of the array.

In summary, it is found that the effectiveness of such a post-filter follows that of the beamformer - if the beamformer is effective, the post-filter will further improve the system output. However, in the case where the beamformer is ineffective, the post-filter, being intrinsically linked to the beamformer performance, will be similarly ineffective.

2.9 Overview of Beamforming Techniques

This section summarises the important characteristics of the beamforming techniques discussed in this chapter. For each technique, Table 1 indicates whether or not it is a fixed or adaptive technique, the optimal noise conditions for its use, and whether the technique should be used with a broadside or endfire array configuration. Table 2 lists a number of key advantages and disadvantages of each technique.

Technique	**Fixed / adaptive**	**Noise condition**	**Array configuration**
Delay-sum	fixed	incoherent	broadside
Sub-array delay-sum	fixed	incoherent	broadside
Superdirectivity	fixed	diffuse	endfire
Near-field Superdirectivity	fixed	diffuse	endfire
Generalised Sidelobe Canceler	adaptive	coherent	broadside
AMNOR	adaptive	coherent	broadside
Post-filtering	adaptive	diffuse	either

Table 1: Properties of beamforming techniques

Technique	Advantages	Disadvantages
Delay-sum	simplicity	low frequency performance narrow-band
Sub-array delay-sum	broad-band	low frequency performance
Superdirectivity	optimised array gain	assumes diffuse noise
Near-field Superdirectivity	optimised array gain near-field sources low frequency performance	assumes diffuse noise assumes noise in far-field
Generalised Sidelobe Canceler	adapts to noise conditions minimises output noise power hard constraint on signal	low frequency performance can distort in practice
AMNOR	adapts to noise conditions minimises output noise power soft constraint on signal distortion level controlled	low frequency performance complexity speech-silence detection some distortion
Post-filtering	adapts to noise conditions improves beamformer output	can distort signal

Table 2: Advantages and disadvantages of beamforming techniques

While these tables give a simplistic overview of the different beamforming techniques, they serve to indicate the characteristics that must be considered when choosing a technique for a given application and noise conditions. For example, if the noise is approximately diffuse and there are no localised noise sources, then a superdirective technique is appropriate. If, however, prominent localised noise sources exist, then an adaptive technique would be advantageous. In applications where it is important to minimise distortion to the desired signal, fixed techniques are generally better than adaptive techniques. Also, depending on the location of the desired signal, a technique designed for the near-field may be required.

References

[1] S. Haykin, *Array Signal Processing*. Prentice-Hall, 1985.

[2] L. J. Ziomek, *Fundamentals of Acoustic Field Theory and Space-Time Signal Processing*. CRC Press, 1995.

[3] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Prentice Hall, 1993.

[4] D. C. Moore, "Speech enhancement using microphone arrays," Master's thesis, Queensland University of Technology, 2000.

[5] B. D. Steinberg, *Principles of Aperture and Array System Design*. John Wiley and Sons, 1976.

[6] E. Ifeachor and B. Jervis, *Digital Signal Processing : A Practical Approach*. Addison-Wesley, 1996.

[7] D. Ward, *Theory and Application of Broadband Frequency Invariant Beamforming*. PhD thesis, Australian National University, July 1996.

[8] R. L. Bouquin and G. Faucon, "Using the coherence function for noise reduction," *IEE Proceedings*, vol. 139, pp. 276–280, June 1992.

[9] D. Templeton and D. Saunders, *Acoustic Design*. London: Architectural Press, 1987.

[10] J. Bitzer, K. Kammeyer, and K. U. Simmer, "An alternative implementation of the superdirective beamformer," in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, (New York), pp. 991–994, October 1999.

[11] H. Cox, R. Zeskind, and T. Kooij, "Practical supergain," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-34, pp. 393–397, June 1986.

[12] R. Taylor and G. Dailey, "The super-directional acoustic sensor," in *Proceedings of OCEANS '92 - Mastering the Oceans through Technology*, vol. 1, pp. 386–391, 1992.

[13] H. Cox, R. Zeskind, and M. Owen, "Robust adaptive beamforming," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 35, pp. 1365–1376, October 1987.

[14] W. Täger, "Near field superdirectivity (NFSD)," in *Proceedings of ICASSP '98*, pp. 2045–2048, 1998.

[15] W. Tager, *Etudes en Traitement d'Antenne pour la Prise de Son*. PhD thesis, Universite de Rennes 1, 1998. in french.

[16] O. L. Frost, "An algorithm for linearly constrained adaptive array processing," *Proceedings of the IEEE*, vol. 60, pp. 926–935, August 1972.

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

[18] Y. Kaneda, "Adaptive microphone-array system for noise reduction," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-34, pp. 1391–1400, December 1986.

[19] Y. Kaneda, "Directivity characteristics of adaptive microphone-array for noise reduction (amnor)," *Journal of the Acoutical Society of Japan*, vol. (E) 12, no. 4, pp. 179–187, 1991.

[20] A. Kataoka and Y. Ichinose, "A microphone-array configuration for amnor (adaptive microphone-array system for noise reduction)," *Journal of the Acoutical Society of Japan*, vol. 11, no. 6, pp. 317–325, 1990.

[21] R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," in *Proceedings of ICASSP-88*, vol. 5, pp. 2578 –2581, 1988.

[22] C. Marro, *Traitements de Dereverberation et de Debruitage Pour le Signal de Parole dans des Contextes de Communication Interactive*. PhD thesis, Universite de Rennes 1, 1996. in french.

[23] C. Marro, Y. Mahieux, and K. Uwe Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering," *IEEE Transactions on Speech and Audio Processing*, vol. 6, pp. 240–259, May 1998.