

CHAPTER 3

FILTERING FOR REMOVAL OF ARTIFACTS

Most biomedical signals appear as weak signals in an environment that is teeming with many other signals of various origins. Any signal other than that of interest could be termed as an interference, artifact, or simply *noise*. The sources of noise could be physiological, the instrumentation used, or the environment of the experiment.

This chapter starts with an introduction to the nature of the artifacts that are commonly encountered in biomedical signals. Several illustrations of signals corrupted by various types of artifacts are provided. Details of the design of filters, spanning a broad range of approaches, from linear time-domain and frequency-domain fixed filters to the optimal Wiener filter to adaptive filters, are then described. The chapter concludes with demonstrations of application of the filters described to ECG, ERP, and VAG signals.

(*Note:* A good background in signal and system analysis [1–4] as well as in probability, random variables, and stochastic processes [5–10] is required to follow the procedures and analysis described in this chapter. Familiarity with systems theory and transforms such as the Laplace transform, the Fourier transform in both the continuous and discrete form, and the z -transform is assumed.)

3.1 Problem Statement

Noise is omnipresent! The problems caused by artifacts in biomedical signals are vast in scope and variety; their potential for degrading the performance of the most sophisticated signal processing algorithms is high. The enormity of the problem of noise removal and its importance are reflected by the size of this chapter and its placement as the first one on signal processing techniques. Let us start with a generic statement of the problem and investigate its nature:

Analyze the various types of artifacts that corrupt biomedical signals and explore filtering techniques to remove them without degrading the signal of interest.

During a procedure to acquire the ECG signal, if the subject coughs or squirms, the EMG associated with such activity will create an interference or artifact. In adult patients, such physiological interference may be minimized by strict instructions and self-control; this solution may, however, not be applicable to infants, children, and severely ill patients. An intriguing example of physiological interference is that of the expectant mother's ECG appearing along with that of the fetus, with the latter being of interest. No external control is feasible or desirable in this case, and the investigator is forced to develop innovative solutions to extract the signal of interest.

Due to the weak levels of most biomedical signals at their source, high amplification factors of several hundred to several thousand may be required. Electronic noise in the instrumentation amplifiers also gets amplified along with the desired signal. While it is possible to reduce the thermal component of the noise by cooling the devices to low temperatures, this step may not be practical in most applications; the cost could also be prohibitive. Low-noise power supplies and specialized electronic amplifiers with high input impedance, high common-mode rejection ratio, and high power-supply rejection ratio are desirable for the acquisition of biomedical signals [11].

Our environment is filled with EM waves, both natural and man-made. EM waves broadcast by radio and television (TV) stations and those radiated by fluorescent lighting devices, computer monitors, and other systems used in the laboratory or work environment are picked up by cables, devices, and connectors. The 50 Hz or 60 Hz power-supply waveform is notorious for the many ways in which it can get mixed with and corrupt the signal of interest. Such interference may be termed as being due to the environment of the experiment. Simple EM shielding of cables and grounding of the chassis of equipment reduce EM and power-supply interference in most cases. Experiments dealing with weak signals such as ERPs and EEGs may require a wire-mesh-shielded (Faraday) cage to contain the subject and the instruments.

The ECG is a relatively strong signal with a readily identifiable waveform. Most types of interference that affect ECG signals may be removed by bandpass filters. Other signals of less recognizable waveforms and broader bandwidths may not be amenable to simple filtering procedures. In the case of signals such as ERPs or SEPs, the noise levels could be much higher than the signal levels, rendering the latter

unrecognizable in a single recording. It is important to gain a good understanding of the noise processes involved before one attempts to filter or preprocess a signal.

3.2 Random, Structured, and Physiological Noise

Several types of artifacts or noise could corrupt biomedical signals in various ways. The following sections provide descriptions of various types of noise and methods to characterize them.

3.2.1 Random noise

A *deterministic signal* is one whose value at a given instant of time may be computed using a closed-form mathematical function of time, or predicted from a knowledge of a few past values of the signal. A signal that does not meet this condition may be labeled as a *nondeterministic signal* or a random signal. The term *random noise* refers to an interference that arises from a random process, such as thermal noise in electronic devices.

Test for randomness: Random signals are generally expected to display more excursions about a certain reference level within a specified interval than signals that are predictable. Kendall [170] and Challis and Kitney [171] recommend a test for randomness based upon the number of peaks or troughs in the signal. A peak or a trough is defined by a set of three consecutive samples of the signal, with the central sample being either the maximum or minimum, respectively. As the direction of excursion of the signal changes at peaks and troughs, such points are collectively known as *turning points*. At a turning point, the sign of the first-order difference (derivative) at the current sample of the signal is not equal to that at the preceding sample. Given a signal of N samples, the signal may be labeled as being random if the number of turning points is greater than the threshold $\frac{2}{3}(N - 2)$ [170, 171]. In the case of a signal of varying characteristics, that is, a nonstationary signal, the test would have to be conducted using a running window of N samples. The width of the window should be chosen by taking into consideration the shortest duration over which the signal may remain in a given state. The method as above was used by Mintchev et al. [70] to study the dynamics of the level of randomness in EGG signals.

Figure 3.1 illustrates the variation in the number of turning points in a moving window of 50 ms (400 samples with the sampling frequency $f_s = 8 \text{ kHz}$) for the speech signal of the word “safety.” The limit on the number of turning points for randomness for $N = 400$ according to the threshold mentioned above is 265. It is seen from the figure that the test indicates that the signal is random for the fricatives /S/ (over the interval of 0.2 – 0.4 s, approximately) and /F/ (0.7 – 0.9 s), and not random for the remaining portions, as expected. (See also Section 1.2.12 and Figures 1.50 and 1.51.)

Statistical analysis of random processes: A random process is characterized by its probability density function (PDF) representing the probabilities of occurrence of

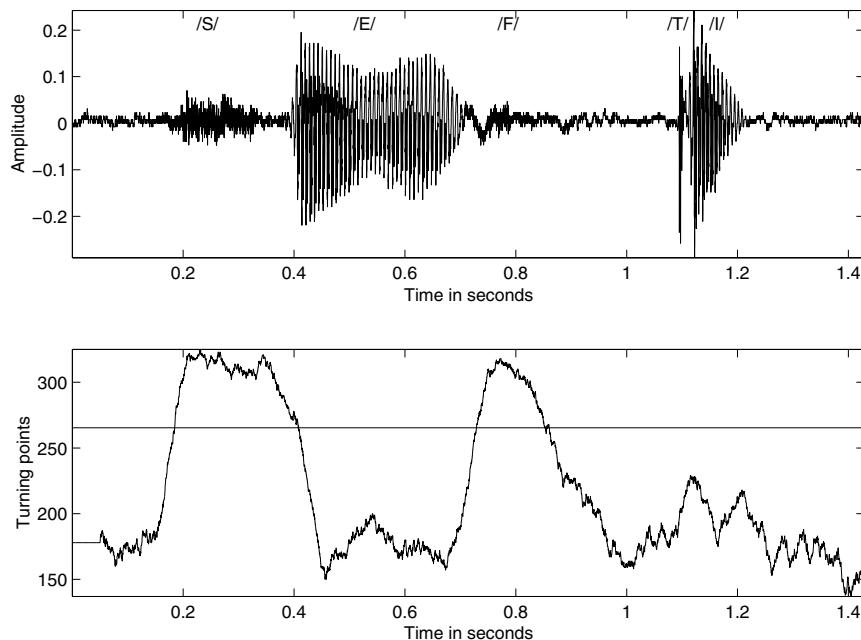


Figure 3.1 Top: Speech signal of the word “safety” uttered by a male speaker. Bottom: Count of turning points in a moving window of 50 ms (400 samples with $f_s = 8 \text{ kHz}$). The threshold for randomness for $N = 400$ is 265.

all possible values of the related random variable. (See Papoulis [5] and Bendat and Piersol [6] for background material on probability, random variables, and stochastic processes.) Consider a random process η that is characterized by the PDF $p_\eta(\eta)$. The mean μ_η of the random process η is given by the first-order moment of the PDF, defined as

$$\mu_\eta = E[\eta] = \int_{-\infty}^{\infty} \eta p_\eta(\eta) d\eta, \quad (3.1)$$

where $E[]$ represents the *statistical expectation operator*. It is common to assume the mean of a random noise process to be zero.

The mean-squared (*MS*) value of the random process η is given by the second-order moment of the PDF, defined as

$$E[\eta^2] = \int_{-\infty}^{\infty} \eta^2 p_\eta(\eta) d\eta. \quad (3.2)$$

The variance σ_η^2 of the process is defined as the second central moment:

$$\sigma_\eta^2 = E[(\eta - \mu_\eta)^2] = \int_{-\infty}^{\infty} (\eta - \mu_\eta)^2 p_\eta(\eta) d\eta. \quad (3.3)$$

The square root of the variance gives the *SD* σ_η of the process. Note that $\sigma_\eta^2 = E[\eta^2] - \mu_\eta^2$. If the mean is zero, it follows that $\sigma_\eta^2 = E[\eta^2]$, that is, the variance and the *MS* values are the same.

The coefficient of variation (*CV*) of a process is defined as the ratio σ/μ . This dimensionless measure places the variability of a process in the context of its mean and is useful in the analysis of the variability of processes with widely different means; however, as $\mu \rightarrow 0$, $CV \rightarrow \infty$. *CV* is applicable in the analysis of nonnegative quantities.

A few other measures used to characterize random processes are skewness, kurtosis, and entropy. Skewness is defined as a normalized version of the third central moment, given by

$$S_\eta = \frac{1}{\sigma_\eta^3} \int_{-\infty}^{\infty} (\eta - \mu_\eta)^3 p_\eta(\eta) d\eta. \quad (3.4)$$

Skewness characterizes the lack of symmetry of a PDF. Symmetric PDFs, such as the Gaussian, have a skewness value of zero. A process with higher probabilities of occurrence of larger values of the associated variable has a PDF with a negative skewness. On the other hand, a process with higher probabilities of occurrence of smaller values of the associated variable has a PDF with a positive skewness.

Kurtosis is defined as a normalized version of the fourth central moment, given by

$$K_\eta = \frac{1}{\sigma_\eta^4} \int_{-\infty}^{\infty} (\eta - \mu_\eta)^4 p_\eta(\eta) d\eta. \quad (3.5)$$

Kurtosis characterizes the presence of a long tail in the PDF. The Gaussian PDF has a kurtosis value of 3. The value $K' = K - 3$ (referred to as kurtosis excess) is used to represent the difference in the kurtosis of a PDF with respect to that of a Gaussian PDF. A positive K' indicates a PDF with a strong peak near the mean that declines with a heavy tail. A PDF that is nearly flat has a negative K' value.

The entropy of a random process is a statistical measure of the information conveyed by the process [18, 172, 173]. It is also a measure of the extent of disorder in the process. The most commonly used measure of entropy of a PDF is defined as

$$H_\eta = - \int_{-\infty}^{\infty} p_\eta(\eta) \log_2[p_\eta(\eta)] d\eta. \quad (3.6)$$

The unit of entropy, defined as above, is *bits (b)*. The entropy of a process is at its maximum when all associated values or events occur with equal probability, that is, the PDF is uniform.

The definitions of the moments of a random process given above refer to a continuous variable and the situation when its PDF is known. In a practical situation, when we have only a certain number, N , of the values of the process observed as $\eta(n)$, $n = 0, 1, 2, \dots, N - 1$, and no knowledge of its PDF, we could estimate the statistical measures from the given samples as

$$\mu_\eta = \frac{1}{N} \sum_{n=0}^{N-1} \eta(n), \quad (3.7)$$

$$MS_\eta = \frac{1}{N} \sum_{n=0}^{N-1} [\eta(n)]^2, \quad (3.8)$$

$$RMS_\eta = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} [\eta(n)]^2}, \quad (3.9)$$

and

$$\sigma_\eta = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} [\eta(n) - \mu_\eta]^2}. \quad (3.10)$$

[Some authors use the notation (nT) , $T = 1/f_s$ being the sampling interval, where f_s is the sampling frequency, to denote the index of a sampled signal; in this book we shall use just (n) , the sample number.]

In the case of a process with a finite number L of discrete (or quantized) values η_l , $l = 0, 1, 2, \dots, L - 1$, the entropy is defined as

$$H_\eta = - \sum_{l=0}^{L-1} p_\eta(\eta_l) \log_2[p_\eta(\eta_l)], \quad (3.11)$$

where $p_\eta(\eta_l)$ is the probability of occurrence of the l^{th} quantized value of η . Each probability value in the equation given above may be estimated from a large number of observations of the values of the process. This form of entropy is known as Shannon entropy.

When the values of a random process η form a time series or a function of time, we have a random signal (or a stochastic process) $\eta(t)$. The statistical measures described above then have physical meanings: The mean represents the DC component, the MS value represents the average power, and the RMS value gives the average noise magnitude or level. The measures are useful in calculating the SNR , which is commonly defined as the ratio of the peak-to-peak amplitude range of the signal to the RMS value of the noise, or as the ratio of the average power of the signal to that of the noise.

Observe the use of the same symbol η to represent the random variable, the random process, and the random signal as a function of time. The subscript of the PDF or the statistical parameter derived indicates the random process of concern. The context of the discussion or expression should make the meaning of the symbol clear.

A biomedical signal of interest $x(t)$ may also, for the sake of generality, be considered to be a realization of a random process x . For example, although a normal heart sound signal is heard as the same comforting *lub-dub* sound over every cycle, the corresponding PCG vibration waveforms are not precisely the same from one cycle to another. The PCG signal may be represented as a random process exhibiting certain characteristics *on the average*.

When a (random) signal $x(t)$ is observed in an environment with random noise, the measured signal $y(t)$ may be treated as a realization of another random process

y. In most cases, the noise is additive, and the observed signal is expressed as

$$y(t) = x(t) + \eta(t). \quad (3.12)$$

Each of the random processes x and y is characterized by its own PDF $p_x(x)$ and $p_y(y)$, respectively.

In most practical applications, the random processes representing a signal of interest and the noise affecting the signal may be assumed to be *statistically independent processes*. Two random processes x and η are said to be statistically independent if their joint PDF $p_{x,\eta}(x, \eta)$ is equal to the product of their individual PDFs given as $p_x(x) p_\eta(\eta)$.

The first-order moment and second-order central moment of the signals $x(t)$ and $y(t)$ in Equation 3.12 are related as

$$E[y] = \mu_y = \mu_x + \mu_\eta; \quad (3.13)$$

if $\mu_\eta = 0$, then $\mu_y = \mu_x$; and

$$E[(y - \mu_y)^2] = \sigma_y^2 = \sigma_x^2 + \sigma_\eta^2, \quad (3.14)$$

where μ represents the mean and σ^2 represents the variance of the random process indicated by the subscript.

Ensemble averages: When the PDFs of the random processes of concern are not known, it is common to approximate the statistical expectation operation by averages computed using a collection or *ensemble* of sample observations of the random process. Such averages are known as *ensemble averages*. Suppose we have M observations of the random process x as functions of time: $x_1(t), x_2(t), \dots, x_M(t)$. We may estimate the mean of the process at a particular instant of time t_1 as

$$\mu_x(t_1) = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{k=1}^M x_k(t_1). \quad (3.15)$$

Figure 3.2 illustrates ten sample acquisitions of flash visual ERPs (see also Figure 3.41). The vertical lines at $t = t_1$ and $t = t_2 = t_1 + \tau$ represent the ensemble averaging process at two different instants of time.

The autocorrelation function (ACF) $\phi_{xx}(t_1, t_1 + \tau)$ of a random process x that is a time series is given by

$$\begin{aligned} \phi_{xx}(t_1, t_1 + \tau) &= E[x(t_1) x(t_1 + \tau)] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(t_1) x(t_1 + \tau) p_{x_1, x_2}(x_1, x_2) dx_1 dx_2, \end{aligned} \quad (3.16)$$

where x_1 and x_2 represent the random variables corresponding to the processes $x(t_1)$ and $x(t_1 + \tau)$, respectively, and $p_{x_1, x_2}(x_1, x_2)$ is the joint PDF of the two processes. The ACF may be estimated as

$$\phi_{xx}(t_1, t_1 + \tau) = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{k=1}^M x_k(t_1) x_k(t_1 + \tau), \quad (3.17)$$

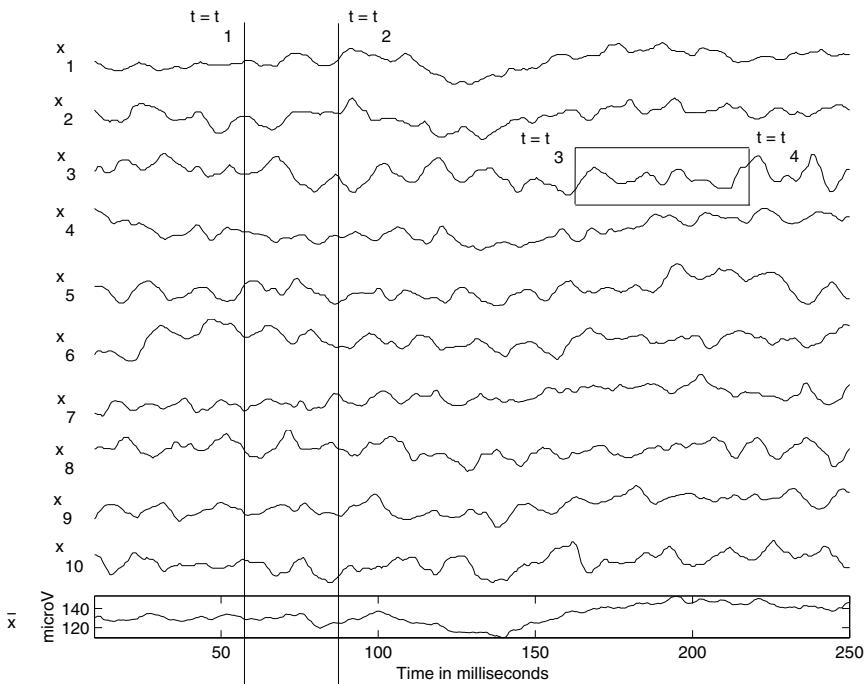


Figure 3.2 Ten sample acquisitions (x_1 to x_{10}) of individual flash visual ERPs from the occipital midline (oz) position of a normal adult male (the author of this book!). The earlobes were used to form the reference lead (a1a2), and the left forehead was used as the reference (see Figure 1.37). The signals may be treated as ten realizations of a random process in the form of time series or signals. The vertical lines at $t = t_1$ and $t = t_2 = t_1 + \tau$ represent the ensemble averaging process at two different instants of time. The last plot (framed) gives the ensemble average or prototype $\bar{x}(t)$ of the ten individual signals. The horizontal box superimposed on the third trace represents the process of computing temporal statistics over the duration $t = t_3$ to $t = t_4$ of the sample ERP $x_3(t)$. See also Figure 3.41. Data courtesy of L. Alfaro and H. Darwish, Alberta Children's Hospital, Calgary.

where τ is the delay parameter. If the signals are complex, one of the functions in the expression above should be conjugated; in this book, we shall deal with physiological signals that are always real. The two vertical lines at $t = t_1$ and $t = t_2 = t_1 + \tau$ in Figure 3.2 represent the ensemble averaging process to compute $\phi_{xx}(t_1, t_2)$. The ACF indicates how the values of a signal at a particular instant of time are statistically related to (or have characteristics in common with) the values of the same signal at another instant of time.

When dealing with random processes that are observed as functions of time (or stochastic processes), it becomes possible to compute ensemble averages at every

point of time. Then, we obtain an averaged function of time $\bar{x}(t)$ as

$$\bar{x}(t) = \mu_x(t) = \frac{1}{M} \sum_{k=1}^M x_k(t) \quad (3.18)$$

for all time t . The signal $\bar{x}(t)$ may be used to represent the random process x as a prototype; see the last trace (framed) in Figure 3.2.

Time averages: When we have a sample observation of a random process $x_k(t)$ as a function of time, it is possible to compute *time averages* or *temporal statistics* by integrating along the time axis, such as

$$\mu_x(k) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x_k(t) dt. \quad (3.19)$$

The integral would be replaced by a summation in the case of sampled or discrete-time signals. The time-averaged ACF $\phi_{xx}(\tau, k)$ is given by

$$\phi_{xx}(\tau, k) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x_k(t) x_k(t + \tau) dt. \quad (3.20)$$

(See Section 6.3 for details on estimation of the ACF of finite-length data sequences.) The horizontal box superimposed on the third trace in Figure 3.2 represents the process of computing temporal statistics over the duration $t = t_3$ to $t = t_4$ of the sample ERP $x_3(t)$ selected from the ensemble of ERPs illustrated in the figure.

Random noise may thus be characterized in terms of ensemble and/or temporal statistics. The mean does not play an important role: it is usually assumed to be zero, or may be subtracted out if it is not zero. The ACF plays an important role in the characterization of random processes. The Fourier transform of the ACF is the power spectral density (PSD) function, which is useful in spectral analysis and filter design.

Covariance and cross-correlation: When two random processes x and y need to be compared, we could compute the covariance between them as

$$C_{xy} = E[(x - \mu_x)(y - \mu_y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) p_{x,y}(x, y) dx dy, \quad (3.21)$$

where $p_{x,y}(x, y)$ is the joint PDF of the two processes. The covariance parameter may be normalized to get the correlation coefficient, defined as

$$\rho_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y}, \quad (3.22)$$

with $-1 \leq \rho_{xy} \leq +1$. A high covariance indicates that the two processes have similar statistical variability or behavior. The processes x and y are said to be uncorrelated if $\rho_{xy} = 0$. Two processes that are statistically independent are also uncorrelated; the converse of this property is, in general, not true.

When dealing with random processes x and y that are functions of time, the cross-correlation function (CCF) between them is defined as

$$\theta_{xy}(t_1, t_1 + \tau) = E[x(t_1)y(t_1 + \tau)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(t_1) y(t_1 + \tau) p_{x,y}(x, y) dx dy. \quad (3.23)$$

Correlation functions are useful in analyzing the nature of variability and spectral bandwidth of signals, as well as for detection of events by template matching.

3.2.2 Structured noise

Power-line interference at 50 Hz or 60 Hz is an example of structured noise: The typical waveform of the interference is known in advance. It should, however, be noted that the phase of the interfering waveform will not usually be known. Furthermore, the interfering waveform may not be an exact sinusoid; this is indicated by the presence of harmonics of the fundamental 50 Hz or 60 Hz component. Analysis of the power spectrum of a given noisy signal can reveal the presence of structured or periodic noise in the form of peaks or spikes at the fundamental frequency and its harmonics.

3.2.3 Physiological interference

The human body is a complex conglomeration of several systems and processes. Several physiological processes could be active at a given instant of time, each one producing many signals of different types. A patient or experimental subject may not be able to exercise control on all physiological processes and systems. The appearance of signals from systems or processes other than those of interest may be termed as physiological interference; several examples are listed below.

- EMG related to coughing, breathing, or squirming affecting the ECG
- EGG interfering with precordial ECG
- Maternal ECG getting added to the fetal ECG of interest
- ECG interfering with the EEG
- Ongoing EEG in ERPs and SEPs
- Breath, lung, or bowel sounds contaminating the heart sounds (PCG)
- Heart sounds getting mixed with breath or lung sounds
- Muscle sound (VMG) interference in joint sounds (VAG)
- Needle-insertion activity appearing at the beginning of a needle-EMG recording

Physiological interference may not be characterized by any specific waveform or spectral content, and is typically dynamic and nonstationary (varying with the level

of the activity of relevance and hence with time; see the next section for a discussion on stationarity). Thus, simple linear bandpass filters will usually not be effective in removing physiological interference.

3.2.4 Stationary, nonstationary, and cyclostationary processes

We saw in Section 3.2.1 that random processes may be characterized in terms of their ensemble and/or temporal statistics. A random process is said to be *stationary in the strict sense* or *strongly stationary* if its statistics are not affected by a shift in the origin of time. In practice, only first-order and second-order averages are used. A random process is said to be *weakly stationary* or *stationary in the wide sense* if its mean is a constant and its ACF depends only upon the difference (or shift) in time. Then, from Equations 3.15 and 3.17, we have $\mu_x(t_1) = \mu_x$ and $\phi_{xx}(t_1, t_1 + \tau) = \phi_{xx}(\tau)$. The ACF is now a function of the delay or shift parameter τ only; the PSD of the process does not vary with time.

A stationary process is said to be *ergodic* if the temporal statistics computed are independent of the sample observed; that is, the same result is obtained for any sample observation $x_k(t)$. The time averages in Equations 3.19 and 3.20 are then independent of k : $\mu_x(k) = \mu_x$ and $\phi_{xx}(\tau, k) = \phi_{xx}(\tau)$. All ensemble statistics may be replaced by temporal statistics when analyzing ergodic processes. Ergodic processes are an important type of stationary random processes because their statistics may be computed from a single observation as a function of time. The use of ensemble and temporal averages for noise filtering is illustrated in Sections 3.5.1 and 3.5.2, respectively.

Signals or processes that do not meet the conditions described above may be, in general, called *nonstationary processes*. A nonstationary process possesses statistics that vary with time. It is readily seen in Figure 1.28 (see also Figure 3.6) that the mean level (baseline) of the signal is varying over the duration of the signal. Therefore, the signal is nonstationary in the mean, a first-order statistical measure. Figure 3.3 illustrates the variance of the speech signal of the word “safety” computed in a moving window of 50 ms (400 samples with $f_s = 8 \text{ kHz}$). Because the variance changes substantially from one portion of the signal to another, it should be concluded that the signal is nonstationary in its second-order statistics (variance, *SD*, or *RMS*). While the speech signal is stationary in the mean, this is not an important characteristic as the mean is typically removed from speech signals. (A DC signal bears no information related to vibration or sound.)

Note that variance displays a behavior that is almost the opposite of that of turning points in Figure 3.1. Variance is sensitive to changes in amplitude, with large swings about the mean leading to large variance values. The procedure to detect turning points examines the presence of peaks and troughs with no consideration of their relative amplitudes; the low-amplitude ranges of the fricatives in the signal have resulted in low variance values, even though their counts of turning points are high.

Most biomedical systems are dynamic and produce nonstationary signals (for example, EMG, EEG, VMG, PCG, VAG, and speech signals). However, a physical or physiological system has limitations in the rate at which it can change its character-

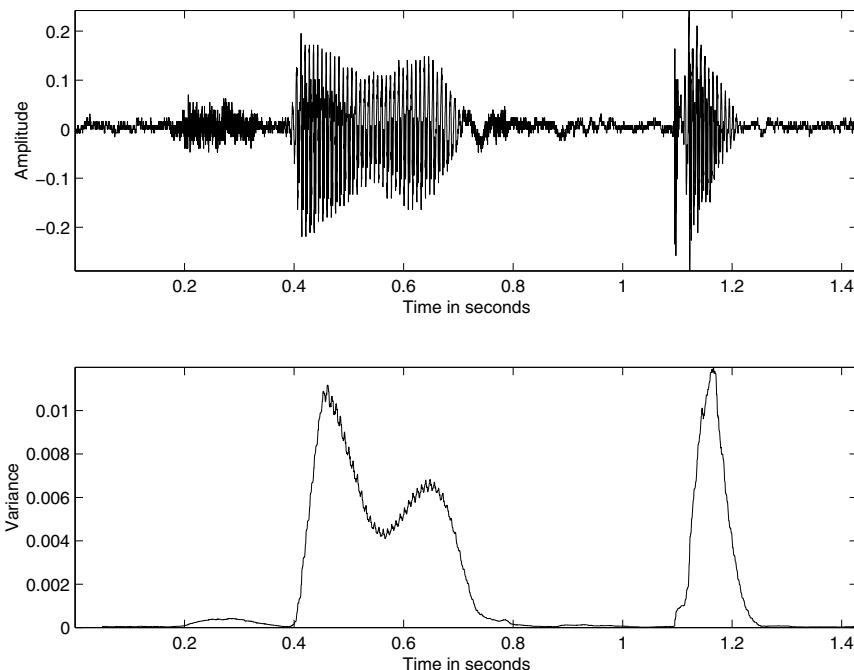


Figure 3.3 Top: Speech signal of the word “safety” uttered by a male speaker. Bottom: Variance computed in a moving window of 50 ms (400 samples with $f_s = 8 \text{ kHz}$).

istics. This limitation facilitates breaking a signal into segments of short duration (typically a few tens of milliseconds), over which the statistics of interest are not varying, or may be assumed to remain the same. The signal is then referred to as a *quasistationary process*; the approach is known as *short-time analysis*. Figure 3.4 illustrates the spectrogram of the speech signal of the word “safety.” The spectrogram was computed by computing an array of magnitude spectra of segments of the signal of duration 64 ms; an overlap of 32 ms was permitted between successive segments. It is evident that the spectral characteristics of the signal vary over its duration: The fricatives demonstrate more high-frequency content than the vowels and also lack formant (resonance) structure. The signal is, therefore, nonstationary in terms of its PSD; because the PSD is related to the ACF, the signal is also nonstationary in the second-order statistical measure of the ACF. However, it can be observed that the PSD shows the same or similar spectral content over the duration of each of the phonemes present in the signal. Thus, over such intervals, the signal may be considered to be quasistationary.

Further discussion and examples of techniques of this nature are presented in Sections 8.4.1 and 8.5. Adaptive signal processing techniques may also be designed to detect changes in certain statistical measures of an observed signal; the signal may then be broken into quasistationary segments of variable duration that meet the

specified conditions of stationarity. Methods for analysis of nonstationary signals are discussed in Chapter 8. Adaptive segmentation of the EEG, VAG, and PCG signals is discussed in Sections 8.5, 8.6, 8.9, and 8.10.

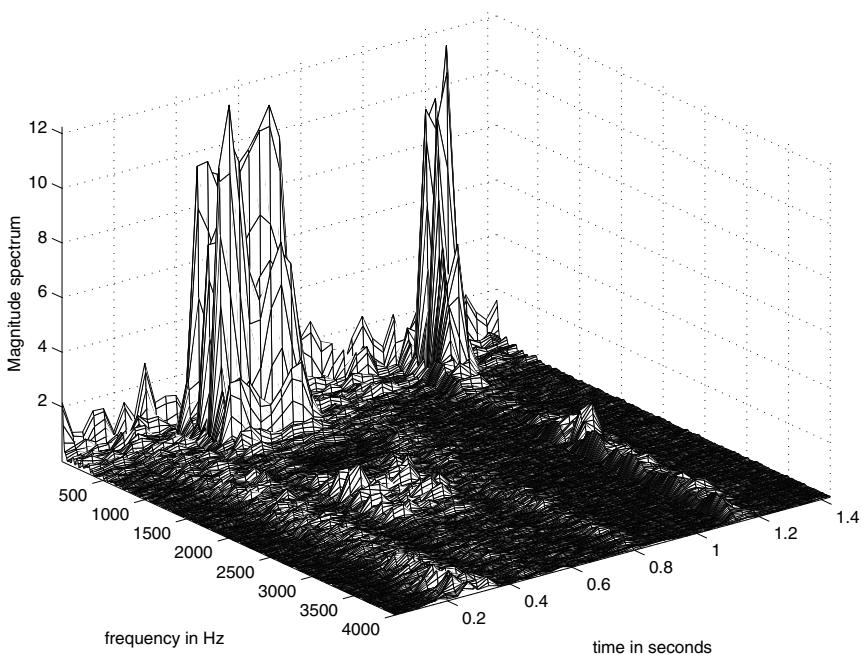


Figure 3.4 Spectrogram of the speech signal of the word “safety” uttered by a male speaker. (The signal is also illustrated in Figures 1.50, 3.1, and 3.3.) Each function or plot along a line parallel to the frequency axis represents the magnitude spectrum of the signal in a moving window of duration 64 ms (512 samples with $f_s = 8 \text{ kHz}$), with the window advance interval being 32 ms. The spectrogram is plotted on a linear scale to display better the major differences between the voiced and unvoiced sounds.

Certain systems, such as the cardiac system, normally perform rhythmic operations. The resulting signal, such as the ECG, PCG, or carotid pulse, is then almost periodic and may be referred to as a *cyclostationary signal*. The statistics of the PCG signal vary within the duration of a cardiac cycle, especially when murmurs are present, but repeat themselves at regular intervals. The cyclic repetition of the process facilitates ensemble averaging, using epochs or events extracted from an observation of the signal over many cycles (which is, strictly speaking, a single function of time). Exploitation of the cyclic nature of the ECG signal for synchronized averaging to reduce noise is illustrated in Section 3.5.1. Application of the same concept to estimate the envelopes of PCG signals is described in Section 5.5.2. Further extensions of the approach to extract A2 from S2 in PCG signals are demonstrated in Section 4.11. Procedures to estimate the PSDs of PCG segments in systole and diastole are presented in Section 6.3.5.

