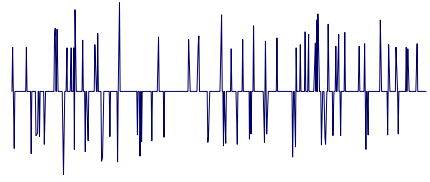


2



NOISE AND DISTORTION

- | | |
|-----------------------------------|----------------------------------|
| 2.1 Introduction | 2.6 Thermal Noise |
| 2.2 White Noise | 2.7 Shot Noise |
| 2.3 Coloured Noise | 2.8 Electromagnetic Noise |
| 2.4 Impulsive Noise | 2.9 Channel Distortions |
| 2.5 Transient Noise Pulses | 2.10 Modelling Noise |

Noise can be defined as an unwanted signal that interferes with the communication or measurement of another signal. A noise itself is a signal that conveys information regarding the source of the noise. For example, the noise from a car engine conveys information regarding the state of the engine. The sources of noise are many, and vary from audio frequency acoustic noise emanating from moving, vibrating or colliding sources such as revolving machines, moving vehicles, computer fans, keyboard clicks, wind, rain, etc. to radio-frequency electromagnetic noise that can interfere with the transmission and reception of voice, image and data over the radio-frequency spectrum. Signal distortion is the term often used to describe a systematic undesirable change in a signal and refers to changes in a signal due to the non-ideal characteristics of the transmission channel, reverberations, echo and missing samples.

Noise and distortion are the main limiting factors in communication and measurement systems. Therefore the modelling and removal of the effects of noise and distortion have been at the core of the theory and practice of communications and signal processing. Noise reduction and distortion removal are important problems in applications such as cellular mobile communication, speech recognition, image processing, medical signal processing, radar, sonar, and in any application where the signals cannot be isolated from noise and distortion. In this chapter, we study the characteristics and modelling of several different forms of noise.

2.1 Introduction

Noise may be defined as any unwanted signal that interferes with the communication, measurement or processing of an information-bearing signal. Noise is present in various degrees in almost all environments. For example, in a digital cellular mobile telephone system, there may be several variety of noise that could degrade the quality of communication, such as acoustic background noise, thermal noise, electromagnetic radio-frequency noise, co-channel interference, radio-channel distortion, echo and processing noise. Noise can cause transmission errors and may even disrupt a communication process; hence noise processing is an important part of modern telecommunication and signal processing systems. The success of a noise processing method depends on its ability to characterise and model the noise process, and to use the noise characteristics advantageously to differentiate the signal from the noise. Depending on its source, a noise can be classified into a number of categories, indicating the broad physical nature of the noise, as follows:

- (a) Acoustic noise: emanates from moving, vibrating, or colliding sources and is the most familiar type of noise present in various degrees in everyday environments. Acoustic noise is generated by such sources as moving cars, air-conditioners, computer fans, traffic, people talking in the background, wind, rain, etc.
- (b) Electromagnetic noise: present at all frequencies and in particular at the radio frequencies. All electric devices, such as radio and television transmitters and receivers, generate electromagnetic noise.
- (c) Electrostatic noise: generated by the presence of a voltage with or without current flow. Fluorescent lighting is one of the more common sources of electrostatic noise.
- (d) Channel distortions, echo, and fading: due to non-ideal characteristics of communication channels. Radio channels, such as those at microwave frequencies used by cellular mobile phone operators, are particularly sensitive to the propagation characteristics of the channel environment.
- (e) Processing noise: the noise that results from the digital/analog processing of signals, e.g. quantisation noise in digital coding of speech or image signals, or lost data packets in digital data communication systems.

Depending on its frequency or time characteristics, a noise process can be classified into one of several categories as follows:

- (a) Narrowband noise: a noise process with a narrow bandwidth such as a 50/60 Hz ‘hum’ from the electricity supply.
- (b) White noise: purely random noise that has a flat power spectrum. White noise theoretically contains all frequencies in equal intensity.
- (c) Band-limited white noise: a noise with a flat spectrum and a limited bandwidth that usually covers the limited spectrum of the device or the signal of interest.
- (d) Coloured noise: non-white noise or any wideband noise whose spectrum has a non-flat shape; examples are pink noise, brown noise and autoregressive noise.
- (e) Impulsive noise: consists of short-duration pulses of random amplitude and random duration.
- (f) Transient noise pulses: consists of relatively long duration noise pulses.

