1

Introduction

Signal enhancement is a process to either restore a signal of interest or boost the relevant information embedded in the signal of interest and suppress less relevant information from the observation signals. Today, there is almost no field of technical endeavor that is not impacted in some way by this process.

Array signal processing manipulates the signals picked up by the sensors that form an array in order to estimate some specific parameters, enhance a signal of interest, or make a particular decision. The main purpose of this chapter is to:

- define the scope of the field that we call signal enhancement
- present a brief historic overview of this topic
- give some examples of fields where signal enhancement is needed and used
- discuss briefly the principal approaches to signal enhancement
- explain how array signal processing works.

1.1 Signal Enhancement

We human beings rely on our senses to sense the environment around us. Based on this information we build and expand intelligence in our brain to help make decisions and take actions. Similarly, we strive to build systems to help us "see" or "hear" distant events that cannot be reached by our senses. For example, nowadays, sonar systems can hear ships across hundreds of miles of ocean, radar devices can see airplanes from a thousand miles over the horizon, telecommunication systems can connect two or more users from different corners of the world, and high-definition cameras can see events happening on our planet from space. These systems use sensors to measure the physical environment of interest. Signal processing is then applied to extract as much relevant information as possible from the sensors' outputs. Generally, sensors' outputs consist of the signal of interest, which carries very important information, and also a composition of unwanted signals, which is generally termed "noise". This does not contain useful information but interferes with the desired signal. To extract the useful information in the presence of noise, signal enhancement is needed, the objective of which is to:

- enhance the signal-to-noise ratio (SNR)
- restore the signal of interest

Fundamentals of Signal Enhancement and Array Signal Processing, First Edition.
Jacob Benesty, Israel Cohen, and Jingdong Chen.

Companion website: www.wiley.com/go/benesty/arraysignalprocessing

- boost the relevant information while suppressing less relevant information
- improve the performance of signal detection and parameter estimation.

Signal enhancement is a specialized branch of signal processing that has been around for many decades and has profound impact on many fields. In the following subsections, we describe a few areas that routinely use signal enhancement techniques, particularly those developed in the following chapters of this text. Note that we can only cover a few applications, but this should leave the reader with no doubt as to the importance and breadth of application of signal enhancement techniques.

1.1.1 Speech Enhancement and Noise Reduction

In applications related to speech acquisition, processing, recognition, and communications, the speech signal of interest (generally called the "desired speech") can never be recorded in a pure form; it is always immersed in noise. The noise can come from very different sources. For example, microphones that we use to convert acoustic pressure into electronic signals have self-noise, even though the noise floor of popularly used capacitor microphones has been dropping significantly over the years. The associated digital signal processing boards, including preamplifiers, analog-to-digital (A/D) converters, and processors for processing the signals, may also generate noise. Most importantly, noise comes from ambient sources; the environment where we live is full of different kinds of sounds. While the sensors' self and circuit noise is generally white in spectrum, the noise from sound sources in the surrounding environment can vary significantly from one application scenario to another.

Commonly, noise from acoustic environments can be divided into the following four basic categories depending on how the noise is generated:

- *Additive noise* can come from various sources, such as cooling fans, air conditioners, slamming doors, and passing traffic.
- *Echoes* occur due to the coupling between loudspeakers and microphones.
- *Reverberation* is the result of multipath propagation and is introduced by reflections from enclosure surfaces.
- *Interference* comes from concurrent sound sources. In some communication applications, such as teleconferencing, it is possible that each communication site has multiple participants and loudspeakers, so there can be multiple competing sound sources.

Combating these four categories of noise has led to the development of diverse acoustic signal processing techniques. They include noise reduction (or speech enhancement), echo cancellation and suppression, speech dereverberation, and source separation, each of which is a rich subject of research [1–3]. This text presents many methods, algorithms, and techniques that are useful in dealing with additive noise, reverberation, and interference while its major focus, particularly the signal enhancement part from Chapter 2 to Chapter 6, is on reducing additive noise.

Additive noise and the desired speech signal are in general statistically independent. While the noise does not modify the speech characteristics directly, the characteristics of the observation signal are very different from those of the desired speech since it is a mixture of the desired speech signal and noise. Figure 1.1 plots a speech signal recorded in an anechoic (quiet and non-reflective) environment and the same speech

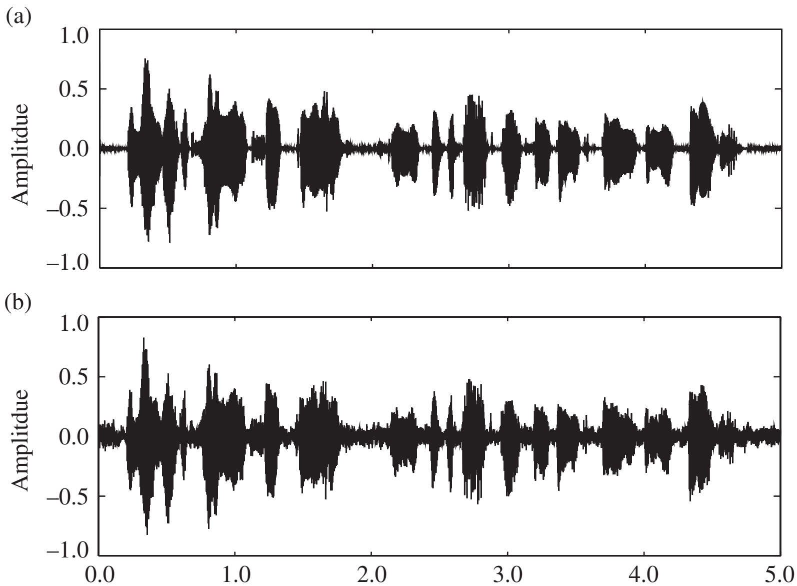

Figure 1.1 (a) A speech signal recorded by a microphone in an anechoic environment and (b) the same speech signal recorded by the same microphone but in a conference room.

signal but recorded in a conference room. The spectrograms of these two signals are shown in Figure 1.2. As can be seen, both the waveform and the spectrogram of the noisy signal are dramatically different from those of the clean speech. The effect of noise may dramatically affect the listener's perception and also machine processing of the observed speech. It is therefore generally required to "clean" the observation signal before it is stored, transmitted, or played (through a loudspeaker, for example). This problem is generally referred to as either noise reduction or speech enhancement.