3.3 Illustration of the Problem with Case Studies

The following case studies present several examples of various types of interference in biomedical signals of different origins. The aim of this section is to gain familiarity with the various possibilities of interference and their general characteristics. Filtering techniques to remove various types of interference are described in sections to follow.

3.3.1 Noise in event related potentials

An ERP is a signal obtained in response to a stimulus. The response is usually of small amplitude (of the order of $10 \mu V$), and it is submerged in ambient EEG activity and noise that could be larger than the ERP. The waveform of a single response may be barely recognizable against the background activity. Figure 3.2 shows ten individual flash visual ERP signals. The signals were recorded at the occipital midline position, with the left and right earlobes combined to form the reference lead. The left forehead was used as the reference. The ERP signals are buried in ongoing EEG and power-line (60 Hz) interference, and they cannot be analyzed using the individual acquisitions shown in the figure.

3.3.2 High frequency noise in the ECG

Figure 3.5 shows a segment of an ECG signal with high-frequency noise. The noise could be due to the instrumentation amplifiers, the recording system, pickup of ambient EM signals by the cables, and other sources. The signal illustrated has also been corrupted by power-line interference at 60 Hz and its harmonics, which may also be considered as a part of high-frequency noise relative to the low-frequency nature of the ECG signal.

3.3.3 Motion artifact in the ECG

Low-frequency artifacts and baseline drift may be caused in chest-lead ECG signals by coughing or breathing with large movement of the chest, or when an arm or leg is moved in the case of limb-lead ECG acquisition. The EGG is a common source of artifact in chest-lead ECG. Poor contact and polarization of the electrodes may also cause low-frequency artifacts. Baseline drift may sometimes be caused by variations in temperature and bias in the instrumentation and amplifiers as well. Figure 3.6 shows an ECG signal with low-frequency artifact. Baseline drift makes analysis of isoelectricity of the ST segment difficult. A large baseline drift may cause the positive or negative peaks in the ECG to be clipped by the amplifiers or the ADC.

3.3.4 Power line interference in ECG signals

The most commonly encountered periodic artifact in biomedical signals is the power-line interference at 50 Hz or 60 Hz . If the power-line waveform is not a pure sinu-

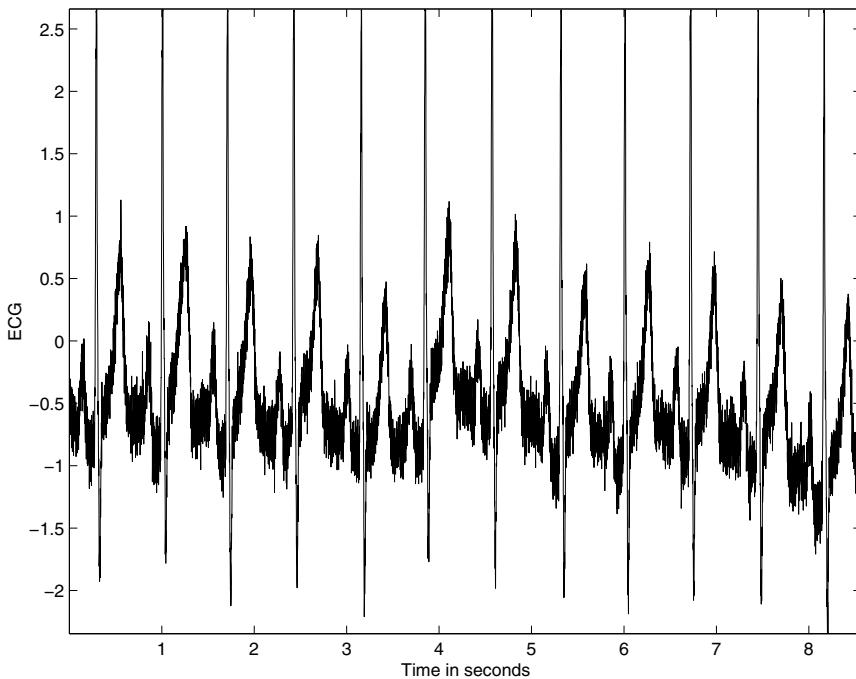


Figure 3.5 ECG signal with high-frequency noise.

soid due to distortions, harmonics of the fundamental frequency could also appear. Harmonics will also appear if the interference is a periodic waveform that is not a sinusoid (such as rectangular pulses).

Power-line interference may be difficult to detect visually in signals having non-specific waveforms such as the PCG or EMG; however, the interference is easily visible if present on well-defined signal waveforms such as the ECG or carotid pulse signals. In either case, the power spectrum of the signal should provide a clear indication of the presence of power-line interference as an impulse or spike at 50 Hz or 60 Hz; harmonics, if present, will appear as additional spikes at integral multiples of the fundamental frequency.

Figure 3.7 shows a segment of an ECG signal with 60 Hz interference. Observe the regular or periodic structure of the interference, which rides on top of the ECG waves. Figure 3.8 shows the power spectrum of the signal. The periodic interference is clearly displayed not only as a spike at its fundamental frequency of 60 Hz, but also as spikes at 180 Hz and 300 Hz, which represent the third and fifth harmonics, respectively. (The recommended sampling rate for ECG signals is 500 Hz; the higher rate of 1,000 Hz was used in this case because the ECG was recorded as a reference signal with the PCG. The larger bandwidth also permits better illustration of artifacts and filtering.)

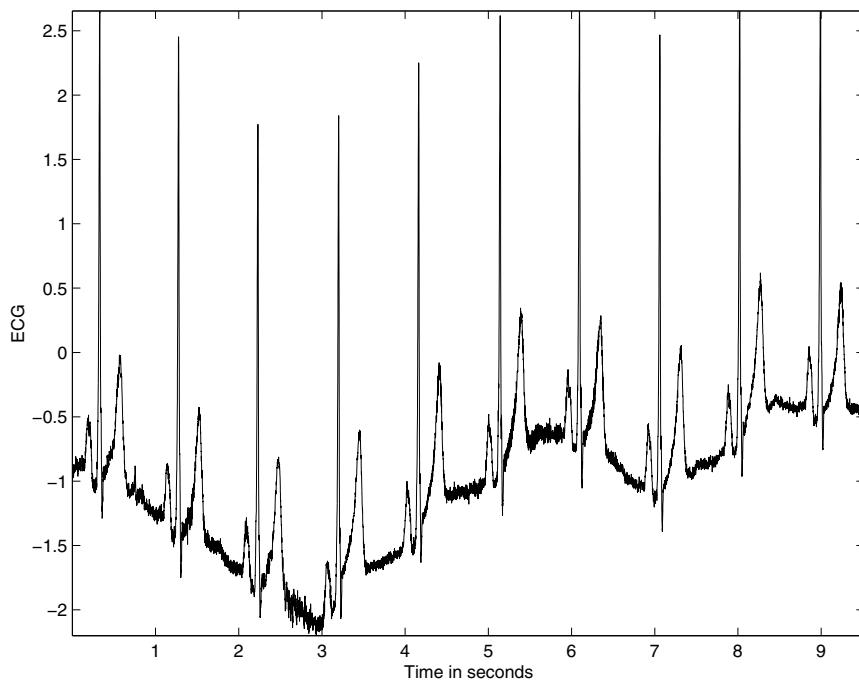


Figure 3.6 ECG signal with low-frequency artifact.

The bandwidth of interest of the ECG signal, which is usually in the range $0.05 - 100\text{ Hz}$, includes the 60 Hz component; hence simple lowpass filtering will not be appropriate for removal of power-line interference. Lowpass filtering of the ECG to a bandwidth lower than 60 Hz could smooth and blur the QRS complex as well as affect the PQ and ST segments. The ideal solution would be to remove the 60 Hz component without sacrificing any other component.

3.3.5 Maternal interference in fetal ECG

Figure 3.9 shows an ECG signal recorded from the abdomen of a pregnant woman. Shown also is a simultaneously recorded ECG from the woman's chest. Comparing the two, we see that the abdominal ECG demonstrates multiple peaks (QRS complexes) corresponding to the maternal ECG (occurring at the same time instants as the QRS complexes in the chest lead) as well as several others at weaker levels and a higher repetition rate. The QRS complexes that are not of the expectant mother represent the ECG of the fetus. Observe that the QRS complex shapes of the maternal ECG from the chest and abdominal leads have different shapes due to the projection of the cardiac electrical vector on to different axes. Given that the two signals being combined have almost the same bandwidth, how would we be able to separate them and obtain the fetal ECG of interest?

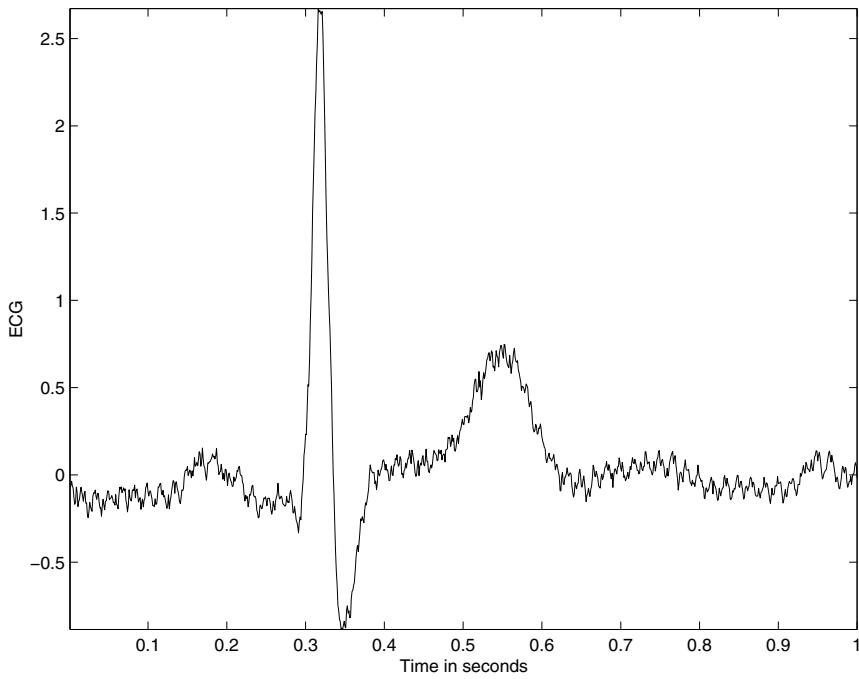


Figure 3.7 ECG signal with power-line (60 Hz) interference.

3.3.6 Muscle contraction interference in VAG signals

Figure 3.10 shows the recording setup used by Zhang et al. [125] to study the possibility of VMG signals appearing as muscle-contraction interference in VAG signals. The left-hand column in Figure 3.11 shows VMG signals recorded using accelerometers placed at the distal rectus femoris (thigh), midpatella (knee cap), tibial tuberosity, and midtibial shaft positions of a subject during isometric contraction of the rectus femoris muscle (with no leg or knee movement). The right-hand column of the figure shows vibration signals recorded at the same positions using the same accelerometers, but during isotonic contraction (swinging movement of the leg). The top signal (a) in the right-hand column indicates the VMG signal generated at the rectus femoris during acquisition of the VAG signals; parts (b)–(d) of the right-hand column show the VAG signals.

VAG signals are difficult to analyze because they have no predefined or recognizable waveforms; it is even more difficult to identify any noise or interference that may be present in VAG signals. The signals shown in Figure 3.11 indicate that a transformed version of the VMG could get added to the VAG, especially during extension of the leg when the rectus femoris muscle is active [the second halves of the VAG signals in parts (b)–(d) of the right-hand column]. The left-hand column of VMG signals in Figure 3.11 illustrates that the VMG generated at the distal rec-

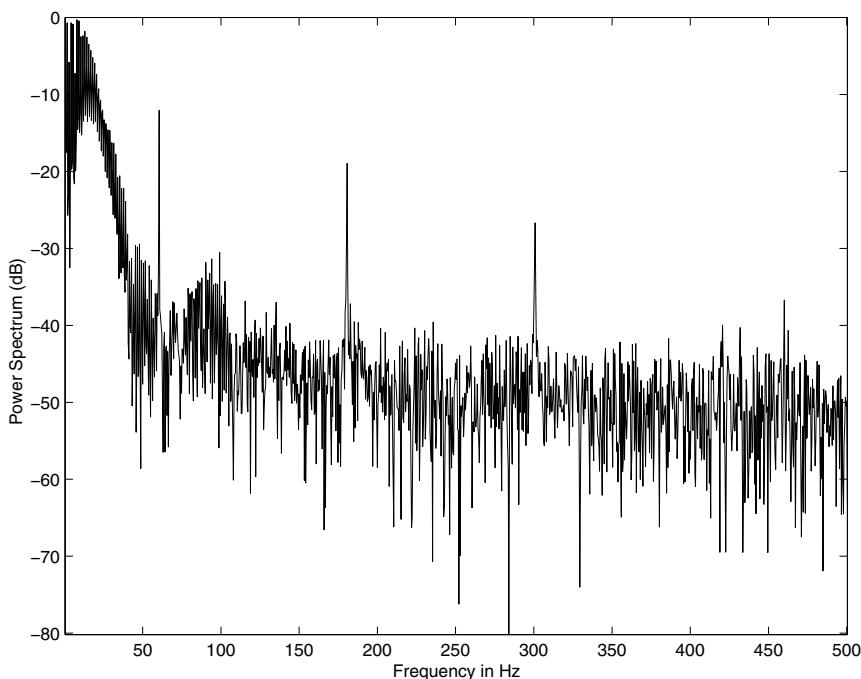


Figure 3.8 Power spectrum of the ECG signal in Figure 3.7 with power-line interference. The spectrum illustrates peaks at the fundamental frequency of 60 Hz as well as the third and fifth harmonics at 180 Hz and 300 Hz , respectively.

tus femoris gets transmitted well down the leg and appears at the other recording positions. It may be observed from the VAG signals in the right-hand column that vibration signals comparable to the VMG are present in the VAG channels (b)–(d) during extension (second halves) but are not as prominent in flexion (first halves). Interestingly, the knee-joint “crepitus” and click signals that appear in the first half of the VAG signal at the midpatella position [right (b)] have been transmitted downward along the leg to the tibial tuberosity [right (c)] and midtibial shaft [right (d)] positions farther down the leg, presumably along the tibia, but not upwards to the distal rectus femoris position [right (a)].

It should also be noted that the VAG signal cannot be expected to be the same during the extension and flexion parts of a swing cycle: Extension causes more stress or force per unit area on the patellofemoral joint than flexion. Furthermore, the VAG and VMG signals are nonstationary: Characteristics of the VAG vary with the quality of the cartilage surfaces that come into contact at different joint angles, while the VMG varies in accordance with the level of contraction of the muscles involved. To make the problem even more difficult, the bandwidths of the two signals overlap in the range of about $0 - 100\text{ Hz}$. These factors make removal of the VMG or muscle-contraction interference from VAG signals a challenge.

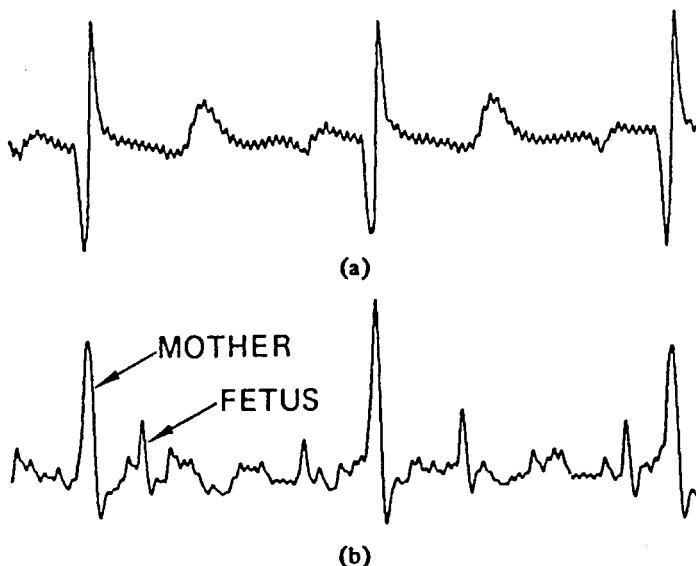


Figure 3.9 ECG signals of a pregnant woman from abdominal and chest leads: (a) chest-lead ECG and (b) abdominal-lead ECG; the former presents the maternal ECG whereas the latter is a combination of the maternal and fetal ECG signals. (See also Figure 3.101.) Reproduced with permission from B. Widrow, J.R. Glover, Jr., J.M. McCool, J. Kaunitz, C.S. Williams, R.H. Hearn, J.R. Zeidler, E. Dong, Jr., and R.C. Goodlin, Adaptive noise cancelling: Principles and applications, *Proceedings of the IEEE*, 63(12):1692–1716, 1975. ©IEEE.

3.3.7 Potential solutions to the problem

Now that we have gained an understanding of a few sources of artifacts in biomedical signals and their nature, we are prepared to look at specific problems and develop effective filtering techniques to solve them. The following sections present studies of artifacts of various types and demonstrate increasingly complex signal processing techniques to remove them. The problem statement at the beginning of each section defines the nature of the problem in as general terms as possible, sets the terms and conditions, and defines the scope of the investigation to follow. The solution proposed provides the details of an appropriate filtering technique. Each solution is demonstrated with an illustration of its application. Further examples of application of the techniques studied are provided at the end of the chapter. Comparative evaluation of filtering techniques is also provided where applicable. Examples of both success and failure of filtering methods are presented to facilitate learning.

A practical problem encountered by an investigator in the field may not precisely match a specific problem considered in this chapter. However, it is expected that the knowledge of several techniques and an appreciation of the results of their ap-

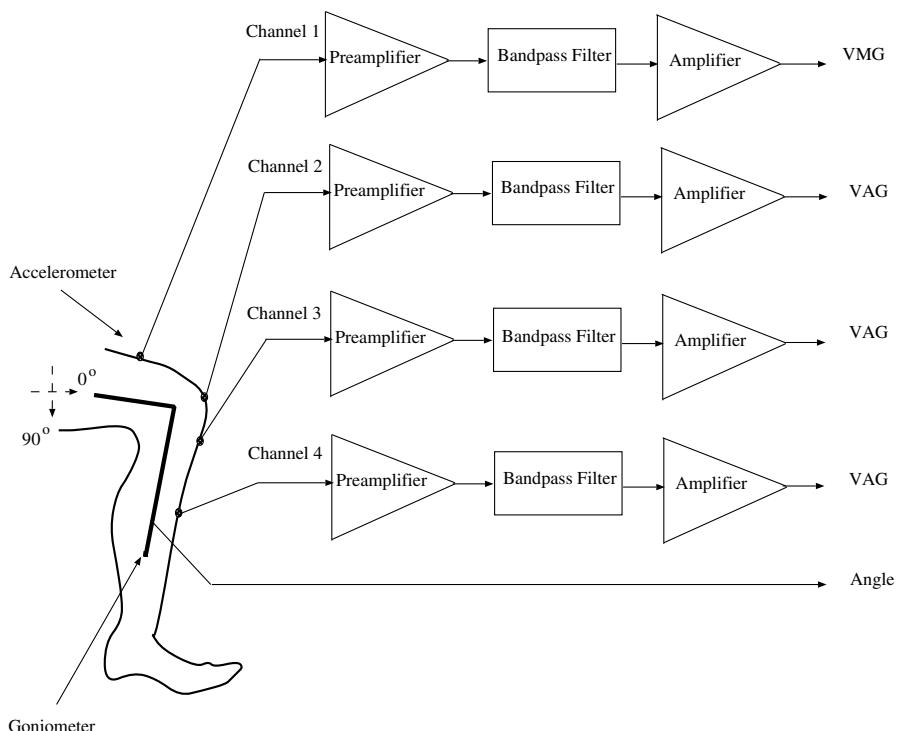


Figure 3.10 Experimental setup to measure VMG and VAG signals at various positions along the leg [125].

plication gained from this chapter will help in designing innovative and appropriate solutions to new problems.

3.4 Fundamental Concepts of Filtering

A filter is a signal processing system, algorithm, or method, realized in hardware or software, that is used to modify a given signal in a particular manner. Quite often, a filtering operation is performed on a signal to remove undesired components that are referred to as noise or artifacts. Regardless of the nature of the operations performed or their effects, filters may be categorized as

- linear or nonlinear,
- stationary or nonstationary,
- fixed (time-invariant) or adaptive (time-variant),
- active or passive, or

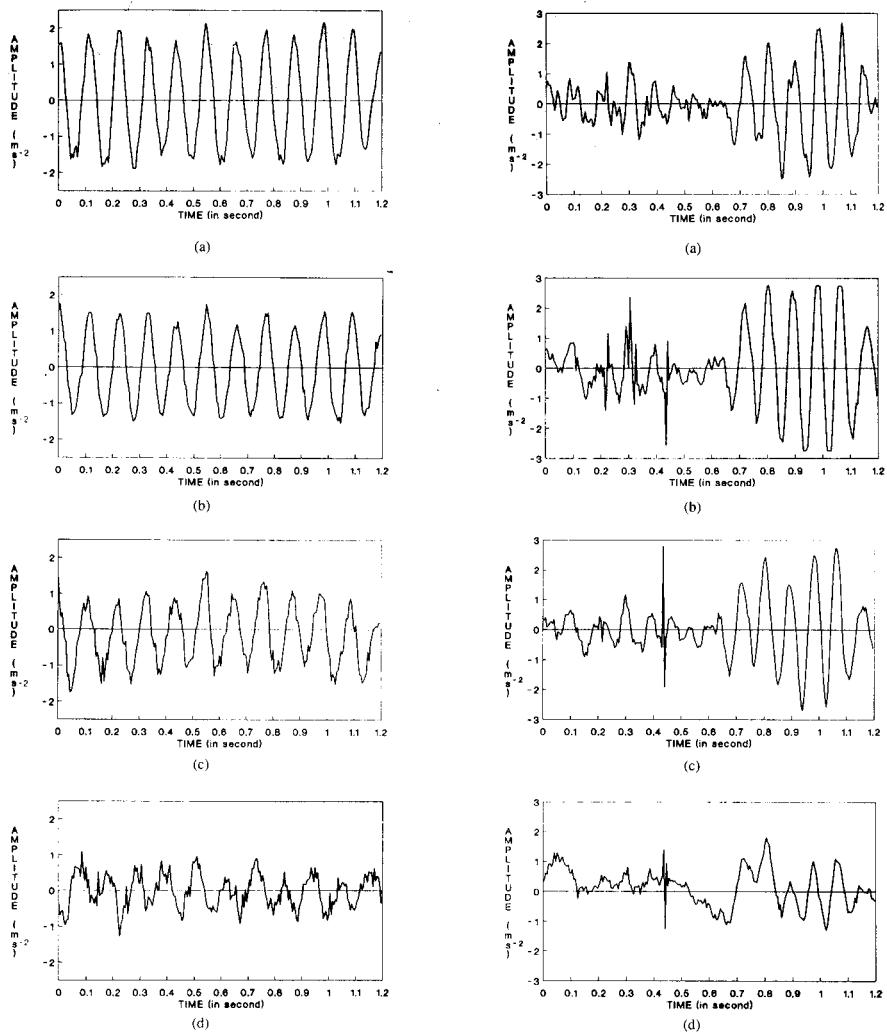


Figure 3.11 Left-hand column: VMG signals recorded simultaneously at (top-to-bottom) (a) the distal rectus femoris, (b) midpatella, (c) tibial tuberosity, and (d) midtibial shaft positions during isometric contraction (no leg or knee movement). Right-hand column: Vibration signals recorded simultaneously at the same positions as above during isotonic contraction (swinging movement of the leg). Observe the muscle-contraction interference appearing in the extension parts (second halves) of each of the VAG signals [plots (b)–(d)] in the right-hand column [125]. The recording setup is shown in Figure 3.10. Reproduced with permission from Y.T. Zhang, R.M. Rangayyan, C.B. Frank, and G.D. Bell, Adaptive cancellation of muscle-contraction interference from knee joint vibration signals, *IEEE Transactions on Biomedical Engineering*, 41(2):181–191, 1994. ©IEEE.

- statistical or deterministic.

In this chapter, we shall study several filters that span almost all of the categories mentioned above. Before studying specific filters, it would be beneficial to remind ourselves of a few fundamental notions associated with filters and their characteristics, as presented in the following sections.

3.4.1 Linear shift invariant filters

Linear time-invariant (LTI) or shift-invariant (LSI) filters form an important category of filters that have a long and well-established history in the field of signal processing [1–3, 15]. The following presentation summarizes important notions and characteristics of LSI filters and filtering operations. Both continuous-time and discrete-time notations are used, as appropriate or convenient. The proofs of the various properties stated are left as exercises for the reader. Readers unfamiliar with the material are referred to Lathi [1, 2], Oppenheim et al. [3], and Oppenheim and Schafer [15].

A fundamental characteristic of an LSI system is its impulse response, which is the output of the system when the input is a Dirac delta or impulse function. Before looking at the details of the impulse response, let us review a few important definitions related to the delta or impulse function. In continuous time, the Dirac delta function is defined as [1–3, 15]

$$\delta(t) = \begin{cases} \text{undefined} & \text{at } t = 0, \\ 0 & \text{otherwise,} \end{cases} \quad (3.24)$$

and

$$\int_{t=-\infty}^{\infty} \delta(t) dt = 1. \quad (3.25)$$

The delta function may be visualized as the limit of a function whose integral over the full extent of time or an independent variable is maintained equal to unity while its duration or extent is compressed toward zero. Figure 3.12 demonstrates this definition graphically using a rectangular pulse function. Another example of this phenomenon is given by

$$\delta(t) = \frac{1}{2} \lim_{a \rightarrow 0} a|t|^{(a-1)}. \quad (3.26)$$

Figure 3.13 illustrates three plots of the function defined above for $a = 0.8, 0.4$, and 0.2 . The emergence of the delta function as $a \rightarrow 0$ is evident.

The delta function is the derivative of the unit step function $u(t)$, which is defined as

$$u(t) = \begin{cases} 1 & \text{for } t > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (3.27)$$

The delta function is also defined in terms of its action within an integral over a certain interval $[T_1, T_2]$:

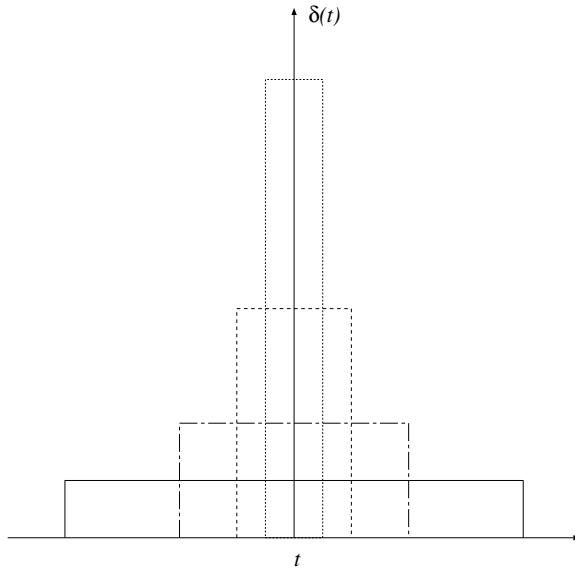


Figure 3.12 Schematic representation of the delta function as a limit of a rectangular pulse function. The time duration of the pulse is reduced while maintaining unit area under the function.

$$\int_{T_1}^{T_2} x(t) \delta(t - t_o) dt = \begin{cases} x(t_o) & \text{if } T_1 < t_o < T_2, \\ 0 & \text{otherwise,} \end{cases} \quad (3.28)$$

where $x(t)$ is a function that is continuous at t_o . This is known as the *sifting property* of the delta function, because the value of the function $x(t)$ at the location t_o of the delta function is sifted or selected from all of its values. The expression may be extended to all t as

$$x(t) = \int_{\alpha=-\infty}^{\infty} x(\alpha) \delta(t - \alpha) d\alpha, \quad (3.29)$$

where α is a temporary variable, which may also be interpreted as resolving the arbitrary signal $x(t)$ into a weighted combination of mutually orthogonal delta functions.

Let us consider a continuous-time signal, $x(t)$, that is being processed by an LSI system, as shown schematically in Figure 3.14. An LSI system is completely characterized or specified by its impulse response, $h(t)$, which is the output of the system when the input is a delta function. The output of the system, $y(t)$, is given by the convolution of the input, $x(t)$, with the impulse response, $h(t)$, defined as

$$y(t) = \int_{\tau=-\infty}^{\infty} x(\tau) h(t - \tau) d\tau, \quad (3.30)$$

where τ is a temporary variable of integration. An equivalent result is given by

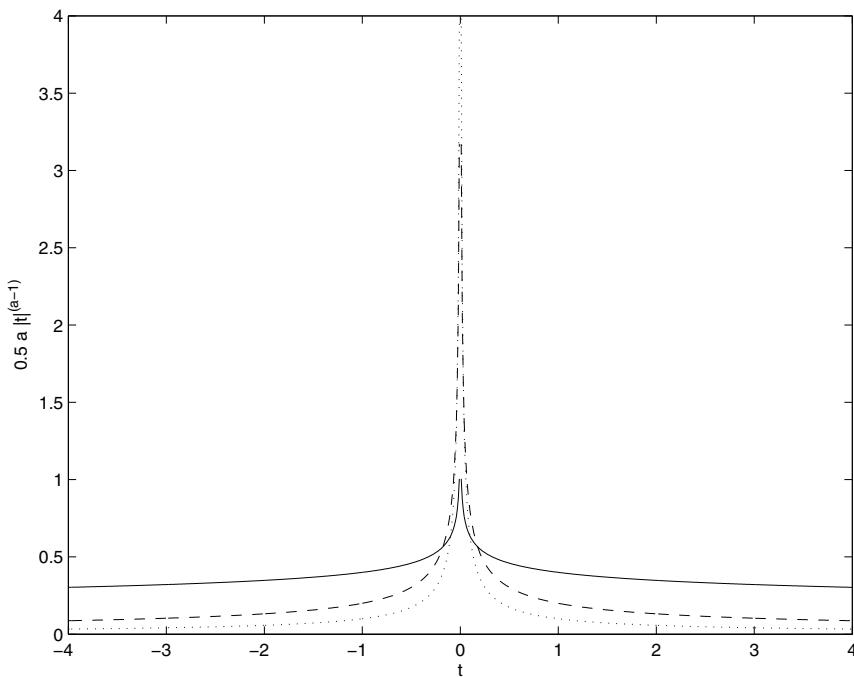


Figure 3.13 The delta function as the limit of $0.5 a |t|^{(a-1)}$ as $a \rightarrow 0$. The function is plotted for $a = 0.8$ (solid line), $a = 0.4$ (dashed line), and $a = 0.2$ (dotted line).

$$y(t) = \int_{\tau=-\infty}^{\infty} h(\tau) x(t-\tau) d\tau. \quad (3.31)$$

The convolution operation given above is *linear convolution*; another version of convolution known as periodic or circular convolution is defined in Equation 3.85. When a causal system or causality is considered, the lower limit of the integrals given above for convolution may be changed to zero and the upper limit changed to t , the current instant of time, resulting in

$$y(t) = \int_{\tau=0}^t x(\tau) h(t-\tau) d\tau \quad (3.32)$$

or

$$y(t) = \int_{\tau=0}^t h(\tau) x(t-\tau) d\tau. \quad (3.33)$$

The discrete-time unit impulse function or delta function is defined as [1–3]

$$\delta(n) = \begin{cases} 1 & \text{if } n = 0, \\ 0 & \text{otherwise.} \end{cases} \quad (3.34)$$

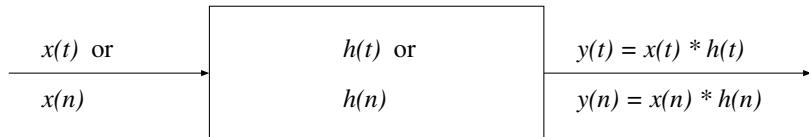


Figure 3.14 A schematic representation of a continuous-time or discrete-time LSI filter.