2.2 White Noise

White noise is defined as an uncorrelated noise process with equal power at all frequencies (Figure 2.1). A noise that has the same power at all frequencies in the range of $\pm\infty$ would necessarily need to have infinite power, and is therefore only a theoretical concept. However a band-limited noise process, with a flat spectrum covering the frequency range of a band-limited communication system, is to all intents and purposes from the point of view of the system a white noise process. For example, for an audio system with a bandwidth of 10 kHz, any flat-spectrum audio noise with a bandwidth greater than 10 kHz looks like a white noise.

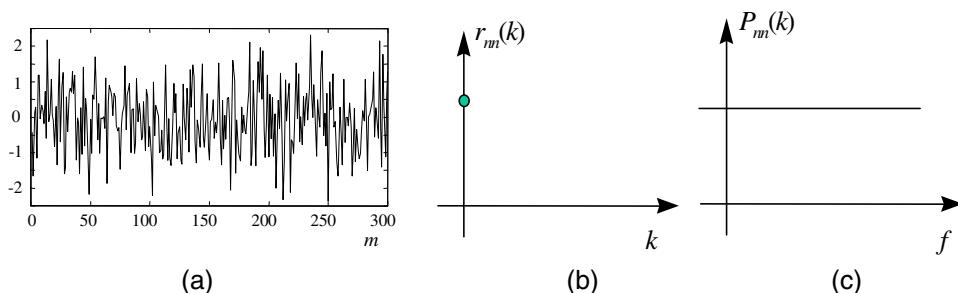


Figure 2.1 Illustration of (a) white noise, (b) its autocorrelation, and (c) its power spectrum.

The autocorrelation function of a continuous-time zero-mean white noise process with a variance of σ^2 is a delta function given by

$$r_{NN}(\tau) = \mathcal{E}[N(t)N(t+\tau)] = \sigma^2 \delta(\tau) \quad (2.1)$$

The power spectrum of a white noise, obtained by taking the Fourier transform of Equation (2.1), is given by

$$P_{NN}(f) = \int_{-\infty}^{\infty} r_{NN}(t) e^{-j2\pi ft} dt = \sigma^2 \quad (2.2)$$

Equation (2.2) shows that a white noise has a constant power spectrum.

A pure white noise is a theoretical concept, since it would need to have infinite power to cover an infinite range of frequencies. Furthermore, a discrete-time signal by necessity has to be band-limited, with its highest frequency less than half the sampling rate. A more practical concept is band-limited white noise, defined as a noise with a flat spectrum in a limited bandwidth. The spectrum of band-limited white noise with a bandwidth of B Hz is given by

$$P_{NN}(f) = \begin{cases} \sigma^2, & |f| \leq B \\ 0, & \text{otherwise} \end{cases} \quad (2.3)$$

Thus the total power of a band-limited white noise process is $2B\sigma^2$. The autocorrelation function of a discrete-time band-limited white noise process is given by

$$r_{NN}(T_s k) = 2B\sigma^2 \frac{\sin(2\pi B T_s k)}{2\pi B T_s k} \quad (2.4)$$

where T_s is the sampling period. For convenience of notation T_s is usually assumed to be unity. For the case when $T_s = 1/2B$, i.e. when the sampling rate is equal to the Nyquist rate, Equation (2.4) becomes

$$r_{NN}(T_s k) = 2B\sigma^2 \frac{\sin(\pi k)}{\pi k} = 2B\sigma^2 \delta(k) \quad (2.5)$$

In Equation (2.5) the autocorrelation function is a delta function.

2.3 Coloured Noise

Although the concept of white noise provides a reasonably realistic and mathematically convenient and useful approximation to some predominant noise processes encountered in telecommunication systems, many other noise processes are non-white. The term coloured noise refers to any broadband noise with a non-white spectrum. For example most audio-frequency noise, such as the noise from moving cars, noise from computer fans, electric drill noise and people talking in the background, has a non-white predominantly low-frequency spectrum. Also, a white noise passing through a channel is “coloured” by the shape of the channel spectrum. Two classic varieties of coloured noise are so-called pink noise and brown noise, shown in Figures 2.2 and 2.3.