1.1.2 Underwater Acoustic Signal Enhancement

Over the last few decades, ocean exploration activity for both military and civilian interests has been steadily increasing. As a result, there has been growing demand for underwater communication and signal detection and estimation technologies. Electromagnetic and light waves do not propagate over long distances under water (particularly sea water). In contrast, acoustic waves may propagate across tens or even hundreds of miles under the sea. Therefore, acoustic waves have played an important role in underwater communication and signal detection and estimation. For example, passive sonar systems can detect a submarine from tens of miles away by listening to the sound produced by the submarine, such as from the propellers, engine, and pumps; active sonars transmit sound pulses into the water and listen to the echoes, thereby detecting underwater features such as the location of fish, sunken objects, vessels, and submarines. Underwater wireless communication systems modulate useful information on acoustic carriers with frequencies between a few kilohertz and a few tens of kilohertz and transmit the modulated signal from one end to another through underwater acoustic channels.

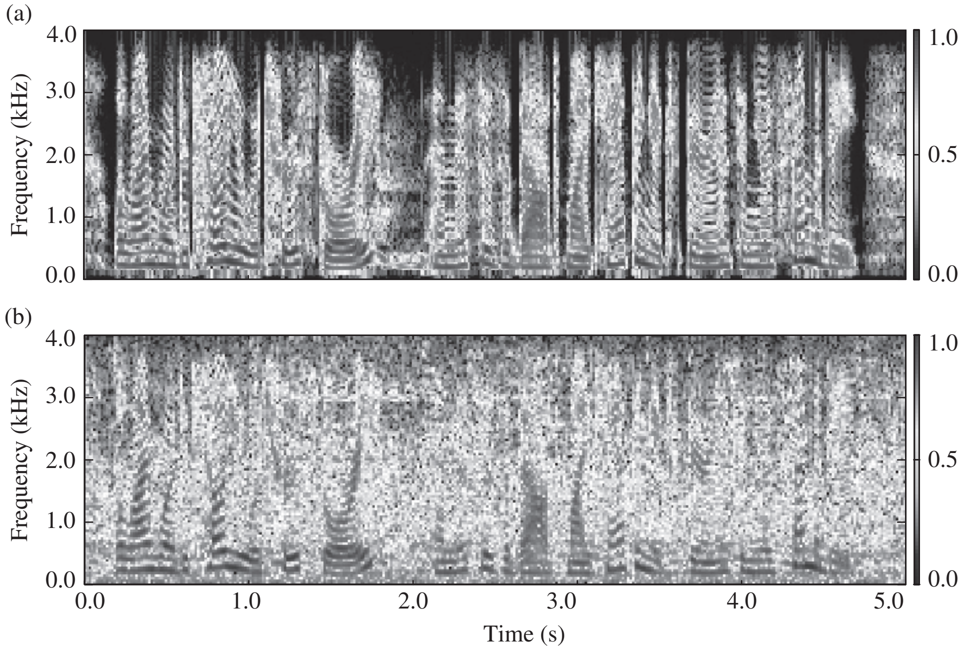

Figure 1.2 (a) The spectrogram of the speech signal in Figure 1.1a; (b) the spectrogram of the speech signal in Figure 1.1b.

However, processing underwater acoustic signals is by no means an easy task. First of all, underwater acoustic channels are generally known as one of the most difficult communication media in use today. Underwater acoustic propagation suffers from the time-varying multipath effect (due to sound reflection at the surface, bottom, and any objects in the vicinity, and also sound refraction in the water), frequency-dependent attenuation (due to absorption and signal spreading loss), and a severe Doppler effect (due to the low speed of sound and motion of the transmitter or receiver or the objects to be detected). Secondly, the ocean is filled with sounds, which interfere with the acoustic signal we are interested in. Underwater sounds are generated by both natural sources, such as marine animals, breaking waves, rain, cracking sea ice, and undersea earthquakes, as well as man-made sources, such as ships, submarines, and military sonars.

Marine animals use sound to obtain detailed information about their surroundings. They rely on sound to communicate, navigate, and feed. For example, dolphins can detect individual prey and navigate around objects underwater by emitting short pulses of sound and listening to the echo. Marine mammal calls can increase ambient noise levels by 20–25 dB in some locations at certain times of year. Blue and fin whales produce low-frequency moans at frequencies of 10–25 Hz, with estimated source levels of up to 190 dB at 1 m. Sounds generated by human activities are also an important part of the total ocean noise. Undersea sound is used for many valuable purposes, including communication, navigation, defense, research and exploration, and fishing.

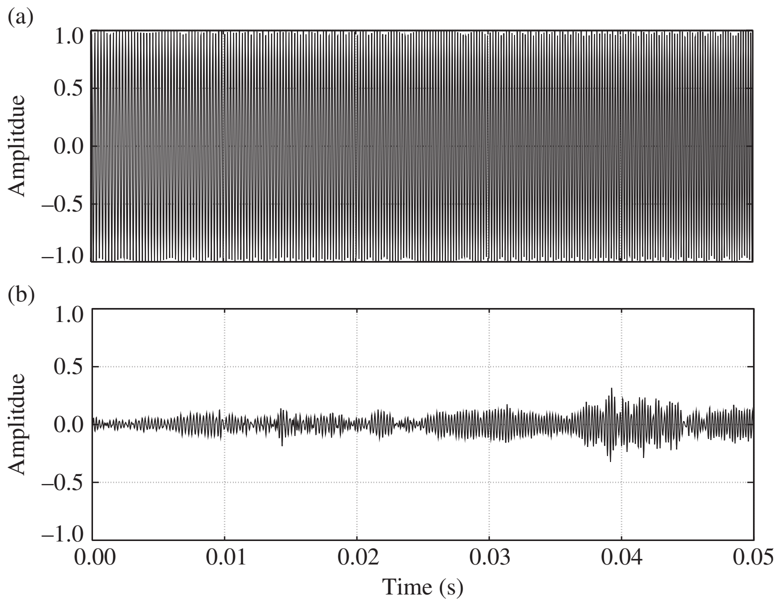

Figure 1.3 A linear frequency modulated signal: (a) emitted by a transmitter of an underwater acoustic communication system and (b) received by a hydrophone six miles away from the transmitter in an underwater environment.

Sounds generated by human activities cover a wide range of frequencies, from a few hertz up to several hundred kilohertz, and a wide range of source levels.