Figure 3.15 illustrates a few versions of the discrete-time delta function.

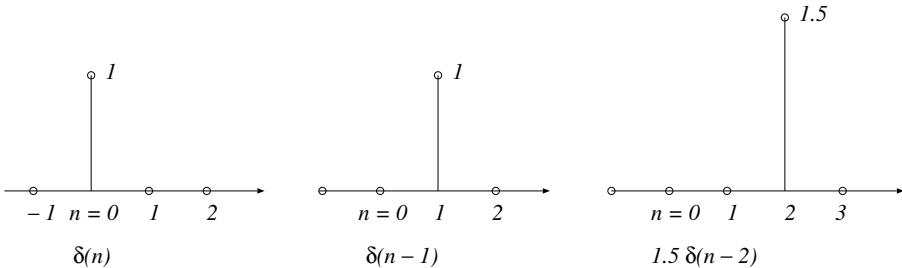


Figure 3.15 A schematic representation of the discrete-time unit impulse function or delta function as well as shifted and scaled versions of the same.

The discrete-time unit step function is defined as

$$u(n) = \begin{cases} 1 & \text{for } n \geq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (3.35)$$

Figure 3.16 illustrates a few versions of the discrete-time unit step function.

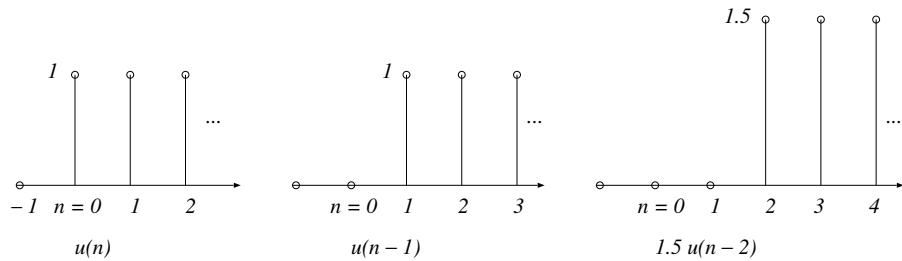


Figure 3.16 A schematic representation of the discrete-time unit step function as well as shifted and scaled versions of the same.

A discrete-time LSI system is shown schematically in Figure 3.17, displaying its impulse response, $h(n)$. The output, $y(n)$, of the system is given by the linear convolution of the input, $x(n)$, with the impulse response, $h(n)$, as

$$y(n) = \sum_{k=0}^n x(k) h(n - k), \quad (3.36)$$

where k is a temporary variable of summation. An equivalent result is given by

$$y(n) = \sum_{k=0}^n h(k) x(n - k). \quad (3.37)$$

In the two equations given above, causality has been assumed.

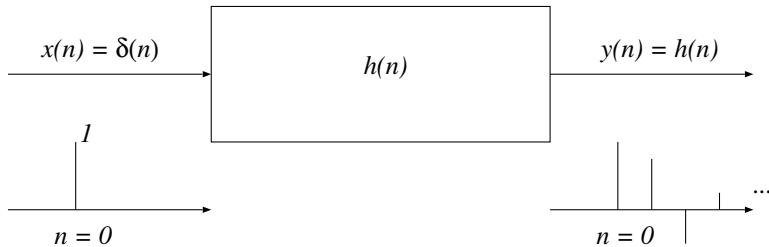


Figure 3.17 A schematic representation of the impulse response, $h(n)$, of a discrete-time LSI filter.

To understand convolution, let us consider the discrete-time version in Equation 3.36. In addition to the change of the independent (time) variable from n to k , there are two important points to note:

- $h(-k)$ represents a reversal in time of the function or signal $h(k)$; and
- $h(n - k)$ represents a shift of the reversed signal $h(-k)$ by n samples.

Multiplication of $h(n - k)$ by $x(k)$ can be viewed as scaling. The summation represents accumulation of the results or integration of $x(k) h(n - k)$ over the interval $k = 0$, the origin of time, to n , the present instant of time. A simple illustration of shifting or adding a delay in time, as well as scaling, is shown in Figure 3.15. A numerical illustration of shifting a signal is shown in Figure 3.18. An illustration of reversing as well as shifting a signal is shown in Figure 3.19.

The linear convolution of two discrete-time signals is illustrated numerically in Figure 3.20. To facilitate understanding of the procedure, we could expand and modify Equation 3.36 as follows:

$$\begin{aligned} y(n) &= \sum_{k=0}^n x(k) h(n - k), \\ y(0) &= \sum_{k=0}^0 x(k) h(0 - k) \end{aligned}$$

$n:$	0	1	2	3	4	5	6	7
$x(n):$	4	5	3	1				
$x(n-1):$		4	5	3	1			
$x(n-2):$			4	5	3	1		
$x(n-3):$				4	5	3	1	

Figure 3.18 A numerical illustration of shifted versions of signal. Blank spaces indicate samples of the signal that are undefined or zero in value. This type of shifting is known as linear shifting. See Figure 3.37 for an example of circular or periodic shifting.

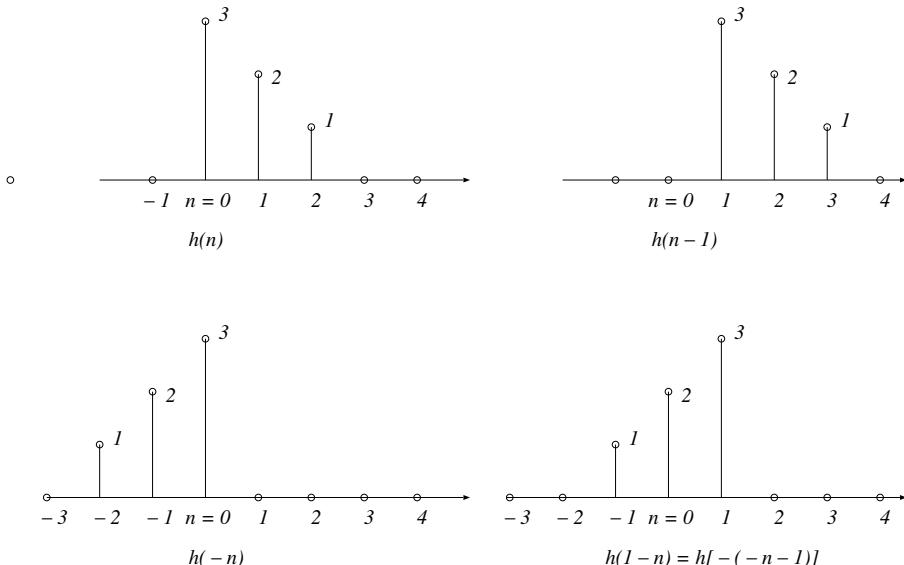


Figure 3.19 A schematic illustration of reversing and shifting of a signal.

$$\begin{aligned}
 &= x(0)h(0). \\
 y(1) &= \sum_{k=0}^1 x(k)h(1-k) \\
 &= x(0)h(1) + x(1)h(0). \\
 y(2) &= \sum_{k=0}^2 x(k)h(2-k) \\
 &= x(0)h(2) + x(1)h(1) + x(2)h(0).
 \end{aligned}$$

$$\begin{aligned}
 y(3) &= \sum_{k=0}^3 x(k) h(3-k) \\
 &= x(0)h(3) + x(1)h(2) + x(2)h(1) + x(3)h(0).
 \end{aligned} \tag{3.38}$$

The procedure is stopped and the output ceases to exist (or has only zero values) when the shift exceeds a certain amount such that $x(k)$ and $h(n - k)$ do not overlap in time any more. It is evident that the linear convolution of two discrete-time signals with durations of N_1 and N_2 samples leads to a result with the duration of $N_1 + N_2 - 1$ samples.

$$\begin{array}{cccccccc}
 n: & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
 x(n): & 4 & 1 & 3 & 1 & & & & \\
 h(n): & 3 & 2 & 1 & & & & & \\
 k: & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
 x(k): & 4 & 1 & 3 & 1 & 0 & 0 & 0 & 0 \\
 \\
 h(0 - k): & 1 & 2 & 3 & & & & & \\
 h(1 - k): & & 1 & 2 & 3 & & & & \\
 h(2 - k): & & & 1 & 2 & 3 & & & \\
 h(3 - k): & & & & 1 & 2 & 3 & & \\
 h(4 - k): & & & & & 1 & 2 & 3 & \\
 h(5 - k): & & & & & & 1 & 2 & 3 \\
 h(6 - k): & & & & & & & 1 & 2 & 3 \\
 \\
 y(n): & 12 & 11 & 15 & 10 & 5 & 1 & 0 & 0 \\
 n: & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7
 \end{array}$$

Figure 3.20 A numerical illustration of the convolution of two discrete-time signals. The notations used in the figure agree with Equation 3.36. For each shift, the corresponding output sample value is obtained by multiplying each pair of samples of $x(k)$ and $h(n - k)$ that overlap (exist at the same instant of time k) and then adding the results.

Taking a different approach, we could expand and modify Equation 3.36 as follows:

$$\begin{aligned}
 y(n) &= \sum_{k=0}^n x(k) h(n - k), \\
 &= x(0) h(n) + x(1) h(n - 1) + x(2) h(n - 2) + x(3) h(n - 3) + \dots
 \end{aligned} \tag{3.39}$$

This result may be interpreted as the sum of several delayed and weighted versions of the impulse response of the system, the weights being provided by the samples of the input signal. Just as the system produces the impulse response, $h(n)$, when the input is $\delta(n)$, when the system is triggered by the input $x(0)$ at $n = 0$, it produces the output $x(0)h(n)$, which lasts for the duration of $h(n)$. When the input is $x(1)$, occurring at $n = 1$, the output is $x(1)h(n - 1)$, recognizing the fact that the system is causal and the corresponding output can only start from $n = 1$. Note that the part of the output signal that started at $n = 0$ will, in general, continue past $n = 1$, and hence the parts of the output that start at $n = 0$ and $n = 1$ will overlap. The process continues, as illustrated in Figure 3.21.

$n:$	0	1	2	3	4	5	6	7
$x(n):$	4	1	3	1				
$h(n):$	3	2	1					
$x(0) h(n - 0):$	12	8	4	0	0	0	0	0
$x(1) h(n - 1):$	0	3	2	1	0	0	0	0
$x(2) h(n - 2):$	0	0	9	6	3	0	0	0
$x(3) h(n - 3):$	0	0	0	3	2	1	0	0
$y(n):$	12	11	15	10	5	1	0	0
$n:$	0	1	2	3	4	5	6	7

Figure 3.21 A numerical illustration of the convolution of two discrete-time signals. The output, $y(n)$, is obtained by adding the corresponding values in the four rows labeled as $x(0)h(n - 0)$ through $x(3)h(n - 3)$. The notations used in the figure agree with Equation 3.39. Compare this procedure with that shown in Figure 3.20. Although the view is different, the end result is the same.

Illustrations of application: To illustrate the effects of random noise being added to a deterministic signal, the following signal was created:

$$x(t) = 5 \sin(2\pi 2t) + 2 \cos(2\pi 3t), \quad (3.40)$$

using a sampling frequency of 2 kHz. Random noise with a Gaussian PDF was simulated and added to the signal x for an effective SNR of 10 dB. The original signal, the noise samples generated, and the noisy signal are shown in Figure 3.22. The distinction between the deterministic nature of the original signal and the random nature of the noise are clearly evident. The histogram of a realization of the noise process used is shown in Figure 3.23, which approximates a Gaussian PDF with zero mean.

The noisy signal was filtered by computing the mean of each sample and the preceding 10 samples. The equation of the corresponding filter could be expressed as

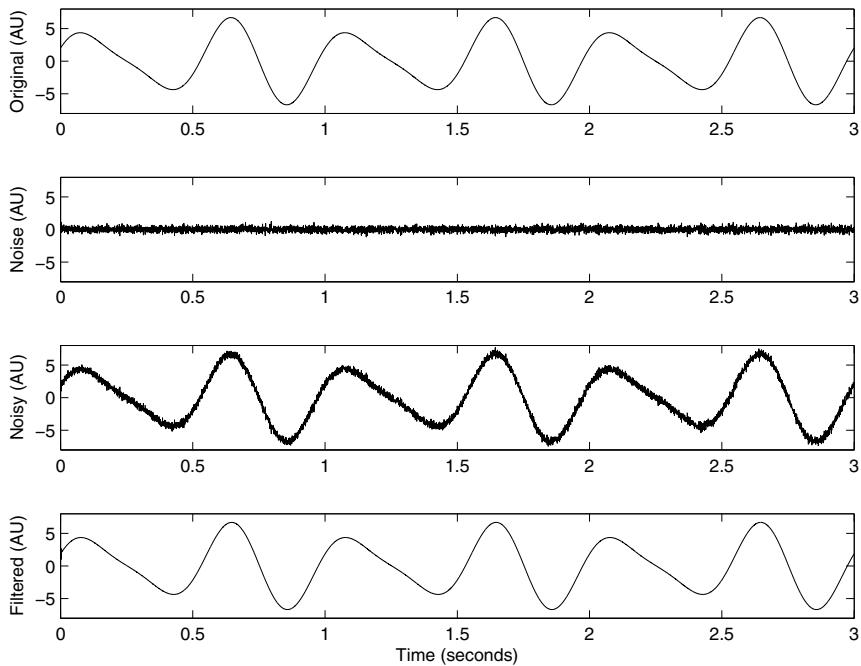


Figure 3.22 Top to bottom: original signal $x(t)$ as in Equation 3.40; Gaussian-distributed random noise; noisy signal; result of filtering using the mean of the signal values in a sliding window of width 11 samples or 5.5 ms. See Figure 3.23 for the histogram of the noise process.

$$y(n) = \frac{1}{11} \sum_{k=0}^{10} x(n-k), \quad (3.41)$$

for $n = 10, 11, \dots, N - 1$, where N is the number of samples in the signal. A filter as above is known as a moving-average (MA) filter: average values of the input signal are computed in a moving temporal window and used to define the output signal. (See Figure 3.2 for a related illustration.) Note that the operation as defined in Equation 3.41 cannot be started until the first 11 samples are available in the input data stream. The operation may also be viewed as a convolution of $x(n)$ with the impulse response of the filter, as in Equation 3.37, with $h(k) = 1/11$, $k = 0, 1, \dots, 10$. Due to the random nature of the noise process (with the mean of the process being zero), the mean of a number of samples of the process tends to zero as the number of samples used increases. Therefore, the values of the filtered result, also shown in Figure 3.22, approach the corresponding values in the original signal. Whereas a small number of samples used in computing the mean would not suppress much of the noise, a large number of samples could overly smooth the signal and remove fine details.

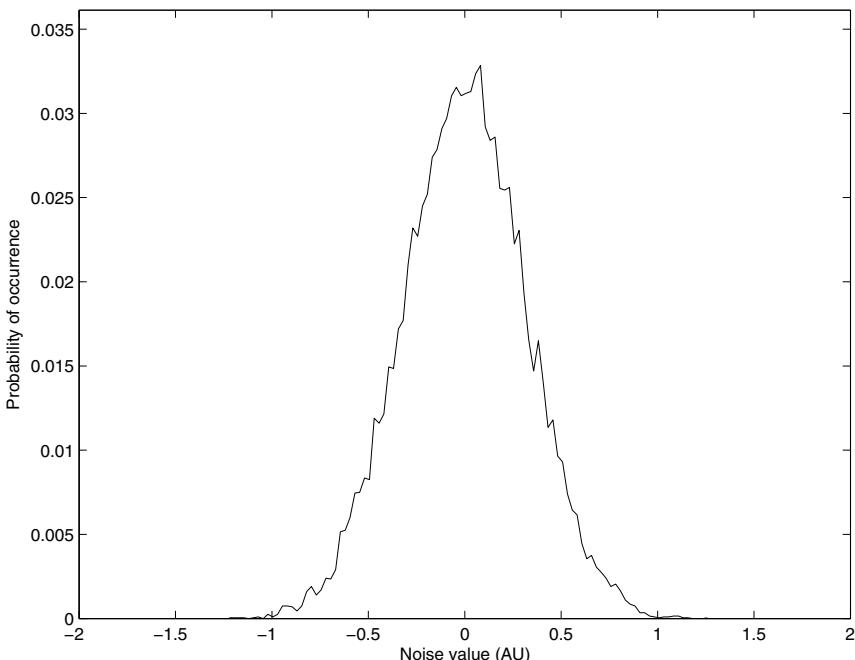


Figure 3.23 Histogram of a realization of the noise process used in the example in Figure 3.22.

Figure 3.24 shows another version of a noisy signal generated using the same processes as in the preceding example. The noisy signal was filtered with a filter having its impulse response as a linearly decreasing function or a ramp of duration 0.25 s, defined in the continuous-time notation as

$$h(t) = 10(0.25 - t), \quad 0 \leq t \leq 0.25 \text{ s}, \quad (3.42)$$

with the sampling frequency of 2 kHz. The output was divided by the sum of all of the values of $h(n)$. The result is a weighted average of the corresponding values of the input signal. Figure 3.25 shows the filter function in Equation 3.42 superimposed on the noisy signal being filtered. The impulse response of the filter has been reversed in time, as required in Equation 3.36 for convolution, and placed at $t = 2.3$ s. The output at $t = 2.3$ s is given by the area under the product of the signal and the impulse response; equivalently, it is given by the sum of the products of all of the overlapping samples of the signal and the impulse response. It is seen that the filtered output is smooth and free of noise; however, some of the minor details in the original signal have been suppressed by the lengthy filter. Furthermore, the delay introduced by the filter is clearly seen by comparing the peaks in the original signal with the corresponding peaks in the filtered output.

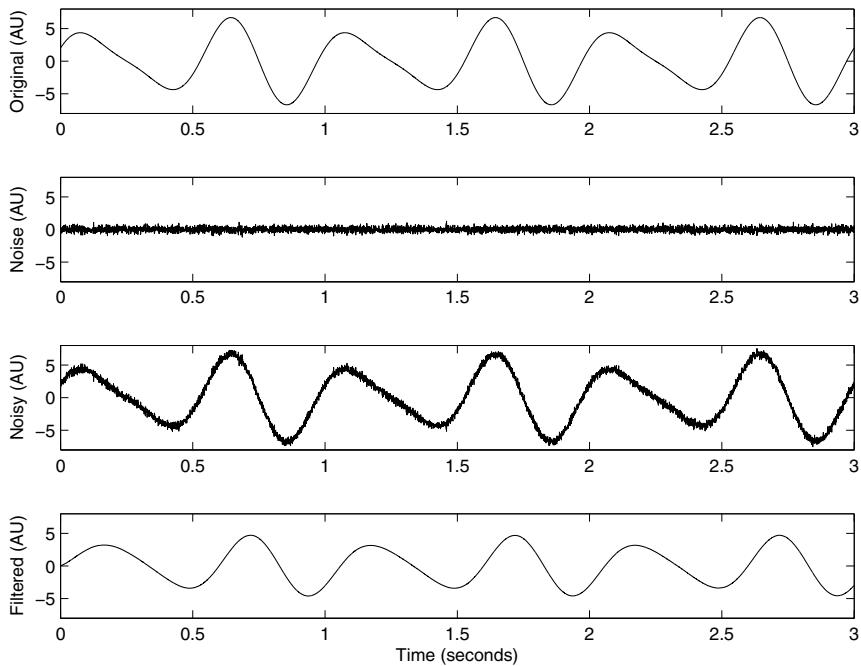


Figure 3.24 Top to bottom: original signal $x(t)$ as in Equation 3.40; Gaussian-distributed random noise; noisy signal; result of filtering using the ramp function in Equation 3.42. See Figure 3.23 for the histogram of the noise process. See also Figure 3.25.

LSI systems in series or parallel: When a number of LSI systems are used in series or in parallel, their effects may be combined into a single equivalent system or operation. Figure 3.26 shows two LSI systems in series (cascade). The first system, with the impulse response $h_1(n)$, operates upon the input, $x(n)$, to give the output as

$$s(n) = x(n) * h_1(n). \quad (3.43)$$

The second system, with the impulse response $h_2(n)$, operates upon $s(n)$ to produce the output

$$\begin{aligned} y(n) &= s(n) * h_2(n) \\ &= x(n) * h_1(n) * h_2(n) \\ &= x(n) * h(n), \end{aligned} \quad (3.44)$$

where

$$h(n) = h_1(n) * h_2(n) \quad (3.45)$$

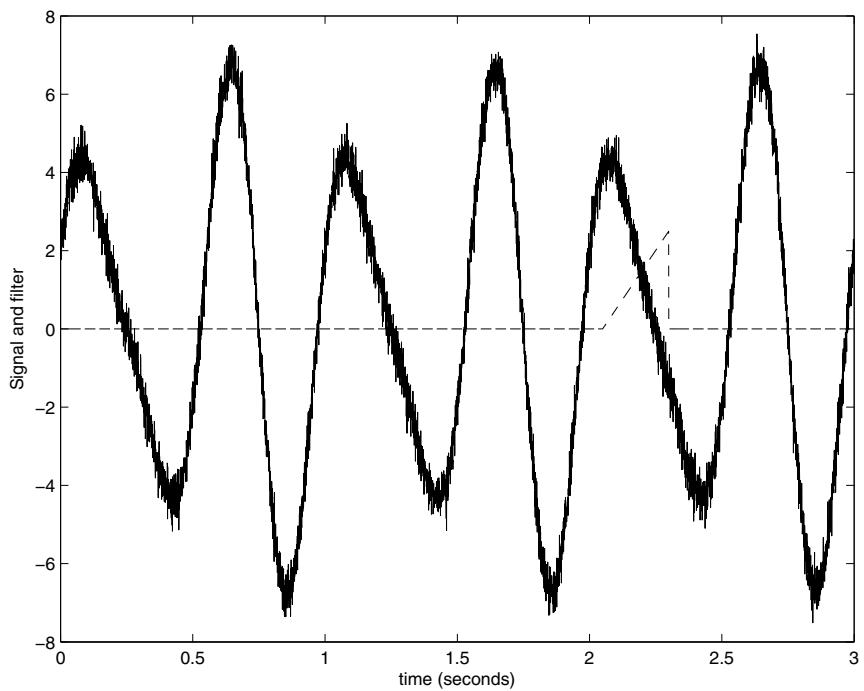


Figure 3.25 The linear ramp filter in Equation 3.42 is shown superimposed (in dashed line) on the noisy signal in the example shown in Figure 3.24. The impulse response of the filter has been reversed in time and placed at $t = 2.3\text{ s}$. The output at $t = 2.3\text{ s}$ is given by the area under the product of the signal and the impulse response; equivalently, it is given by the sum of the products of all samples of the signal and the impulse response that overlap.

is the impulse response of the combination of the two LSI systems in series.

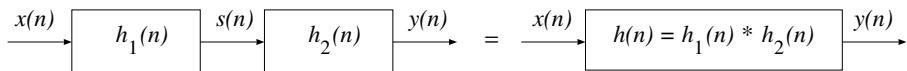


Figure 3.26 Two LSI systems in series and the equivalent system.

Figure 3.27 shows two LSI systems in parallel. For the first system, we have the output as

$$s_1(n) = x(n) * h_1(n). \quad (3.46)$$

Similarly, the second system produces the output

$$s_2(n) = x(n) * h_2(n). \quad (3.47)$$

The combined result is

$$\begin{aligned}
 y(n) &= s_1(n) + s_2(n) \\
 &= x(n) * h_1(n) + x(n) * h_2(n) \\
 &= x(n) * [h_1(n) + h_2(n)] \\
 &= x(n) * h(n),
 \end{aligned} \tag{3.48}$$

where

$$h(n) = h_1(n) + h_2(n) \tag{3.49}$$

is the impulse response of the combination of the two LSI systems in parallel. Relationships such as above are useful in the analysis of sophisticated signal processing systems designed by combining several LSI systems.

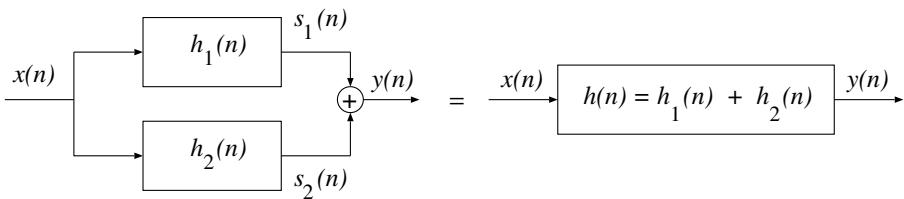


Figure 3.27 Two LSI systems in parallel and the equivalent system.

3.4.2 Transform domain analysis of signals and systems

Quite often, it is convenient to analyze the behavior, characteristics, and performance of an LTI or LSI system in a transform domain. The commonly used transforms for the analysis of continuous-time systems and signals are the Laplace and Fourier transforms [1–3]. The Laplace transform, $H(s)$, of the impulse response, $h(t)$, of an LTI system is defined as

$$H(s) = \int_{-\infty}^{\infty} h(t) \exp(-st) dt, \tag{3.50}$$

where $s = \sigma + j\omega$ is a complex variable representing the transform domain. Here, $\omega = 2\pi f$ represents the frequency variable in radians per second, with f being the frequency variable in Hz; the unit for σ is neper or Napier. If the signal $h(t)$ is causal and of finite duration, existing only over the interval of time $[0, T]$, the limits of the integral could be changed as

$$H(s) = \int_0^T h(t) \exp(-st) dt. \tag{3.51}$$

The function $H(s)$ is known as the system transfer function, or simply the transfer function of the filter.

Figure 3.28 shows a schematic representation of the s -plane. When $H(s)$ is evaluated on the imaginary axis in the s -plane, with $s = j\omega$, we get the frequency response of the system as

$$H(\omega) = H(s)|_{s=j\omega} = \int_0^T h(t) \exp(-j\omega t) dt, \quad (3.52)$$

which is the Fourier transform of the impulse response, $h(t)$. [Note: Some authors express the function above as $H(e^{j\omega})$ or $H(j\omega)$.] In general, $H(\omega)$ is a complex quantity even when $h(t)$ is a real-valued signal. The magnitude response of the system or filter is given by $|H(\omega)|$ and the phase response is given by $\angle H(\omega)$.

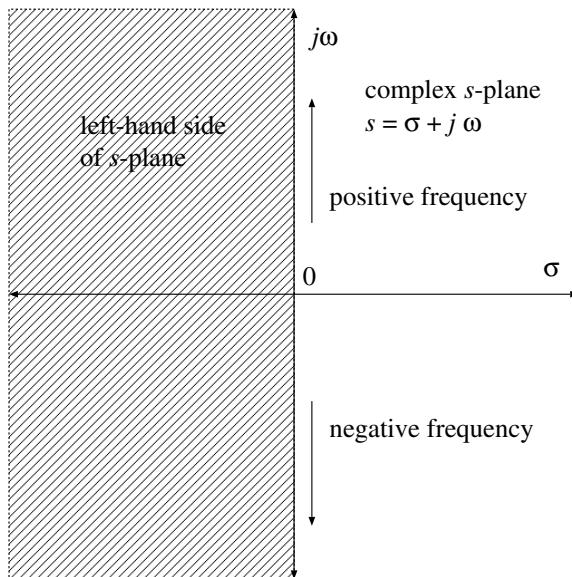


Figure 3.28 Schematic representation of the s -plane or Laplace transform domain.

It is common to express the transfer function, $H(s)$, of an LTI system as a ratio of two polynomials in s . The roots of the polynomial in the numerator give the zeros of the system whereas the roots of the polynomial in the denominator give the poles of the system. For a stable LTI system, all of the poles should be located on the left-hand side (LHS) of the s -plane with $\sigma < 0$. For a real-valued impulse response or signal, poles and zeros should occur in complex-conjugate pairs or with real values. The pole-zero plot of a system is adequate to derive its impulse response, transfer function, and output (for a given input signal) except for a scale factor or gain.

The Laplace transform is a linear and reversible transform. An important property of the Laplace transform when dealing with LTI systems is that the convolution of two signals in the time domain is converted to the product of their individual Laplace transforms in the s -domain, expressed as

$$\begin{aligned} & \text{if } y(t) = x(t) * h(t), \\ & \text{then } Y(s) = X(s) H(s), \\ & Y(\omega) = X(\omega) H(\omega). \end{aligned} \quad (3.53)$$

Figure 3.29 shows the input–output relationship for an LTI system in the s -domain.

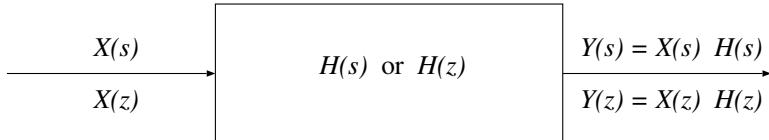


Figure 3.29 Input–output relationship for an LTI system in the s -domain or an LSI system in the z -domain.

It follows that, when two LTI systems are in parallel, as in Figure 3.27, we have the output as $Y(s) = [H_1(s) + H_2(s)]X(s)$, with the equivalent transfer function of $H(s) = H_1(s) + H_2(s)$. When two LTI systems are in series, as in Figure 3.26, we have the output as $Y(s) = H_1(s) H_2(s) X(s)$, with the equivalent transfer function of $H(s) = H_1(s) H_2(s)$. For further details regarding the properties of the Laplace transform, the inverse Laplace transform, and examples of signals and their transforms, refer to Lathi [1, 2] and Oppenheim et al. [3].

The discrete-time counterpart of the Laplace transform is the z -transform, which is useful for the analysis of discrete-time LSI systems and signals. The z -transform of a signal $x(n)$ is defined as

$$X(z) = \sum_{n=-\infty}^{\infty} x(n) z^{-n}, \quad (3.54)$$

where z is a complex variable. Lathi [1, 2] and Oppenheim et al. [3] provide detailed discussions on the properties of the z -transform, its region of convergence (ROC), and the inverse z -transform.

If we consider the z -transform of the impulse response, $h(n)$, of a causal, LSI, finite-impulse response (FIR) system, with $h(n)$ existing only for $n = 0, 1, 2, \dots, N-1$, then we have

$$H(z) = \sum_{n=0}^{N-1} h(n) z^{-n}, \quad (3.55)$$

which represents the transfer function of the system.

An important property of the z -transform when dealing with LSI systems is that the convolution of two signals in the time domain is converted to the product of their individual z -transforms, expressed as

$$\begin{aligned} \text{if } y(n) &= x(n) * h(n), \\ \text{then } Y(z) &= X(z) H(z). \end{aligned} \quad (3.56)$$

Figure 3.29 shows the input–output relationship for an LSI system in the z -domain. The relationship expressed above may be derived in detail as follows:

$$\begin{aligned} Y(z) &= \sum_{n=-\infty}^{\infty} y(n) z^{-n} \\ &= \sum_{n=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} x(k) h(n-k) z^{-n} \\ &= \sum_{k=-\infty}^{\infty} x(k) \sum_{n=-\infty}^{\infty} h(n-k) z^{-n}. \end{aligned} \quad (3.57)$$

Here, the definition of convolution relationship $y(n) = x(n) * h(n)$ has been expanded as in Equation 3.36. The range of summation has been indicated as $[-\infty, \infty]$ for the sake of generality and ease of manipulation of the equations. In the last step above, the order of the two summation operations has been switched and the terms have been rearranged. Now, let $m = n - k$. Then, $n = m + k$, and we have

$$\begin{aligned} Y(z) &= \sum_{k=-\infty}^{\infty} x(k) \sum_{m=-\infty}^{\infty} h(m) z^{-(m+k)} \\ &= \sum_{k=-\infty}^{\infty} x(k) \sum_{m=-\infty}^{\infty} h(m) z^{-m} z^{-k} \\ &= \sum_{k=-\infty}^{\infty} x(k) z^{-k} \sum_{m=-\infty}^{\infty} h(m) z^{-m} \\ &= X(z) H(z). \end{aligned} \quad (3.58)$$

The effects of combining two LSI systems in series or in parallel, as seen in the z -domain, are shown in Figures 3.30 and 3.31. For the combination in series, we have

$$S(z) = X(z) H_1(z) \quad (3.59)$$

and

$$\begin{aligned} Y(z) &= S(z) H_2(z) \\ &= X(z) H_1(z) H_2(z) \\ &= X(z) H(z), \end{aligned} \quad (3.60)$$

where

$$H(z) = H_1(z) H_2(z). \quad (3.61)$$

For the combination in parallel, we have

$$S_1(z) = X(z) H_1(z), \quad (3.62)$$

$$S_2(z) = X(z) H_2(z), \quad (3.63)$$

and

$$\begin{aligned} Y(z) &= S_1(z) + S_2(z) \\ &= X(z) H_1(z) + X(z) H_2(z) \\ &= X(z) [H_1(z) + H_2(z)] \\ &= X(z) H(z), \end{aligned} \quad (3.64)$$

where

$$H(z) = H_1(z) + H_2(z). \quad (3.65)$$

Relationships such as these are useful in the analysis of sophisticated signal processing systems that are designed by combining several LSI systems.