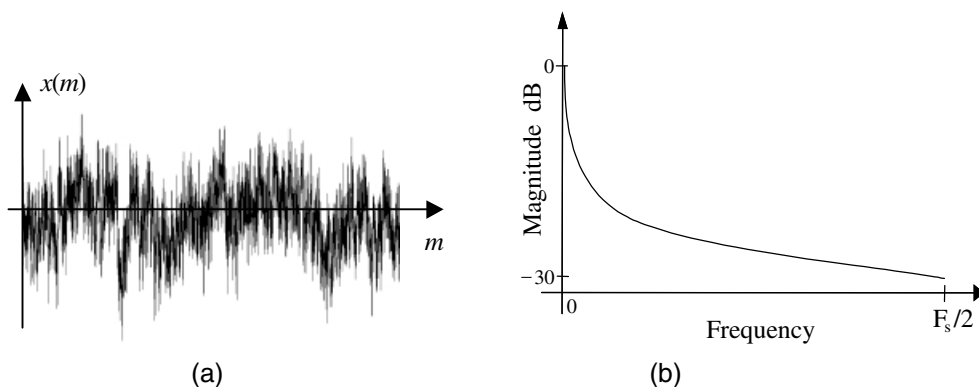


Figure 2.2 (a) A pink noise signal and (b) its magnitude spectrum.

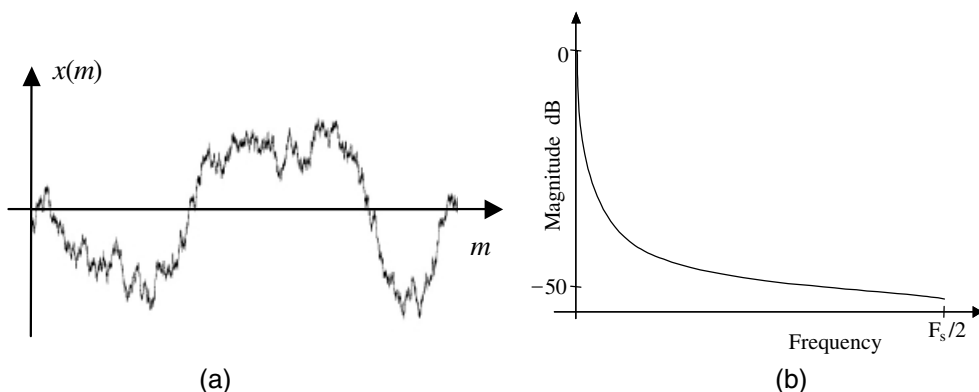


Figure 2.3 (a) A brown noise signal and (b) its magnitude spectrum.

2.4 Impulsive Noise

Impulsive noise consists of short-duration “on/off” noise pulses, caused by a variety of sources, such as switching noise, adverse channel environment in a communication system, drop-outs or surface degradation of audio recordings, clicks from computer keyboards, etc. Figure 2.4(a) shows an ideal impulse and its frequency spectrum. In communication systems, a real impulsive-type noise has a duration that is normally more than one sample long. For example, in the context of audio signals, short-duration, sharp pulses, of up to 3 milliseconds (60 samples at a 20 kHz sampling rate) may be considered as impulsive noise. Figures 2.4(b) and (c) illustrate two examples of short-duration pulses and their respective spectra.

In a communication system, an impulsive noise originates at some point in time and space, and then propagates through the channel to the receiver. The received noise is time-dispersed and shaped by the channel, and can be considered as the channel impulse response. In general, the characteristics of a communication channel may be linear or non-linear, stationary or time varying. Furthermore, many communication systems, in response to a large-amplitude impulse, exhibit a non-linear characteristic.