The underwater channel condition and noise sets the ultimate limit on the minimum detectable signal in detection and communication systems. To illustrate how challenging it is to process underwater signals for extracting the useful information, Figure 1.3 plots a linear frequency modulation chirp signal transmitted by an acoustic antenna and the signal received by a hydrophone placed six miles away from the transmitter. The magnitude spectra of these two signals are plotted in Figure 1.4. As seen, the transmitted signal is dramatically distorted by the acoustic channel and noise. Sophisticated signal enhancement techniques, such as those developed in this text, are needed to extract the important parameters or information embedded in the transmitted signal from the received signal.

1.1.3 Signal Enhancement in Radar Systems

A radar system has a transmitter that emits electromagnetic waves (called radar signals) in look directions. When these waves come into contact with an object, they are usually reflected or scattered in many directions. Receivers (usually, but not always, in the same location as the transmitter) are then used to receive the echoes. Through processing the echoes, the radar can determine the range, angle, or velocity of the objects of interest.

The invention of the radar dates back to the late 19th century. Such systems are now used in a broad range of applications, including air defense, traffic control, aircraft anticollision, ocean surveillance, geological observations, meteorological precipitation monitoring, and autonomous cars. In order to estimate the range, angle, or velocity of

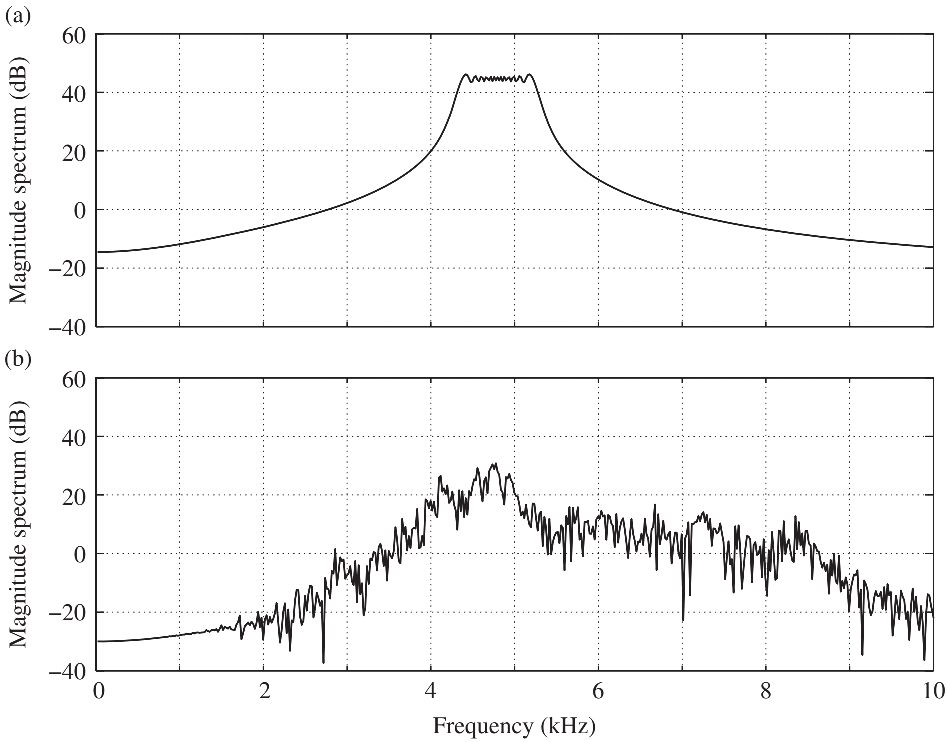

Figure 1.4 The power spectrum of the signal in Figure 1.3a; (b) the power spectrum of the signal in Figure 1.3b.

the objects of interest, radar systems must overcome unwanted signals, which can be divided into the following three categories.

- *Additive noise* is generated by both internal sources (electronics) and external sources (the natural thermal radiation of the background surrounding the target of interest). In modern radar systems, the internal noise is generally lower than the external noise.
- *Clutter* is a term used for echoes returned from targets that are not useful to the radar system user. Clutters can be generated by irrelevant targets, natural objects such as the ground, sea, atmospheric turbulence, ionospheric reflections, and man-made objects such as buildings, as illustrated in Figure 1.5.
- *Jamming* refers to signals received by the radar on its own frequency band but emitted from outside sources. Jamming may be intentional, as with an electronic warfare tactic, or unintentional, as with friendly forces' using equipment that transmits using the same frequency range. It is problematic to radar since the jamming signal only needs to travel one way – from the jammer to the radar receiver – whereas the radar echoes travel two ways – from radar to target and to radar – and are therefore significantly reduced in power by the time they return to the radar receiver. Therefore, jammers can effectively mask targets along the line of sight from the jammer to the radar, even when they are much less powerful than the jammed radars.

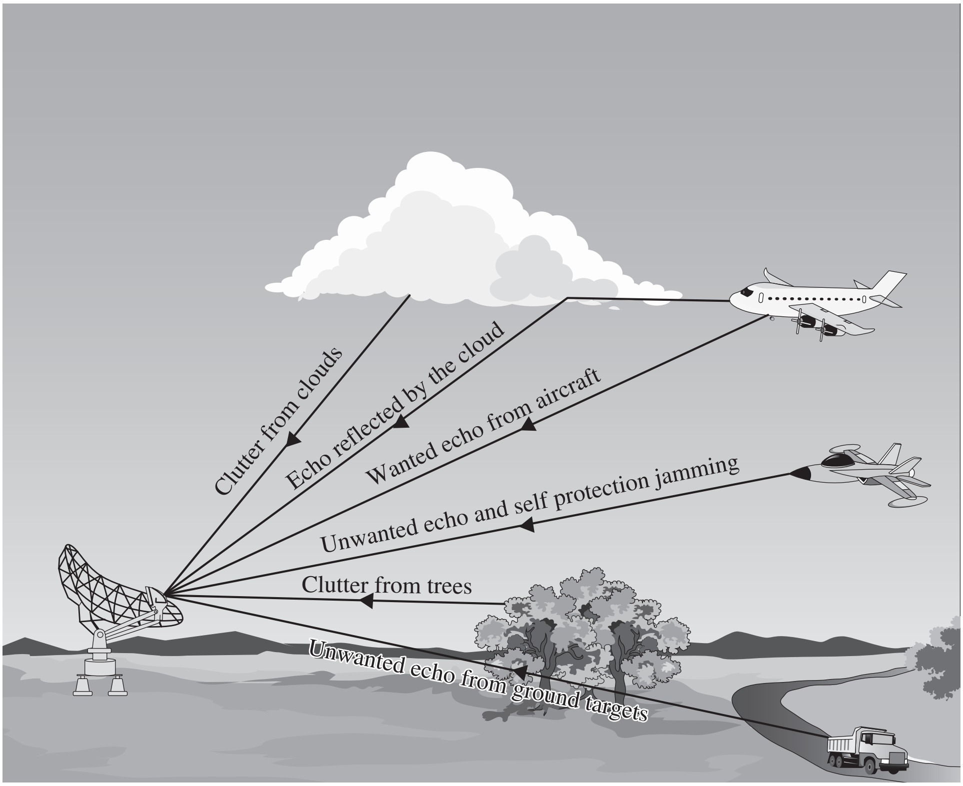

Figure 1.5 An illustration of a radar system and its environments.