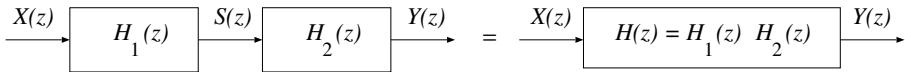


Figure 3.30 Two LSI systems in series and the equivalent system in the z -domain.

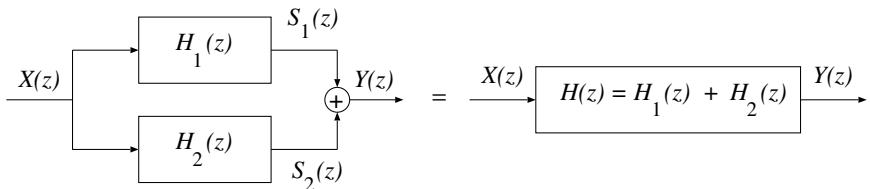


Figure 3.31 Two LSI systems in parallel and the equivalent system in the z -domain.

A few of the important properties of and examples related to the z -transform are as follows [1–3]:

- The z -transform of $\delta(n)$ is 1. The ROC is all z .
- The z -transform of the unit step function, $u(n)$, is $\frac{1}{1-z^{-1}}$. The ROC is $|z| > 1$.

- The z -transform of $a^n u(n)$ is $\frac{1}{1-a z^{-1}}$. The ROC is $|z| > |a|$.
- The z -transform of a causal signal $x(n)$ shifted by k samples, that is, $x(n - k)$, is $z^{-k} X(z)$, where $X(z)$ is the z -transform of $x(n)$.
- If $y(n) = x(n) * h(n)$, then $Y(z) = X(z) H(z)$.

Figure 3.32 illustrates the relationship between the Laplace domain and the z -domain with $z = \exp(sT)$, where $T = 1/f_s$ is the sampling interval and f_s is the sampling frequency. (The maximum frequency permitted in the signal without aliasing errors is $f_m = f_s/2$; this is also referred to as the folding frequency or the Nyquist frequency.) The entire LHS of the s -plane is mapped to the region within the unit circle in the z -plane. Thus, all of the poles of a stable system should be located within the unit circle in the z -plane. See Oppenheim and Schafer [174] for a detailed discussion on this relationship. The bilinear transformation is another method of mapping from the s -plane to the z -plane [1–3].

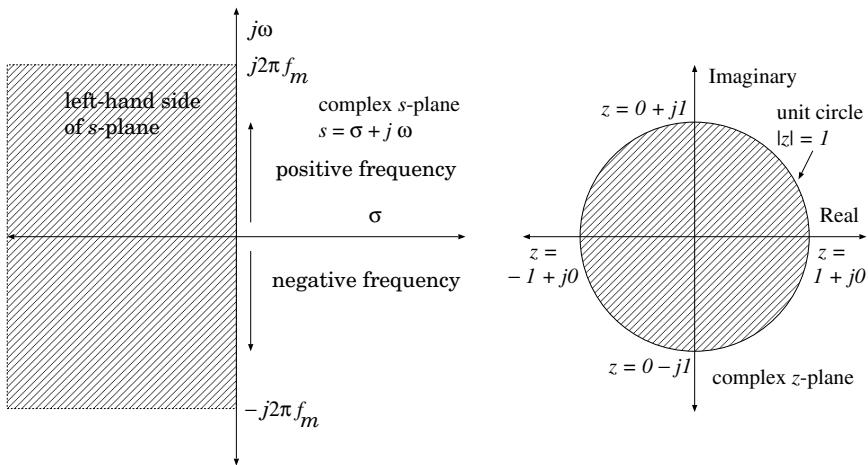
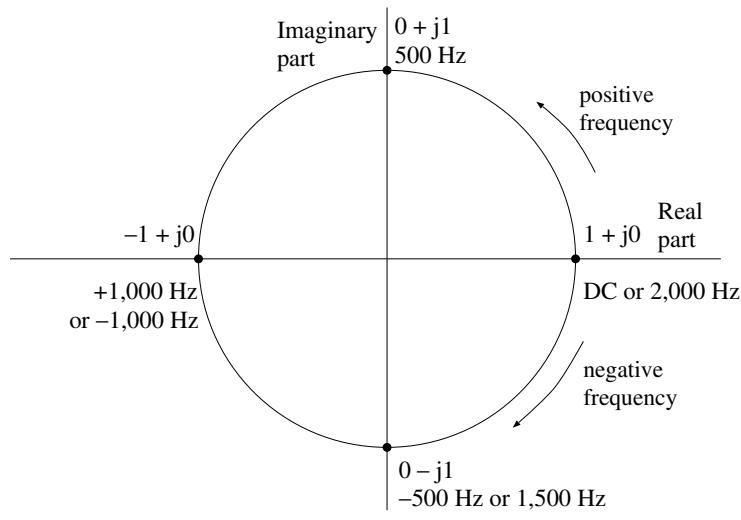
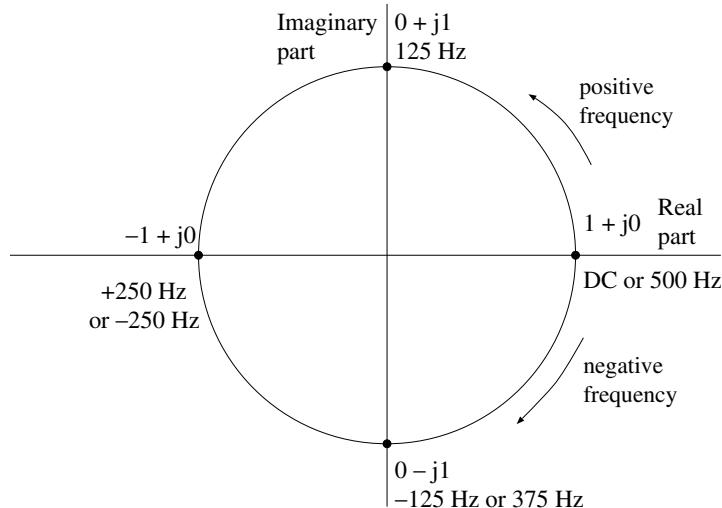


Figure 3.32 Transformation from the Laplace domain to the z -domain with $z = \exp(j\omega T)$. $f_m = f_s/2$.

Figure 3.33 provides interpretation of the frequency variable around the unit circle in the z -domain for two different sampling frequencies, $f_s = 2,000$ Hz and $f_s = 500$ Hz. The unit circle is given by $z = \exp(j\omega T)$, and represents the frequency axis in a circular and periodic manner instead of the linear vertical axis in the s -plane. (The same symbol ω is being used in the present discussion for the frequency variable in both cases of continuous-time and discrete-time analysis. Different symbols will be used for the two cases when the two variables need to be distinguished.) The notions of a limited bandwidth of $\pm f_s/2$ or $[0, f_s]$ and the introduction of aliasing errors if the Nyquist sampling rate [1–3] is not satisfied are demonstrated by the circular and periodic nature of the frequency axis.

(a) unit circle in the z -plane with $f_s = 2,000$ Hz(b) unit circle in the z -plane with $f_s = 500$ Hz**Figure 3.33** Interpretation of the frequency variable with uniform sampling along the unit circle in the z -domain for two different sampling frequencies.

Just as the Fourier transform may be seen as the Laplace transform evaluated on the imaginary axis with $s = j\omega$, evaluation of the z -transform on the unit circle in the z -plane gives us the Fourier transform, as

$$\begin{aligned} X(\omega) &= \sum_{n=0}^{N-1} x(n) z^{-n}|_{z=\exp(j\omega T)} \\ &= \sum_{n=0}^{N-1} x(n) \exp(-j\omega nT). \end{aligned} \quad (3.66)$$

Using $\omega = 2\pi f$ and $T = 1/f_s$, we have the argument of the \exp function above as $-j\omega nT = -j2\pi n f/f_s$. We may now consider the ratio f/f_s to represent a normalized frequency variable in the range $[0, 1]$, with zero corresponding to DC and unity corresponding to f_s . The result of multiplication of this ratio with 2π may be seen as division of the range $[0, 2\pi]$ into the frequency axis spanning the range $[0, f_s]$. Thus, the variable T may be dropped.

3.4.3 The pole–zero plot

Consider a signal processing system or filter with the transfer function specified as a ratio of two polynomials in z or a rational function of z :

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{k=0}^N b_k z^{-k}}{1 + \sum_{k=1}^M a_k z^{-k}}. \quad (3.67)$$

The equivalent time-domain difference equation of the filter is

$$y(n) = \sum_{k=0}^N b_k x(n-k) - \sum_{k=1}^M a_k y(n-k). \quad (3.68)$$

The equations given above do not include a gain factor that could be specified separately.

The polynomials in the numerator and denominator of Equation 3.67 may be solved for their roots. Let the roots of the numerator be z_k , $k = 1, 2, \dots, N$, which are the zeros of the transfer function $H(z)$. Let the roots of the denominator be p_k , $k = 1, 2, \dots, M$, which are the poles of the transfer function $H(z)$. A plot of the roots of $H(z)$ in the z -plane is the pole–zero plot of the system; see Figure 3.34. The transfer function may be expressed in terms of the poles and zeros as

$$H(z) = \frac{\prod_{k=1}^N (1 - z_k z^{-1})}{\prod_{k=1}^M (1 - p_k z^{-1})}, \quad (3.69)$$

which may be modified to

$$H(z) = z^{(M-N)} \frac{\prod_{k=1}^N (z - z_k)}{\prod_{k=1}^M (z - p_k)}. \quad (3.70)$$

The term $(z - z_k)$ represents the vector from an arbitrary point z to the zero z_k . Similarly, the term $(z - p_k)$ represents the vector from an arbitrary point z to the pole p_k .

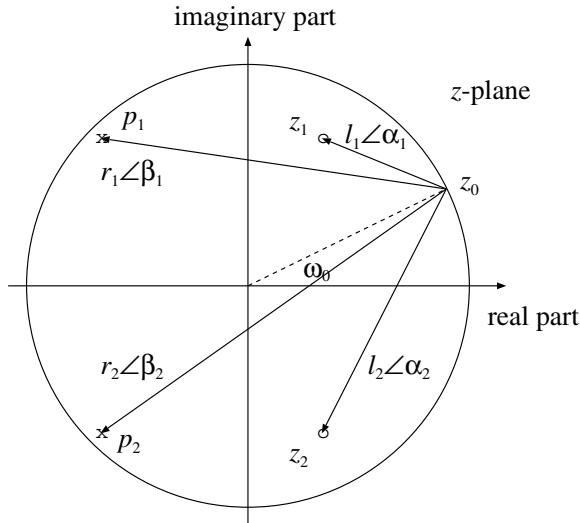


Figure 3.34 Derivation of the frequency response of a system from its pole–zero plot.

Now, consider the evaluation of $H(z)$ for a point z_0 on the unit circle in the z -plane, as shown in Figure 3.34. This gives the frequency response of the system for the corresponding radian frequency ω_0 , as

$$H(\omega_0)|_{z=z_0} = z_0^{(M-N)} \frac{\prod_{k=1}^N (z_0 - z_k)}{\prod_{k=1}^M (z_0 - p_k)}. \quad (3.71)$$

Following the illustration in Figure 3.34, let us represent the vector from z_0 to the zero z_1 as $(z_0 - z_1) = l_1\angle\alpha_1$, the vector from z_0 to the pole p_1 as $(z_0 - p_1) = r_1\angle\beta_1$, and apply the same procedure to the remaining poles and zeros. Then, we have

$$|H(\omega_0)| = \frac{\prod_{k=1}^N l_k}{\prod_{k=1}^M r_k} \quad (3.72)$$

and

$$\angle H(\omega_0) = \angle z_0^{(M-N)} + \sum_{k=1}^N \alpha_k - \sum_{k=1}^M \beta_k. \quad (3.73)$$

Thus, the magnitude of the response of a system at a particular frequency is given by the ratio of the product of the distances from the corresponding frequency point in the z -plane to all of the zeros of the system to the product of the distances from the same

frequency point to all of the poles of the system (except for a gain factor). Similarly, the phase response is given by the difference between the sum of the angles of the vectors from the frequency point to all of the zeros of the system and the sum of the angles of the vectors from the same frequency point to all of the poles of the system (except for any additional linear phase factor).

It is evident from the equations and the related discussions given above that, as the point z_0 approaches a zero of the system, one of the distances l_k will become small, leading to a low value of the response. If a zero is present on the unit circle, the corresponding l_k will be zero, resulting in a zero-valued response at the related frequency. Thus, a zero on the unit circle is associated with a spectral null of the system. As the frequency variable crosses from one side of a zero on the unit circle to the other side, there will be an associated 180° change in the phase response. On the contrary, as the point z_0 approaches a pole of the system that is close to the unit circle, the corresponding distance r_k will become small, leading to a high value of the response at the related frequency. Thus, a pole close to the unit circle is associated with a spectral peak or resonance of the system. In this manner, the pole-zero plot of a system may be inspected to elicit information related to its frequency response in an immediate, albeit qualitative, manner.

3.4.4 The discrete Fourier transform

If we consider evaluation of the Fourier transform for only certain sampled values of the frequency, we could span the range $[0, 2\pi]$ with K samples, with even steps of $2\pi/K$. This is equivalent to spanning the range $[0, 1]$ of the normalized frequency with K samples. The discrete Fourier transform (DFT) could then be expressed as

$$X(k) = \sum_{n=0}^{N-1} x(n) \exp\left(-j\frac{2\pi}{K}nk\right), \quad (3.74)$$

with $k = 0, 1, 2, \dots, K - 1$. It can be shown that, when the given signal has only N nonzero samples, we need only N samples of the DFT evenly spaced over the unit circle in the z -plane [15] for exact recovery of the original signal from the DFT. Then, we have

$$X(k) = \sum_{n=0}^{N-1} x(n) \exp\left(-j\frac{2\pi}{N}nk\right), \quad (3.75)$$

for $k = 0, 1, 2, \dots, N - 1$, and

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) \exp\left(+j\frac{2\pi}{N}nk\right), \quad (3.76)$$

for $n = 0, 1, 2, \dots, N - 1$, as the forward and inverse DFT relationships, respectively.

If we define a complex variable

$$W_N = \exp\left(-j\frac{2\pi}{N}\right), \quad (3.77)$$

which can also be seen as a vector or a phasor, we can write the DFT relationship as

$$X(k) = \sum_{n=0}^{N-1} x(n) W_N^{nk}, \quad (3.78)$$

for $k = 0, 1, 2, \dots, N - 1$. Considering a simple case with $N = 8$, Figure 3.35 shows the vectors (or phasors) representing the $N = 8$ roots of unity, with W_8^k , $k = 0, 1, 2, \dots, 7$, where $W_8 = \exp(-j\frac{2\pi}{8})$. Given the relationship

$$W_8^{nk} = \exp\left(-j\frac{2\pi}{N}nk\right) = \cos\left(\frac{2\pi}{N}nk\right) - j \sin\left(\frac{2\pi}{N}nk\right), \quad (3.79)$$

the sinusoidal nature of the basis functions used in the DFT becomes clear. Figure 3.36 shows stem plots of the sin functions as above, for $k = 0, 1, 2, \dots, 7$. For the sake of illustration, each signal was created with 64 samples, with $n = 0, 1, 2, \dots, 63$.

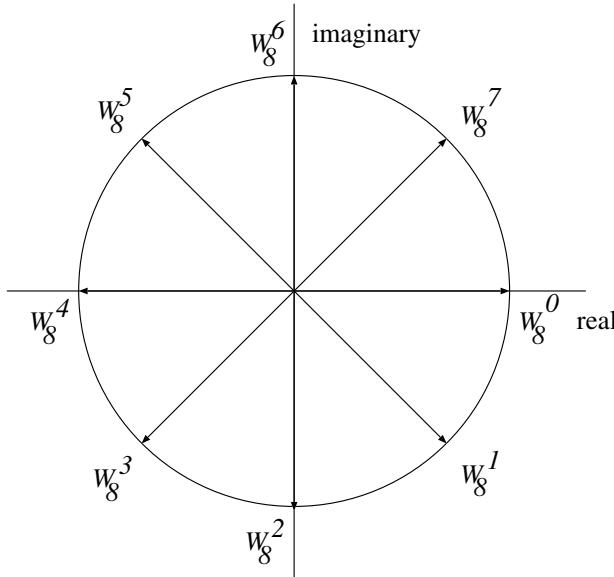


Figure 3.35 Vectors (or phasors) representing the $N = 8$ roots of unity, with W_8^k , $k = 0, 1, 2, \dots, 7$, where $W_8 = \exp(-j\frac{2\pi}{8})$. Based upon a similar figure by Hall [175].

Let us consider the DFT relationship again, with the basis functions written in terms of the sin and cos functions as follows:

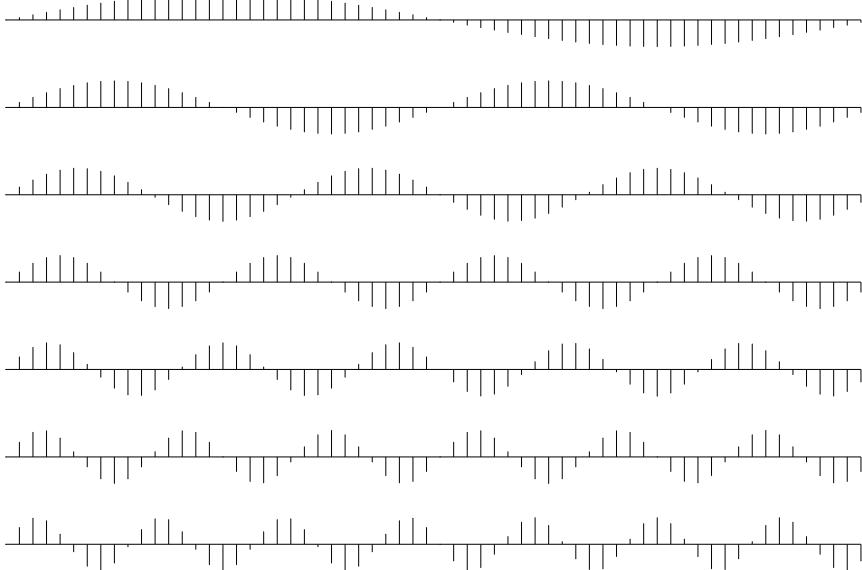


Figure 3.36 Sinusoidal basis functions used in the DFT. The eight plots (top to bottom) show the function $\sin(\frac{2\pi}{N}nk)$, for $k = 0, 1, 2, \dots, 7$, with $N = 64$. Each sinusoidal function (except for $k = 0$) varies over the range $[-1, 1]$. The axis labels have been suppressed for improved visualization.

$$\begin{aligned}
 X(k) &= \sum_{n=0}^{N-1} x(n) \exp\left(-j\frac{2\pi}{N}nk\right) \\
 &= \sum_{n=0}^{N-1} x(n) \left\{ \cos\left(\frac{2\pi}{N}nk\right) - j \sin\left(\frac{2\pi}{N}nk\right) \right\} \\
 &= \sum_{n=0}^{N-1} x(n) \cos\left(\frac{2\pi}{N}nk\right) - j \sum_{n=0}^{N-1} x(n) \sin\left(\frac{2\pi}{N}nk\right). \quad (3.80)
 \end{aligned}$$

It is evident that the real part of $X(k)$ is given by the dot product of the given signal, $x(n)$, with the k^{th} cos basis function, $\cos(\frac{2\pi}{N}nk)$. Similarly, the imaginary part of $X(k)$ is given by the dot product with the corresponding sin function. The dot product represents the projection of one vector or series of values on to another and indicates the extent of commonality between them. Thus, it is clear that the DFT coefficients indicate the amount or strength of each sinusoid present in the given signal. This is the essence of the analysis of a signal in the frequency domain.

Along similar lines, we can view the inverse DFT relationship as a regeneration or synthesis of the original signal as a weighted combination of sinusoids of various frequencies, the weights being the corresponding DFT coefficients, as follows:

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) \cos\left(\frac{2\pi}{N}nk\right) + j \frac{1}{N} \sum_{k=0}^{N-1} X(k) \sin\left(\frac{2\pi}{N}nk\right), \quad (3.81)$$

for $n = 0, 1, 2, \dots, N - 1$.

The periodic nature of the frequency variable and the range of the basis functions used in the DFT relationships leads to an important property: The results of the DFT and inverse DFT operations are periodic signals. Even when we start with a discrete-time signal of finite duration that is not periodic, its spectrum obtained via the DFT is a periodic signal; furthermore, the result of application of the inverse DFT to the spectrum so obtained (and further manipulated as desired) is a periodic discrete-time signal. Unless proper attention is paid to this property, errors comparable to aliasing due to inadequate sampling of a continuous-time signal arise in the final result. Of particular interest are implications of the following property:

$$\begin{aligned} \text{if } y(n) &= x(n) * h(n), \\ \text{then } Y(k) &= X(k) H(k). \end{aligned} \quad (3.82)$$

The symmetry and periodicity of the basis functions used in the DFT can be exploited to reduce the computational requirements for the DFT, which has led to the creation of several versions of the fast Fourier transform (FFT) [15]. The two properties are expressed as follows:

$$W_N^{-nk} = (W_N^{nk})^* \quad (3.83)$$

and

$$W_N^{nk} = W_N^{n(k+N)} = W_N^{(n+N)k}. \quad (3.84)$$

The advantages provided by FFT algorithms are so substantial that it is usually computationally less expensive to use the second line of Equation 3.82 to realize convolution than the original definition of convolution itself. The algorithm to achieve convolution via the DFT is as follows:

1. Compute the DFTs, $X(k)$ and $H(k)$, of the two signals given, $x(n)$ and $h(n)$.
2. Multiply the two DFTs, sample by sample, to get the DFT of the output, $Y(k) = X(k) H(k)$.
3. Compute the inverse DFT of $Y(k)$. The output signal, $y(n)$, is given by the real part of the result.

However, it should be noted that the convolution achieved in this manner is circular or periodic convolution; the results, in general, will not be the same as those

provided by linear convolution. This arises due to the fact that the results of DFT and inverse DFT operations are periodic entities. When we are interested in computing the output of an LSI system, we need to perform linear convolution. To understand the differences between the two types of convolution, let us examine the following examples in detail.

Figure 3.37 shows a numerical illustration of periodic or circular shift applied to a signal. Comparing this figure with Figure 3.18 facilitates understanding the differences between circular and linear shifting.

Figure 3.37 A numerical illustration of circularly shifted versions of a periodic signal. The window shown in dashed lines on the lower set of signals over four cycles represents one period. Shifted versions of the signal are shown for one period in the upper part of the figure. The circular nature of the shift is evident in the upper illustration, even if the shift in the lower illustration is considered to be linear. This type of shifting is known as circular or periodic shifting. A sample that gets shifted out of the observation window on the right reenters the window from the left. See Figure 3.18 for an example of linear shifting.

In view of circular shifting of periodic signals, we can express the convolution of two periodic signals, having the same period of N samples, as

$$y_p(n) = \sum_{k=0}^{N-1} x_p(k) h_p[(n - k) \bmod N], \quad (3.85)$$

where the subscript p indicates that the signals are periodic, and $[(n - k) \bmod N]$ returns values within the range $[0, N - 1]$, by adding or subtracting N , as required. The result is also periodic, having the same period of N samples. Figure 3.38 shows a numerical example of circular or periodic convolution of two periodic discrete-time signals.

$n:$	0	1	2	3
$x(n):$	4	1	3	1
$h(n):$	4	3	2	1
$k:$	0	1	2	3
$x(k):$	4	1	3	1
$h(0 - k):$	4	1	2	3
$h(1 - k):$	3	4	1	2
$h(2 - k):$	2	3	4	1
$h(3 - k):$	1	2	3	4
$y(n):$	26	21	24	19
$n:$	0	1	2	3

Figure 3.38 A numerical illustration of the circular or periodic convolution of two periodic discrete-time signals.

Given two signals, $x(n)$, $n = 0, 1, 2, \dots, N_1$, and $h(n)$, $n = 0, 1, 2, \dots, N_2$, we know that the result of linear convolution of the two signals should have $N_1 + N_2 - 1$ samples. The application of periodic convolution in place of linear convolution leads to erroneous results; compare the numerical example in Figure 3.39 with that in Figure 3.20.

$n:$	0	1	2	3
$x(n):$	4	1	3	1
$h(n):$	3	2	1	0
$k:$	0	1	2	3
$x(k):$	4	1	3	1
$h(0 - k):$	3	0	1	2
$h(1 - k):$	2	3	0	1
$h(2 - k):$	1	2	3	0
$h(3 - k):$	0	1	2	3
$y(n):$	17	12	15	10
$n:$	0	1	2	3

Figure 3.39 A numerical illustration of the undesired effect of periodic convolution of two discrete-time signals when linear convolution is desired. Compare this with the illustration of linear convolution of the same signals in Figure 3.20.

In order to gain the computational benefits provided by the FFT, the DFT approach to convolution may be modified by padding two given nonperiodic signals of unequal duration with zeros so as to have the same extended duration as that of the desired result, and then considering them to represent one period each of their periodic extensions. The algorithm to achieve linear convolution via the DFT is as follows:

1. Extend the two signals given, $x(n)$ and $h(n)$, to have N samples each, with $N \geq N_1 + N_2 - 1$, by appending or padding with zeros.
2. Compute the DFTs, $X(k)$ and $H(k)$, of the two extended signals, for $k = 0, 1, 2, \dots, N - 1$.
3. Multiply the two DFTs, sample by sample, to get the DFT of the output, $Y(k) = X(k) H(k)$, $k = 0, 1, 2, \dots, N - 1$.
4. Compute the inverse DFT of $Y(k)$. The output signal, $y(n)$, for $n = 0, 1, 2, \dots, N_1 + N_2 - 1$, is given by the real part of the result.

Note that the sample-by-sample product given as $Y(k) = X(k) H(k)$ cannot be computed if $X(k)$ and $H(k)$ are of different lengths (or periods). The final result $y(n)$ will, in general be complex, and may have extra values for $N_1 + N_2 - 1 < n < N$ that are not of interest. Figure 3.40 shows a numerical example of circular or periodic convolution of two nonperiodic discrete-time signals. By comparing the illustration in Figure 3.40 with the related case of linear convolution of the same signals in Figure 3.20, it is evident that the final results are the same.

3.4.5 Properties of the Fourier transform

The Fourier transform is the most commonly used transform to study the frequency-domain characteristics of signals [1–3, 15, 174]. This is mainly because the Fourier transform uses sinusoidal functions as its basis functions. Projections are computed of the given signal $x(t)$ on to the complex exponential basis function of frequency ω radians/s, given by $\exp(j\omega t) = \cos(\omega t) + j \sin(\omega t)$, as

$$X(\omega) = \int_{-\infty}^{\infty} x(t) \exp(-j\omega t) dt, \quad (3.86)$$

or in the frequency variable f in Hz as

$$X(f) = \int_{-\infty}^{\infty} x(t) \exp(-j2\pi ft) dt. \quad (3.87)$$

[The complex exponential function is conjugated in computing the projection. In some fields, the forward Fourier transform is defined with $\exp(+j\omega t)$ in the integral.] Equations 3.86 and 3.87 represent *analysis* of the signal $x(t)$ with reference to the complex exponential basis functions. The lower limit of the integral will be 0 if the

$n:$	0	1	2	3	4	5
$x(n):$	4	1	3	1	0	0
$h(n):$	3	2	1	0	0	0
$k:$	0	1	2	3	4	5
$x(k):$	4	1	3	1	0	0
$h(0 - k):$	3	0	0	0	1	2
$h(1 - k):$	2	3	0	0	0	1
$h(2 - k):$	1	2	3	0	0	0
$h(3 - k):$	0	1	2	3	0	0
$h(4 - k):$	0	0	1	2	3	0
$h(5 - k):$	0	0	0	1	2	3
$y(n):$	12	11	15	10	5	1
$n:$	0	1	2	3	4	5

Figure 3.40 A numerical illustration of the use of periodic convolution of two discrete-time and nonperiodic signals to achieve linear convolution by adequate zero padding. Compare this with the illustration of linear convolution of the same signals in Figure 3.20. With the expected number of samples in the result being six, the two input signals, with four and three samples each, have been padded with zeros to the same length of six samples. The effective period used in periodic convolution is six samples.

signal is causal; the upper limit will be equal to the duration of the signal in the case of a finite-duration signal. The value of $X(\omega)$ or $X(f)$ at each frequency of interest $\omega = 2\pi f$ represents the amount or strength of the corresponding cosine and sine functions present in the signal $x(t)$. Note that, in general, $X(\omega)$ is complex for real signals, and includes the magnitude and phase of the corresponding complex exponential.

The inverse transformation, representing *synthesis* of the signal $x(t)$ as a weighted combination of the complex exponential basis functions, is given as

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) \exp(j\omega t) d\omega = \int_{-\infty}^{\infty} X(f) \exp(j2\pi ft) df. \quad (3.88)$$

The second version of the equation given above with the frequency variable f in Hz may be more convenient in some situations than the first one with ω in radians/s, due to the absence of the $\frac{1}{2\pi}$ factor. [If the forward Fourier transform is defined with $\exp(+j\omega t)$, the inverse Fourier transform will have $\exp(-j\omega t)$ in the integral.]

In the case of a discrete-time signal $x(n)$, we may still compute the Fourier transform with a continuous frequency variable ω as

$$X(\omega) = \sum_{n=-\infty}^{\infty} x(n) \exp(-j\omega n), \quad (3.89)$$

with the normalized-frequency range $0 \leq \omega \leq 2\pi$ (equivalent to $0 \leq f \leq 1$). The lower limit of the summation will be 0 if the signal is causal. The upper limit of the summation will be equal to the index ($N - 1$) of the last sample in the case of a finite-duration signal with N samples. Various relationships between the Laplace transform, the Fourier transform, and the z -transform, as well as sampling of the z -transform to obtain the DFT, are described in Section 3.4.2.

Some of the important properties of the DFT and their implications are summarized below [1–3, 15, 174].

- A signal $x(n)$ and its DFT $X(k)$ are both periodic sequences.
- If a signal $x(n)$ has N samples, its DFT $X(k)$ must be computed with at least N samples equally spaced over the normalized frequency range $0 \leq \omega \leq 2\pi$ (or, equivalently, around the unit circle in the z -plane) for complete representation and determination of $X(\omega)$, and hence exact reconstruction of $x(n)$ via the inverse DFT of $X(k)$. One may use more than N samples to compute $X(k)$ in order to employ an FFT algorithm with $L = 2^M \geq N$ samples, where M is an integer, or to obtain $X(\omega)$ with finer frequency sampling than $2\pi/N$.
- The DFT is linear: the DFT of $ax(n) + by(n)$ is $aX(k) + bY(k)$, where $X(k)$ and $Y(k)$ are the DFTs of $x(n)$ and $y(n)$, respectively.
- The DFT of $x(n - n_o)$ is $\exp(-j\frac{2\pi}{N}kn_o)X(k)$, where $X(k)$ is the DFT of $x(n)$. A time shift leads to a linear component being added to the phase of the original signal. As all sequences in DFT relationships are periodic, the shift operation should be defined as a circular or periodic shift. If at least n_o zeros are present or are padded at the end of the signal before the shift operation, a circular shift will be equivalent to a linear shift.
- The DFT of $x(n) * h(n)$ is $X(k)H(k)$, where $X(k)$ and $H(k)$ are the DFTs of $x(n)$ and $h(n)$, respectively. The inverse DFT of $X(k)H(k)$ is $x(n) * h(n)$. Similarly, $x(n)h(n)$ and $X(k) * H(k)$ form a DFT pair. Convolution in one domain is equivalent to multiplication in the other. It is necessary for all of the signals in the relationships mentioned above to have the same number of samples N .

