1 Introduction

Signal processing provides the theory, the methods and the tools for such purposes as the analysis and modelling of signals, classification and recognition of patterns, extraction of information from signals, synthesis and morphing of signals – morphing is creating a new voice or image out of the existing samples. Signal processing is concerned with the modelling, extraction, communication and utilisation of information patterns and structures in a signal process.

Applications of signal processing methods are very wide and include audio hi-fi, TV and radio, cellular mobile phones, voice recognition, vision, antenna arrays, radar, sonar, geophysical exploration, medical electronics, bio-medical signal processing, physics and in general any system that is concerned with the communication or processing and retrieval of information. Signal processing plays a central role in the development of the new generations of mobile telecommunication and intelligent automation systems and in the efficient transmission, reception, decoding, organisation and retrieval of information content in search engines.

This chapter begins with a definition of signals, and a brief introduction to various signal processing methodologies. We consider several key applications of digital signal processing in biomedical signal processing, adaptive noise reduction, channel equalisation, pattern classification/recognition, audio signal coding, signal detection, spatial processing for directional reception of signals, Dolby noise reduction, radar and watermarking.

1.1 Signals and Information

A signal is the variation of a quantity such as air pressure waves of sounds, colours of an image, depths of a surface, temperature of a body, current/voltage in a conductor or biological system, light, electromagnetic radio waves, commodity prices or volume and mass of an object. A signal conveys information regarding one or more attributes of the source such as the state, the characteristics, the composition, the trajectory, the evolution or the intention of the source. Hence, a signal is a *means to convey information* regarding the past, the current or the future states of a variable.

For example, astrophysicists analyse the spectrum of signals, the light and other electromagnetic waves, emitted from distant stars or galaxies to deduce information about their movements, origins

Multimedia Signal Processing: Theory and Applications in Speech, Music and Communications Saeed V. Vaseghi

and evolution. Imaging radars calculate the round trip delay of reflected light or radio waves bouncing from the surface of the earth to produce maps of the earth.

A signal can be a function of one dimension, that is a function of one variable, such as speech or music whose amplitude fluctuations are a function of the time variable, or a signal can be multi-dimensional such as an image (i.e. reflected light intensity) which is a function of two-dimensional plane or video which is a function of two-dimensional plane and time. Note that a photograph effectively projects a view of objects in three-dimensional space onto a two-dimensional image plane where depth information can be deduced from the shadows and gradients of colours.

The information conveyed in a signal may be used by humans or machines (e.g. computers or robots) for communication, forecasting, decision-making, control, geophysical exploration, medical diagnosis, forensics etc.

The types of signals that signal processing systems deal with include text, image, audio, video, ultrasonic, subsonic, electromagnetic waves, medical, biological, thermal, financial or seismic signals.

Figure 1.1 illustrates a simplified overview of a communication system composed of an information source $I(t)$ followed by a signalling system $T[\cdot]$ for transformation of the information into variation of a signal $x(t)$ that carries the information, a communication channel $h[\cdot]$ for modelling the propagation of the signal from the transmitter to the receiver, additive channel and background noise $n(t)$ that exists in every real-life system and a signal processing unit at the receiver for extraction of the information from the received signal.

In general, there is a mapping operation (e.g. modulation) that maps the output $I(t)$of an information source to the physical variations of a signal $x(t)$ that carries the information, this mapping operator may be denoted as $T[\cdot]$ and expressed as

$$x(t) = T[I(t)] \tag{1.1}$$

The information source $I(t)$ is normally discrete-valued whereas the signal $x(t)$ that carries the information to a receiver may be continuous or discrete. For example, in multimedia communication the information from a computer, or any other digital communication device, is in the form of a sequence of binary numbers (ones and zeros) which would need to be transformed into a physical quantity such as voltage or current and modulated to the appropriate form for transmission in a communication channel such as a radio channel, telephone line or cable.

As a further example, in human speech communication the voice-generating mechanism provides a means for the speaker to map each discrete word into a distinct pattern of modulation of the acoustic vibrations of air that can propagate to the listener. To communicate a word w, the speaker generates an acoustic signal realisation of the word $x(t)$; this acoustic signal may be contaminated by ambient noise and/or distorted by a communication channel or room reverberations, or impaired by the speaking abnormalities of the talker, and received as the noisy, distorted and/or incomplete signal $y(t)$ modelled as

$$y(t) = h[x(t)] + n(t) \tag{1.2}$$

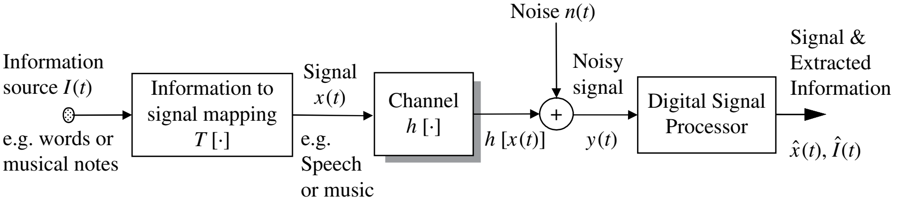

Figure 1.1 Illustration of a communication and signal processing system.

In addition to conveying the spoken word, the acoustic speech signal conveys information on the prosody (i.e. pitch intonation and stress patterns) of speech and the speaking characteristic, accent and the emotional state of the talker. The listener extracts this information by processing the signal $y(t)$.

In the past few decades, the theory and applications of digital signal processing have evolved to play a central role in the development of modern telecommunication and information technology systems.

Signal processing methods are central to efficient mobile communication, and to the development of intelligent man/machine interfaces in such areas as speech and visual pattern recognition for multimedia systems. In general, digital signal processing is concerned with two broad areas of information theory:

(a) Efficient and reliable coding, transmission, reception, storage and representation of signals in communication systems such as mobile phones, radio and TV.
(b) The extraction of information from noisy and/or incomplete signals for pattern recognition, detection, forecasting, decision-making, signal enhancement, control, automation and search engines.

In the next section we consider four broad approaches to signal processing.

1.2 Signal Processing Methods

Signal processing methods provide a variety of tools for modelling, analysis, coding, synthesis and recognition of signals. Signal processing methods have evolved in algorithmic complexity aiming for optimal utilisation of the available information in order to achieve the best performance. In general the computational requirement of signal processing methods increases, often exponentially, with the algorithmic complexity. However, the implementation costs of advanced signal processing methods have been offset and made affordable by the consistent trend in recent years of a continuing increase in the performance, coupled with a simultaneous decrease in the cost, of signal processing hardware.

Depending on the method used, digital signal processing algorithms can be categorised into one or a combination of four broad categories. These are transform-based signal processing, model-based signal processing, Bayesian statistical signal processing and neural networks, as illustrated in Figure 1.2. These methods are briefly described in the following.

1.2.1 Transform-Based Signal Processing

The purpose of a transform is to express a signal or a system in terms of a combination of a set of elementary simple signals (such as sinusoidal signals, eigenvectors or wavelets) that lend themselves to relatively easy analysis, interpretation and manipulation. Transform-based signal processing methods include Fourier transform, Laplace transform, z-transform, and wavelet transforms.

The most widely applied signal transform is the Fourier transform (introduced in Chapter 2) which is effectively a form of vibration analysis; a signal is expressed in terms of a combination of the sinusoidal vibrations that make up the signal. Fourier transform is employed in a wide range of applications including popular music coders, noise reduction and feature extraction for pattern recognition. The Laplace transform, and its discrete-time version the z-transform (introduced in Chapter 3), are generalisations of the Fourier transform and describe a signal or a system in terms of a set of transient sinusoids with exponential amplitude envelops.

In Fourier, Laplace and z-transform, the different sinusoidal basis functions of each transform all have the same duration and differ in terms of their frequency of vibrations and the amplitude envelopes.