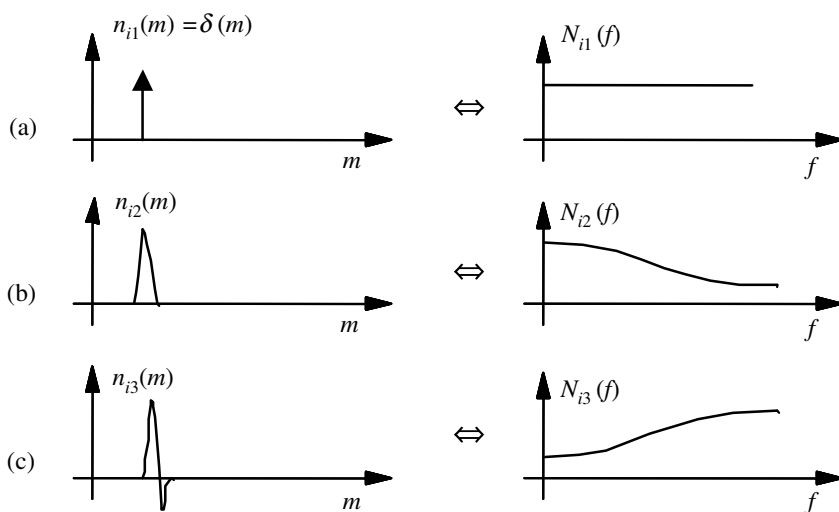


Figure 2.4 Time and frequency sketches of: (a) an ideal impulse, (b) and (c) short-duration pulses.

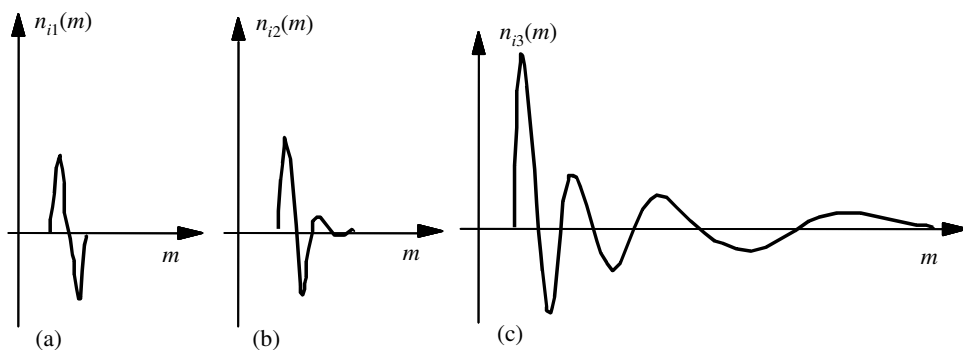


Figure 2.5 Illustration of variations of the impulse response of a non-linear system with the increasing amplitude of the impulse.

Figure 2.5 illustrates some examples of impulsive noise, typical of those observed on an old gramophone recording. In this case, the communication channel is the playback system, and may be assumed to be time-invariant. The figure also shows some variations of the channel characteristics with the amplitude of impulsive noise. For example, in Figure 2.5(c) a large impulse excitation has generated a decaying transient pulse. These variations may be attributed to the non-linear characteristics of the playback mechanism.

2.5 Transient Noise Pulses

Transient noise pulses often consist of a relatively short sharp initial pulse followed by decaying low-frequency oscillations as shown in Figure 2.6. The initial pulse is usually due to some external or internal impulsive interference, whereas the oscillations are often due to the resonance of the

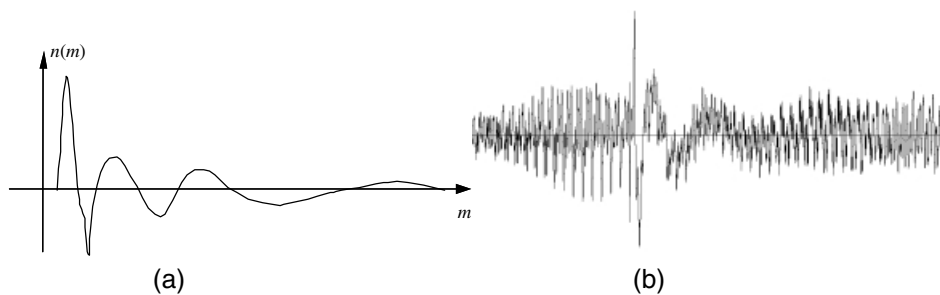


Figure 2.6 (a) A scratch pulse and music from a gramophone record. (b) The averaged profile of a gramophone record scratch pulse.

communication channel excited by the initial pulse, and may be considered as the response of the channel to the initial pulse. In a telecommunication system, a noise pulse originates at some point in time and space, and then propagates through the channel to the receiver. The noise pulse is shaped by the channel characteristics, and may be considered as the channel pulse response. Thus we should be able to characterize the transient noise pulses with a similar degree of consistency as in characterizing the channels through which the pulses propagate.