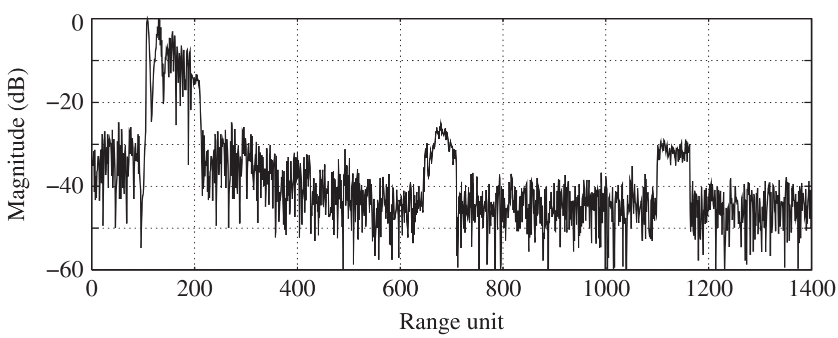

Figure 1.6 Normalized magnitude of an echo received by a pulse radar.

Figure 1.6 plots the magnitude of a signal received by a pulse radar where the transmitted signal is a short pulse. The received signal is composed of an echo returned from a target of interest, two unwanted echoes, and some noise. Figure 1.7 shows a radar image directly mapped from the received signals without using any signal enhancement techniques. Without signal enhancement, it is difficult to determine the position of one target, let alone track multiple targets with high resolution. Therefore, signal

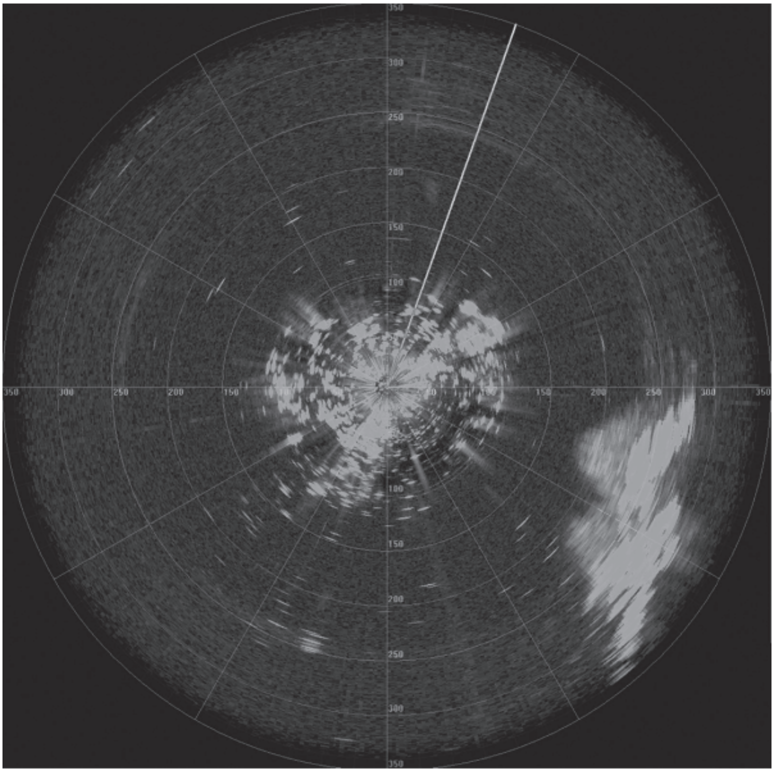

Figure 1.7 Illustration of an image displayed in a radar screen without using signal enhancement.

enhancement techniques, particularly those developed in Chapters 7–11, are needed to deal with additive noise, clutter, and jamming in radar systems. This is done by estimating the important parameters embedded in the radar echo signals.

1.1.4 Signal Enhancement in Ultrasound Systems

Ultrasound refers to sound waves with frequencies greater than 20 kHz, a level which is commonly accepted to be the upper limit of human hearing. This type of high-frequency sound wave is used in many fields for a wide range of applications, such as non-intrusive testing of products and structures, invisible flaw detection, distance measurement, and medical diagnosis, to name just a few. One of the best known ultrasound systems is the sonography instrument that is used in medicine to examine many of the body's internal organs, including – but not limited to – the heart and blood vessels, liver, gallbladder, pancreas, kidneys, and uterus, as well as unborn children (fetus) in pregnant patients.

Typically, a sonography device consists of an array of transmitters, which send short, high-frequency (generally between 2 and 20 MHz) sound pulses into the body. Beamforming is applied to the transmitted pulses so that the ultrasound waves are focused towards a particular point. As the beamformed waves travel toward the desired focal point, they propagate through materials with different densities. With each change in density, reflected waves are produced, some of which propagate back. They are then collected by an array of receivers (the transmitters typically become sensors to receive signals once they have finished generating their respective sound waves). The signal received by each receiver is composed of the wanted echoes and noise. Commonly, noise in sonography is one of two major types:

- *Additive noise* is generated from the sensors, amplifiers, A/D converters, and other electronic system components. It can also come from sources such as background tissues, other organs and anatomical influences, and breathing motion. Generally, this type of noise is independent (or weakly dependent) on the echo signals and is often modeled mathematically as a white Gaussian noise process.
- *Speckle* is the result of three sound scattering effects: specular, diffusive, and diffractive. Specular scattering occurs when the scattering object is large compared to the sound wavelength; diffusive scattering happens when the scattering object is small relative to the wavelength; diffractive scattering occurs mostly for medium-size scattering objects. Unlike additive noise, speckles are generally correlated with the wanted echo signals.