In contrast wavelets are multi-resolution transforms in which a signal is described in terms of a combination of elementary waves of different dilations. The set of basis functions in a wavelet

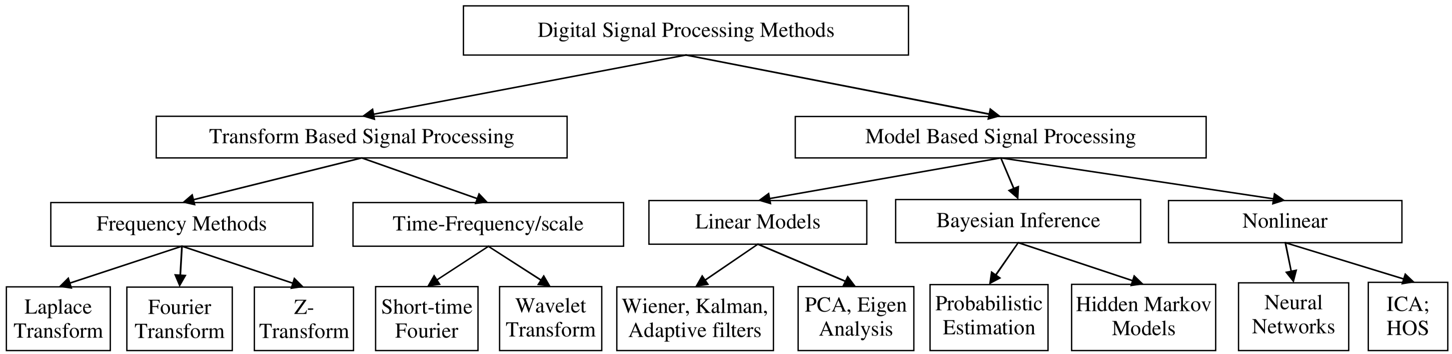

Figure 1.2 A broad categorisation of some of the most commonly used signal processing methods. ICA = Independent Component Analysis, HOS = Higher order statistics. Note that there may be overlap between different methods and also various methods can be combined.

is composed of contractions and dilations of a single elementary wave. This allows non-stationary events of various durations in a signal to be identified and analysed. Wavelet analysis is effectively a tree-structured filter bank analysis in which a set of high pass and low filters are used repeatedly in a binary-tree structure to split the signal progressively into sub-bands as explained in Chapter 4 on digital filters.

1.2.2 Source-Filter Model-Based Signal Processing

Model-based signal processing methods utilise a parametric model of the signal generation process. The parametric model normally describes the predictable structures and the expected patterns in the signal process, and can be used to forecast the future values of a signal from its past trajectory.

Model-based methods normally outperform non-parametric methods, since they utilise more information in the form of a model of the signal process. However, they can be sensitive to the deviations of a signal from the class of signals characterised by the model.

The most widely used parametric model is the linear prediction model, described in Chapter 10. Linear prediction models have facilitated the development of advanced signal processing methods for a wide range of applications such as low-bit-rate speech coding in cellular mobile telephony, digital video coding, high-resolution spectral analysis, radar signal processing and speech recognition.

1.2.3 Bayesian Statistical Model-Based Signal Processing

Statistical signal processing deals with random processes; this includes all information-bearing signals and noise. The fluctuations of a random signal, or the distribution of a class of random signals in the signal space, cannot be entirely modelled by a predictive equation, but it can be described in terms of the statistical average values, and modelled by a probability distribution function in a multi-dimensional signal space. For example, as described in Chapter 10, a linear prediction model driven by a random signal can provide a source-filter model of the acoustic realisation of a spoken word. However, the random input signal of the linear prediction model, or the variations in the characteristics of different acoustic realisations of the same word across the speaking population, can only be described in statistical terms and in terms of probability functions.

Bayesian inference theory provides a generalised framework for statistical processing of random signals, and for formulating and solving estimation and decision-making problems. Bayesian methods are used for pattern recognition and signal estimation problems in applications such as speech processing, communication, data management and artificial intelligence. In recognising a pattern or estimating a signal, from noisy and/or incomplete observations, Bayesian methods combine the evidence contained in the incomplete signal observation with the prior information regarding the distributions of the signals and/or the distributions of the parameters associated with the signals. Chapter 7 describes Bayesian inference methodology and the estimation of random processes observed in noise.

1.2.4 Neural Networks

Neural networks are combinations of relatively simple non-linear adaptive processing units, arranged to have a structural resemblance to the transmission and processing of signals in biological neurons. In a neural network several layers of parallel processing elements are interconnected with a hierarchically structured connection network. The connection weights are trained to ‘memorise patterns’ and perform a signal processing function such as prediction or classification.

Neural networks are particularly useful in non-linear partitioning of a signal space, in feature extraction and pattern recognition, and in decision-making systems. In some hybrid pattern recognition systems neural networks are used to complement Bayesian inference methods. Since the main objective of this book is to provide a coherent presentation of the theory and applications of statistical signal processing, neural networks are not discussed here.

1.3 Applications of Digital Signal Processing

In recent years, the development and commercial availability of increasingly powerful and affordable digital computers has been accompanied by the development of advanced digital signal processing algorithms for a wide variety of applications such as noise reduction, telecommunication, radar, sonar, video and audio signal processing, pattern recognition, geophysics explorations, data forecasting, and the processing of large databases for the identification, extraction and organisation of unknown underlying structures and patterns. Figure 1.3 shows a broad categorisation of some DSP applications. This section provides a review of several key applications of digital signal processing methods.

Part III of this book covers the applications of DSP to speech processing, music processing and communications. In the following an overview of some applications of DSP is provided. Note that these applications are by no means exhaustive but they represent a useful introduction.

1.3.1 Digital Watermarking

Digital watermarking is the embedding of a signature signal, i.e. the digital watermark, underneath a host image, video or audio signal. Although watermarking may be visible or invisible, the main challenge in digital watermarking is to make the watermark secret and imperceptible (meaning invisible or inaudible). Watermarking takes its name from the watermarking of paper or money for security and authentication purposes.

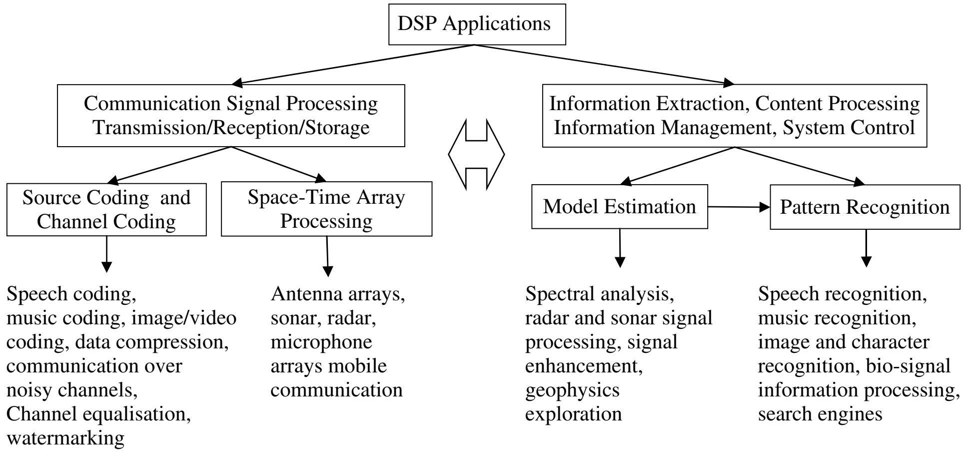

Figure 1.3 A classification of the applications of digital signal processing.

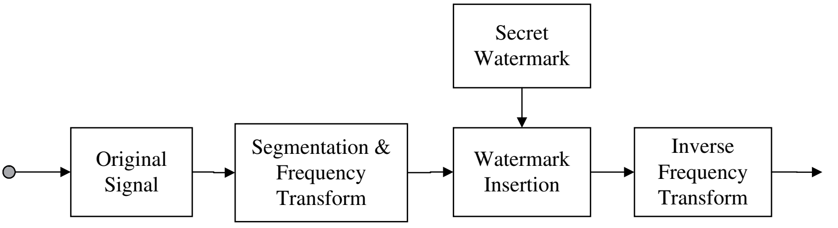

Figure 1.4 A simplified illustration of frequency domain watermarking.