As an illustration of the shape of a transient noise pulse, consider the scratch pulses from a damaged gramophone record shown in Figures 2.6(a) and (b). Scratch noise pulses are acoustic manifestations of the response of the stylus and the associated electro-mechanical playback system to a sharp physical discontinuity on the recording medium. Since scratches are essentially the impulse response of the playback mechanism, it is expected that for a given system, various scratch pulses exhibit a similar characteristics. As shown in Figure 2.6(b), a typical scratch pulse waveform often exhibits two distinct regions:

- (a) the initial high-amplitude pulse response of the playback system to the physical discontinuity on the record medium, followed by;
- (b) decaying oscillations that cause additive distortion. The initial pulse is relatively short and has a duration on the order of 1–5 ms, whereas the oscillatory tail has a longer duration and may last up to 50 ms or more.

Note in Figure 2.6(b) that the frequency of the decaying oscillations decreases with time. This behaviour may be attributed to the non-linear modes of response of the electro-mechanical playback system excited by the physical scratch discontinuity. Observations of many scratch waveforms from damaged gramophone records reveals that they have a well-defined profile, and can be characterised by a relatively small number of typical templates. Scratch pulse modelling and removal is considered in detail in Chapter 13.

2.6 Thermal Noise

Thermal noise, also referred to as Johnson noise (after its discoverer J. B. Johnson), is generated by the random movements of thermally energised particles. The concept of thermal noise has its roots in thermodynamics and is associated with the temperature-dependent random movements of free

particles such as gas molecules in a container or electrons in a conductor. Although these random particle movements average to zero, the fluctuations about the average constitute the thermal noise. For example, the random movements and collisions of gas molecules in a confined space produce random fluctuations about the average pressure. As the temperature increases, the kinetic energy of the molecules and the thermal noise increase.

Similarly, an electrical conductor contains a very large number of free electrons, together with ions that vibrate randomly about their equilibrium positions and resist the movement of the electrons. The free movement of electrons constitutes random spontaneous currents, or thermal noise, that average to zero since in the absence of a voltage electrons move in all different directions. As the temperature of a conductor, provided by its surroundings, increases, the electrons move to higher-energy states and the random current flow increases. For a metallic resistor, the mean square value of the instantaneous voltage due to the thermal noise is given by

$$\overline{v^2} = 4kTRB \quad (2.6)$$

where $k=1.38 \times 10^{-23}$ joules per degree Kelvin is the Boltzmann constant, T is the absolute temperature in degrees Kelvin, R is the resistance in ohms and B is the bandwidth. From Equation (2.6) and the preceding argument, a metallic resistor sitting on a table can be considered as a generator of thermal noise power, with a mean square voltage $\overline{v^2}$ and an internal resistance R . From circuit theory, the maximum available power delivered by a “thermal noise generator”, dissipated in a matched load of resistance R , is given by

$$P_N = \overline{i^2} R = \left(\frac{v_{\text{rms}}}{2R} \right)^2 R = \frac{\overline{v^2}}{4R} = kTB \quad (\text{W}) \quad (2.7)$$

where v_{rms} is the root mean square voltage. The spectral density of thermal noise is given by

$$P_N(f) = \frac{kT}{2} \quad (\text{W/Hz}) \quad (2.8)$$

From Equation (2.8), the thermal noise spectral density has a flat shape, i.e. thermal noise is a white noise. Equation (2.8) holds well up to very high radio frequencies of 10^{13} Hz.

2.7 Shot Noise

The term shot noise arose from the analysis of random variations in the emission of electrons from the cathode of a vacuum tube. Discrete electron particles in a current flow arrive at random times, and therefore there will be fluctuations about the average particle flow. The fluctuations in the rate of particle flow constitutes the shot noise. Other instances of shot noise are the flow of photons in a laser beam, the flow and recombination of electrons and holes in semiconductors, and the flow of photoelectrons emitted in photodiodes. The concept of randomness of the rate of emission or arrival of particles implies that shot noise can be modelled by a Poisson distribution. When the average number of arrivals during the observing time is large, the fluctuations will approach a Gaussian distribution. Note that whereas thermal noise is due to “unforced” random movement of particles, shot noise happens in a forced directional flow of particles.