As all sequences in DFT relationships are periodic, the convolution operations in the relationships mentioned here are *periodic convolution* and not linear convolution. Note that circular or periodic convolution is defined for periodic signals having the same period, and that the result will also be periodic with the same period as that of the individual input signals.

The result of linear convolution of two signals $x(n)$ and $h(n)$ with different durations N_x and N_h samples, respectively, will have a duration of $N_x + N_h - 1$ samples. If linear convolution is desired via the inverse DFT of $X(k)H(k)$, the DFTs must be computed with $L \geq N_x + N_h - 1$ samples. The individual signals should be padded with zeros at the end to make their effective durations equal

for the sake of DFT computation and multiplication. All signals and their DFTs are then periodic with the augmented period of L samples.

- The DFT of a real signal $x(n)$ possesses conjugate symmetry, that is, $X(-k) = X^*(k)$. As a consequence, the real part and the magnitude of $X(k)$ are even-symmetric sequences, whereas the imaginary part and the phase of $X(k)$ are odd-symmetric sequences.
- According to Parseval's theorem, the total energy of the signal must remain the same before and after Fourier transformation. We then have the following equalities:

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |X(\omega)|^2 d\omega, \quad (3.90)$$

$$\sum_{n=0}^{N-1} |x(n)|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |X(k)|^2.$$

Because the integral of $|X(\omega)|^2$ over all ω or the sum of $|X(k)|^2$ over all k represents the total energy of the signal (or average power, if the quantity is divided by the duration of the signal), $|X(\omega)|^2$ and $|X(k)|^2$ represent the spread or *density* of the power of the signal along the frequency axis.

Note: A signal or function $x(n)$ possesses even symmetry if $x(-n) = x(n)$ or odd symmetry if $x(-n) = -x(n)$. An arbitrary signal $x(n)$ may be expressed as a combination of a part with even symmetry, $x_e(n)$, and a part with odd symmetry, $x_o(n)$, with the even part given by

$$x_e(n) = \frac{1}{2}[x(n) + x(-n)] \quad (3.91)$$

and the odd part given by

$$x_o(n) = \frac{1}{2}[x(n) - x(-n)]. \quad (3.92)$$

Then, we have

$$x(n) = x_e(n) + x_o(n). \quad (3.93)$$

See Figure 3.8 for an illustration of the PSD of the ECG signal in Figure 3.7 computed via the FFT. Several examples of Fourier spectra and PSDs are given in the sections that follow.

The remaining sections in the present chapter provide details of application of the various concepts described in the present section to filtering of biomedical signals for several purposes.

3.5 Time domain Filters

Certain types of noise may be filtered directly in the time domain using signal processing techniques or digital filters. An advantage of time-domain filtering is that spectral characterization of the signal and noise may not be required (at least in a direct manner). Time-domain processing may also be faster in most cases than frequency-domain filtering.

3.5.1 Synchronized averaging

Problem: Propose a time-domain technique to remove random noise given the possibility of acquiring multiple realizations of the signal or event of interest.

Solution: Linear filters fail to perform well or are not applicable when the signal and noise spectra overlap. Synchronized signal averaging can separate a repetitive signal from noise without distorting the signal [46, 171]. ERP or SEP epochs may be obtained a number of times by repeated application of the stimulus; they may then be averaged by using the stimulus as a trigger for aligning the epochs. ECG signals may be filtered by detecting the QRS complexes and using their positions to align the waveforms for synchronized averaging. If the noise is random with zero mean and is uncorrelated with the signal, averaging will improve the *SNR*.

Let $y_k(n)$ represent one realization of a signal, with $k = 1, 2, \dots, M$ representing the ensemble index, and $n = 0, 1, 2, \dots, N - 1$ representing the time-sample index. Here, M is the number of copies (events, epochs, observations, or realizations) of the signal available, and N is the number of time samples in each copy of the signal (event). We may express the observed signal as

$$y_k(n) = x_k(n) + \eta_k(n), \quad (3.94)$$

where $x_k(n)$ represents the original uncorrupted signal and $\eta_k(n)$ represents the noise in the k^{th} observed signal. Now, if for each instant of time n we add the M copies of the signal, we get

$$\sum_{k=1}^M y_k(n) = \sum_{k=1}^M x_k(n) + \sum_{k=1}^M \eta_k(n); \quad n = 0, 1, 2, \dots, N - 1. \quad (3.95)$$

If the repetitions of the signal are identical and aligned, $\sum_{k=1}^M x_k(n) = Mx(n)$. If the noise is random and has zero mean and variance σ_η^2 , $\sum_{k=1}^M \eta_k(n)$ will tend to zero as M increases, with a variance of $M\sigma_\eta^2$. The *RMS* value of the noise in the summed signal is $\sqrt{M}\sigma_\eta$. Thus, the *SNR* of the signal will increase by a factor of $\frac{M}{\sqrt{M}}$ or \sqrt{M} . The larger the number of epochs or realizations that are averaged, the better will be the *SNR* of the result. Note that synchronized averaging is a type of ensemble averaging.

An algorithmic description of synchronized averaging is as follows:

1. Obtain a number of realizations of the signal or event of interest.

2. Determine a reference point for each realization of the signal. This is directly given by the trigger if the signal is obtained by external stimulation (such as ERPs or SEPs), or may be obtained by detecting the repetitive events in the signal if it is quasiperiodic (such as the QRS complex in the ECG or S1 and S2 in the PCG).
3. Extract parts of the signal corresponding to the events and add them to a buffer. Note that it is possible for the various parts to be of different durations. Alignment of the copies at the trigger point is important; the tail ends of all parts may not be aligned.
4. Divide the result in the buffer by the number of events added.

Illustrations of application: Figure 3.41 illustrates two single-flash ERPs in the upper two traces. The results of averaging over 10 and 20 flashes are shown in the third and fourth plots, respectively, in the same figure. The averaging process has facilitated identification of the first positivity and the preceding and succeeding troughs (marked on the fourth trace) with certainty; the corresponding features are not reliably seen in the single acquisitions (see also the single-flash ERPs in Figure 3.2). Visual ERPs are analyzed in terms of the latencies of the first major peak or positivity, labeled as P120 due to the fact that the normal expected latency for adults is 120 ms; the trough or negativity before P120, labeled as N80; and the trough following P120, labeled as N145. The N80, P120, and N145 latencies measured from the averaged signal in Trace 4 of Figure 3.41 are 85.7, 100.7, and 117 ms, respectively, which are considered to be within the normal range for adults.

The upper trace in Figure 3.42 illustrates a noisy ECG signal over several beats. In order to obtain trigger points, a sample QRS complex of 86 ms duration (86 samples at a sampling rate of 1,000 Hz) was extracted from the first beat in the signal and used as a template. Template matching was performed using a normalized correlation coefficient defined as [171]

$$\gamma_{xy}(k) = \frac{\sum_{n=0}^{N-1} [x(n) - \bar{x}][y(k - N + 1 + n) - \bar{y}_k]}{\sqrt{\sum_{n=0}^{N-1} [x(n) - \bar{x}]^2 \sum_{n=0}^{N-1} [y(k - N + 1 + n) - \bar{y}_k]^2}}, \quad (3.96)$$

where

$$\bar{y}_k = \frac{1}{N} \sum_{n=0}^{N-1} y(k - N + 1 + n),$$

that is, \bar{y}_k is the average of the part of the signal y being used in the template matching procedure. The operation given above is causal, and it is valid from $k = N$ to the last sample of the signal y . x is the template, and \bar{x} is the average value of x . k is the time index of the signal y at which the template is placed. {Jenkins et al. [132] used a measure similar to $\gamma_{xy}(k)$ but without subtraction of the mean and without the shift parameter k to match segmented ECG cycles with a template.} The lower trace in Figure 3.42 shows $\gamma_{xy}(k)$, where it is seen that the cross-correlation result peaks to values near unity at the locations of the QRS complexes in the signal. Averaging

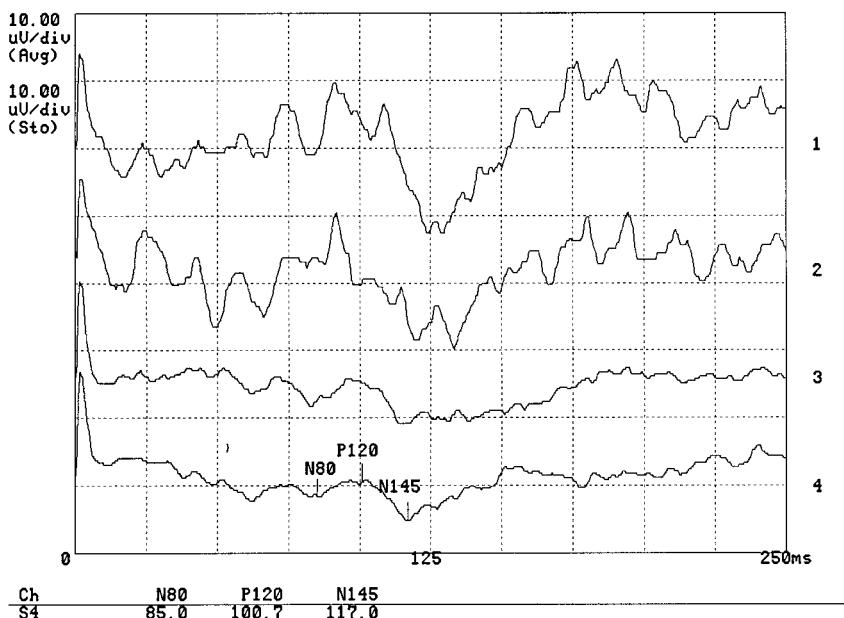


Figure 3.41 Traces 1 and 2: Two sample acquisitions of individual flash visual ERPs from the occipital midline (oz) position of a normal adult male. The earlobes were used to form the ground lead (a1a2), and the left forehead was used as the reference (see Figure 1.37). Trace 3: Average of 10 ERPs. Trace 4: Average of 20 ERPs. The latencies of interest have been labeled on Trace 4 by an EEG technologist as N80, 85.0 ms; P120, 100.7 ms; and N145, 117.0 ms. The height of each grid division (box) is 10 μ V and the width is 25 ms. See also Figure 3.2. Data courtesy of L. Alfaro and H. Darwish, Alberta Children's Hospital, Calgary.

inherent in the cross-correlation formula (over N samples) has reduced the effect of noise on template matching.

By choosing an appropriate threshold, it becomes possible to obtain a trigger point to extract the QRS complex locations in the ECG signal. (*Note:* The QRS template matches with the P and T waves with cross-correlation values of about 0.5; wide QRS complexes may yield high cross-correlation values with tall P and T waves. The threshold has to be chosen so as to detect only the QRS complexes.) A threshold of 0.9 was applied to $\gamma_{xy}(k)$, and the QRS positions of all of the 12 beats in the signal were detected.

Figure 3.43 illustrates two ECG cycles extracted using the trigger points obtained by thresholding the cross-correlation function, as well as the result of averaging the first 11 cycles in the signal. It is seen that the noise has been effectively suppressed by synchronized averaging. The low-level baseline variation and power-line interference present in the signal have caused minor artifacts in the result, which are negligible in this illustration.

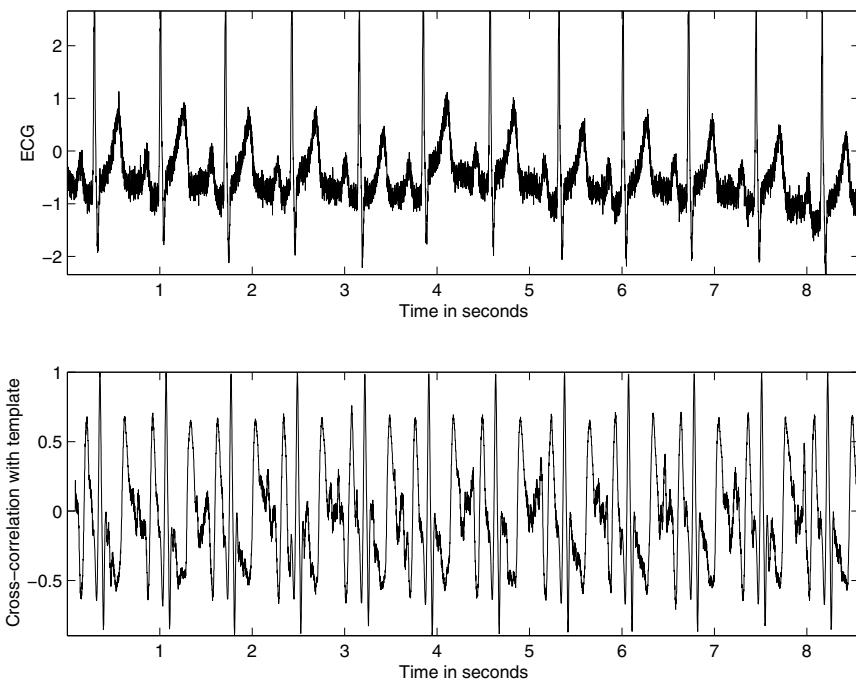


Figure 3.42 An ECG signal with noise (upper trace) and the result of cross-correlation (lower trace) with the QRS template selected from the first cycle. The cross-correlation coefficient is normalized to the range $[-1, 1]$.

The most important requirement in synchronized averaging is indicated by the first word in the name of the process: The realizations of the signal that are added for averaging *must be synchronized or aligned* such that the repetitive part of the signal appears at exactly the same instant in each realization of the signal. If this condition is not met, the waveform of the event in the signal will be blurred or smudged along the time axis.

A major advantage of synchronized averaging is that no frequency-domain filtering is performed — either explicitly or implicitly. No spectral content of the signal is lost as is the case with frequency-domain (lowpass) filters or other time-domain filters such as moving-window averaging filters.

Structured noise such as power-line interference may be suppressed by synchronized averaging if the phase of the interference in each realization is different. To facilitate this feature, the repetition rate of the stimulus should be set so that it is not directly related to the power-line frequency (for example, the flashes used to acquire the averaged ERPs in Figure 3.41 were delivered at 2.1 pps). Physiological interference such as background EEG in ERPs and SEPs may also be suppressed by synchronized averaging, as such activity may bear no interrelationship from one

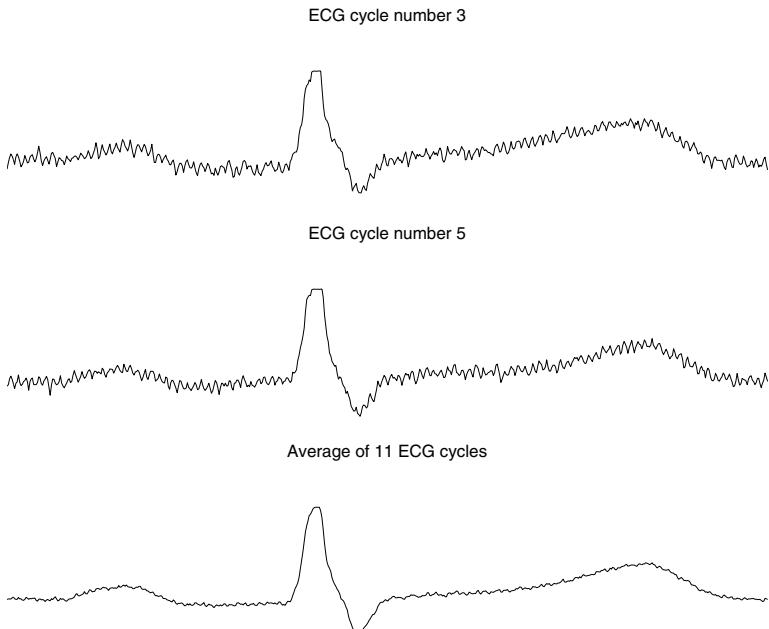


Figure 3.43 Upper two traces: two cycles of the ECG extracted from the signal in Figure 3.42. Bottom trace: the result of synchronized averaging of 11 cycles from the same ECG signal.

epoch of the desired signal to another. See Sections 3.11, 4.11, 5.5.2, and 6.3.5 for further applications and discussions on synchronized averaging.

3.5.2 MA filters

Problem: Propose a time-domain technique to remove random noise given only one realization of the signal or event of interest.

Solution: When an ensemble of several realizations of an event is not available, synchronized averaging will not be possible. We are then forced to consider temporal averaging for noise removal, with the assumption that the processes involved are ergodic, that is, temporal statistics may be used instead of ensemble statistics. As temporal statistics are computed using a few samples of the signal along the time axis and the temporal window of samples is moved to obtain the output at various points of time, such a filtering procedure is called a moving-window averaging filter in general; the term MA filter is commonly used. See Figures 3.22, 3.24, and 3.25 as well as the related discussions.

The general form of the difference equation of an MA filter is

$$y(n) = \sum_{k=0}^N b_k x(n - k), \quad (3.97)$$

where x and y are the input and output of the filter, respectively. The b_k values are the filter coefficients or tap weights, $k = 0, 1, 2, \dots, N$, where N is the order of the filter. The effect of division by the number of samples used ($N + 1$) is included in the values of the filter coefficients. The signal-flow diagram of a generic MA filter is shown in Figure 3.44.

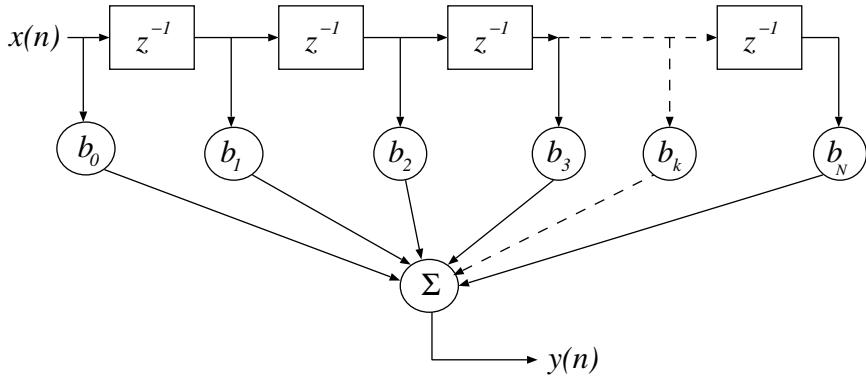


Figure 3.44 Signal-flow diagram of an MA filter of order N . Each block with the symbol z^{-1} represents a delay of one sample, and it serves as a memory unit for the corresponding signal sample value.

Applying the z -transform, we get the transfer function $H(z)$ of the filter as

$$H(z) = \frac{Y(z)}{X(z)} = \sum_{k=0}^N b_k z^{-k} = b_0 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_N z^{-N}, \quad (3.98)$$

where $X(z)$ and $Y(z)$ are the z -transforms of $x(n)$ and $y(n)$, respectively.

A simple MA filter for filtering noise is the von Hann (also known as Hann or Hanning) filter [46], given by

$$y(n) = \frac{1}{4}[x(n) + 2x(n-1) + x(n-2)]. \quad (3.99)$$

The signal-flow diagram of the Hann filter is shown in Figure 3.45. The impulse response of the filter is obtained by letting $x(n) = \delta(n)$, resulting in

$$h(n) = \frac{1}{4}[\delta(n) + 2\delta(n-1) + \delta(n-2)]; \quad (3.100)$$

see Figure 3.46.

Applying the z -transform to Equation 3.99, we get

$$Y(z) = \frac{1}{4}[X(z) + 2z^{-1}X(z) + z^{-2}X(z)], \quad (3.101)$$

which leads to the derivation of the transfer function of the Hann filter as

$$H(z) = \frac{Y(z)}{X(z)} = \frac{1}{4}[1 + 2z^{-1} + z^{-2}]. \quad (3.102)$$

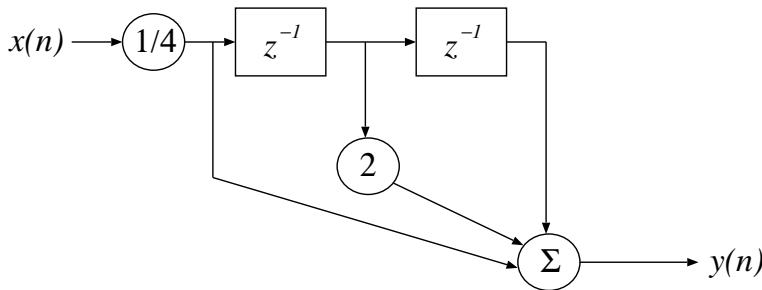


Figure 3.45 Signal-flow diagram of the Hann filter.

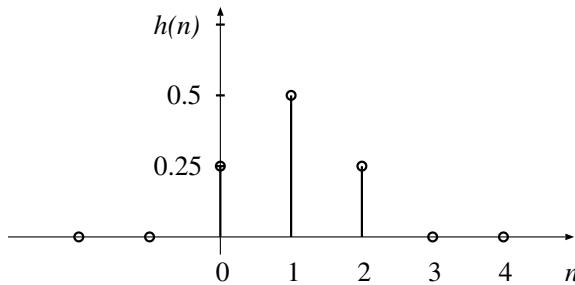


Figure 3.46 Impulse response of the Hann filter.

The transfer function has a double-zero at $z = -1$.

An MA filter is an FIR filter with the following attributes and advantages:

- The impulse response $h(k)$ has a finite number of terms: $h(k) = b_k$, $k = 0, 1, 2, \dots, N$.
- An FIR filter may be realized nonrecursively (with no feedback).
- The output depends only on the present input sample and a few past input samples.
- The filter is defined by a set of tap weights of the delay stages, as illustrated in Figure 3.44.
- The filter transfer function has no poles except at $z = 0$: the filter is inherently stable.
- The filter has linear phase if the series of tap weights is symmetric or antisymmetric.

(*Note:* A recursive filter uses previous values of the output to compute the current output value.)

The frequency response of a filter is obtained by substituting $z = e^{j\omega T}$ in the expression for $H(z)$, where T is the sampling interval in seconds and ω is the radian frequency ($\omega = 2\pi f$, where f is the frequency in Hz). Note that we may set $T = 1$ and deal with normalized frequency in the range $0 \leq \omega \leq 2\pi$ or $0 \leq f \leq 1$; then, $f = 1$ or $\omega = 2\pi$ represents the sampling frequency, with lower frequency values being represented as a normalized fraction of the sampling frequency.

The frequency response of the Hann filter is obtained as

$$H(\omega) = H(z)|_{z=\exp(j\omega)} = \frac{1}{4}[1 + 2e^{-j\omega} + e^{-j2\omega}]. \quad (3.103)$$

This expression can be simplified as follows:

$$\begin{aligned} H(\omega) &= \frac{1}{4}[1 + 2e^{-j\omega} + e^{-j2\omega}] \\ &= \frac{1}{4}e^{-j\omega}[e^{+j\omega} + 2 + e^{-j\omega}] \\ &= \frac{1}{4}e^{-j\omega}[2\cos(\omega) + 2] \\ &= \frac{1}{2}[1 + \cos(\omega)]e^{-j\omega}. \end{aligned} \quad (3.104)$$

Note that $e^{\pm j\omega} = \cos(\omega) \pm j\sin(\omega)$, $e^{+j\omega} + e^{-j\omega} = 2\cos(\omega)$, and $e^{+j\omega} - e^{-j\omega} = 2j\sin(\omega)$. The magnitude and phase of the frequency response are

$$|H(\omega)| = \left| \frac{1}{2}[1 + \cos(\omega)] \right| \quad (3.105)$$

and

$$\angle H(\omega) = -\omega. \quad (3.106)$$

The magnitude and phase responses of the Hann filter are plotted in Figure 3.47. It is clear that the filter is a lowpass filter with linear phase.

Note that, although we started with a description of the Hann filter in the time domain, subsequent analysis of the filter was performed in the frequency domain using the z -transform and the frequency response. System analysis is easier to perform in the z domain in terms of the poles and zeros of the transfer function and in the frequency domain in terms of the magnitude and phase responses. The magnitude and phase responses assist in understanding the effect of the filter on the frequency components of the signal (and noise).

It is seen from the magnitude response of the Hann filter (Figure 3.47) that components beyond about 20% of the sampling frequency of 1,000 Hz are reduced in amplitude by more than 3 dB, that is, to less than one-half of their levels in the input. High-frequency components beyond 40% of the sampling frequency are suppressed to less than 20 dB below their input levels. The filter will perform adequate filtering of ECG signals sampled at 200 Hz, with the gain being lower than -20 dB beyond 80 Hz. However, if the signal is sampled at 1,000 Hz (as in the present example),

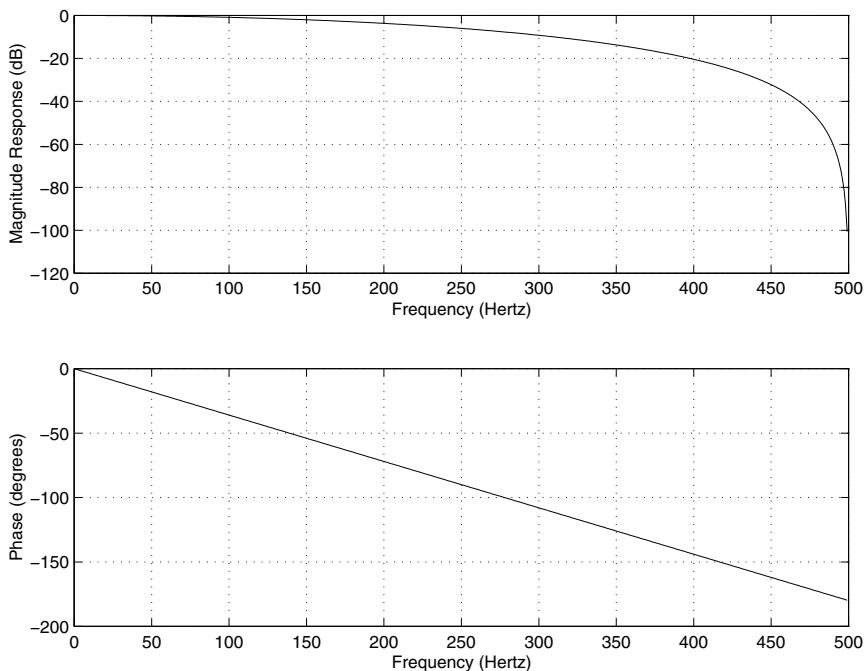


Figure 3.47 Magnitude and phase responses of the Hann (smoothing) filter.

the gain remains above -20 dB for frequencies up to 400 Hz ; such a lowpass filter may not be adequate for filtering ECG signals, but may be appropriate for other signals, such as the PCG and the EMG.

Increased smoothing may be achieved by averaging signal samples over longer time windows, at the expense of increased filter delay. If the signal samples over a window of eight samples are averaged, we get the 8-point MA filter with the output defined as

$$\begin{aligned} y(n) &= \frac{1}{8} \sum_{k=0}^7 x(n-k) \\ &= \frac{1}{8}[x(n) + x(n-1) + x(n-2) + x(n-3) + x(n-4) \\ &\quad + x(n-5) + x(n-6) + x(n-7)]. \end{aligned} \tag{3.107}$$

The impulse response of the filter is

$$\begin{aligned} h(n) &= \frac{1}{8}[\delta(n) + \delta(n-1) + \delta(n-2) + \delta(n-3) + \delta(n-4) \\ &\quad + \delta(n-5) + \delta(n-6) + \delta(n-7)]. \end{aligned} \tag{3.108}$$

The transfer function of the filter is

$$H(z) = \frac{1}{8} \sum_{k=0}^7 z^{-k}, \quad (3.109)$$

and the frequency response is given by

$$\begin{aligned} H(\omega) &= \frac{1}{8} \sum_{k=0}^7 \exp(-j\omega k) \\ &= \frac{1}{8} \{1 + \exp(-j4\omega) \\ &\times [1 + 2 \cos(\omega) + 2 \cos(2\omega) + 2 \cos(3\omega)]\}. \end{aligned} \quad (3.110)$$

The frequency response of the 8-point MA filter is shown in Figure 3.48; the pole-zero plot of the filter is depicted in Figure 3.49. It is seen that the filter has zeros at $\frac{f_s}{8} = 125 \text{ Hz}$, $\frac{f_s}{4} = 250 \text{ Hz}$, $\frac{3f_s}{8} = 375 \text{ Hz}$, and $\frac{f_s}{2} = 500 \text{ Hz}$. Comparing the frequency response of the 8-point MA filter with that of the Hann filter in Figure 3.47, we see that the former provides increased attenuation in the range $90 - 400 \text{ Hz}$ over the latter. Note that the attenuation provided by the filter after about 100 Hz is nonuniform, which may not be desirable in certain applications. Furthermore, the phase response of the filter is not linear, although it is piecewise linear.

Relationship of MA filtering to integration: Disregarding the $\frac{1}{8}$ scale factor for a moment, the operation in Equation 3.107 may be interpreted as the summation or integration of the signal over the duration $n - 7$ to n . A comparable integration of a continuous-time signal $x(t)$ over the interval $t - \tau$ to t is expressed as

$$y(t) = \int_{t-\tau}^t x(t) dt. \quad (3.111)$$

The general definition of the integral of a signal is

$$y(t) = \int_{-\infty}^t x(t) dt, \quad (3.112)$$

or, if the signal is causal,

$$y(t) = \int_0^t x(t) dt. \quad (3.113)$$

The Fourier transforms of the signals in the relationship above are related as [1–3]

$$Y(\omega) = \frac{1}{j\omega} X(\omega) + \pi X(0)\delta(\omega). \quad (3.114)$$

The frequency response of the integration operator is

$$H(\omega) = \frac{1}{j\omega}, \quad (3.115)$$

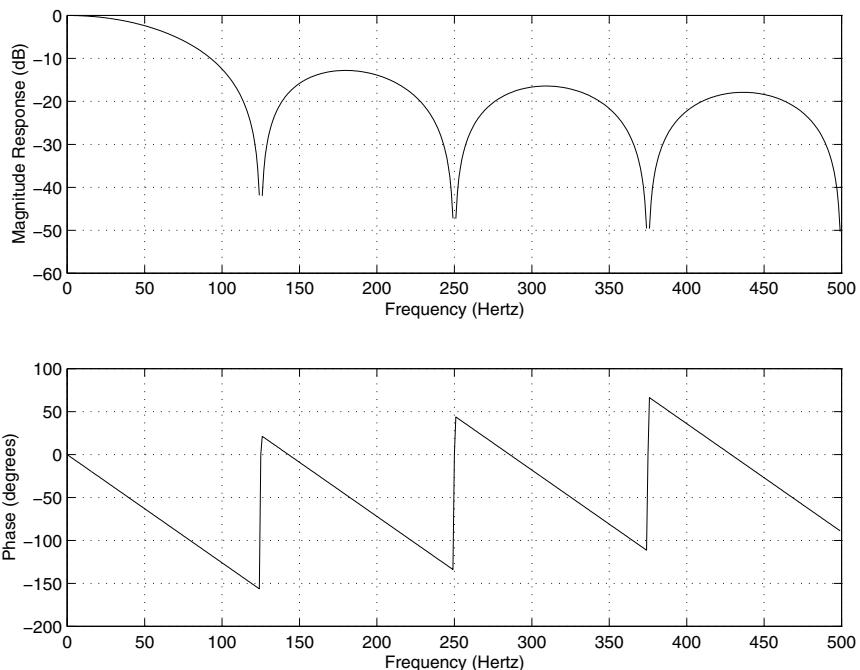


Figure 3.48 Magnitude and phase responses of the 8-point MA (smoothing) filter.

with the magnitude response

$$|H(\omega)| = \left| \frac{1}{\omega} \right| \quad (3.116)$$

and phase response

$$\angle H(\omega) = -\frac{\pi}{2}. \quad (3.117)$$

It is seen from the frequency response that the gain of the filter reduces (nonlinearly) as the frequency is increased; therefore, the corresponding filter has lowpass characteristics.