Watermarking is used in digital media for the following purposes:

(1) Authentication of digital image and audio signals. The watermark may also include owner information, a serial number and other useful information.
(2) Protection of copyright/ownership of image and audio signals from unauthorised copying, use or trade.
(3) Embedding of audio or text signals into image/video signals for subsequent retrieval.
(4) Embedding a secret message into an image or audio signal.

Watermarking has to be robust to intended or unintended degradations and resistant to attempts at rendering it ineffective. In particular watermarking needs to survive the following processes:

(1) Changes in the sampling rate, resolution and format of the signal.
(2) Changes in the orientation of images or phase of the signals.
(3) Noise and channel distortion.
(4) Non-linear imperceptible changes of time/space scales. For example non-linear time-warping of audio or non-linear warping of the dimensions of an image.
(5) Segmentation and cropping of the signals.

The simplest forms of watermarking methods, Figure 1.4, exploit the time–frequency structure of the signal together with the audio-visual perceptual characteristics of humans. The watermark signal is hidden in the parts of the host signal spectrum, where it is invisible or inaudible. For example, a simple way to embed a watermark into an audio signal is to transform the watermark such that it closely follows the envelope of the time-varying spectrum of the audio signal. The transformed watermark is then added to the audio signal.

An example of invisible watermarking is shown in Figure 1.5. The figure shows a host image and another image acting as the watermark together with the watermarked image and the retrieved watermark.

1.3.2 Bio-medical Signal Processing

Bio-medical signal processing is concerned with the analysis, denoising, synthesis and classification of bio-signals such as magnetic resonance images (MRI) of brain or electrocardiograph (ECG) signals of heart or electroencephalogram (EEG) signals of brain neurons.

An electrocardiograph signal is produced by recording the electrical voltage signals of the heart. It is the main tool in cardiac electrophysiology, and has a prime function in the screening and diagnosis of cardiovascular diseases.

Figure 1.5 Illustration of invisible watermarking of an image, clockwise from top-left: a picture of my son, the watermark, watermarked image and retrieved watermark. The watermark may be damaged due to modifications such as a change of image coding format.

Electroencephalography is the neurophysiologic measurement of the electrical activity of the neurons in brain picked up by electrodes placed on the scalp or, in special cases, on the cortex. The resulting signals are known as an electroencephalograph and represent a mix of electrical signals and noise from a large number of neurons.

The observations of ECG or EEG signals are often a noisy mixture of electrical signals generated from the activities of several different sources from different parts of the body. The main issues in the processing of bio-signals, such as EEG or ECG, are the denoising, separation and identification of the signals from different sources.

An important bio-signal analysis tool, considered in Chapter 12, is known as independent componentanalysis (ICA). ICA is primarily used for separation of mixed signals in multi-source multi-sensor applications such as in ECG and EEG. ICA is also used for beam forming in multiple-input multiple-output (MIMO) telecommunication.

The ICA problem is formulated as follows. The observed signal vector $\boldsymbol{x}$ is assumed to be a linear mixture of M independent source signals $\boldsymbol{s}$. In a linear matrix form the mixing operation is expressed as

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s} \tag{1.3}$$

The matrix $\boldsymbol{A}$ is known as the *mixing matrix* or the observation matrix. In many practical cases of interest all we have is the sequence of observation vectors $[\boldsymbol{x}(0), \boldsymbol{x}(1), \ldots, \boldsymbol{x}(N-1)]$. The mixing matrix $\boldsymbol{A}$ is unknown and we wish to estimate a demixing matrix $\boldsymbol{W}$ to obtain an estimate of the original signal $\boldsymbol{s}$.

This problem is known as *blind source separation* (BSS); the term blind refers to the fact that we have no other information than the observation $\boldsymbol{x}$ and an assumption that the source signals are independent of each other. The demixing problem is the estimation of a matrix $\boldsymbol{W}$ such that

$$\hat{\boldsymbol{s}} = \boldsymbol{W}\boldsymbol{x} \tag{1.4}$$

The details of the derivation of the demixing matrix are discussed in Chapter 12. Figure 1.6 shows an example of ECG signal mixture of the hearts of a pregnant mother and foetus plus other noise and interference. Note that application of ICA results in separation of the mother and foetus heartbeats. Also note that the foetus heartbeat rate is about 25% faster than the mother's heartbeat rate.

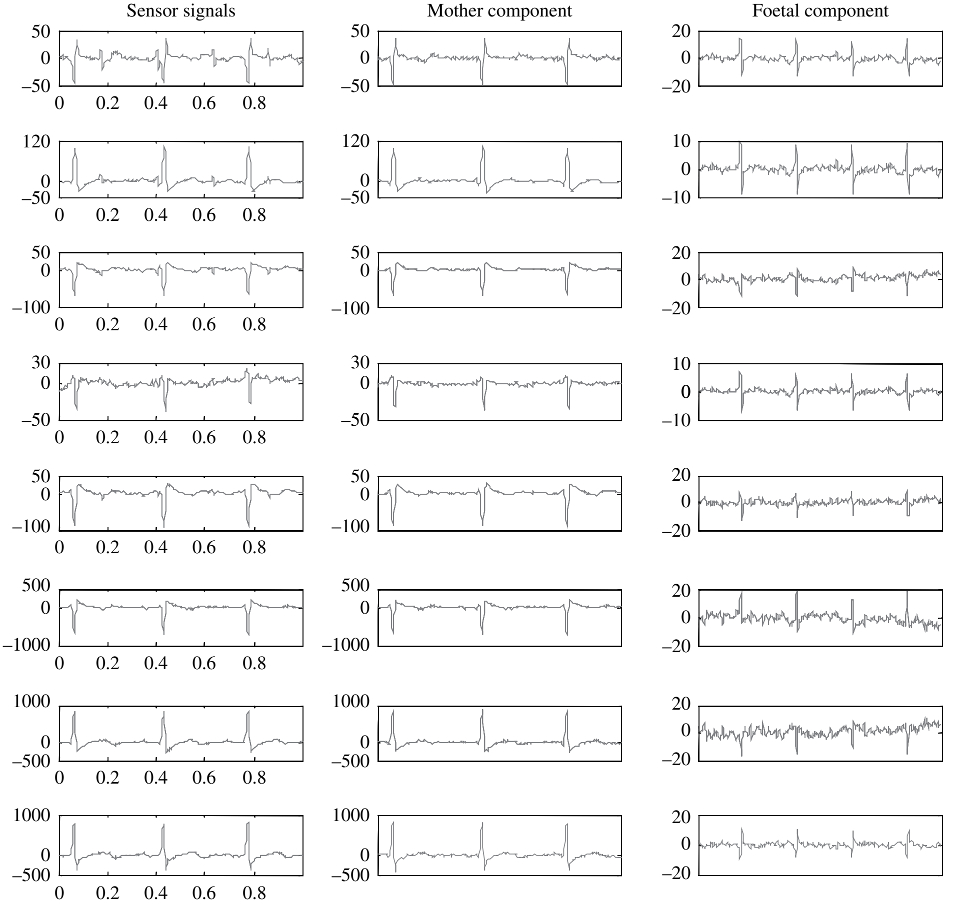

Figure 1.6 Application of ICA to separation of mother and foetus ECG. Note that signals from eight sensors are used in this example (see Chapter 12).

1.3.3 Adaptive Noise Cancellation

In speech communication from a noisy acoustic environment such as a moving car or train, or over a noisy telephone channel, the speech signal is observed in an additive random noise. In signal measurement systems the information-bearing signal is often contaminated by noise from its surrounding environment. The noisy observation $y(m)$ can be modelled as

$$y(m) = x(m) + n(m) \tag{1.5}$$

where $x(m)$ and $n(m)$ are the signal and the noise, and m is the discrete-time index. In some situations, for example when using a mobile telephone in a moving car, or when using a radio communication device in an aircraft cockpit, it may be possible to measure and estimate the instantaneous amplitude of the ambient noise using a directional microphone. The signal $x(m)$ may then be recovered by subtraction of an estimate of the noise from the noisy signal.