Now consider an electric current as the flow of discrete electric charges. If the charges act independently of each other the fluctuating current is given by

$$I_{\text{Noise(rms)}} = (2eI_{\text{dc}}B)^{1/2} \quad (2.9)$$

where $e = 1.6 \times 10^{-19}$ coulomb is the electron charge, and B is the measurement bandwidth. For example, a “steady” current I_{dc} of 1 amp in a bandwidth 1 MHz has an rms fluctuation of 0.57 microamps. Equation (2.9) assumes that the charge carriers making up the current act independently. That is the case for charges crossing a barrier, as for example the current in a junction diode, where the charges move by diffusion; but it is not true for metallic conductors, where there are long-range correlations between charge carriers.

2.8 Electromagnetic Noise

Virtually every electrical device that generates, consumes or transmits power is a potential source of electromagnetic noise and interference for other systems. In general, the higher the voltage or the current level, and the closer the proximity of electrical circuits/devices, the greater will be the induced noise. The common sources of electromagnetic noise are transformers, radio and television transmitters, mobile phones, microwave transmitters, ac power lines, motors and motor starters, generators, relays, oscillators, fluorescent lamps, and electrical storms.

Electrical noise from these sources can be categorized into two basic types: electrostatic and magnetic. These two types of noise are fundamentally different, and thus require different noise-shielding measures. Unfortunately, most of the common noise sources listed above produce combinations of the two noise types, which can complicate the noise reduction problem.

Electrostatic fields are generated by the presence of voltage, with or without current flow. Fluorescent lighting is one of the more common sources of electrostatic noise. Magnetic fields are created either by the flow of electric current or by the presence of permanent magnetism. Motors and transformers are examples of the former, and the Earth's magnetic field is an instance of the latter. In order for noise voltage to be developed in a conductor, magnetic lines of flux must be cut by the conductor. Electric generators function on this basic principle. In the presence of an alternating field, such as that surrounding a 50/60 Hz power line, voltage will be induced into any stationary conductor as the magnetic field expands and collapses. Similarly, a conductor moving through the Earth's magnetic field has a noise voltage generated in it as it cuts the lines of flux.

2.9 Channel Distortions

On propagating through a channel, signals are shaped and distorted by the frequency response and the attenuating characteristics of the channel. There are two main manifestations of channel distortions: magnitude distortion and phase distortion. In addition, in radio communication, we have the

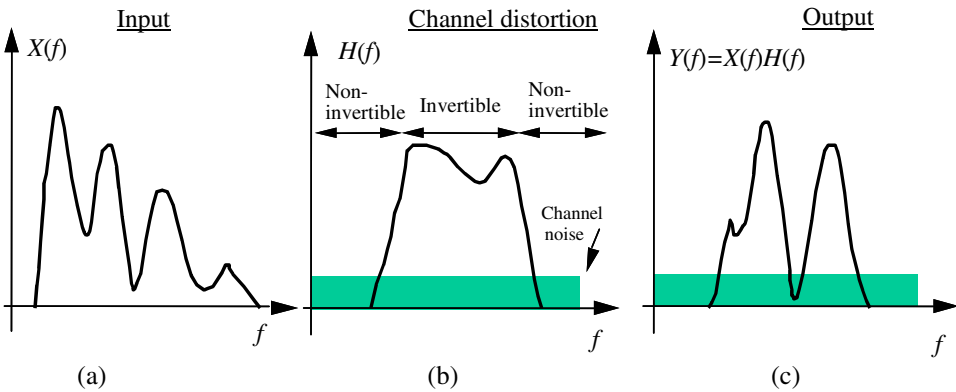


Figure 2.7 Illustration of channel distortion: (a) the input signal spectrum, (b) the channel frequency response, (c) the channel output.