Integration or accumulation of a discrete-time signal for all samples up to the present sample results in the transfer function $H(z) = \frac{1}{1-z^{-1}}$ [1-3]. Such an operation is seldom used in practice. Instead, a moving-window sum is computed, such as the 8-point MA filter in Equation 3.107. It follows from Equation 3.107 that

$$\begin{aligned} y(n-1) &= \frac{1}{8}[x(n-1) + x(n-2) + x(n-3) + x(n-4) + x(n-5) \\ &\quad + x(n-6) + x(n-7) + x(n-8)]. \end{aligned} \quad (3.118)$$

By combining Equations 3.107 and 3.118, we get the relationship

$$y(n) = y(n-1) + \frac{1}{8}x(n) - \frac{1}{8}x(n-8). \quad (3.119)$$

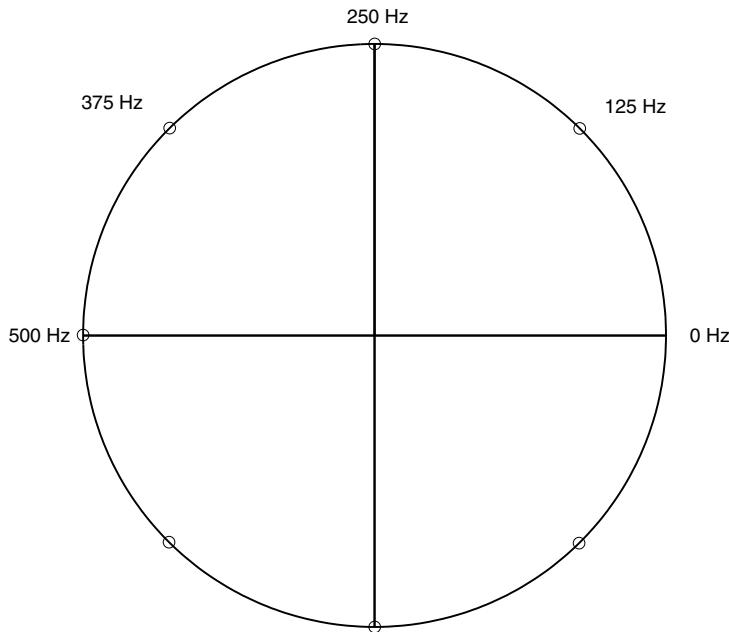


Figure 3.49 Pole–zero plot of the 8-point MA (smoothing) filter.

The recursive form as above clearly depicts the integration aspect of the filter. The transfer function of this expression is easily derived to be

$$H(z) = \frac{1}{8} \left[\frac{1 - z^{-8}}{1 - z^{-1}} \right]. \quad (3.120)$$

The frequency response of the filter is given by

$$H(\omega) = \frac{1}{8} \left[\frac{1 - e^{-j8\omega}}{1 - e^{-j\omega}} \right] = \frac{1}{8} e^{-j\frac{7}{2}\omega} \left[\frac{\sin(4\omega)}{\sin(\frac{\omega}{2})} \right], \quad (3.121)$$

which is equivalent to that in Equation 3.110. Summation over a limited discrete-time window results in a frequency response having sinc-type characteristics, as illustrated in Figure 3.48. See Tompkins [46] for a discussion on other types of integrators.

Illustration of application: Figure 3.50 shows a segment of an ECG signal with high-frequency noise. Figure 3.51 shows the result of filtering the signal with the 8-point MA filter described above. Although the noise level has been reduced, some noise is still present in the result. This is due to the fact that the attenuation of the simple 8-point MA filter is not more than -20 dB at most frequencies (except near the zeros of the filter).

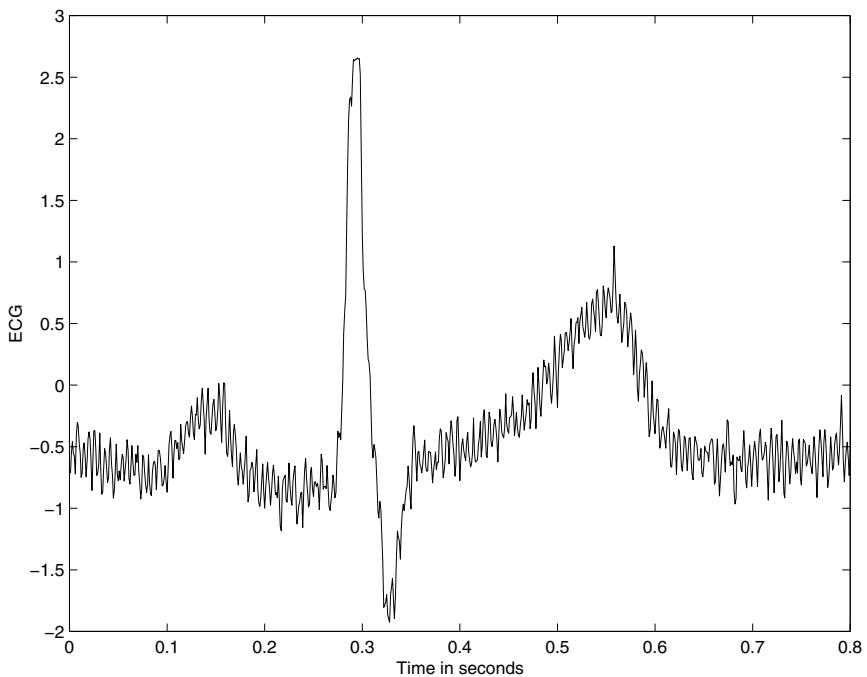


Figure 3.50 ECG signal with high-frequency noise; $f_s = 1,000 \text{ Hz}$.

3.5.3 Derivative based operators to remove low frequency artifacts

Problem: Develop a time-domain technique to remove baseline drift in the ECG signal.

Solution: The derivative operator in the time domain removes the parts of the input that are constant (the output is zero). Large changes in the input lead to high values in the output of the derivative operator. Improved understanding of the derivative operation may be obtained by studying its transform in the frequency domain.

The ideal $\frac{d}{dt}$ operator in the time domain results in multiplication of the Fourier transform of the original signal by $j\omega = j2\pi f$ in the frequency domain. If $X(f)$ represents the Fourier transform of the signal $x(t)$, then the Fourier transform of $\frac{dx}{dt}$ is $j2\pi f X(f)$ or $j\omega X(\omega)$. The frequency response of the operation is $H(\omega) = j\omega$. It is seen that the magnitude of the frequency response increases linearly with frequency, starting with $H(\omega) = 0$ at $\omega = 0$. Thus, the DC component is removed by the derivative operator, and higher frequencies receive linearly increasing gain: the operation represents a highpass filter. The derivative operator may be used to remove DC and suppress low-frequency components (and boost high-frequency components).

It follows readily that the second-order derivative operator $\frac{d^2}{dt^2}$ has the frequency response $H(\omega) = (j\omega)(j\omega) = -\omega^2$, with a quadratic increase in gain for higher

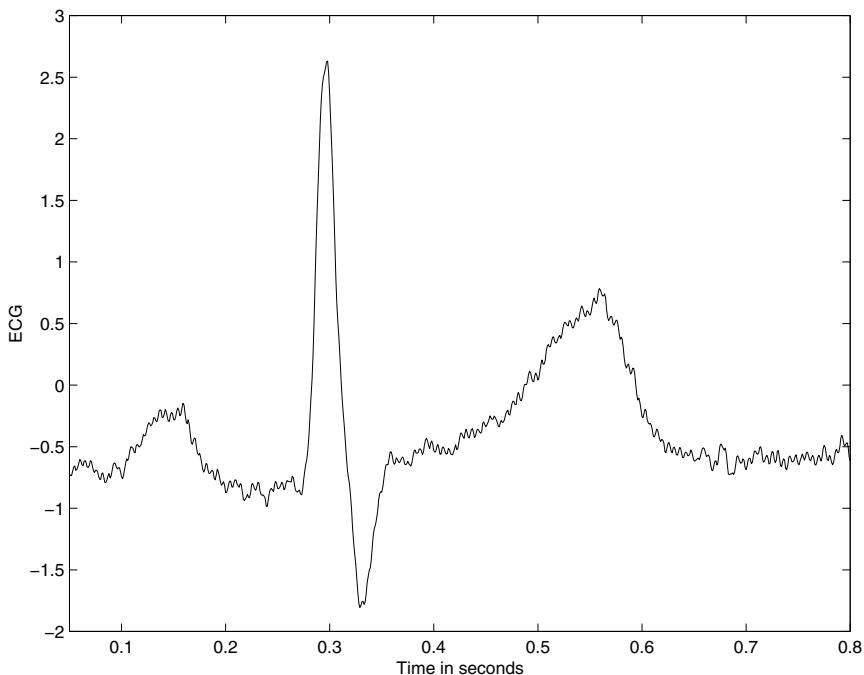


Figure 3.51 The ECG signal with high-frequency noise in Figure 3.50 after filtering by the 8-point MA filter shown in Figure 3.48 and Equation 3.107.

frequencies. The second-order derivative operator may be used to obtain even higher gain for higher frequencies than the first-order derivative operator; the former may be realized as a cascade of two of the latter.

In digital signal processing, the basic derivative is given by the first-order difference operator [46]

$$y(n) = \frac{1}{T} [x(n) - x(n-1)]. \quad (3.122)$$

The scale factor including the sampling interval T is required in order to obtain the rate of change of the signal with respect to the true time. The transfer function of the operator is

$$H(z) = \frac{1}{T} (1 - z^{-1}). \quad (3.123)$$

The filter has a zero at $z = 1$, the DC point.

The frequency response of the operator is

$$H(\omega) = \frac{1}{T} [1 - \exp(-j\omega)] = \frac{1}{T} \exp\left(-j\frac{\omega}{2}\right) \left[2j \sin\left(\frac{\omega}{2}\right)\right], \quad (3.124)$$

which leads to

$$|H(\omega)| = \frac{2}{T} \left|\sin\left(\frac{\omega}{2}\right)\right| \quad (3.125)$$

and

$$\angle H(\omega) = \frac{\pi}{2} - \frac{\omega}{2}. \quad (3.126)$$

The magnitude and phase responses of the first-order difference operator are plotted in Figure 3.52. The gain of the filter increases for higher frequencies up to the folding frequency $f_s/2$. The gain may be taken to approximate that of the ideal derivative operator, that is, $|\omega|$, for low values of ω . Any high-frequency noise present in the signal will be amplified significantly: therefore, the result could be noisy.

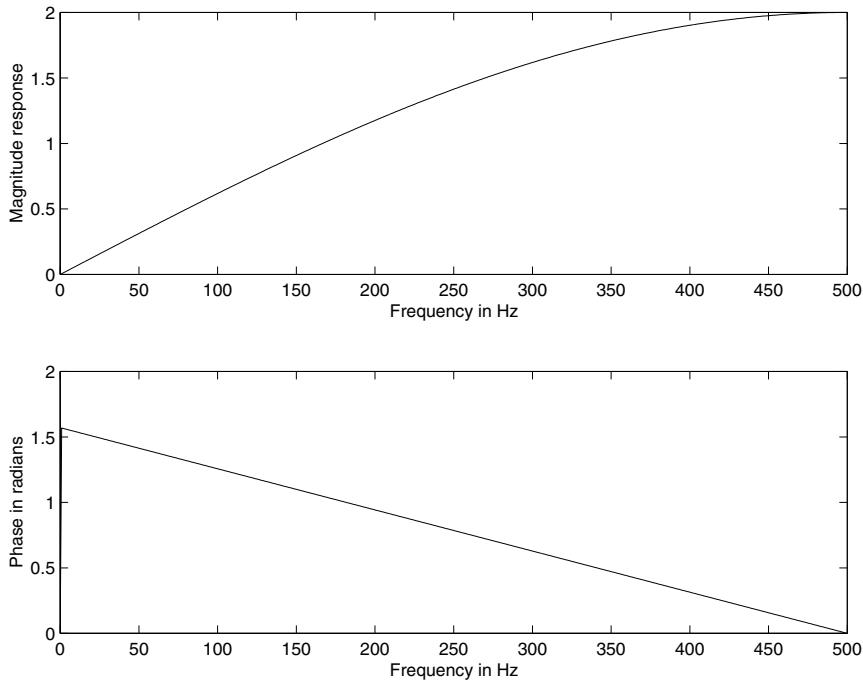


Figure 3.52 Magnitude and phase responses of the first-order difference operator in Equation 3.122. The magnitude response is shown on a linear scale in order to illustrate better its proportionality to frequency.

The noise-amplification problem with the first-order difference operator in Equation 3.122 may be controlled by taking the average of two successive output values:

$$\begin{aligned} y_3(n) &= \frac{1}{2} [y(n) + y(n-1)] \\ &= \frac{1}{2T} \{[x(n) - x(n-1)] + [x(n-1) - x(n-2)]\} \\ &= \frac{1}{2T} [x(n) - x(n-2)]. \end{aligned} \quad (3.127)$$

The transfer function of the operator above, known as the three-point central difference [46], is

$$H(z) = \frac{1}{2T} (1 - z^{-2}) = \left[\frac{1}{T} (1 - z^{-1}) \right] \left[\frac{1}{2} (1 + z^{-1}) \right]. \quad (3.128)$$

Observe that the transfer function of the three-point central-difference operator is the product of the transfer functions of the simple first-order difference operator and a two-point MA filter. The three-point central-difference operation may, therefore, be performed by the simple first-order difference operator and a two-point MA filter in series (cascade).

The magnitude and phase responses of the three-point central-difference operator are given by

$$|H(\omega)| = \frac{1}{T} |\sin(\omega)| \quad (3.129)$$

and

$$\angle H(\omega) = \frac{\pi}{2} - \omega, \quad (3.130)$$

and are plotted in Figure 3.53. The transfer function has zeros at $z = 1$ and $z = -1$, with the latter pulling the gain at the folding frequency ($f_s/2$) to zero: The operator is a bandpass filter. Although the operator does not have the noise-amplification problem of the first-order difference operator, the approximation of the $\frac{d}{dt}$ operation is poor after about $f_s/10$ [46].

Illustration of application: Figures 3.54 and 3.55 show the results of filtering the ECG signal with low-frequency noise shown in Figure 3.6, using the first-order difference and three-point central-difference operators, respectively. It is seen that the baseline drift has been removed, with the latter being less noisy than the former. However, it is obvious that the highpass and high-frequency emphasis effects inherent in both operators have removed the slow P and T waves, and have altered the QRS complexes to such an extent as to make the resulting waveforms look unlike ECG signals. (We shall see in Section 4.3 that, although the derivative operators are not useful in the present application, they are useful in detecting the QRS complex and the dicrotic notch.)

Problem: How can we improve the performance of the basic first-order difference operator as a filter to remove low-frequency noise or baseline wander without distorting the QRS complex?

Solution: The drawback of the first-order difference and the three-point central-difference operators lies in the fact that their magnitude responses remain low for a significant range of frequencies well beyond the band related to baseline wander. The zero of the first-order difference operator at $z = 1$ is desired in order to reject the DC component and low frequencies. However, we would like to maintain the levels of the components present in the signal beyond about 0.5 Hz, that is, we would like the gain of the filter to be close to unity after about 0.5 Hz.

The gain of a filter at specific frequencies may be boosted by placing poles at the related locations around the unit circle in the z -plane. For the sake of stability of the

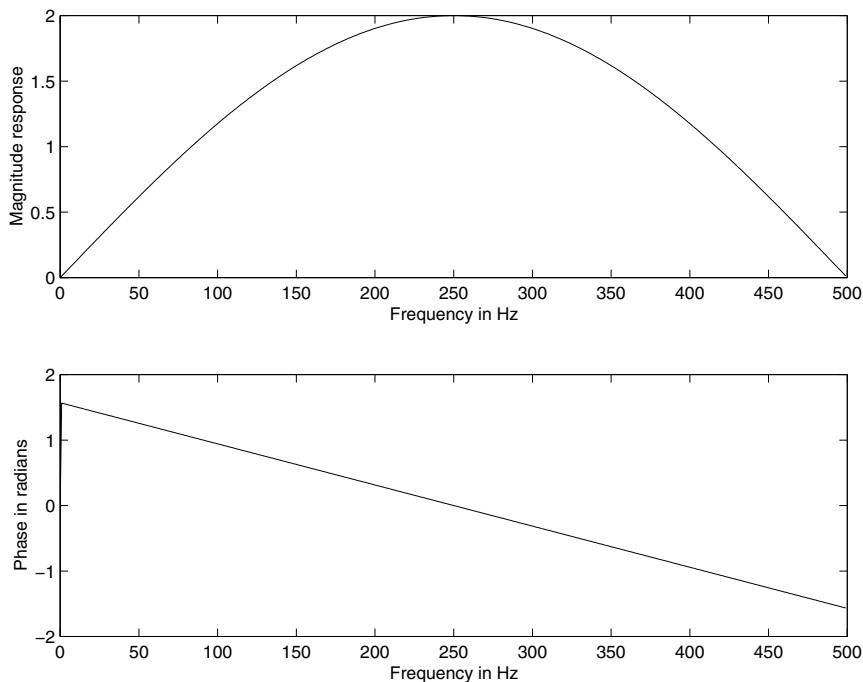


Figure 3.53 Magnitude and phase responses of the three-point central-difference operator in Equation 3.127. The magnitude response is shown on a linear scale.

filter, the poles should be placed within the unit circle. Since we are interested in maintaining a high gain at low frequencies, we could place a pole on the real axis (zero frequency), for example, at $z = 0.995$ [176]. The transfer function of the modified first-order difference filter is then

$$H(z) = \frac{1}{T} \left[\frac{1 - z^{-1}}{1 - 0.995 z^{-1}} \right] \quad (3.131)$$

or, equivalently,

$$H(z) = \frac{1}{T} \left[\frac{z - 1}{z - 0.995} \right]. \quad (3.132)$$

The time-domain input–output relationship is given as

$$y(n) = \frac{1}{T} [x(n) - x(n - 1)] + 0.995 y(n - 1). \quad (3.133)$$

Two equivalent signal-flow diagrams of the filter are shown in Figure 3.56. (Note: The filter is no longer an FIR filter; details on infinite impulse response or IIR filters are presented in Section 3.6.1.)

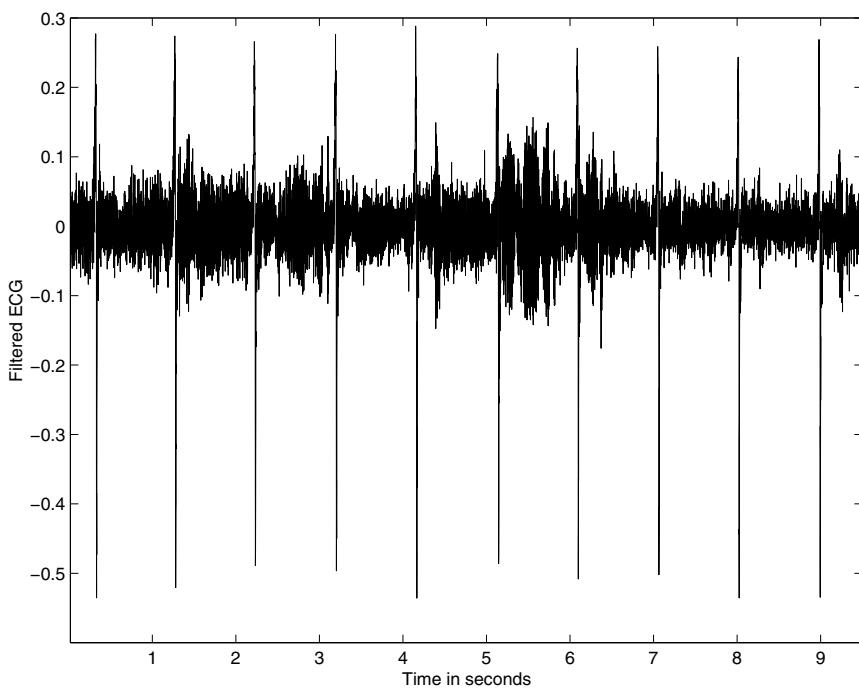


Figure 3.54 Result of filtering the ECG signal with low-frequency noise shown in Figure 3.6 using the first-order difference operator in Equation 3.122.

The form of $H(z)$ in Equation 3.132 in terms of z helps in understanding the graphical method for the evaluation of the frequency response of discrete-time filters [1–3, 46] shown in Figure 3.34. The frequency response of a system is obtained by evaluating its transfer function at various points on the unit circle in the z -plane, that is, by letting $z = \exp(j\omega)$ and evaluating $H(z)$ for various values of the frequency variable ω of interest. The numerator in Equation 3.132 expresses the vectorial distance between a specified point in the z -plane and the zero at $z = 1$; the denominator gives the distance to the pole at $z = 0.995$. The magnitude transfer function of a system for a particular value of z is given by the product of the distances from the corresponding point in the z -plane to all of the zeros of the system's transfer function, divided by the product of the distances to all of its poles. The phase response is given by the sum of the angles of the vectors joining the point to all of the zeros, minus the sum of the angles to all of the poles [1–3, 46]. It is obvious that the magnitude response of the filter in Equations 3.131 and 3.132 is zero at $z = 1$, due to the presence of a zero at that point. Furthermore, for values of z away from $z = 1$, the distances to the zero at $z = 1$ and the pole at $z = 0.995$ will be almost equal; therefore, the gain of the filter will be close to unity for frequencies greater than about 5 Hz. The magnitude and phase responses of the filter shown in Figure 3.57 confirm these observations: The filter is a highpass filter with nonlinear phase.

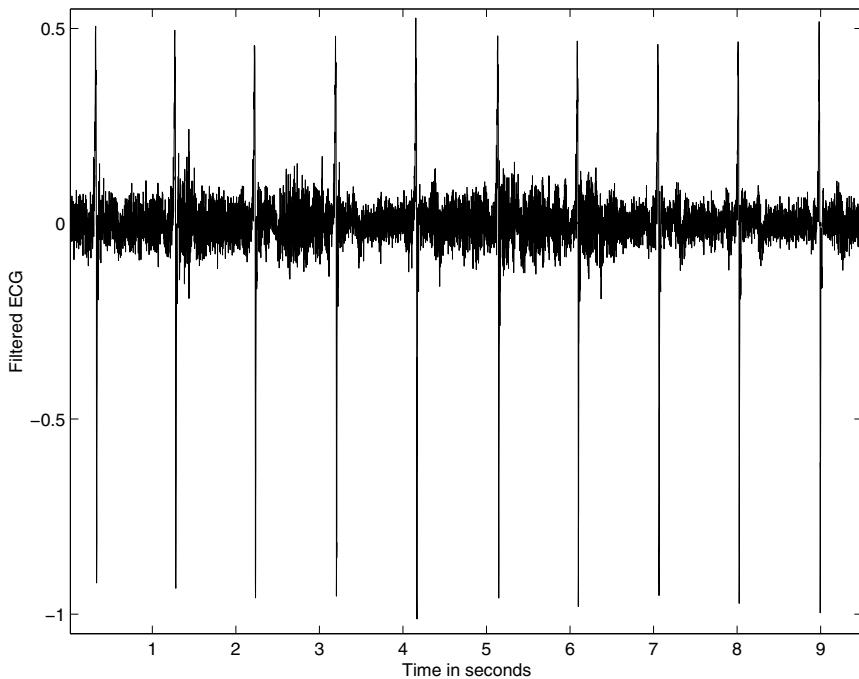


Figure 3.55 Result of filtering the ECG signal with low-frequency noise shown in Figure 3.6 using the three-point central-difference operator in Equation 3.127.

The result of application of the filter in Equation 3.133 to the ECG signal with low-frequency noise shown in Figure 3.6 is displayed in Figure 3.58. It is evident that the low-frequency baseline artifact has been removed without any significant distortion of the ECG waveforms, as compared with the results of differentiation in Figures 3.54 and 3.55. Close inspection, however, reveals that the S wave has been enhanced (made deeper) and that negative undershoots have been introduced after the P and T waves. Removal of the low-frequency baseline artifact has been achieved at the cost of distortion of the ECG waves due to the use of a derivative-based filter and its nonlinear phase response.

3.5.4 Various specifications of a filter

Although we started the present section with the design of filters in the time domain, we used several ways to specify the parameters and analyze the characteristics of filters that are not limited to the time domain. Some of the several possible specifications of a filter are listed below.

- Difference equation.
- Signal-flow diagram.

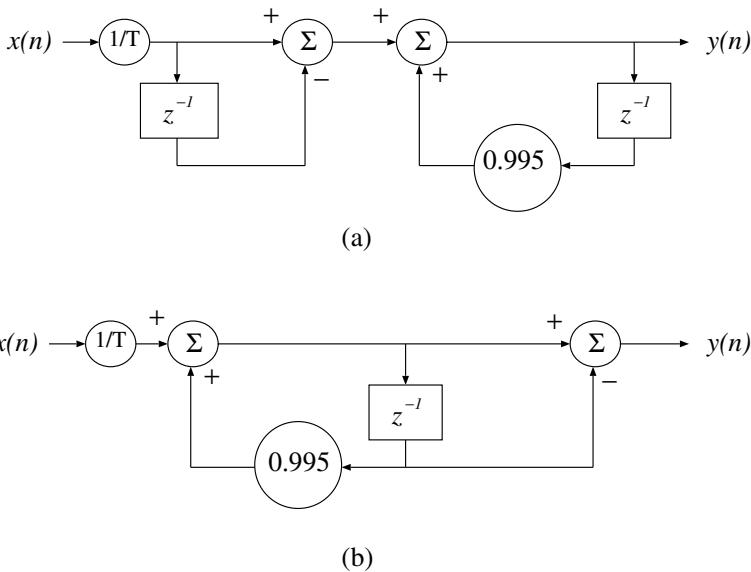


Figure 3.56 Two equivalent signal-flow diagrams of the filter to remove low-frequency noise or baseline wander. The form in (a) uses two delays, whereas that in (b) uses only one delay.

- Tap-weight or filter coefficients.
- Impulse response, $h(n)$.
- Transfer function, $H(z)$.
- Frequency response, $H(\omega)$, including its magnitude and phase parts.
- Pole-zero diagram and a gain factor.

The items listed above are interrelated, and any one of them may be used to derive any of the others, as seen in the several illustrations provided in the present and previous sections; the following sections provide more examples.

3.6 Frequency domain Filters

The filters described in the previous section perform relatively simple operations in the time domain; although their frequency-domain characteristics were explored, the operators were not specifically designed to possess any particular frequency response at the outset. The frequency response of the MA filter, in particular, is not attractive: the attenuation in the stopband is not high and is not uniform, with the gain falling below -20 dB only around the zeros of the transfer function.

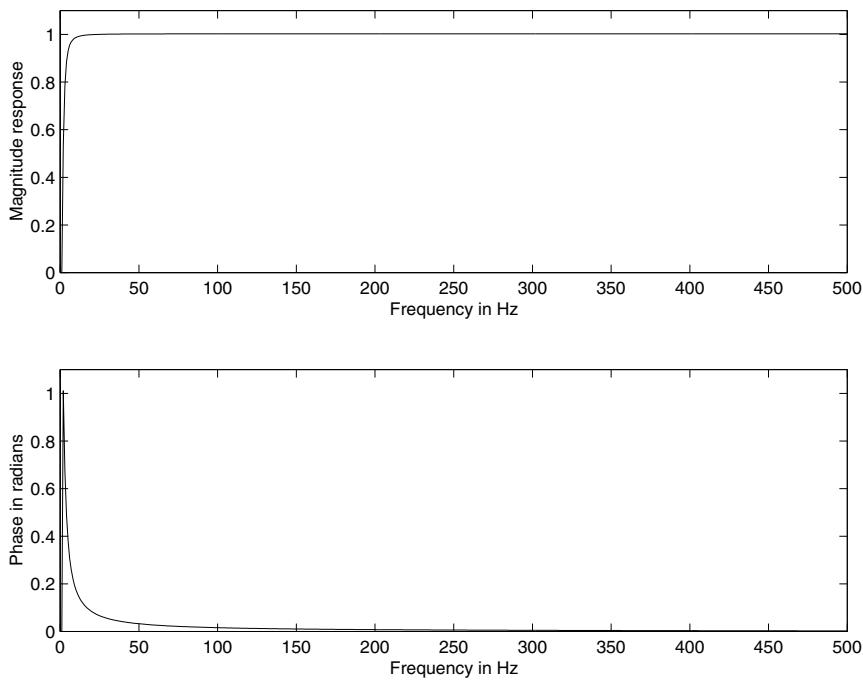


Figure 3.57 Normalized magnitude and phase responses of the filter to remove baseline wander as in Equation 3.131. The magnitude response is shown on a linear scale.

Filters may be designed in the frequency domain to provide specific lowpass, highpass, bandpass, or band-reject (notch) characteristics. Frequency-domain filters may be implemented in software after obtaining the Fourier transform of the input signal, or converted into equivalent time-domain filters and applied directly upon the signal samples.

Many design procedures are available in the literature to design various types of filters: The most-commonly used designs are the Butterworth, Chebyshev, elliptic, and Bessel filters [15, 40, 177–181]. Since these filters have been well-established in the analog-filter domain, it is common to commence with an analog design and apply the bilinear transformation to obtain a digital filter in the z -domain. It is also common to design a lowpass filter with the desired passband, transition, and stopband characteristics on a normalized-frequency axis, and then transform it to the desired lowpass, highpass, bandpass, or band-reject characteristics [15, 177]. Frequency-domain filters may also be specified directly in terms of the values of the desired frequency response at certain frequency samples only, and then transformed into the equivalent time-domain filter coefficients via the inverse Fourier transform.

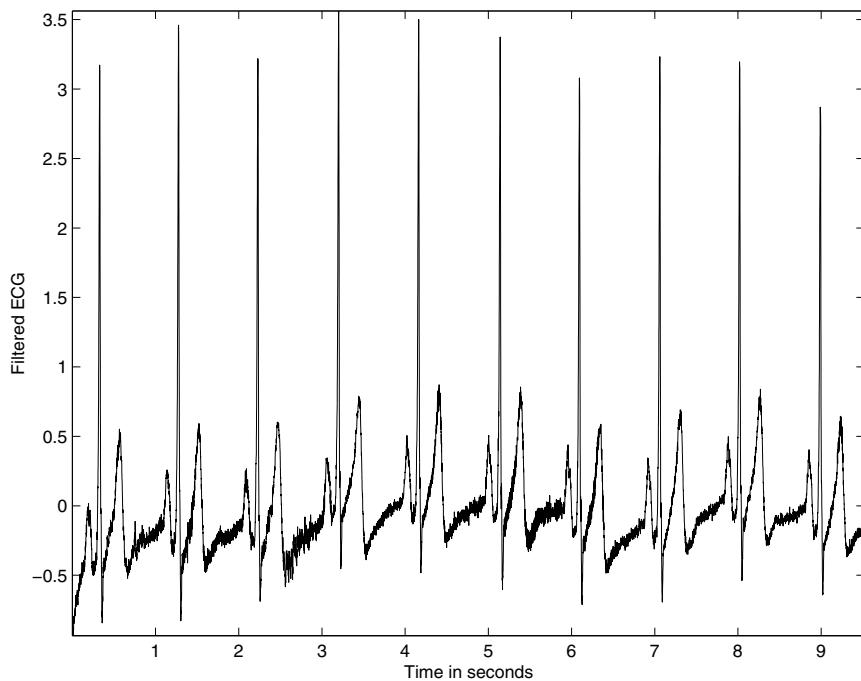


Figure 3.58 Result of processing the ECG signal with low-frequency noise shown in Figure 3.6 using the filter to remove baseline wander as in Equation 3.133. (Compare with the results in Figures 3.54 and 3.55.)

3.6.1 Removal of high frequency noise: Butterworth lowpass filters

Problem: Design a frequency-domain filter to remove high-frequency noise with minimal loss of signal components in the specified passband.

Solution: The Butterworth filter is a commonly used frequency-domain filter due to its simplicity and the property of a maximally flat magnitude response in the passband. For a Butterworth lowpass filter of order N , the first $2N - 1$ derivatives of the squared magnitude response are zero at $\Omega = 0$, where Ω represents the analog radian frequency. The Butterworth filter response is monotonic in the passband as well as in the stopband.

The basic Butterworth lowpass filter function is given as [15, 174]

$$|H_a(j\Omega)|^2 = \frac{1}{1 + \left(\frac{j\Omega}{j\Omega_c}\right)^{2N}}, \quad (3.134)$$

where H_a is the frequency response of the analog filter and Ω_c is the cutoff frequency (in radians/s). A Butterworth filter is completely specified by its cutoff frequency Ω_c and order N . As the order N increases, the filter response becomes more

flat in the passband, and the transition to the stopband becomes faster or sharper. $|H_a(j\Omega_c)|^2 = \frac{1}{2}$ for all N .