Figure 1.7 shows a two-input adaptive noise cancellation system for enhancement of noisy speech. In this system a directional microphone takes as input the noisy signal $x(m) + n(m)$, and a second directional microphone, positioned some distance away, measures the noise $\alpha n(m+\tau)$. The attenuation factor α and the time delay τ provide a rather over-simplified model of the effects of propagation of the noise to different positions in the space where the microphones are placed. The noise from the second microphone is processed by an adaptive digital filter to make it equal to the noise contaminating the speech signal, and then subtracted from the noisy signal to cancel out the noise. The adaptive noise canceller is more effective in cancelling out the low-frequency part of the noise, but generally suffers from the non-stationary character of the signals, and from the over-simplified assumption that a linear filter can model the diffusion and propagation of the noise sound in the space.

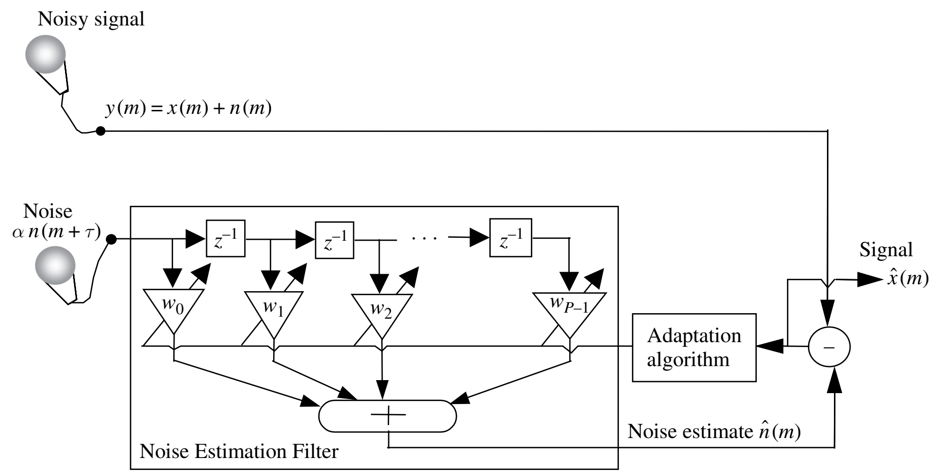

Figure 1.7 Configuration of a two-microphone adaptive noise canceller. The adaptive filter delay elements (z^{-1}) and weights wi model the delay and attenuation that signals undergo while propagating in a medium.

1.3.4 Adaptive Noise Reduction

In many applications, for example at the receiver of a telecommunication system, there is no access to the instantaneous value of the contaminating noise, and only the noisy signal is available. In such cases the noise cannot be cancelled out, but it may be reduced, in an average sense, using the statistics of the signal and the noise process. Figure 1.8 shows a bank of Wiener filters for reducing additive noise when only the noisy signal is available. The filter bank coefficients attenuate each noisy signal frequency in inverse proportion to the signal-to-noise ratio at that frequency. The Wiener filter bank coefficients, derived in Chapter 8, are calculated from estimates of the power spectra of the signal and the noise processes.

1.3.5 Blind Channel Equalisation

Channel equalisation is the recovery of a signal distorted in transmission through a communication channel with a non-flat magnitude or a non-linear phase response. When the channel response is unknown the process of signal recovery is called blind equalisation. Blind equalisation has a wide range of applications, for example in digital telecommunications for removal of inter-symbol interference due to non-ideal channel and multi-path propagation, in speech recognition for removal of the effects of the microphones and the communication channels, in correction of distorted images, analysis of seismic data, de-reverberation of acoustic gramophone recordings etc.

In practice, blind equalisation is feasible only if some useful statistics of the channel input are available. The success of a blind equalisation method depends on how much is known about the characteristics of the input signal and how useful this knowledge can be in the channel identification

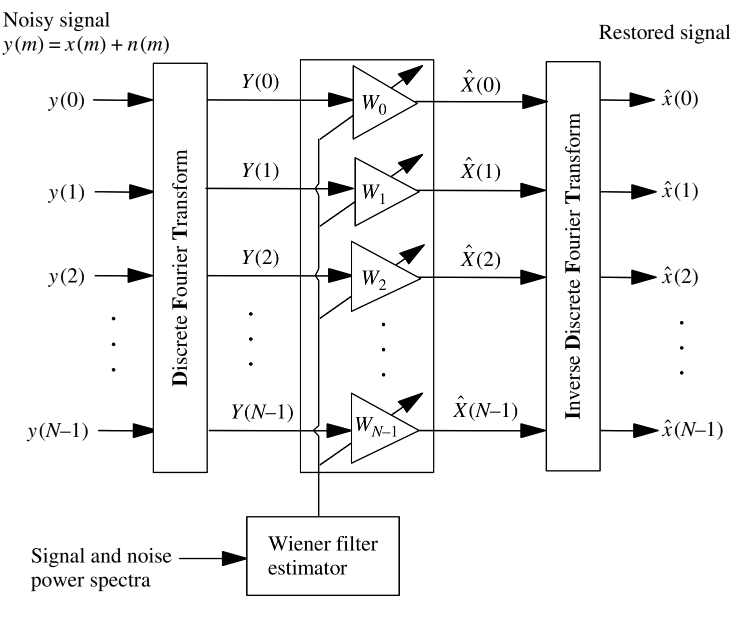

Figure 1.8 A frequency-domain Wiener filter for reducing additive noise.

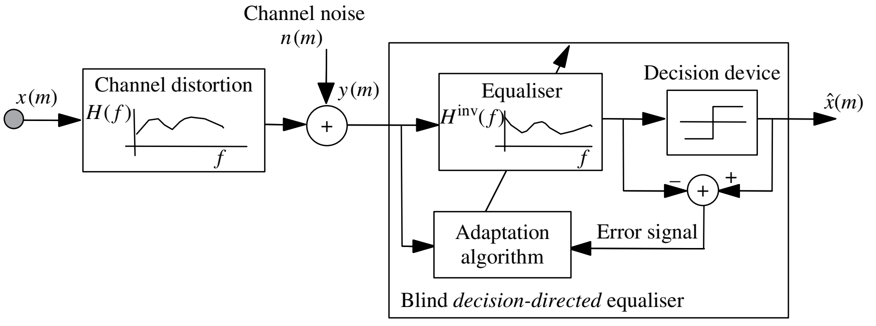

Figure 1.9 Configuration of a decision-directed blind channel equaliser.

and equalisation process. Figure 1.9 illustrates the configuration of a decision-directed equaliser. This blind channel equaliser is composed of two distinct sections: an adaptive equaliser that removes a large part of the channel distortion, followed by a non-linear decision device for an improved estimate of the channel input. The output of the decision device is the final estimate of the channel input, and it is used as the desired signal *to direct* the equaliser adaptation process. Blind equalisation is covered in detail in Chapter 17.

1.3.6 Signal Classification and Pattern Recognition

Signal classification is used in detection, pattern recognition and decision-making systems. For example, a simple binary-state classifier can act as the detector of the presence, or the absence, of a known waveform in noise. In signal classification, the aim is to design a minimum-error system for *labelling* a signal with one of a number of likely classes of signal.

To design a classifier, a set of models are trained for the classes of signals that are of interest in the application. The simplest form that the models can assume is a bank, or codebook, of waveforms, each representing the prototype for one class of signals. A more complete model for each class of signals takes the form of a probability distribution function. In the classification phase, a signal is labelled with the nearest or the most likely class. For example, in communication of a binary bit stream over a band-pass channel, the binary phase-shift keying (BPSK) scheme signals the bit '1' using the waveform $A_c \sin \omega_c t$ and the bit '0' using $-A_c \sin \omega_c t$.

At the receiver, the decoder has the task of classifying and labelling the received noisy signal as a '1' or a '0'. Figure 1.10 illustrates a correlation receiver for a BPSK signalling scheme. The receiver has two correlators, each programmed with one of the two symbols representing the binary states for the bit '1' and the bit '0'. The decoder correlates the unlabelled input signal with each of the two candidate symbols and selects the candidate that has a higher correlation with the input.