To deal with additive noise and speckles in sonography, beamforming, noise reduction, speckle reduction, and many other enhancement processes are applied to the received signals before high-resolution two-dimensional images are formed to display the distances and intensities of the echoes on the screen.

1.2 Approaches to Signal Enhancement

Signal enhancement is one of the most interesting and appealing yet challenging areas of signal processing. Its objective is generally problem oriented, ranging from simply improving the SNR, boosting relevant information, restoring the signal of interest, to improving measures of which only human subjects can judge the quality. As a result, there is no general rule as what method is optimal and it is quite common that a method that produces the best enhancement result for one application may not be very useful for another. In general, signal enhancement techniques can be classified into one of four broad categories, depending on how the information embedded in the signal and noise are used:

- *time-domain methods*, which directly use temporal information
- *frequency-domain approaches*, which operate on spectra (obtained using the Fourier transform or other time-to-frequency-domain transformations) of a signal
- *spatial-domain techniques*, which acquire and process a signal of interest using an array of sensors
- *combinational methods*, which use temporal, spectral, and spatial information.

Time-domain methods typically achieve signal enhancement by applying a finite-impulse-response filter to the noisy signal that is observed at a sensor. So, the core problem of enhancement is converted to one of designing an optimal filter that can attenuate noise as much as possible while keeping the signal of interest relatively unchanged. The history of this class of methods dates back to the seminal work by Wiener [4], in which the optimal filter from a second-order-statistics viewpoint is achieved through the optimization of the classical mean-squared error (MSE) criterion. The Wiener filter is well known, and has been intensively investigated for signal enhancement [5–8]. However, while it is optimal from the MSE point of view, it introduces signal distortion if the signal to be enhanced is broadband. The amount of signal distortion may not be acceptable for some applications. If this is the case, one may consider using some suboptimal filters that minimize the MSE criterion under certain constraints [5, 6].

The Wiener technique, by its assumption, can only deal with stationary signals. One popularly used approach to extending the Wiener filter to deal with nonstationary signals is to relax the stationarity assumption to one of short-time stationarity. Then, the Wiener filter is computed using signals within only a short-time, sliding window. In this case, the length of the short-time window plays an important role on the tradeoff between the nonstationarity and performance within the short-time window. There are, of course, other ways to deal with signal enhancement of nonstationary signals, for example, combining the Kalman filter and the linear-prediction-coding method [9]. Comprehensive coverage of signal enhancement using temporal information will be given in Chapter 2.

Frequency-domain methods, as the name indicates, explicitly operate on the spectra of the signal to be processed. The root of this class of methods can be traced back to the 1920s, when low-pass, high-pass, and band-pass filters were invented to filter out noise that occupies different frequency bands to the signal of interest. Today, the basic principle of band-pass filtering is still widely used in signal enhancement, but often in more complicated forms, such as comb filters [10] and binary masking [11]. Band-pass filtering, comb filtering, and binary masking are hard-decision methods in the sense that, given a narrow frequency band, they either completely remove the signal component or keep it unchanged. In comparison, a soft decision can be achieved through the spectral enhancement method, which was first developed in the 1960s using analog circuits [12]. A digital-domain version of this method, which is called "spectral subtraction", was then developed in the late 1970s [13]. While it is very useful and is often used as a benchmark against which other techniques are compared, the spectral substraction method has no optimality properties associated with it. An optimal spectral enhancement framework was developed in the early 1980s [14]. This unified a broad class of enhancement algorithms, including spectral substraction, frequency-domain Wiener filtering, and maximum likelihood envelope estimator. Following this work, an optimal spectral amplitude estimator using statistical estimation theory was developed in the early 1980s. Following this work, many statistical spectral estimators were developed, including the minimum mean-squared error (MMSE) estimator [15], the MMSE log-spectral amplitude estimator, the maximum-likelihood (ML) spectral amplitude estimator, the ML spectral power estimator, and the maximum a posteriori (MAP) spectral amplitude estimator. Today, there are still tremendous efforts to find better spectral amplitude estimators, inspired by the work of McAulay and Malpass [14] and Ephraim and Malah [15]. A broad coverage of frequency-domain methods will be given in Chapter 3.

When multiple sensors are used, the spatial information embedded in the sensors' outputs can be exploited to enhance the signal of interest and reduce unwanted noise. This can be done in a straightforward way by extending the single-channel methods of Chapters 2 and 3 to the multichannel cases (see, respectively, Chapters 4 and 5). It can also be done in a different way through array beamforming, which will be discussed in the next section.

1.3 Array Signal Processing

An array consists of a set of sensors positioned at known locations with reference to a common point. The sensors collect signals from sources in their own field of view and the output of each sensor is composed of these source components as well as noise. By processing the sensors' outputs, two groups of functionalities can be achieved: estimation of important parameters of sources and enhancement of some signals of interest.

The history of array signal processing dates back to World War II. Early efforts in this field were mainly focused on parameter estimation: estimating the range, angle, and velocity of the sources of interest. A wide range of processing methods were developed to this end, including fixed beamforming (or spatial filtering), matched filtering, Capon's adaptive beamforming, the MUSIC (Multiple SIgnal Classification) method and its varieties, and the ESPRIT (Estimation of Signal Parameters by Rotational Invariance Techniques) algorithm, to name but a few. The reader is referred to the literature for an in-depth consideration of the problem of array parameter estimation and the associated methods [16–23]. In this book, we choose to focus on the signal enhancement problem with the use of sensor arrays.