multi-path effect, in which the transmitted signal may take several different routes to the receiver, with the effect that multiple versions of the signal with different delay and attenuation arrive at the receiver. Channel distortions can degrade or even severely disrupt a communication process, and hence channel modelling and equalization are essential components of modern digital communication systems. Channel equalization is particularly important in modern cellular communication systems, since the variations of channel characteristics and propagation attenuation in cellular radio systems are far greater than those of the landline systems. Figure 2.7 illustrates the frequency response of a channel with one invertible and two non-invertible regions. In the non-invertible regions, the signal frequencies are heavily attenuated and lost to the channel noise. In the invertible region, the signal is distorted but recoverable. This example illustrates that the channel inverse filter must be implemented with care in order to avoid undesirable results such as noise amplification at frequencies with a low SNR. Channel equalization is covered in detail in Chapter 15.

2.10 Modelling Noise

The objective of modelling is to characterise the structures and the patterns in a signal or a noise process. To model a noise accurately, we need a structure for modelling both the temporal and the spectral characteristics of the noise. Accurate modelling of noise statistics is the key to high-quality noisy signal classification and enhancement. Even the seemingly simple task of signal/noise classification is crucially dependent on the availability of good signal and noise models, and on the use of these models within a Bayesian framework. Hidden Markov models described in Chapter 5 are good structure for modelling signals or noise.

One of the most useful and indispensable tools for gaining insight into the structure of a noise process is the use of Fourier transform for frequency

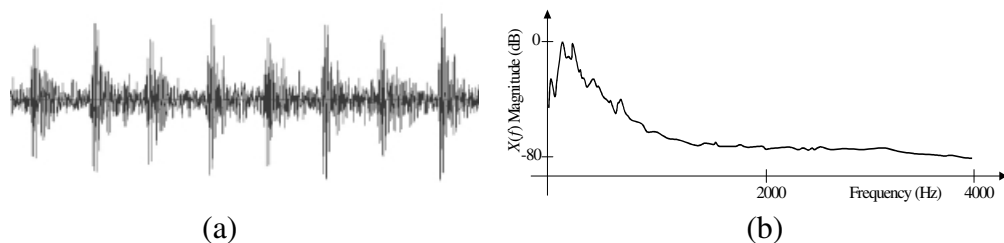


Figure 2.8 Illustration of: (a) the time-waveform of a drill noise, and (b) the frequency spectrum of the drill noise.

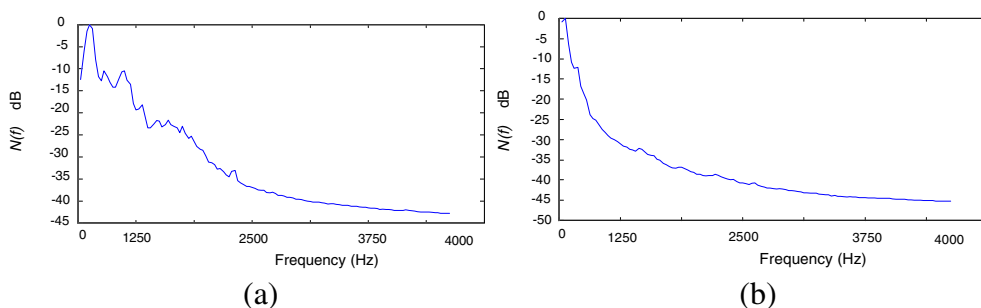


Figure 2.9 Power spectra of car noise in (a) a BMW at 70 mph, and (b) a Volvo at 70 mph.

analysis. Figure 2.8 illustrates the noise from an electric drill, which, as expected, has a periodic structure. The spectrum of the drilling noise shown in Figure 2.8(a) reveals that most of the noise energy is concentrated in the lower-frequency part of the spectrum. In fact, it is true of most audio signals and noise that they have a predominantly low-frequency spectrum. However, it must be noted that the relatively lower-energy high-frequency part of audio signals plays an important part in conveying sensation and quality. Figures 2.9(a) and (b) show examples of the spectra of car noise recorded from a BMW and a Volvo respectively. The noise in a car is nonstationary, and varied, and may include the following sources:

- (a) quasi-periodic noise from the car engine and the revolving mechanical parts of the car;
- (b) noise from the surface contact of wheels and the road surface;
- (c) noise from the air flow into the car through the air ducts, windows, sunroof, etc;
- (d) noise from passing/overtaking vehicles.