Figure 1.11 illustrates the use of a classifier in a limited-vocabulary, isolated-word speech recognition system. Assume there are V words in the vocabulary. For each word a model is trained, on many different examples of the spoken word, to capture the average characteristics and the statistical variations of the word. The classifier has access to a bank of $V+1$ models, one for each word in the vocabulary and an additional model for the silence periods. In the speech recognition phase, the task is to decode and label an acoustic speech feature sequence, representing an unlabelled spoken word, as one of the V likely words or silence. For each candidate word the classifier calculates a probability score and selects the word with the highest score.

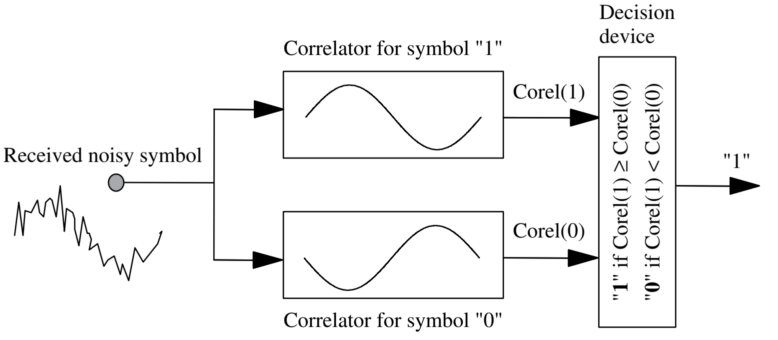

Figure 1.10 Block diagram illustration of the classifier in a binary phase-shift keying demodulation.

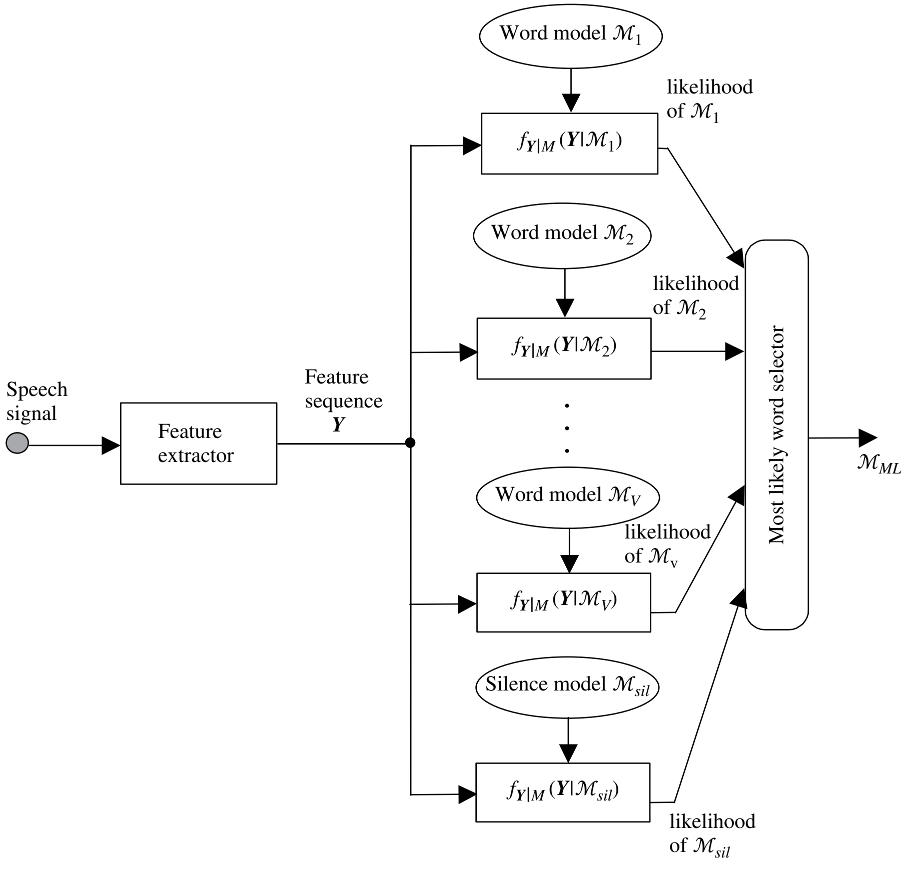

Figure 1.11 Configuration of speech recognition system, $f(\boldsymbol{Y}|\mathcal{M}_i)$ is the likelihood of the model $\mathcal{M}_i$ given an observation sequence $\boldsymbol{Y}$.

1.3.7 Linear Prediction Modelling of Speech

Linear predictive models (introduced in Chapter 12) are widely used in speech processing applications such as low-bit-rate speech coding in cellular telephony, speech enhancement and speech recognition. Speech is generated by inhaling air into the lungs, and then exhaling it through the vibrating glottis cords and the vocal tract. The random, noise-like, air flow from the lungs is spectrally shaped and amplified by the vibrations of the glottal cords and the resonance of the vocal tract. The effect of the vibrations of the glottal cords and the resonance of the vocal tract is to shape the frequency spectrum of speech and introduce a measure of correlation and predictability on the random variations of the air from the lungs. Figure 1.12 illustrates a source-filter model for speech production. The source models the lungs and emits a random excitation signal which is filtered, first by a pitch filter model of the glottal cords and then by a model of the vocal tract.

The main source of correlation in speech is the vocal tract modelled by a linear predictor. A linear predictor is an adaptive filter that forecasts the amplitude of the signal at time m, $x(m)$, using a linear combination of P previous samples $[x(m-1), \mathrm{L}, x(m-P)]$ as

$$\hat{x}(m) = \sum_{k=1}^{P} a_k x(m-k) \tag{1.6}$$

where $\hat{x}(m)$ is the prediction of the signal $x(m)$, and the vector $\boldsymbol{a}^{\mathrm{T}} = [a_1, \ldots, a_P]$ is the coefficients vector of a predictor of order P. The prediction error $e(m)$, i.e. the difference between the actual sample $x(m)$ and its predicted value $\hat{x}(m)$, is defined as

$$e(m) = x(m) - \sum_{k=1}^{P} a_k x(m-k) \tag{1.7}$$

In speech processing, the prediction error $e(m)$ may also be interpreted as the random excitation or the so-called innovation content of $x(m)$. From Equation ((1.7)) a signal generated by a linear predictor can be synthesised as

$$x(m) = \sum_{k=1}^{P} a_k x(m-k) + e(m) \tag{1.8}$$

Linear prediction models can also be used in a wide range of applications to model the correlation or the movements of a signal such as the movements of scenes in successive frames of video.

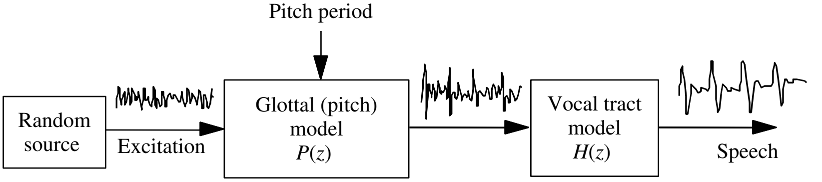

Figure 1.12 Linear predictive model of speech.

1.3.8 Digital Coding of Audio Signals

In digital audio, the memory required to record a signal, the bandwidth and power required for signal transmission and the signal-to-quantisation-noise ratio are all directly proportional to the number of bits per sample. The objective in the design of a coder is to achieve high fidelity with as few bits per sample as possible, at an affordable implementation cost.

Audio signal coding schemes utilise the statistical structures of the signal, and a model of the signal generation, together with information on the psychoacoustics and the masking effects of hearing. In general, there are two main categories of audio coders: model-based coders, used for low-bit-rate speech coding in applications such as cellular telephony; and transform-based coders used in high-quality coding of speech and digital hi-fi audio. Figure 1.13 shows a simplified block diagram configuration of a speech coder–decoder of the type used in digital cellular telephones. The speech signal is modelled as the output of a filter excited by a random signal. The random excitation models the air exhaled through the lungs, and the filter models the vibrations of the glottal cords and the vocal tract. At the transmitter, speech is segmented into blocks of about 20 ms long during which speech parameters can be assumed to be stationary. Each block of speech samples is analysed to extract and transmit a set of excitation and filter parameters that can be used to synthesise the speech. At the receiver, the model parameters and the excitation are used to reconstruct the speech.