The basic principle of signal enhancement using an array of sensors can be illustrated in Figure 1.8. Consider a simple example with a uniformly spaced linear array of M sensors and assume that there is a single source in the farfield such that its spherical wavefront appears planar at the array. If we neglect the propagation attenuation, the signals received at the M sensors can be written as

$$\begin{aligned} Y_m(f) &= X_m(f) + V_m(f) \\ &= X(f)e^{-\jmath 2\pi f(t+\tau_{m-1})} + V_m(f) \\ &= X_1(f)e^{-\jmath 2\pi f\tau_{m-1}} + V_m(f),\ m = 1, 2, \dots, M, \end{aligned} \tag{1.1}$$

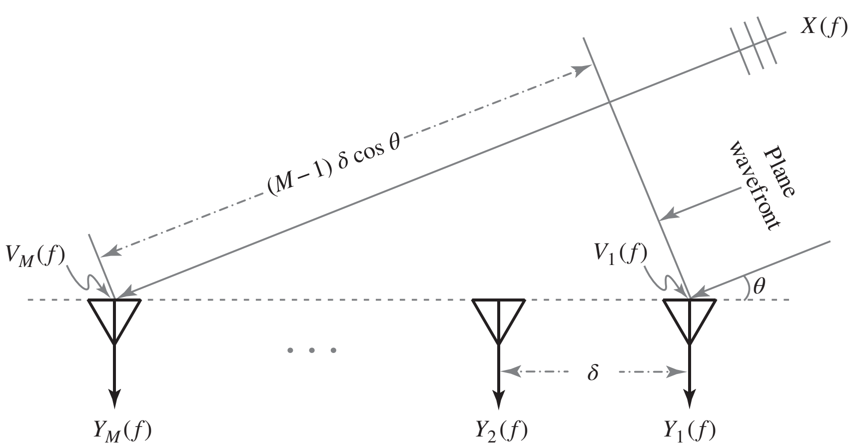

Figure 1.8 Illustration of an array system for signal enhancement.

where f is frequency, t is the propagation time from the source $X(f)$ to sensor 1 (the reference sensor), $X_1(f) = X(f)e^{-j2\pi ft}$ is the signal component at sensor 1, τ_{m-1} is the relative time delay between the mth sensor and the reference one, and $j = \sqrt{-1}$ is the imaginary unit. It is assumed that $X(f)$ is uncorrelated with $V_m(f)$, $m = 1, 2, \dots, M$. With a uniform linear array (ULA) and a farfield source, the delay τ_{m-1} can be expressed in the following form according to the geometry shown in Figure 1.8:

$$\tau_{m-1} = (m-1)\delta \cos\theta / c, \; m = 1, 2, \dots, M, \tag{1.2}$$

where δ is the spacing between two neighboring sensors, c represents velocity of wave propagation, and θ is the signal incidence angle.

Now let us consider processing the M signals $Y_m(f)$, $m = 1, 2, \dots, M$, to extract the source signal $X(f)$ (up to a delay) and reduce the effect of $V_m(f)$. A straightforward way of doing this is to multiply $Y_m(f)$ by $e^{j2\pi f \tau_{m-1}}$, and then average the results. This gives an output:

$$\begin{aligned} Z(f) &= \frac{1}{M} \sum_{m=1}^{M} Y_m(f) e^{j2\pi f \tau_{m-1}} \\ &= X_1(f) + \frac{1}{M} \sum_{m=1}^{M} V_m(f) e^{j2\pi f \tau_{m-1}}. \end{aligned} \tag{1.3}$$

To check whether the output $Z(f)$ is less noisy than the input, let us compare the input and output SNRs. The input SNR, according to the signal model given in (1.1), is defined as the SNR at the reference sensor:

$$\mathrm{iSNR}(f) = \frac{\phi_{X_1}(f)}{\phi_{V_1}(f)}, \tag{1.4}$$

where $\phi_{X_1}(f) = E\left[|X_1(f)|^2\right]$ and $\phi_{V_1}(f) = E\left[|V_1(f)|^2\right]$ are the variances of $X_1(f)$ and $V_1(f)$ respectively, and $E[\cdot]$ denotes mathematical expectation.

The output SNR – the SNR of the $Z(f)$ signal – is written as

$$\mathrm{oSNR}(f) = \frac{\phi_{X_1}(f)}{\frac{1}{M^2} E\left[\left|\sum_{m=1}^{M} V_m(f) e^{j2\pi f \tau_{m-1}}\right|^2\right]}. \tag{1.5}$$

Now, if all the noise signals $V_m(f)$, $m = 1, 2, \dots, M$, are uncorrelated with each other and have the same variance, it is easy to check that

$$E\left[\left|\sum_{m=1}^{M} V_m(f) e^{j2\pi f \tau_{m-1}}\right|^2\right] = M \times E\left[|V_1(f)|^2\right]. \tag{1.6}$$

It follows immediately that $\mathrm{oSNR}(f) = M \times \mathrm{iSNR}(f)$. Thus, a simple phase shifting and averaging operation of the sensors' outputs results in an SNR improvement by a factor of M, the number of sensors. The underlying physical principle behind the

SNR improvement can be explained as follows. Through appropriate phase shifting, the signal components from the source of interest have been coherently combined while the noise signals from different sensors add only incoherently as they are uncorrelated with each other, yielding a gain for the overall output signal compared to the noise.

Of course, the above example is only a particular case. More generally, an estimate of the source signal $X(f)$ can be obtained through weighted linear combination of the sensors' outputs:

$$Z(f) = \sum_{m=1}^{M} H_m^*(f) Y_m(f), \tag{1.7}$$

where $H_m(f)$, $m = 1, 2, \ldots, M$, are complex weighting coefficients. The process of finding the appropriate values of $H_m(f)$ so that $Z(f)$ is a good estimate of $X(f)$ is called beamforming. Therefore, the coefficients $H_m(f)$ are also called beamforming coefficients and the vector that consists of all the coefficients is called the beamforming filter or beamformer.

Beamforming has been the central problem of array signal processing ever since sensor arrays were invented, and a large number of algorithms has been developed and described in the literature. By and large, the developed algorithms fall into two major categories – fixed or adaptive beamforming – depending on whether the noise or signal statistics are considered in forming the beamforming filters.

In fixed beamforming, the beamforming filters are designed explicitly using the array geometry information as well as the assumed knowledge of the look direction and noise statistics. Once computed, the coefficients of the beamforming filters will be fixed regardless of the particular application environment. It is for this reason that this design process is called fixed beamforming. The representative algorithms include the delay-and-sum beamformer, the maximum directivity factor beamformer, and the superdirective beamformer. The reader may find a discussion of these algorithms in different contexts and applications in the literature [3, 16–18]. The basic theory and methods for designing fixed beamformers from a narrowband perspective will be covered in Chapter 7. While the principles presented in this chapter are rather general, in designing optimal fixed beamformers, the resulting beamformers may be insufficient to deal with broadband signals as their beampatterns may vary with frequency. Chapters 9 and 10 are also concerned with fixed beamforming, but with focus on processing broadband signals where beampatterns are expected to be the same across a band of frequencies.