Changing to the Laplace variable s , we get

$$H_a(s)H_a(-s) = \frac{1}{1 + \left(\frac{s}{j\Omega_c}\right)^{2N}}. \quad (3.135)$$

The poles of the squared transfer function are located with equal spacing around a circle of radius Ω_c in the s -plane, distributed symmetrically on either side of the imaginary axis $s = j\Omega$. No pole will lie on the imaginary axis itself; poles will appear on the real axis for odd N . The angular spacing between the poles is $\frac{\pi}{N}$. If $H_a(s)H_a(-s)$ has a pole at $s = s_p$, it will have a pole at $s = -s_p$ as well. Furthermore, for the filter coefficients to be real, complex poles must appear in conjugate pairs. In order to obtain a stable and causal filter, we need to form $H_a(s)$ with only the N poles on the LHS of the s -plane. The pole positions in the s -plane are given by

$$s_k = \Omega_c \exp \left[j\pi \left(\frac{1}{2} + \frac{(2k-1)}{2N} \right) \right], \quad (3.136)$$

$k = 1, 2, \dots, 2N$ [177].

Once the pole positions are obtained in the s -plane, they may be combined to obtain the transfer function in the analog Laplace domain as

$$H_a(s) = \frac{G}{(s - p_1)(s - p_2)(s - p_3) \cdots (s - p_N)}, \quad (3.137)$$

where p_k , $k = 1, 2, \dots, N$, are the N poles of the transfer function in the left-half of the s -plane, and G is a gain factor specified as needed or calculated to normalize the gain at DC ($s = 0$) to be unity.

If we use the bilinear transformation

$$s = \frac{2}{T} \left[\frac{1 - z^{-1}}{1 + z^{-1}} \right], \quad (3.138)$$

the Butterworth circle in the s -plane maps to a circle in the z -plane with its real-axis intercepts at $z = \frac{2-\Omega_c T}{2+\Omega_c T}$ and $z = \frac{2+\Omega_c T}{2-\Omega_c T}$. The poles at $s = s_p$ and $s = -s_p$ in the s -plane map to the locations $z = z_p$ and $z = 1/z_p$, respectively. The poles in the z -plane are not uniformly spaced around the transformed Butterworth circle. For stability, all poles of $H(z)$ must lie within the unit circle in the z -plane.

Consider the unit circle in the z -plane given by $z = e^{j\omega}$. For points on the unit circle, we have

$$s = \sigma + j\Omega = \frac{2}{T} \left(\frac{1 - e^{-j\omega}}{1 + e^{-j\omega}} \right) = \frac{2j}{T} \tan \left(\frac{\omega}{2} \right). \quad (3.139)$$

For the unit circle, $\sigma = 0$; therefore, we get the relationships between the continuous-time (s -domain) frequency variable Ω and the discrete-time (z -domain) frequency variable ω as

$$\Omega = \frac{2}{T} \tan \left(\frac{\omega}{2} \right) \quad (3.140)$$

and

$$\omega = 2 \tan^{-1} \left(\frac{\Omega T}{2} \right). \quad (3.141)$$

This is a nonlinear relationship that warps the frequency values as they are mapped from the imaginary (vertical) axis in the s -plane to the unit circle in the z -plane (or viceversa), and it should be taken into account in specifying cutoff frequencies.

The transfer function $H_a(s)$ may be mapped to the z -domain by applying the bilinear transformation, that is, by substituting $s = \frac{2}{T} \frac{1-z^{-1}}{1+z^{-1}}$ in Equation 3.137. The transfer function $H(z)$ may then be simplified to the form

$$H(z) = \frac{G' (1 + z^{-1})^N}{\sum_{k=0}^N a_k z^{-k}}, \quad (3.142)$$

where a_k , $k = 0, 1, 2, \dots, N$, are the filter coefficients or tap weights (with $a_0 = 1$), and G' is the gain factor [usually calculated so as to obtain $|H(z)| = 1$ at DC, that is, at $z = 1$]. Observe that the filter has N zeros at $z = -1$ due to the use of the bilinear transformation. The filter is now in the familiar form of an IIR filter. Two forms of realization of a generic IIR filter are illustrated as signal-flow diagrams in Figures 3.59 and 3.60: The former represents a direct realization using $2N$ delays and $2N + 1$ multipliers (with $a_0 = 1$), whereas the latter uses only N delays and $2N + 1$ multipliers.

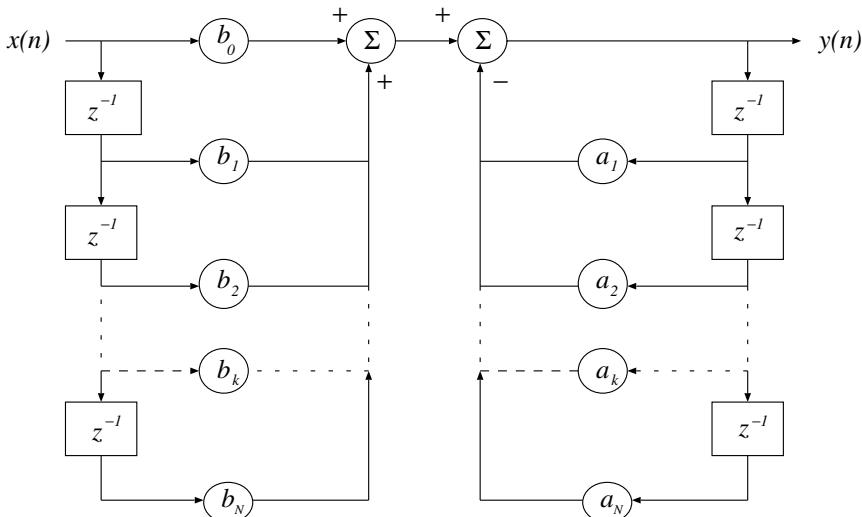


Figure 3.59 Signal-flow diagram of a direct realization of a generic IIR filter. This form uses $2N$ delays and $2N + 1$ multipliers for a filter of order N .

A time-domain representation of the filter will be required if the filter is to be applied to data samples directly in the time domain. From the filter transfer function

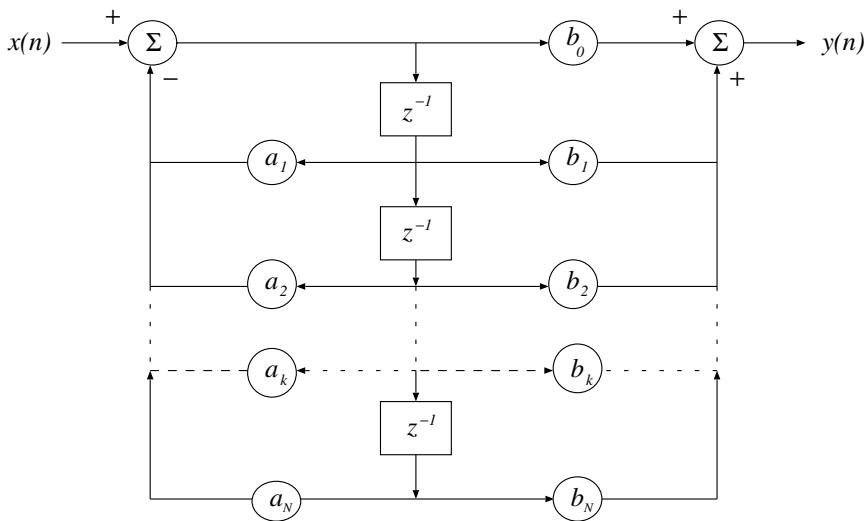


Figure 3.60 Signal-flow diagram of a realization of an IIR filter that uses only N delays and $(2N + 1)$ multipliers for a filter of order N .

$H(z)$ in Equation 3.142, it becomes easy to represent the filter in the time domain with the difference equation

$$y(n) = \sum_{k=0}^N b_k x(n-k) - \sum_{k=1}^N a_k y(n-k). \quad (3.143)$$

The coefficients b_k are given by the coefficients of the expansion of $G'(1 + z^{-1})^N$. (The MATLAB® [182] command *butter* provides Butterworth filters.)

It is also possible to directly specify the Butterworth filter as

$$|H(\omega)|^2 = \frac{1}{1 + \left(\frac{\omega}{\omega_c}\right)^{2N}}, \quad (3.144)$$

with ω normalized to the range $(0, 2\pi)$ for sampled or discrete-time signals; in such a case, the equation is valid only for the range $(0, \pi)$, with the function in the range $(\pi, 2\pi)$ being a reflection of that over $(0, \pi)$. The cutoff frequency ω_c should be specified in the range $(0, \pi)$.

If the DFT is used to compute the Fourier transforms of the signals being filtered, Equation 3.144 may be modified to

$$|H(k)|^2 = \frac{1}{1 + \left(\frac{k}{k_c}\right)^{2N}}, \quad (3.145)$$

where k is the index of the DFT array standing for discretized frequency. With K being the number of points in the DFT array, k_c is the array index corresponding to

the cutoff frequency ω_c (that is, $k_c = \lceil K \frac{\omega_c}{\omega_s} \rceil$). The equation given above is valid for $k = 0, 1, 2, \dots, \frac{K}{2}$, with the second half over $(\frac{K}{2} + 1, K - 1)$ being a reflection of the first half [that is, $H(k) = H(K - k)$, $k = \frac{K}{2} + 1, \dots, K - 1$]. Note that the DFT includes two unique values: the DC component in $H(0)$ and the folding-frequency component in $H(\frac{K}{2})$. The variable k in the filter equation could also be used to represent normalized frequency in the range $(0, 1)$, with unity standing for the sampling frequency, 0.5 standing for the maximum frequency present in the sampled signal (that is, the folding frequency), and k_c being specified in the range $(0, 0.5)$. (Note: MATLAB® normalizes one-half of the sampling frequency to unity; the maximum normalized frequency present in the sampled signal is then unity. MATLAB® and a few programming languages do not allow an array index to be zero: In such a case, the indices mentioned above must be incremented by one.)

One could compute the DFT of the given signal, multiply the result by $|H(k)|$, and compute the inverse DFT to obtain the filtered signal. The advantage of this procedure is that no phase change is involved: The filter is a strictly magnitude-only transfer function. The time-domain implementation described earlier will include a phase response which may not be desired. However, time-domain implementation will be required in on-line signal processing applications.

Butterworth lowpass filter design example: In order to design a Butterworth lowpass filter, we need to specify two parameters: ω_c and N . The two parameters may be specified based on a knowledge of the characteristics of the filter as well as those of the signal and noise. It is also possible to specify the required minimum gain at a certain frequency in the passband and the required minimum attenuation at another frequency in the stopband. The two values may then be used with Equation 3.134 to obtain two equations in the two unknowns ω_c and N , which may be solved to derive the filter parameters [174].

Given the 3 dB cutoff frequency f_c and order N , the procedure to design a Butterworth lowpass filter is as follows:

1. Convert the specified 3 dB cutoff frequency f_c to radians in the normalized range $(0, 2\pi)$ as $\omega_c = \frac{f_c}{f_s} 2\pi$. Then, $T = 1$. Prewarp the cutoff frequency ω_c by using Equation 3.140 and obtain Ω_c .
2. Derive the positions of the poles of the filter in the s -plane as given by Equation 3.136.
3. Form the transfer function $H_a(s)$ of the Butterworth lowpass filter in the Laplace domain by using the poles in the left-half plane only as given by Equation 3.137.
4. Apply the bilinear transformation as per Equation 3.138 and obtain the transfer function of the filter $H(z)$ in the z -domain as in Equation 3.142.
5. Convert the filter to the series of coefficients b_k and a_k as in Equation 3.143.

Let us now design a Butterworth lowpass filter with $f_c = 40 \text{ Hz}$, $f_s = 200 \text{ Hz}$, and $N = 4$. We have $\omega_c = \frac{40}{200} 2\pi = 0.4\pi \text{ radians/s}$. The prewarped s -domain cutoff frequency is $\Omega_c = \frac{2}{T} \tan\left(\frac{\omega_c}{2}\right) = 1.453085 \text{ radians/s}$.

The poles of $H_a(s)H_a(-s)$ are placed around a circle of radius 1.453085 radians /s with an angular separation of $\frac{\pi}{N} = \frac{\pi}{4}$ radians. The poles of interest are located at angles $\frac{5}{8}\pi$ and $\frac{7}{8}\pi$ and the corresponding conjugate positions. Figure 3.61 shows the positions of the poles of $H_a(s)H_a(-s)$ in the Laplace plane. The coordinates of the poles of interest are $(-0.556072 \pm j 1.342475)$ and $(-1.342475 \pm j 0.556072)$. The transfer function of the filter is

$$H_a(s) = \frac{4.458247}{(s^2 + 1.112143s + 2.111456)(s^2 + 2.684951s + 2.111456)}. \quad (3.146)$$

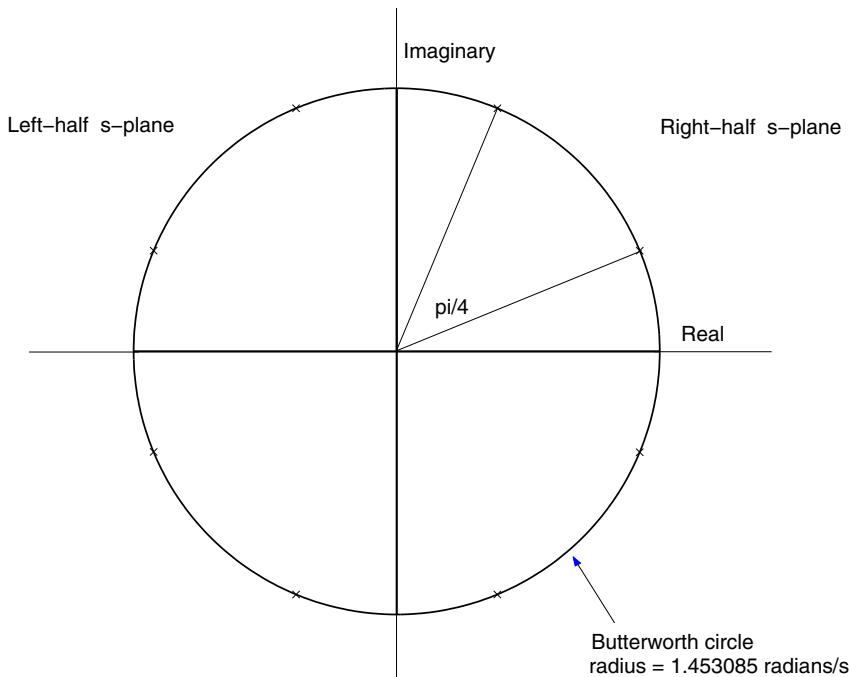


Figure 3.61 Pole positions in the s -plane of the squared magnitude response of the Butterworth lowpass filter with $f_c = 40$ Hz, $f_s = 200$ Hz, and $N = 4$.

Applying the bilinear transformation, we get

$$H(z) = \frac{0.046583(1+z^{-1})^4}{(1 - 0.447765z^{-1} + 0.460815z^{-2})(1 - 0.328976z^{-1} + 0.064588z^{-2})}. \quad (3.147)$$

This filter has four poles at $(0.223882 \pm j 0.640852)$ and $(0.164488 \pm j 0.193730)$, and four zeros at $-1 + j 0$. The b_k coefficients of the filter as in Equation 3.143 are $\{0.0465829, 0.186332, 0.279497, 0.186332, 0.046583\}$, and the a_k coefficients are $\{1, -0.776740, 0.672706, -0.180517, 0.029763\}$. The pole-zero plot and the frequency response of the filter are given in Figures 3.62 and 3.63, respectively. The

frequency response displays the expected monotonic decrease in gain and -3 dB power point or 0.707 gain at 40 Hz.

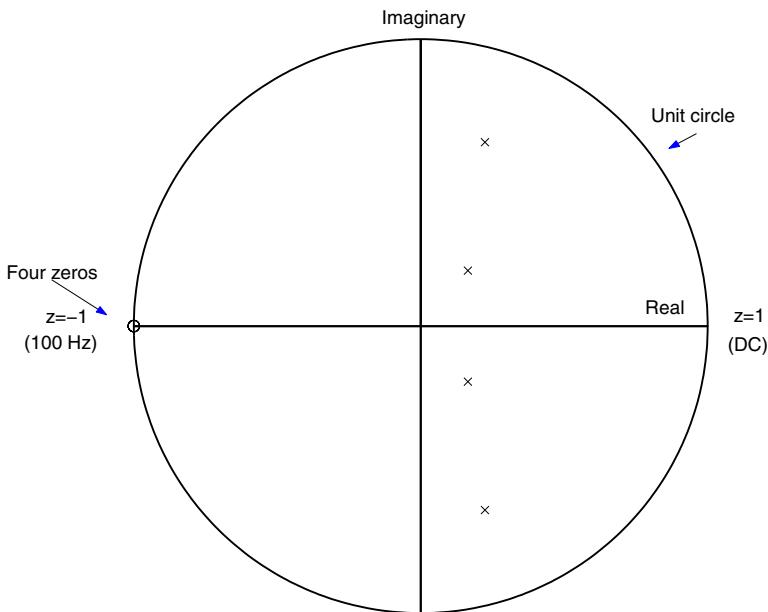


Figure 3.62 Positions of the poles and zeros in the z -plane of the Butterworth lowpass filter with $f_c = 40 \text{ Hz}$, $f_s = 200 \text{ Hz}$, and $N = 4$.

Figure 3.64 compares the magnitude responses of three Butterworth lowpass filters with $f_c = 40 \text{ Hz}$ and $f_s = 200 \text{ Hz}$, with the order increasing from $N = 4$ (dotted) to $N = 8$ (dashed) to $N = 12$ (solid). All three filters have their half-power points (gain = 0.707) at 40 Hz, but the transition band becomes sharper as the order N is increased.

The Butterworth design is popular because of its simplicity, a monotonically decreasing magnitude response, and a maximally flat magnitude response in the passband. Its main disadvantages are a slow (or wide) transition from the passband to the stopband, and a nonlinear phase response. The nonlinear phase may be corrected by passing the filter output again through the same filter but after a reversal in time [178]. This process, however, leads to a magnitude response that is the square of that provided by the initial filter design. The squaring effect may be compensated for in the initial design; however, the approach cannot be applied in real time. The elliptic filter design provides a sharp transition band at the expense of ripples in the passband and the stopband. The Bessel design provides a group delay that is maximally flat at DC, and a phase response that approximates a linear response. Details on the design of Bessel, Chebyshev, elliptic, and other filters may be found in other sources on filter design [15, 40, 177–181].

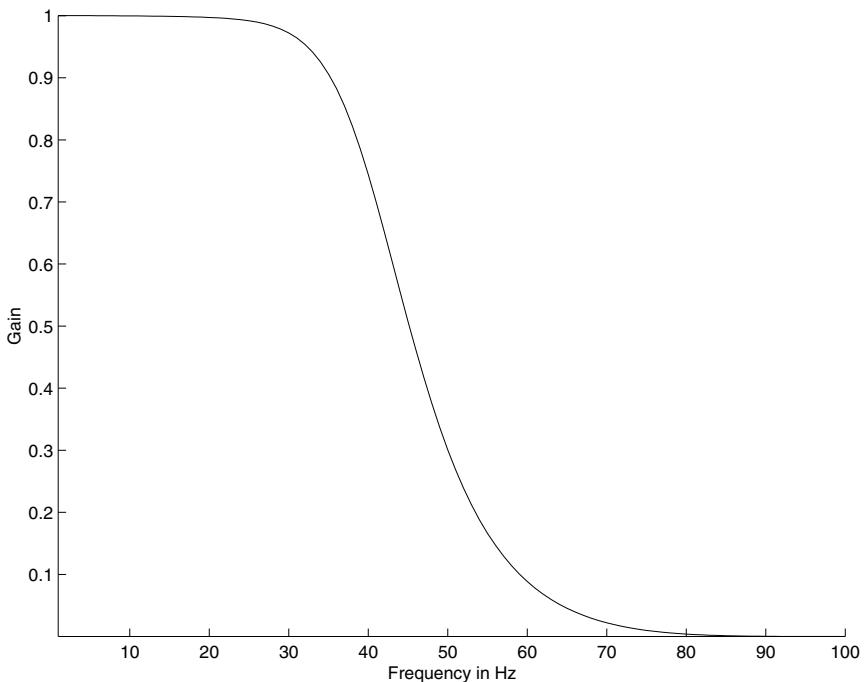


Figure 3.63 Magnitude response of the Butterworth lowpass filter with $f_c = 40 \text{ Hz}$, $f_s = 200 \text{ Hz}$, and $N = 4$.

Illustrations of application: The upper trace in Figure 3.65 illustrates a carotid pulse signal with high-frequency noise. The lower trace in the same figure shows the result of processing with a Butterworth lowpass filter ($f_c = 40 \text{ Hz}$, $f_s = 200 \text{ Hz}$, and $N = 12$). The high-frequency noise has been effectively reduced.

Figure 3.66 shows the result of filtering the noisy ECG signal shown in Figure 3.50 with an eighth-order Butterworth lowpass filter as in Equations 3.144 and 3.145 and a cutoff frequency of 70 Hz . The frequency response $|H(\omega)|$ of the filter is shown in Figure 3.67. It is evident that the high-frequency noise has been suppressed by the filter.

3.6.2 Removal of low frequency noise: Butterworth highpass filters

Problem: Design a frequency-domain filter to remove low-frequency noise with minimal loss of signal components in the passband.

Solution: Highpass filters may be designed on their own, or obtained by transforming a normalized prototype lowpass filter [174, 177]. The latter approach is easier since lowpass filter prototypes with various characteristics are readily available, as are the transformations required to derive highpass, bandpass, and band-

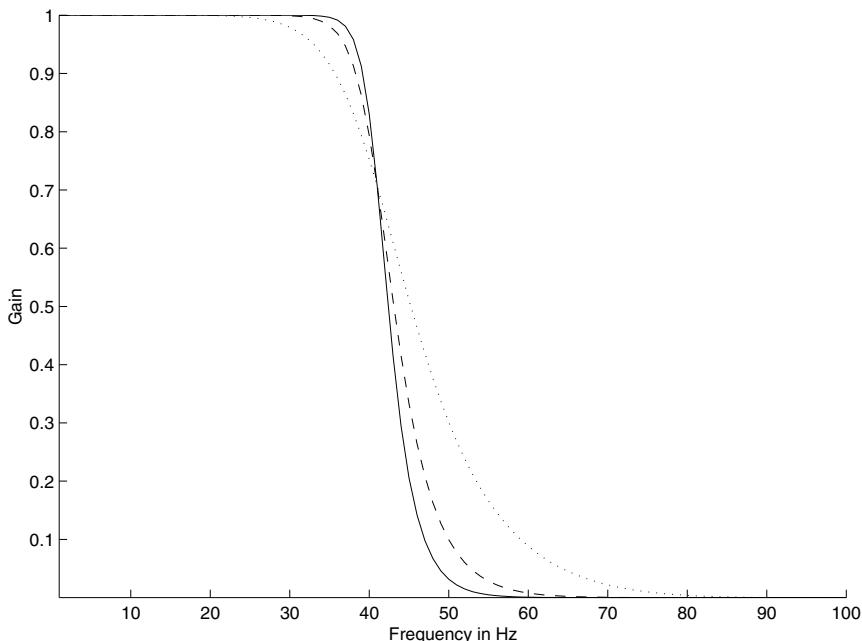


Figure 3.64 Magnitude responses of three Butterworth lowpass filters with $f_c = 40 \text{ Hz}$, $f_s = 200 \text{ Hz}$, and variable order: $N = 4$ (dotted), $N = 8$ (dashed), and $N = 12$ (solid).

stop filters [174, 177]. [MATLAB[®] provides highpass filters with the command `butter(N, fc, 'high')`.]

As in the case of the Butterworth lowpass filter in Equation 3.145, the Butterworth highpass filter may be specified directly in the discrete-frequency domain as

$$|H(k)|^2 = \frac{1}{1 + \left(\frac{k_c}{k}\right)^{2N}}. \quad (3.148)$$

Illustration of application: Figure 3.6 shows a segment of an ECG signal with low-frequency noise appearing in the form of a wandering baseline (baseline drift). Figure 3.68 shows the result of filtering the signal with an eighth-order Butterworth highpass filter as in Equation 3.148 and a cutoff frequency of 2 Hz. The frequency response of the filter is shown in Figure 3.69. While the low-frequency artifact has been removed by the filter, it should be noted that the high-frequency noise present in the signal has not been affected.

Observe that the filtered result retains the characteristics of the QRS complex, unlike the case with the derivative-based time-domain filters (compare Figure 3.68 with Figures 3.54 and 3.55.) This advantage is due to the fact that the Butterworth highpass filter that was used has a gain of almost unity over the frequency range of 3 – 100 Hz; the derivative-based filters severely attenuate these components and hence distort the QRS complex. However, it should be observed that the filter has

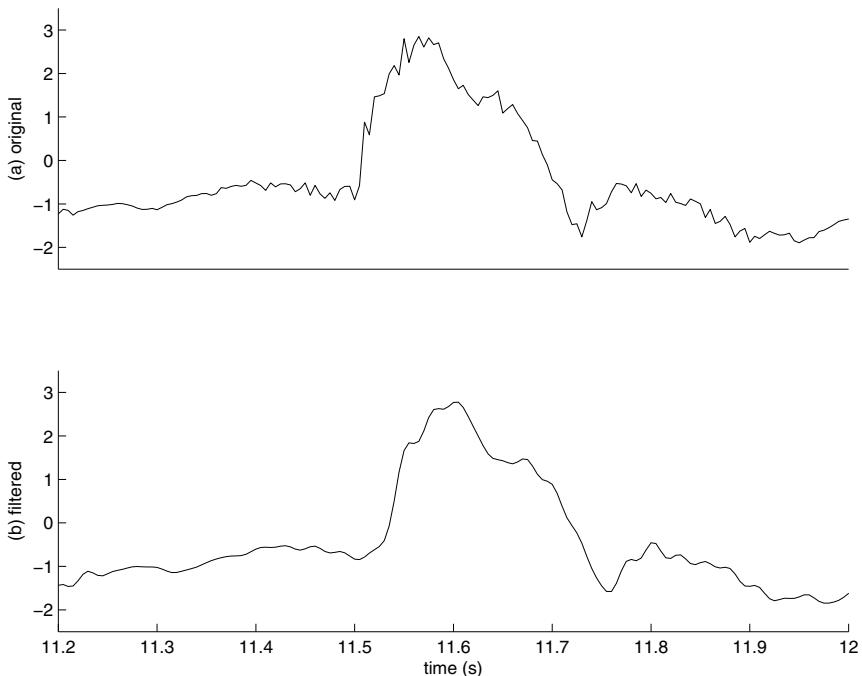


Figure 3.65 Upper trace: a carotid pulse signal with high-frequency noise. Lower trace: result of filtering using a Butterworth lowpass filter with $f_c = 40 \text{ Hz}$, $f_s = 200 \text{ Hz}$, and $N = 12$. The filtering operation was performed in the time domain using the MATLAB[®] *filter* command.

distorted the P and T waves to some extent. The result in Figure 3.68 compares well with that in Figure 3.58, obtained using the simpler IIR filter in Equation 3.133. (Compare the frequency responses in Figures 3.69, 3.52, 3.53, and 3.57.)

3.6.3 Removal of periodic artifacts: Notch and comb filters

Problem: Design a frequency-domain filter to remove periodic artifacts, such as power-line interference.

Solution: The simplest method to remove periodic artifacts is to compute the Fourier transform of the signal, delete the undesired component(s) from the spectrum, and then compute the inverse Fourier transform. The undesired components could be set to zero, or better, to the average level of the signal components over a few frequency samples around the component that is to be removed; the former method will remove the noise components as well as the signal components at the frequencies of concern, whereas the latter assumes that the signal spectrum is smooth in the affected regions.

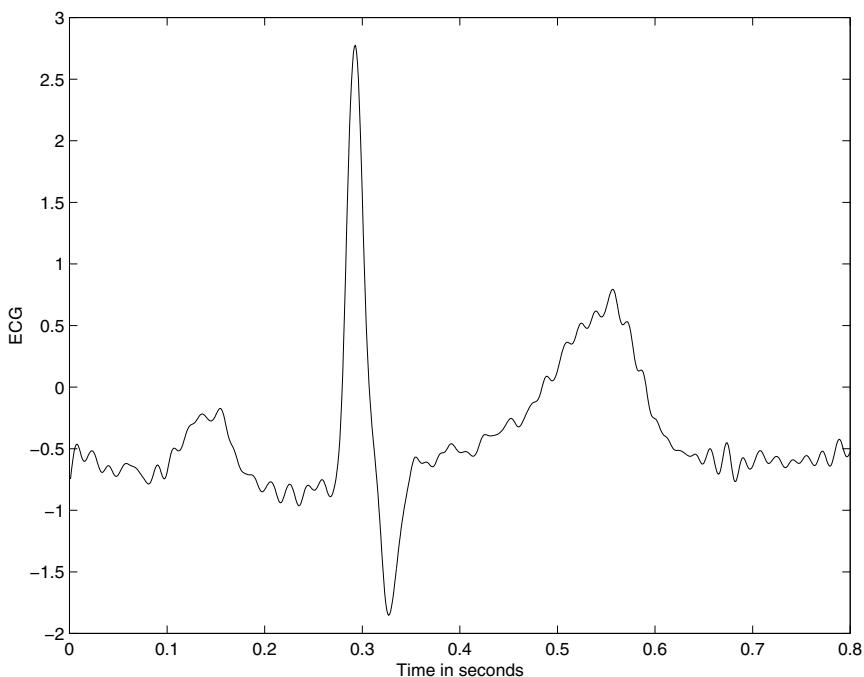


Figure 3.66 Result of frequency-domain filtering of the noisy ECG signal in Figure 3.50 with an eighth-order Butterworth lowpass filter with cutoff frequency = 70 Hz.

Periodic interference may also be removed by notch filters with zeros on the unit circle in the z -domain at the specific frequencies to be rejected. If f_o is the interference frequency, the angles of the (complex conjugate) zeros required will be $\pm \frac{f_o}{f_s}(2\pi)$; the radius of the zeros will be unity. If harmonics are also present, multiple zeros will be required at $\pm \frac{n f_o}{f_s}(2\pi)$, n representing the orders of all of the harmonics present. The angles of the zeros are limited to the range $(-\pi, \pi)$. The filter is then called a “comb” filter. In some situations, higher-order harmonics beyond $\frac{f_s}{2}$ may appear at aliased locations (see Figures 3.8 and 3.100); zeros may then be placed at such frequencies as well.

Notch filter design example: Consider a signal with power-line interference at $f_o = 60$ Hz and sampling rate of $f_s = 1,000$ Hz (see Figures 3.7 and 3.8). The notch filter is then required to have zeros at $\omega_o = \pm \frac{f_o}{f_s}(2\pi) = \pm 0.377$ radians = $\pm 21.6^\circ$. The locations of the zeros are given by $\cos(\omega_o) \pm j \sin(\omega_o)$ or $z_1 = 0.92977 + j0.36812$ and $z_2 = 0.92977 - j0.36812$. The transfer function is

$$H(z) = (1 - z^{-1} z_1)(1 - z^{-1} z_2) = 1 - 1.85955 z^{-1} + z^{-2}. \quad (3.149)$$

Substituting $z = 1$ in the expression above, we get the DC response or gain as $H(1) = 1 - 1.85955 + 1 = 0.14045$. Therefore, if the gain at DC is required to be unity, $H(z)$ as above should be divided by 0.14045.

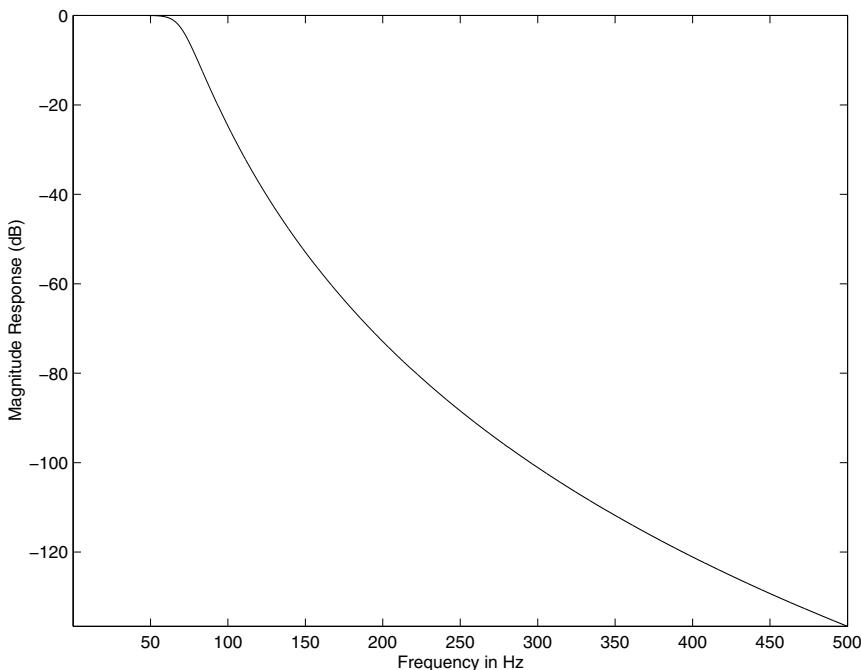


Figure 3.67 Frequency response of the eighth-order Butterworth lowpass filter with cutoff frequency $f_c = 70 \text{ Hz}$ and $f_s = 1,000 \text{ Hz}$.