A transform-based coder is shown in Figure 1.14. The aim of transformation is to convert the signal into a form where it lends itself to a more convenient and useful interpretation and manipulation. In Figure 1.14 the input signal may be transformed to the frequency domain using a discrete Fourier transform or a discrete cosine transform or a filter bank. Three main advantages of coding a signal in the frequency domain are:

(a) The frequency spectrum of a signal has a relatively well-defined structure, for example most of the signal power is usually concentrated in the lower regions of the spectrum.
(b) A relatively low-amplitude frequency would be masked in the near vicinity of a large-amplitude frequency and can therefore be coarsely encoded without any audible degradation.
(c) The frequency samples are orthogonal and can be coded independently with different precisions.

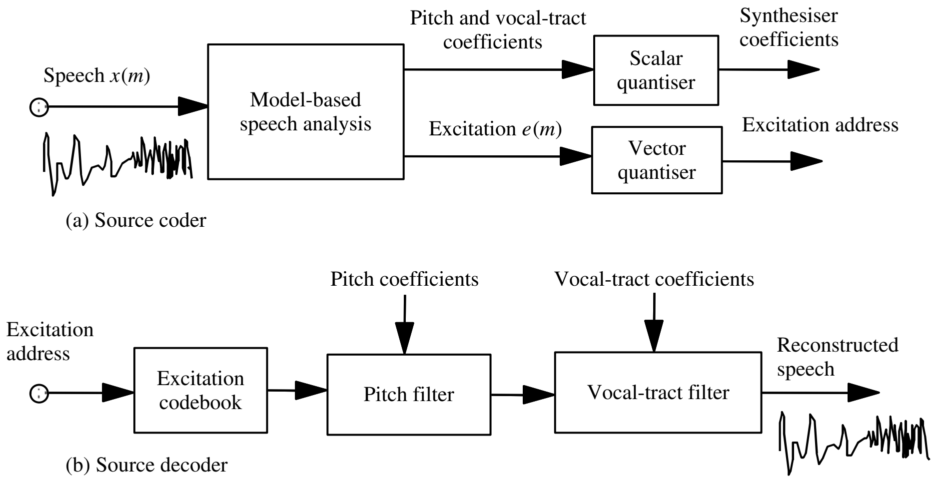

Figure 1.13 Block diagram configuration of a model-based speech (a) coder and (b) decoder.

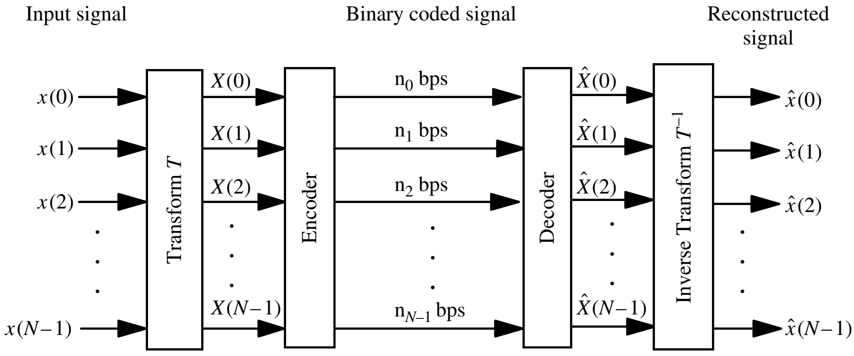

Figure 1.14 Illustration of a transform-based coder.

The number of bits assigned to each frequency of a signal is a variable that reflects the contribution of that frequency to the reproduction of a perceptually high-quality signal. In an adaptive coder, the allocation of bits to different frequencies is made to vary with the time variations of the power spectrum of the signal.

1.3.9 Detection of Signals in Noise

In the detection of signals in noise, the aim is to determine if the observation consists of noise alone, or if it contains a signal. The noisy observation $y(m)$ can be modelled as

$$y(m) = b(m)x(m) + n(m) \tag{1.9}$$

where $x(m)$ is the signal to be detected, $n(m)$ is the noise and $b(m)$ is a binary-valued state indicator sequence such that $b(m) = 1$ indicates the presence of the signal $x(m)$ and $b(m) = 0$ indicates that the signal is absent. If the signal $x(m)$ has a known shape, then a correlator or a matched filter can be used to detect the signal as shown in Figure 1.15. The impulse response $h(m)$ of the matched filter for detection of a signal $x(m)$ is the time-reversed version of $x(m)$ given by

$$h(m) = x(N-1-m) \quad 0 \le m \le N-1 \tag{1.10}$$

where N is the length of $x(m)$. The output of the matched filter is given by

$$z(m) = \sum_{k=0}^{N-1} h(k)y(m-k) \tag{1.11}$$

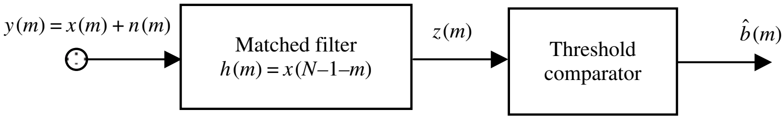

Figure 1.15 Configuration of a matched filter followed by a threshold comparator for detection of signals in noise.

Table 1.1 Four possible outcomes in a signal detection problem.

$\hat{b}(m)$	$b(m)$	Detector decision
0	0	Signal absent (*Correct*)
0	1	Signal absent (*Missed*)
1	0	Signal present (*False alarm*)
1	1	Signal present (*Correct*)

The matched filter output is compared with a threshold and a binary decision is made as

$$\hat{b}(m) = \begin{cases} 1 & \text{if abs}\,(z(m)) \geq \text{ threshold} \\ 0 & \text{otherwise} \end{cases} \tag{1.12}$$

where $\hat{b}(m)$ is an estimate of the binary state indicator sequence $b(m)$, and it may be erroneous in particular if the signal-to-noise ratio is low. Table 1.1 lists four possible outcomes that together $b(m)$ and its estimate $\hat{b}(m)$ can assume. The choice of the threshold level affects the sensitivity of the detector. The higher the threshold, the less the likelihood that noise would be classified as signal, so the false alarm rate falls, but the probability of misclassification of signal as noise increases. The risk in choosing a threshold value θ can be expressed as

$$\mathcal{R}(\text{Threshold} = \theta) = P_{\text{False Alarm}}(\theta) + P_{\text{Miss}}(\theta) \tag{1.13}$$

The choice of the threshold reflects a trade-off between the misclassification rate $P_{\text{Miss}}(\theta)$ and the false alarm rate $P_{\text{FalseAlarm}}(\theta)$.

1.3.10 Directional Reception of Waves: Beam-forming

Beam-forming is the spatial processing of plane waves received by an array of sensors such that the waves' incidents at a particular spatial angle are passed through, whereas those arriving from other directions are attenuated. Beam-forming is used in radar and sonar signal processing (Figure 1.14) to steer the reception of signals towards a desired direction, and in speech processing for reducing the effects of ambient noise.