In comparison with fixed beamforming, adaptive beamforming algorithms consider using either the noise statistics or the statistics of the array observation data to optimize the beamforming filters. The performance of adaptive beamforming can be more optimal than its fixed counterpart as long as the signal statistics are correctly estimated. The representative algorithms in this category include the minimum variance distortionless response (MVDR) beamformer, which is also known as the Capon's beamformer [27], the linearly constrained minimum variance (LCMV) beamformer, which is also called the Frost's beamformer [28], and the generalized sidelobe canceller [29, 30]. Many applications of adaptive beamforming can be found in the literature [1–3, 31, 32]. The fundamental theory and methods for adaptive beamforming from a frequency-domain perspective will be covered in Chapter 8.

While one may see that most discussion on beamforming in both the literature and this text concerns the frequency domain, it is also possible to formulate this problem in the time domain. Chapter 10 is devoted to a time-domain framework for array beamforming, which can be used to design both fixed and adaptive beamformers as well as narrowband and broadband beamformers.

1.4 Organization of the Book

This book attempts to cover the most basic concepts, fundamental principles, and practical methods of signal enhancement and array beamforming. The material discussed occupies ten chapters, which are divided into two parts.

The first part, Signal Enhancement, consists of five chapters: from Chapter 2 to Chapter 6. We start to discuss the signal enhancement problem in the time domain with a single sensor in Chapter 2. With a single sensor, we show how to exploit the temporal information so as to reduce the level of the additive noise from the sensor's output, thereby enhancing the signal of interest. This chapter presents two fundamental approaches: one deals with the problem from the Wiener filtering perspective and the other from a spectral mode perspective based on the joint diagonalization of the correlation matrices of the signal of interest and the noise. In both approaches, we present the problem formulation and performance measures that can be used to evaluate the signal enhancement performance. Different cost functions are also presented, and based on these we discuss how to derive useful optimal enhancement filters.

Chapter 3 continues the investigation of the single-channel signal enhancement problem, but in the frequency domain. The frequency-domain approach is equivalent to the spectral method discussed in Chapter 2 in the sense that the observation signal at each frequency band can be processed independently from the others. The advantage of the algorithms in this chapter is that they can all be implemented efficiently thanks to the use of the fast Fourier transform. Again, we start from problem formulation and then discuss how to perform signal enhancement with just simple gains at each frequency band. Relevant performance measures are defined and we show how to derive several kinds of enhancement gain, some of which can achieve a compromise between distortion of the desired signal and reduction of the additive noise.

Chapter 4 is basically an extension of Chapter 2. The fundamental difference is that in this chapter we consider the signal enhancement problem with the use of multiple sensors, which are located at distinct positions in the space. In this case, every sensor picks up the signal of interest and noise from its own viewpoint. Now, in addition to the temporal information, the spatial information from the multiple sensors can also be exploited to enhance the signal of interest. As a result, either a better enhancement performance or more flexibility to compromise between noise reduction and desired signal distortion can be achieved, as compared to the single-channel scenario. Similar to Chapter 2, we also discuss two approaches: the Wiener filtering one and the one based on the joint diagonalization of the correlation matrices of the signal of interest and noise.

Chapter 5 deals with the problem of signal enhancement with multiple sensors in the frequency domain. As in Chapter 4, the spatial information embedded in the multiple sensors is exploited to enhance the signal of interest, but in a way that is easier to comprehend. Just like the material in the previous chapters, we present the signal model, problem formulation, and performance measures, and show how to derive different optimal filters. We also discuss the problem in a subspace framework, which is an alternative way to approach the enhancement problem.

Chapter 6 is a unification of the material presented from Chapter 2 to Chapter 5. A general framework is presented here so that the signal enhancement problem in either the time or the frequency domain, with either one sensor or multiple sensors, is treated in a unified framework. Within this framework, we derive a class of optimal linear filters, some of which are well known and some can achieve output signal-to-interference-plus-noise ratios (SINRs) that are between those of the conventional maximum SINR filter and the Wiener filter. This chapter also serves as a bridge between the problem of noise reduction in the previous chapters and the following chapters, on beamforming.

The second part, Array Signal Processing, is about beamforming, and also consists of five chapters, from Chapter 7 to Chapter 11. We start in Chapter 7 by discussing the theory and methods of beamforming from fixed beamformers, which are spatial filters that have the ability to form a main beam pointing in the direction of the signal of interest, while placing sidelobes and nulls in directions other than the look direction. By "fixed", we mean that the beamforming filters are designed before the deployment of the array system and the filters' coefficients do not depend on the array output. Generally, the design of fixed beamformers requires the array geometry information, such as the number of sensors or the location of every sensor relative to a reference point, and the look direction. It is helpful if the directions of interference sources are known as well. To simplify the presentation of the main results, we consider in Chapter 7 only ULAs, and study a number of popularly used fixed beamformers. Note that the generalization of the algorithms in this chapter from ULAs to other geometries is not difficult, in general.

Fixed beamformers are generally robust and easy to implement; but they are at best suboptimal in terms of noise and interference rejection, as neither the statistics of the signal of interest nor those of noise and interference are considered in the beamformer design process. One way to improve the performance of fixed beamformers is through using the statistics of the array outputs or *a priori* information about the source or noise signals, leading to the so-called "adaptive" beamformers, which will be studied in Chapter 8. This chapter discusses several interesting adaptive beamformers, including their derivation, underlying principles, as well as their equivalent forms from a theoretical viewpoint.

In processing broadband signals such as audio and speech, it is desirable, if not a must, to use beamformers that have frequency-invariant beampatterns. One way to achieve this is through differential beamforming, which will be studied in Chapter 9. Differential beamforming is a particular kind of fixed beamforming. It differs from those beamformers in Chapter 7 in that it attempts to measure the differential pressure field of different orders, instead of designing a special beampattern. Besides the property of frequency-invariant beampatterns, differential beamforming can achieve the maximum directivity factor, leading to the highest gains in diffuse noise. As a matter of fact, the so-called "superdirective" beamformer is a particular case of the differential beamformer. In this chapter, we present the fundamental principles underlying differential beamforming, as well as approaches to designing differential beamformers of different orders. One main drawback of differential beamforming as compared to those beamformers in Chapter 7 is white noise amplification. We will present a method that can deal with this problem.