The characteristic of car noise varies with the speed, the road surface conditions, the weather, and the environment within the car.

The simplest method for noise modelling, often used in current practice, is to estimate the noise statistics from the signal-inactive periods. In optimal Bayesian signal processing methods, a set of probability models are trained for the signal and the noise processes. The models are then used for the decoding of the underlying states of the signal and noise, and for noisy signal recognition and enhancement.

2.10.1 Additive White Gaussian Noise Model (AWGN)

In communication theory, it is often assumed that the noise is a stationary additive white Gaussian (AWGN) process. Although for some problems this is a valid assumption and leads to mathematically convenient and useful solutions, in practice the noise is often time-varying, correlated and non-Gaussian. This is particularly true for impulsive-type noise and for acoustic noise, which are non-stationary and non-Gaussian and hence cannot be modelled using the AWGN assumption. Non-stationary and non-Gaussian noise processes can be modelled by a Markovian chain of stationary subprocesses as described briefly in the next section and in detail in Chapter 5.

2.10.2 Hidden Markov Model for Noise

Most noise processes are non-stationary; that is the statistical parameters of the noise, such as its mean, variance and power spectrum, vary with time. Nonstationary processes may be modelled using the hidden Markov models (HMMs) described in detail in Chapter 5. An HMM is essentially a finite-state Markov chain of stationary subprocesses. The implicit assumption in using HMMs for noise is that the noise statistics can be modelled by a Markovian chain of stationary subprocesses. Note that a stationary noise process can be modelled by a single-state HMM. For a non-stationary noise, a multistate HMM can model the time variations of the noise process with a finite number of stationary states. For non-Gaussian noise, a mixture Gaussian density model can be used to model the space of the noise within each state. In general, the number of states per model and number of mixtures per state required to accurately model a noise process depends on

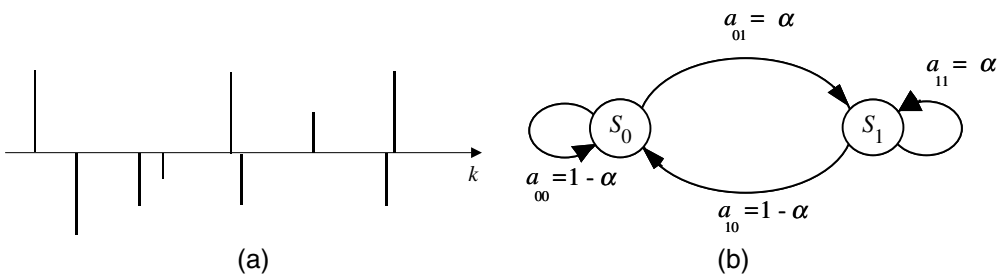


Figure 2.10 (a) An impulsive noise sequence. (b) A binary-state model of impulsive noise.

the non-stationary character of the noise.

An example of a non-stationary noise is the impulsive noise of Figure 2.10(a). Figure 2.10(b) shows a two-state HMM of the impulsive noise sequence: the state S_0 models the “impulse-off” periods between the impulses, and state S_1 models an impulse. In those cases where each impulse has a well-defined temporal structure, it may be beneficial to use a multi-state HMM to model the pulse itself. HMMs are used in Chapter 11 for modelling impulsive noise, and in Chapter 14 for channel equalisation.

Bibliography

- BELL D.A. (1960) *Electrical Noise and Physical Mechanism*. Van Nostrand, London.
- BENNETT W.R. (1960) *Electrical Noise*. McGraw-Hill, New York.
- DAVENPORT W.B. and ROOT W.L. (1958) *An Introduction to the Theory of Random Signals and Noise*. McGraw-Hill, New York.
- GODSILL S.J. (1993) *The Restoration of Degraded Audio Signals*. Ph.D. Thesis, Cambridge University.
- SCHWARTZ M. (1990) *Information Transmission, Modulation and Noise*. 4th Ed., McGraw-Hill, New York.
- EPHRAIM Y. (1992) Statistical Model Based Speech Enhancement Systems. *Proc. IEEE* **80**, **10**, pp. 1526–1555.
- VAN-TREES H.L. (1971) *Detection, Estimation and Modulation Theory*. Parts I, II and III. Wiley, New York.