To explain the process of beam-forming, consider a uniform linear array of sensors as illustrated in Figure 1.16. The term *linear array* implies that the array of sensors is spatially arranged in a straight

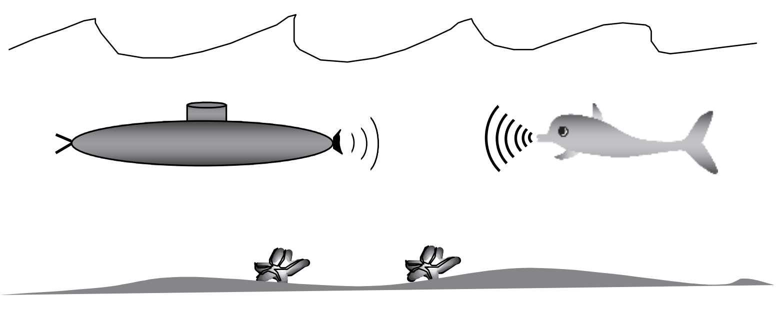

Figure 1.16 Sonar: detection of objects using the intensity and time delay of reflected sound waves.

line and with equal spacing d between the sensors. Consider a sinusoidal far-field plane wave with a frequency F_0 propagating towards the sensors at an incidence angle of θ as illustrated in Figure 1.16. The array of sensors samples the incoming wave as it propagates in space. The time delay for the wave to travel a distance of d between two adjacent sensors is given by

$$\tau = \frac{d \sin\theta}{c} \tag{1.14}$$

where c is the speed of propagation of the wave in the medium. The phase difference corresponding to a delay of τ is given by

$$\varphi = 2\pi \frac{\tau}{T_0} = 2\pi F_0 \frac{d \sin\theta}{c} \tag{1.15}$$

where T_0 is the period of the sine wave. By inserting appropriate corrective time delays in the path of the samples at each sensor, and then averaging the outputs of the sensors, the signals arriving from the direction θ will be time-aligned and coherently combined, whereas those arriving from other directions will suffer cancellations and attenuations. Figure 1.17 illustrates a beam-former as an array of digital filters arranged in space. The filter array acts as a two-dimensional space-time signal processing system. The space filtering allows the beam-former to be steered towards a desired direction, for example

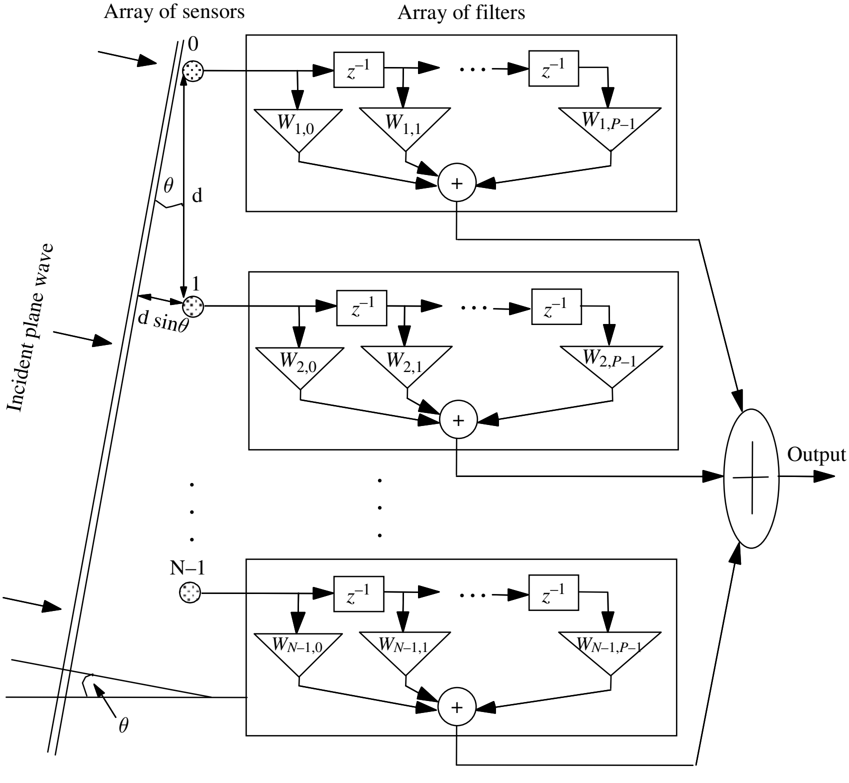

Figure 1.17 Illustration of a beam-former, for directional reception of signals.

towards the direction along which the incoming signal has the maximum intensity. The phase of each filter controls the time delay, and can be adjusted to coherently combine the signals. The magnitude frequency response of each filter can be used to remove the out-of-band noise.

1.3.11 Space-Time Signal Processing

Conventionally transmission resources are shared among subscribers of communication systems through the division of time and frequency leading to such resource-sharing schemes as time division multiple access or frequency division multiple access. Space provides a valuable additional resource that can be used to improve both the communication capacity and quality for wireless communication systems.

Space-time signal processing refers to signal processing methods that utilise simultaneous transmission and reception of signals through multiple spatial routes. The signals may arrive at the destinations at different times or may use different time slots. Space-time signal processing, and in particular the division of space among different users, is an important area of research and development for improving the system capacity in the new generations of high-speed broadband multimedia mobile communication systems.

For example, in mobile communication the multi-path effect, where a radio signal propagates from the transmitter to the receiver via a number of different paths, can be used to advantage in space-time signal processing. The multiple noisy versions of a signal, arriving via different routes with different noise and distortions, are processed and combined such that the signals add up constructively and become stronger compared with the random uncorrelated noise. The uncorrelated fading that the signals suffer in their propagation through different routes can also be mitigated.

The use of transmitter/receiver antenna arrays for beam-forming allows the division of the space into narrow sectors such that the same frequencies, in different narrow spatial sectors, can be used for simultaneous communication by different subscribers and/or different spatial sectors can be used to transmit the same information in order to achieve robustness to fading and interference. In fact combination of space and time can provide a myriad of possibilities, as discussed in Chapter 18 on mobile communication signal processing. Note that the ICA method, described in Section 1.3.2 and Chapter 12, is often used in space-time signal processing for separation of multiple signals at the receiver.

1.3.12 Dolby Noise Reduction

Dolby noise reduction systems work by boosting the energy and the signal-to-noise ratio of the high-frequency spectrum of audio signals. The energy of audio signals is mostly concentrated in the low-frequency part of the spectrum (below 2 kHz). The higher frequencies that convey quality and sensation have relatively low energy, and can be degraded even by a low amount of noise. For example when a signal is recorded on a magnetic tape, the tape ‘hiss’ noise affects the quality of the recorded signal. On playback, the higher-frequency parts of an audio signal recorded on a tape have smaller signal-to-noise ratio than the low-frequency parts. Therefore noise at high frequencies is more audible and less masked by the signal energy. Dolby noise reduction systems broadly work on the principle of emphasising and boosting the low energy of the high-frequency signal components prior to recording the signal. When a signal is recorded it is processed and encoded using a combination of a pre-emphasis filter and dynamic range compression. At playback, the signal is recovered using a decoder based on a combination of a de-emphasis filter and a decompression circuit. The encoder and decoder must be well matched and cancel each other out in order to avoid processing distortion.

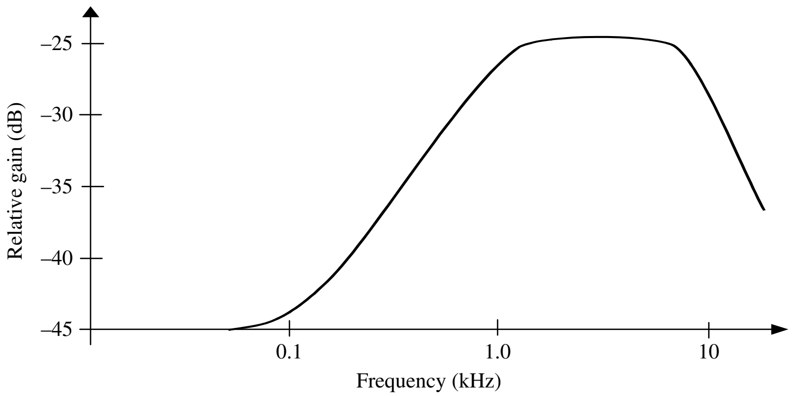

Figure 1.18 Illustration of the pre-emphasis response of Dolby C: up to 20 dB boost is provided when the signal falls 45 dB below maximum recording level.