Beampattern design is the most fundamental and important problem in array beamforming. Chapter 10 is dedicated to this issue. Again, for simplicity of presentation, a ULA is assumed. Since we are interested in frequency-invariant beampatterns, the spacing between neighboring sensors must be small, which is assumed here, as it is in Chapter 9. In this chapter, we revisit the definitions of the beampatterns and show some relationships. We then present different techniques for beampattern design. Note that the beampatterns designed in this chapter are similar to the ones obtained with differential sensor arrays in Chapter 9. This makes sense, as most assumptions used in this chapter are the same as those in Chapter 9.

Finally, in Chapter 11, we address the beamforming problem in the time domain. The approach depicted here is broadband in nature. We first describe the time-domain signal model that we adopt, and explain how broadband beamforming works. Then we define several performance measures, some relevant for fixed beamforming while others are more relevant for adaptive beamforming. Finally we show how to derive in great detail three classes of beamformers: fixed, adaptive, and differential. As the reader can see, the algorithms presented in this chapter are more intuitive than the frequency-domain beamformers discussed in previous chapters.

1.5 How to Use the Book

Signal enhancement and array signal processing is a broad subject that finds applications in many different fields. The background description or even the formulation of the problem in the literature is generally field-oriented, typically starting from the physics of wave propagation. Thus many students have been deterred from approaching the subject as it requires confronting, often for the first time, both the physics of wave propagation and the theory of signal processing and optimization. This book is designed as a textbook and it is written with students and instructors from different backgrounds in mind. To help students from the very different backgrounds to quickly get insight into the problem, we choose to focus on the theory, principles, and methods of signal enhancement and array signal processing from a purely signal processing perspective. Readers can then enjoy studying the fundamentals of the problem instead of enduring the introductory material on wave propagation in different types of media.

The material is designed for both advanced undergraduate students and graduate students. A one-semester advanced graduate course can cover virtually all of the text. However, there are also a few other ways to break the material into short courses for teaching advanced undergraduate or junior graduate students. First, a straightforward way is to break the material into two courses: from Chapter 2 to Chapter 6 for a course on signal enhancement and from Chapter 7 to Chapter 11 for a course on array signal processing. Chapter 2 serves as the basis for comprehending the material in Chapters 4, 6, and 11. A short course on signal enhancement and array beamforming in the time domain can be designed using the material presented in Chapters 2, 4, 6, and 11, and the rest of the material can be considered optional. Alternatively, a short course on signal enhancement and array beamforming can be taught based on all the material related to the frequency-domain theory and methods, in Chapters 3, 5, and 7–10.

References

1 Y. Huang, J. Benesty, and J. Chen, *Acoustic MIMO Signal Processing*. Berlin, Germany: Springer-Verlag, 2006.

2 J. Benesty, M. M. Sondhi, and Y. Huang, Eds., *Springer Handbook of Speech Processing*. Berlin, Germany: Springer-Verlag, 2007.

3 J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Berlin, Germany: Springer-Verlag, 2008.

4 N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. New York: Wiley, 1949.

5 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.

6 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, Jul. 2006.

7 J. Benesty and J. Chen, *Optimal Time-domain Noise Reduction Filters – A Theoretical Study*. Springer Briefs in Electrical and Computer Engineering, 2011.

8 J. Benesty, J. Chen, Y. Huang, and I. Cohen, *Noise Reduction in Speech Processing*. Berlin, Germany: Springer-Verlag, 2009.

9 K. K. Paliwal and A. Basu, "A speech enhancement method based on Kalman filtering," in *Proc. IEEE ICASSP*, 1987, pp. 177–180.

10 J. S. Lim (ed.), *Speech Enhancement*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

11 D. L. Wang, "On ideal binary mask as the computational goal of auditory scene analysis," in *Speech Separation by Humans and Machines*, P. Divenyi (ed.). Norwell, MA: Kluwer, pp. 181–197.

12 M. R. Schroeder, "Apparatus for suppressing noise and distortion in communication signals," U.S. Patent No. 3,180,936, filed 1 Dec. 1960, issued 27 Apr. 1965.

13 S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-27, pp. 113–120, Apr. 1979.

14 R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-28, pp. 137–145, Apr. 1980.

15 Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, pp. 1109–1121, Dec. 1984.

16 H. L. van Trees, *Detection, Estimation, and Modulation Theory, Optimum Array Processing* (Part IV). Hoboken, NJ: Wiley-Interscience, 2002.

17 D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Upper Saddle River, NJ: Prentice Hall, 1993.

18 S. Haykin, *Array Signal Processing*. Upper Saddle River, NJ: Prentice-Hall, 1984.

19 M. Sullivan, *Practical Array Processing*. New York: McGraw-Hill Education, 2008.

20 M. A. Richards, *Fundamentals of Radar Signal Processing*, 2nd edn. New York: McGraw-Hill, 2014.

21 P. S. Naidu, *Sensor Array Signal Processing*, 2nd edn. Boca Raton, FL: CRC Press, 2009.

22 S. Haykin and K. J. R. Liu, *Handbook on Array Processing and Sensor Networks*. Hoboken, NJ: Wiley & Sons, 2010.

23 H. Krim and M. Viberg, “Two decades of array signal processing research: the parametric approach,” *IEEE Sig. Process. Mag.*, vol. 13, pp. 67–94, Jul. 1996.

24 J. Benesty and J. Chen, *Study and Design of Differential Microphone Arrays*. Berlin, Germany: Springer-Verlag, 2013.

25 J. Benesty, J. Chen, and I.Cohen, *Design of Circular Differential Microphone Arrays*. Switzerland: Springer, 2015.

26 M. Brandstein and D. Ward (eds), *Microphone Arrays: Signal Processing Techniques and Applications*. Berlin, Germany: Springer-Verlag, 2001.

27 J. Capon, “High resolution frequency-wavenumber spectrum analysis,” *Proc. IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.

28 O. L. Frost, III, “An algorithm for linearly constrained adaptive array processing,” *Proc. IEEE*, vol. 60, pp. 926–935, Aug. 1972.

29 L. J. Griffiths and C. W. Jim, “An alternative approach to linearly constrained adaptive beamforming,” *IEEE Trans. Antennas Propag.*, vol.AP-30, pp. 27–34, Jan. 1982.

30 H. Cox, R. M. Zeskind, and M. M. Owen, “Robust adaptive beamforming,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, pp. 1365–1376, Oct. 1987.

31 D. G. Manolakis, D. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing*. Norwood, MA: Artech House, 2005.

32 W. Herbordt, *Sound Capture for Human/Machine Interfaces: Practical Aspects of Microphone Array Signal Processing*. Berlin, Germany: Springer, 2008.