Dolby developed a number of noise reduction systems designated Dolby A, Dolby B and Dolby C. These differ mainly in the number of bands and the pre-emphasis strategy that that they employ. Dolby A, developed for professional use, divides the signal spectrum into four frequency bands: band 1 is low-pass and covers 0 Hz to 80 Hz; band 2 is band-pass and covers 80 Hz to 3 kHz; band 3 is high-pass and covers above 3 kHz; and band 4 is also high-pass and covers above 9 kHz. At the encoder the gain of each band is adaptively adjusted to boost low-energy signal components. Dolby A provides a maximum gain of 10 to 15 dB in each band if the signal level falls 45 dB below the maximum recording level. The Dolby B and Dolby C systems are designed for consumer audio systems, and use two bands instead of the four bands used in Dolby A. Dolby B provides a boost of up to 10 dB when the signal level is low (less than 45 dB than the maximum reference) and Dolby C provides a boost of up to 20 dB as illustrated in Figure 1.18.

1.3.13 Radar Signal Processing: Doppler Frequency Shift

Figure 1.19 shows a simple diagram of a radar system that can be used to estimate the range and speed of an object such as a moving car or a flying aeroplane. A radar system consists of a transceiver

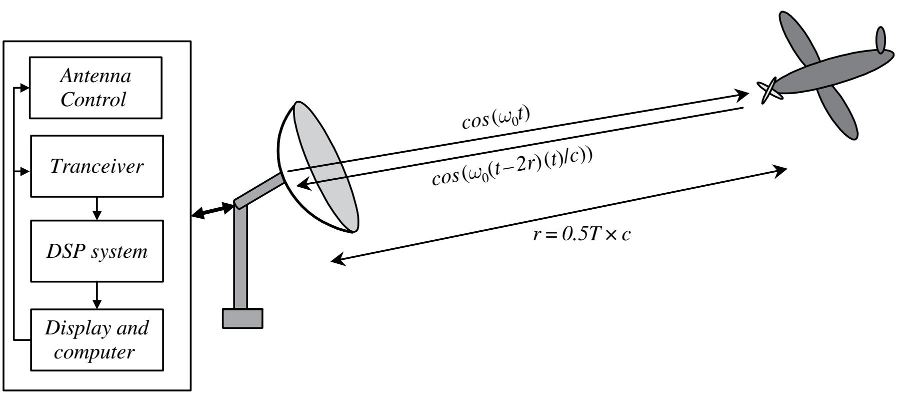

Figure 1.19 Illustration of a radar system.

(transmitter/receiver) that generates and transmits sinusoidal pulses at microwave frequencies. The signal travels with the speed of light and is reflected back from any object in its path. The analysis of the received echo provides such information as range, speed and acceleration. The received signal has the form

$$x(t) = A(t)\cos\{\omega_0[t - 2r(t)/c]\} \tag{1.16}$$

where $A(t)$, the time-varying amplitude of the reflected wave, depends on the position and the characteristics of the target, $r(t)$ is the time-varying distance of the object from the radar and c is the velocity of light. The time-varying distance of the object can be expanded in a Taylor series as

$$r(t) = r_0 + \dot{r}t + \frac{1}{2!}\ddot{r}t^2 + \frac{1}{3!}\dddot{r}\,t^3 + \cdots \tag{1.17}$$

where r_0 is the distance, $\dot{r}$ is the velocity, $\ddot{r}$ is the acceleration etc. Approximating $r(t)$ with the first two terms of the Taylor series expansion we have

$$r(t) \approx r_0 + \dot{r}t \tag{1.18}$$

Substituting Equation (1.18) in Equation (1.16) yields

$$x(t) = A(t)\cos[(\omega_0 - 2\dot{r}\omega_0/c)t - 2\omega_0 r_0/c] \tag{1.19}$$

Note that the frequency of reflected wave is shifted by an amount

$$\omega_d = 2\dot{r}\omega_0/c \tag{1.20}$$

This shift in frequency is known as the Doppler frequency. If the object is moving towards the radar then the distance $r(t)$ is decreasing with time, $\dot{r}$ is negative, and an increase in the frequency is observed. Conversely if the object is moving away from the radar then the distance $r(t)$ is increasing, $\dot{r}$ is positive, and a decrease in the frequency is observed. Thus the frequency analysis of the reflected signal can reveal information on the direction and speed of the object. The distance r_0 is given by

$$r_0 = 0.5T \times c \tag{1.21}$$

where T is the round-trip time for the signal to hit the object and arrive back at the radar and c is the velocity of light.

1.4 Summary

This chapter began with a definition of signal and information and provided a qualitative explanation of their relationship. A broad categorisation of the various signal processing methodologies was provided. We considered several key applications of digital signal processing in biomedical signal processing, adaptive noise reduction, channel equalisation, pattern classification/recognition, audio signal coding, signal detection, spatial processing for directional reception of signals, Dolby noise reduction, radar and watermarking.

Further Reading

Alexander S.T. (1986) Adaptive Signal Processing Theory and Applications. Springer-Verlag, New York.

Davenport W.B. and Root W.L. (1958) An Introduction to the Theory of Random Signals and Noise. McGraw-Hill, New York.

Ephraim Y. (1992) Statistical Model Based Speech Enhancement Systems. Proc. IEEE, 80 (10), pp. 1526–1555.
Gallager R.G. (1968) Information Theory and Reliable Communication. Wiley, New York.
Gauss K.G. (1963) Theory of Motion of Heavenly Bodies. Dover, New York.
Haykin S. (1985) Array Signal Processing. Prentice-Hall, Englewood Cliffs, NJ.
Haykin S. (1991) Adaptive Filter Theory. Prentice-Hall, Englewood Cliffs, NJ.
Kailath T. (1980) Linear Systems. Prentice Hall, Englewood Cliffs, NJ.
Kalman R.E. (1960) A New Approach to Linear Filtering and Prediction Problems. Trans. of the ASME, Series D, Journal of Basic Engineering, 82, pp. 35–45.
Kay S.M. (1993) Fundamentals of Statistical Signal Processing, Estimation Theory. Prentice-Hall, Englewood Cliffs, NJ.
Kung S.Y. (1993) Digital Neural Networks. Prentice-Hall, Englewood Cliffs, NJ.
Lim J.S. (1983) Speech Enhancement. Prentice Hall, Englewood Cliffs, NJ.
Lucky R.W., Salz J. and Weldon E.J. (1968) Principles of Data Communications. McGraw-Hill, New York.
Marple S.L. (1987) Digital Spectral Analysis with Applications. Prentice-Hall, Englewood Cliffs, NJ.
Oppenheim A.V. and Schafer R.W. (1989) Discrete-Time Signal Processing. Prentice-Hall, Englewood Cliffs, NJ.
Proakis J.G., Rader C.M., Ling F. and Nikias C.L. (1992) Advanced Signal Processing. Macmillan, New York.
Rabiner L.R. and Gold B. (1975) Theory and Applications of Digital Processing. Prentice-Hall, Englewood Cliffs, NJ.
Rabiner L.R. and Schafer R.W. (1978) Digital Processing of Speech Signals. Prentice-Hall, Englewood Cliffs, NJ.
Scharf L.L. (1991) Statistical Signal Processing: Detection, Estimation, and Time Series Analysis. Addison Wesley, Reading, MA.
Shannon C.E. (1948) A Mathematical Theory of Communication. Bell Systems Tech. J., 27, pp. 379–423, 623–656.
Therrien C.W. (1992) Discrete Random Signals and Statistical Signal Processing. Prentice-Hall, Englewood Cliffs, NJ.
Van-Trees H.L. (1971) Detection, Estimation and Modulation Theory. Parts I, II and III. Wiley, New York.
Vaseghi S. (2006) Advanced Digital Signal Processing and Noise Reduction, Wiley, New York.
Widrow B. (1975) Adaptive Noise Cancelling: Principles and Applications. Proc. IEEE, 63, pp. 1692–1716.
Wiener N. (1948) Extrapolation, Interpolation and Smoothing of Stationary Time Series. MIT Press, Cambridge, MA.
Wiener N. (1949) Cybernetics. MIT Press, Cambridge, MA.
Wilsky A.S. (1979) Digital Signal Processing, Control and Estimation Theory: Points of Tangency, Areas of Intersection and Parallel Directions. MIT Press, Cambridge, MA.