Figure 3.70 shows a plot of the zeros of the notch filter in the z -plane. Figure 3.71 shows the magnitude and phase responses of the notch filter. Observe that the filter attenuates not only the 60 Hz component but also a band of frequencies around 60 Hz . The sharpness of the notch may be improved by placing a few poles near or symmetrically around the zeros and inside the unit circle [1, 2, 176]. Note also that the gain of the filter is at its maximum at $f_s/2$; additional lowpass filtering in the case of application to ECG signals could be used to reduce the gain at frequencies beyond about 80 Hz .

Comb filter design example: Let us consider the presence of a periodic artifact with the fundamental frequency of 60 Hz and odd harmonics at 180 Hz , 300 Hz , and 420 Hz . Let $f_s = 1,000 \text{ Hz}$, and assume the absence of any aliasing error. Zeros are then desired at 60 Hz , 180 Hz , 300 Hz , and 420 Hz , which translate to $\pm 21.6^\circ$, $\pm 64.8^\circ$, $\pm 108^\circ$, and $\pm 151.2^\circ$, with 360° corresponding to $1,000 \text{ Hz}$. The coordinates of the zeros are $0.92977 \pm j0.36812$, $0.42578 \pm j0.90483$, $-0.30902 \pm j0.95106$, and $-0.87631 \pm j0.48175$. The transfer function of the filter is

$$\begin{aligned} H(z) &= G (1 - 1.85955z^{-1} + z^{-2})(1 - 0.85156z^{-1} + z^{-2}) \\ &\quad \times (1 + 0.61803z^{-1} + z^{-2})(1 + 1.75261z^{-1} + z^{-2}), \end{aligned} \quad (3.150)$$

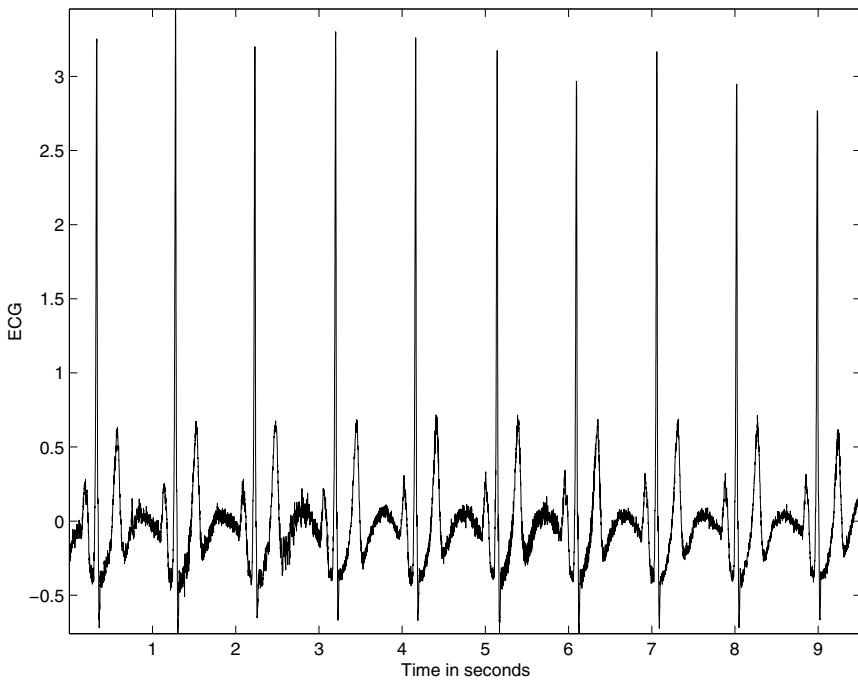


Figure 3.68 Result of frequency-domain filtering of the ECG signal with low-frequency noise in Figure 3.6 with an eighth-order Butterworth highpass filter with cutoff frequency = 2 Hz. (Compare with the results in Figures 3.54, 3.55, and 3.58.)

where G is the desired gain or scaling factor. With G computed so as to set the gain at DC to be unity, the filter transfer function becomes

$$\begin{aligned} H(z) = & 0.6310 - 0.2149z^{-1} + 0.1512z^{-2} - 0.1288z^{-3} + 0.1227z^{-4} \\ & - 0.1288z^{-5} + 0.1512z^{-6} - 0.2149z^{-7} + 0.6310z^{-8}. \end{aligned} \quad (3.151)$$

A plot of the locations of the zeros in the z -plane is shown in Figure 3.72. The frequency response of the comb filter is shown in Figure 3.73. Observe the low gain at not only the notch frequencies but also in the adjacent regions.

Illustrations of application: Figure 3.74 shows an ECG signal with power-line interference at $f_o = 60$ Hz. Figure 3.75 shows the result of applying the notch filter in Equation 3.149 to the signal. The 60 Hz interference has been effectively removed, with no perceptible distortion of the ECG waveform.

Figure 3.76 shows an ECG signal with noise including power-line interference at 60 Hz and $f_s = 200$ Hz. A notch filter was designed with two zeros at 60 Hz with radius equal to unity. The coordinates of the zeros in the z -plane are $-0.3090 \pm j0.9511$. The frequency response of the notch filter is shown in Figure 3.77 (dashed line). The result of filtering the noisy signal in Figure 3.76 is shown in Figure 3.78, which indicates effective removal of the interference.

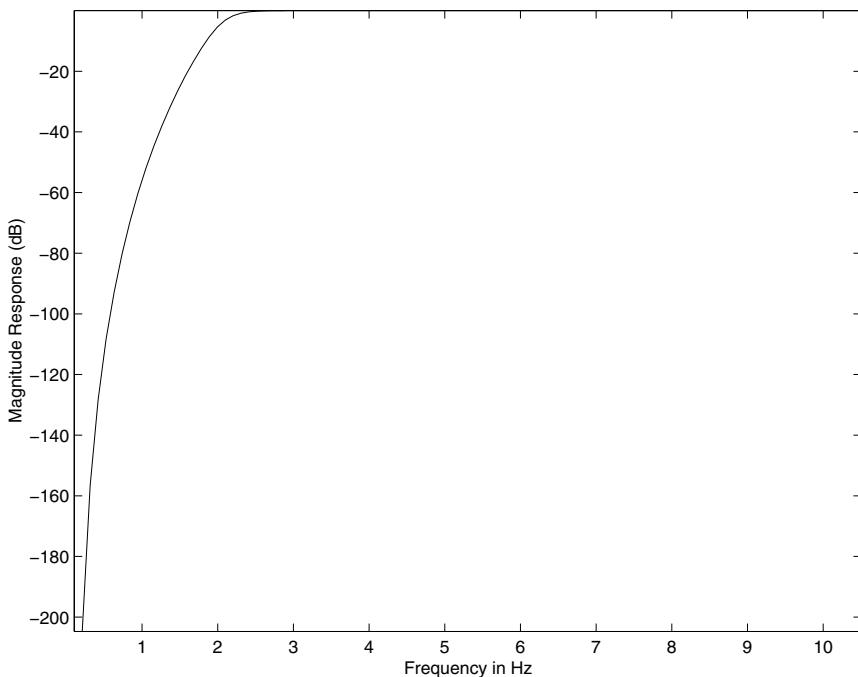


Figure 3.69 Frequency response of an eighth-order Butterworth highpass filter with cutoff frequency = 2 Hz. $f_s = 1,000$ Hz. The frequency response is shown on an expanded scale for the range 0 – 10 Hz only.

In order to improve the notch filter, specifically, to make the notch in the frequency response narrower or sharper, four poles were incorporated in the filter. The frequencies of the poles were set at 59.95 Hz and 60.05 Hz; the radii of the poles were set at 0.4 to place them well within the unit circle in the z -plane. The coordinates of the poles in the z -plane are $-0.1230 \pm j 0.3806$ and $-0.1242 \pm j 0.3802$. The frequency response of the modified notch filter is shown in Figure 3.77 (solid line). It is evident that the notch is sharper or narrower for the modified filter as compared to the response of the filter with only two zeros; however, the gain increases for higher frequencies. As a result, the output of the filter, shown in Figure 3.79, has more noise than the previous result in Figure 3.78.

An illustration of the application of the comb filter is provided in Section 3.12.

3.7 Order statistic filters

The class of filters based on order statistics includes several nonlinear filters that are useful in filtering different types of noise in signals [18, 183, 184]. The first step in order-statistic filtering is to arrange in rank order, usually from the minimum

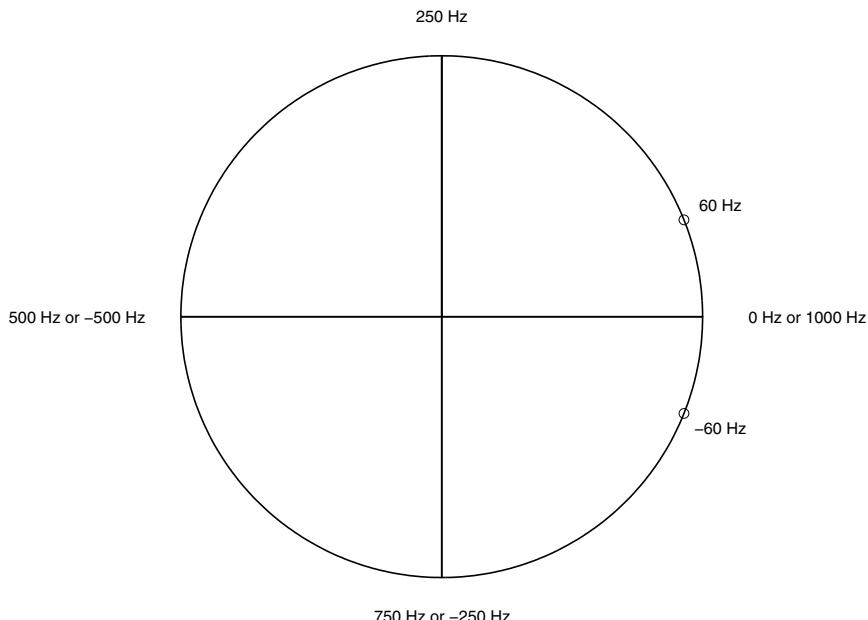


Figure 3.70 Zeros of the notch filter (in the z -domain) to remove 60 Hz interference, the sampling frequency being 1,000 Hz .

to the maximum, the values of the signal in a moving window positioned at the current sample being processed. Then, the i^{th} entry in the list is the output of the i^{th} order-statistic filter. A few commonly used order-statistic filters are defined in the following list.

- *Min filter*: the first entry in the rank-ordered list, useful in removing high-valued impulsive noise.
- *Max filter*: the last entry in the rank-ordered list, useful in removing low-valued impulsive noise.
- *Min/Max filter*: sequential application of the Min and Max filters, useful in removing impulsive noise of both of the types mentioned above.
- *Median filter*: the entry in the middle of the list. The median filter is the most popular and commonly used filter among the order-statistic filters.
- *α -trimmed mean filter*: the mean of a reduced or trimmed list, where the first $\alpha \times 100\%$ and the last $\alpha \times 100\%$ of the entries in the original list are removed, with $0 \leq \alpha < 0.5$. Outliers, which are samples with values substantially different from the rest of the samples in the list, are rejected by the trimming process. A value close to but less than 0.5 for α leads to the rejection of the entire list

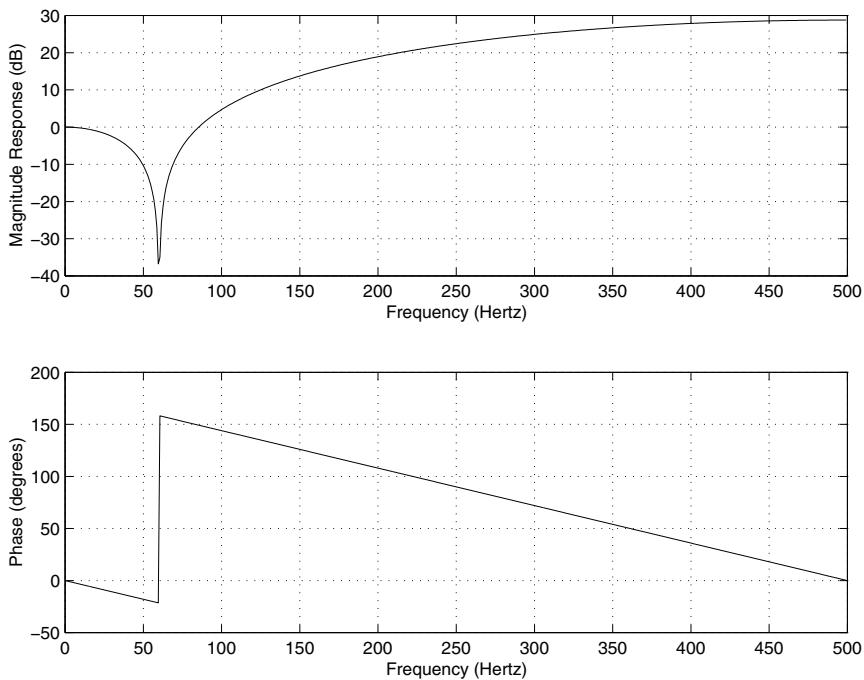


Figure 3.71 Magnitude and phase responses of the 60 Hz notch filter with zeros as shown in Figure 3.70. $f_s = 1,000$ Hz.

except the median or a few values close to the median; the output is then close to or equal to that of the median filter. The mean of the trimmed list provides a compromise between the generic mean and median filters.

- *L-filters:* a weighted combination of all of the elements in the rank-ordered list. Appropriate weights can provide outputs equivalent to those of all of the filters listed above, and facilitate the design of several nonlinear filters based on order statistics.

The methods described above may be used to realize several types of linear, nonlinear, nonstationary, and adaptive filters for the removal of various kinds of noise. It should be noted that nonlinear filters are not amenable to analysis using the Fourier transform.

Illustrations of application: Figure 3.80 (a) shows a synthesized test signal with a rectangular pulse; part (b) of the same figure shows the test signal contaminated with simulated impulsive (shot) noise. The results of filtering the noisy signal using the mean and median with filter length $M = 3$ samples are shown in parts (c) and (d) of Figure 3.80, respectively. The mean filter has blurred the edges of the pulse, which is undesirable; it has also created variations or details within the pulse that are artifacts. Furthermore, the result of the mean filter has retained the noise spikes

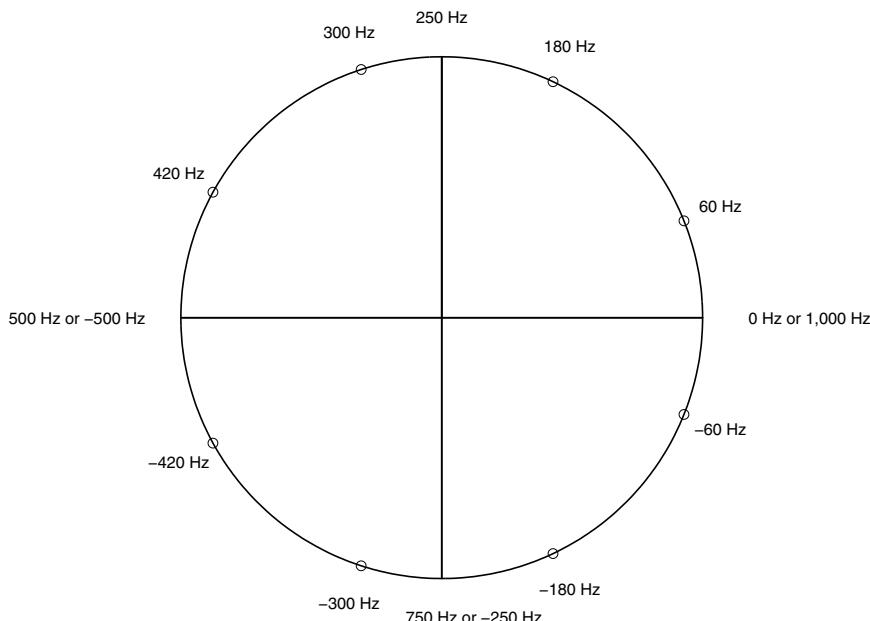


Figure 3.72 Zeros of the comb filter (in the z -domain) to remove 60 Hz interference with odd harmonics; the sampling frequency is 1,000 Hz .

present near the beginning and end of the signal, albeit reduced in amplitude. The median filter has completely eliminated the noise without distorting the pulse.

Figure 3.81 (a) shows another synthesized test signal with two rectangular pulses, the first one having a width of only two samples. Part (b) of the same figure shows the test signal degraded with uniformly distributed noise. The results of filtering the noisy signal using the mean and median with filter length $M = 5$ samples are shown in parts (c) and (d) of Figure 3.81, respectively. The mean filter has reduced the noise, but has also blurred the edges of the pulses; furthermore, the amplitude of the first short pulse has been reduced. Whereas the median filter has removed the noise to some extent without distorting the edges of the long pulse, it has obliterated the short pulse. The examples demonstrate the need to choose an appropriate type and length of the filter in accordance with the nature and strength of the noise as well as the signal.

Figure 3.82 shows a noisy ECG signal ($f_s = 1,000 \text{ Hz}$) and the results of filtering with the mean and median using a causal sliding window including nine samples. Both of the filters have substantially reduced the noise without causing any noticeable distortion.

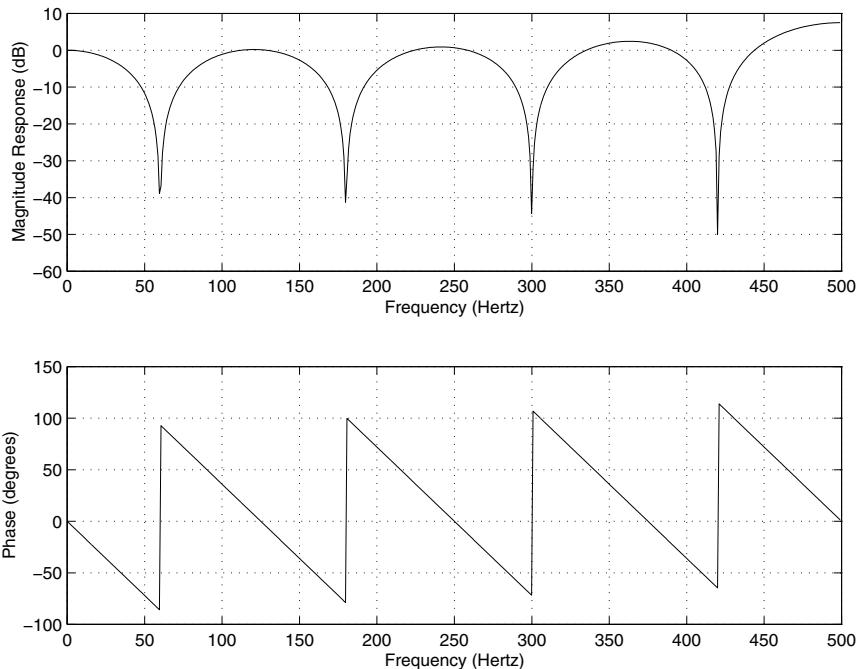


Figure 3.73 Magnitude and phase responses of the comb filter with zeros as shown in Figure 3.72.

3.8 Optimal Filtering: The Wiener Filter

The filters described in the preceding sections can take into account only limited information about the temporal or spectral characteristics of the signal and noise processes. They are often labeled as *ad hoc* filters: One may have to try several filter parameters and settle upon the filter that appears to provide a usable result. The output is not guaranteed to be the best achievable result: It is not optimized in any sense.

Problem: Design an optimal filter to remove noise from a signal, given that the signal and noise processes are independent, stationary, random processes. You may assume that the “desired” or ideal characteristics of the uncorrupted signal are known. The noise characteristics may also be assumed to be known.

Solution: Wiener filter theory provides for *optimal* filtering by taking into account the statistical characteristics of the signal and noise processes. The filter parameters are *optimized* with reference to a *performance criterion*. The output is guaranteed to be the best achievable result under the conditions imposed and the information provided. The Wiener filter is a powerful tool that changed traditional approaches to signal processing.

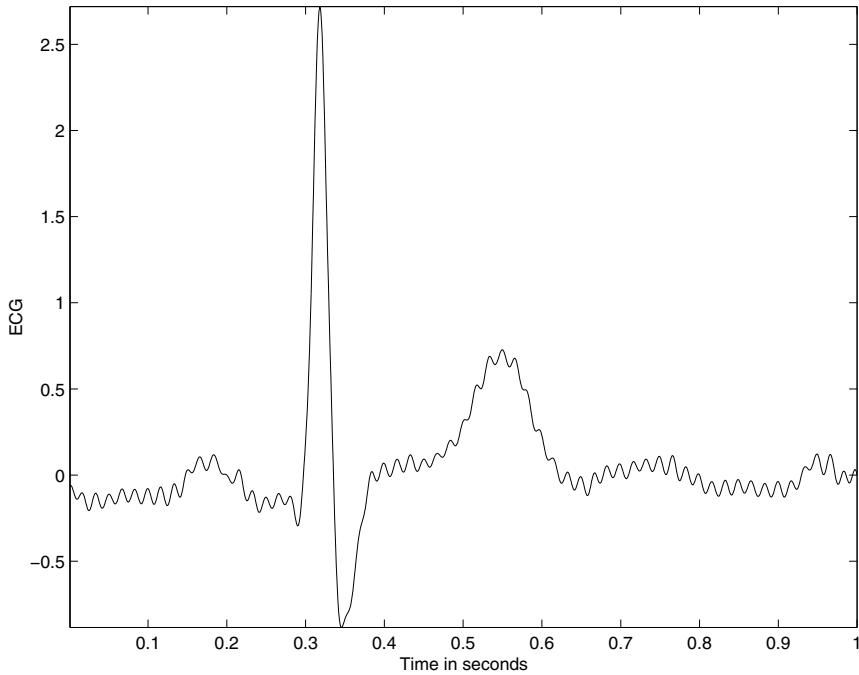


Figure 3.74 ECG signal with 60 Hz interference.

Considering the application of filtering a biomedical signal to remove noise, let us limit ourselves to a single-input, single-output, FIR filter with real input signal values and real coefficients. Figure 3.83 shows the general signal-flow diagram of a transversal filter with coefficients or tap weights w_i , $i = 0, 1, 2, \dots, M - 1$, input $x(n)$, and output $\tilde{d}(n)$ [148]. The output is usually considered to be an estimate of some “desired” signal $d(n)$ that represents the ideal, uncorrupted signal, and is, therefore, indicated as $\tilde{d}(n)$. If we assume, for the moment, that the desired signal is available, we could compute the *estimation error* between the output and the desired signal as

$$e(n) = d(n) - \tilde{d}(n). \quad (3.152)$$

Since $\tilde{d}(n)$ is the output of a linear FIR filter, it can be expressed as the convolution of the input $x(n)$ with the tap-weight sequence w_i (which is also the impulse response of the filter) as

$$\tilde{d}(n) = \sum_{k=0}^{M-1} w_k x(n-k). \quad (3.153)$$

For easier handling of the optimization procedures, the tap-weight sequence may be written as an $M \times 1$ *tap-weight vector*

$$\mathbf{w} = [w_0, w_1, w_2, \dots, w_{M-1}]^T, \quad (3.154)$$

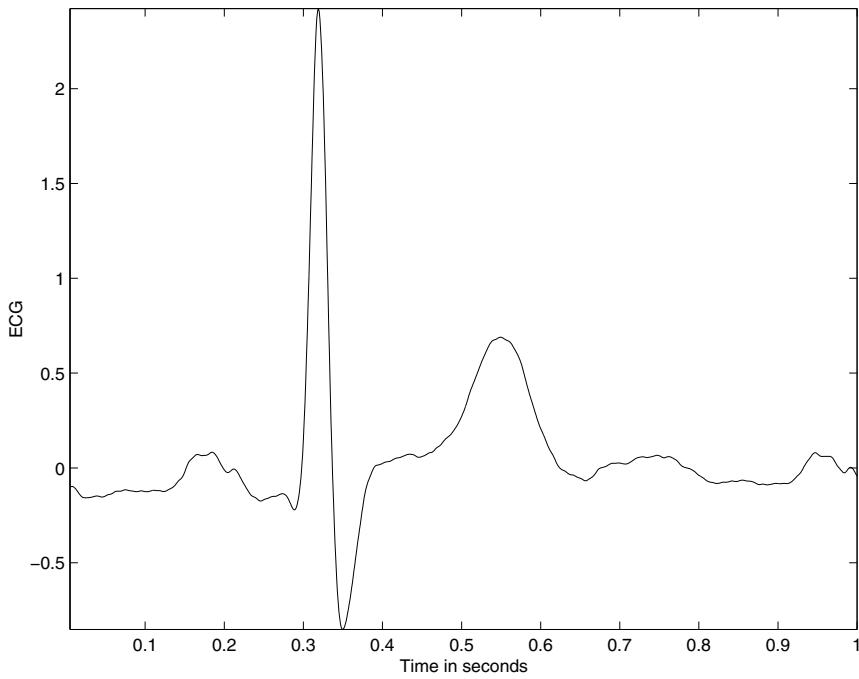


Figure 3.75 The ECG signal in Figure 3.74 after filtering with the 60 Hz notch filter shown in Figures 3.70 and 3.71.

where the boldfaced character \mathbf{w} represents a vector and the superscript T indicates vector transposition. As the tap weights are combined with M values of the input in the convolution expression, we could also write the M input values as an $M \times 1$ vector:

$$\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-M+1)]^T. \quad (3.155)$$

Note that the vector $\mathbf{x}(n)$ varies with time: At a given instant n , the vector contains the current input sample $x(n)$ and the preceding $(M-1)$ input samples from $x(n-1)$ to $x(n-M+1)$. The convolution expression in Equation 3.153 may now be written in a simpler form as the inner or dot product of the vectors \mathbf{w} and $\mathbf{x}(n)$:

$$\tilde{d}(n) = \mathbf{w}^T \mathbf{x}(n) = \mathbf{x}^T(n) \mathbf{w} = \langle \mathbf{x}, \mathbf{w} \rangle. \quad (3.156)$$

The estimation error is then given by

$$e(n) = d(n) - \mathbf{w}^T \mathbf{x}(n). \quad (3.157)$$

Wiener filter theory estimates the tap-weight sequence that minimizes the MS value of the estimation error; the output could then be called the *minimum mean-squared error* (MMSE) estimate of the desired response, the filter then being an

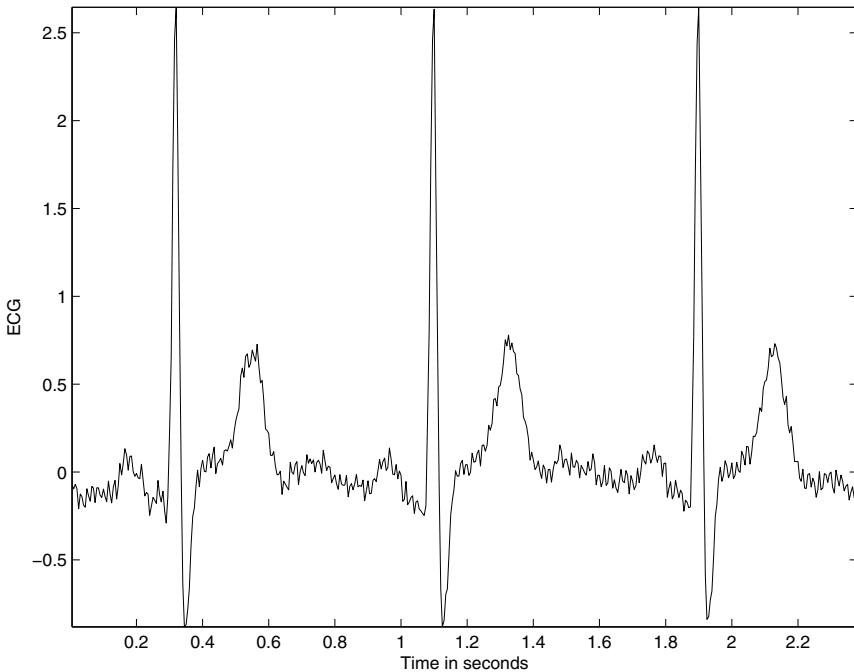


Figure 3.76 An ECG signal with power-line interference at 60 Hz .

optimal filter. The mean-squared error (MSE) is used to define the performance criterion as

$$\begin{aligned}
 J(\mathbf{w}) &= E[e^2(n)] \\
 &= E[\{d(n) - \mathbf{w}^T \mathbf{x}(n)\}\{d(n) - \mathbf{x}^T(n)\mathbf{w}\}] \\
 &= E[d^2(n)] - \mathbf{w}^T E[\mathbf{x}(n)d(n)] - E[d(n)\mathbf{x}^T(n)]\mathbf{w} \\
 &\quad + \mathbf{w}^T E[\mathbf{x}(n)\mathbf{x}^T(n)]\mathbf{w}.
 \end{aligned} \tag{3.158}$$

Note that the expectation operator is not applicable to \mathbf{w} as it is not a random variable.

Under the assumption that the input vector $\mathbf{x}(n)$ and the desired response $d(n)$ are jointly stationary, the statistical expectation expressions in the equation given above have the following interpretations [148]:

- $E[d^2(n)]$ is the variance of $d(n)$, written as σ_d^2 , with the further assumption that the mean of $d(n)$ is zero.
- $E[\mathbf{x}(n)d(n)]$ is the cross-correlation between the input vector $\mathbf{x}(n)$ and the desired response $d(n)$, which is an $M \times 1$ vector:

$$\Theta = E[\mathbf{x}(n)d(n)]. \tag{3.159}$$

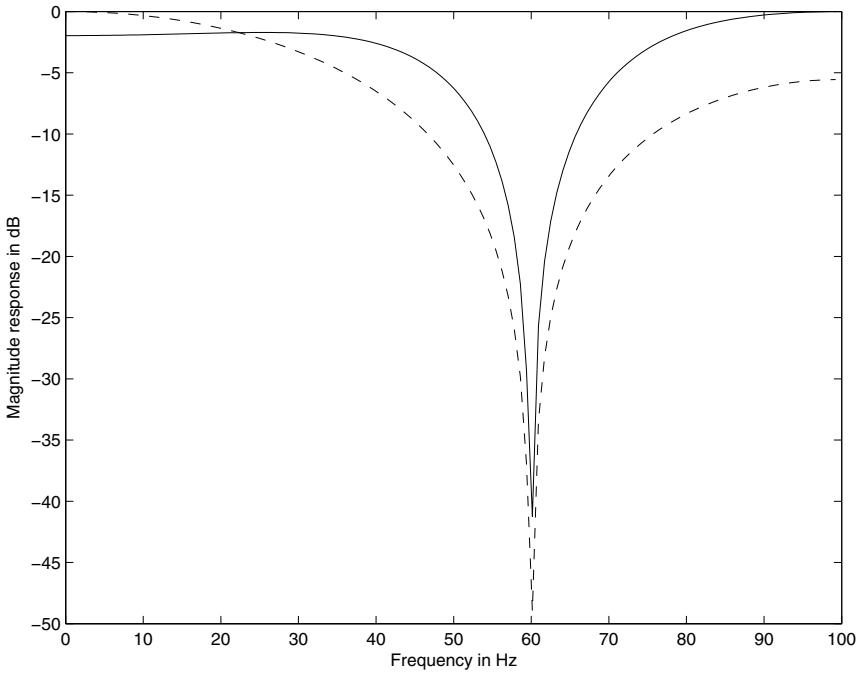


Figure 3.77 Dashed line: frequency response (magnitude, in dB) of a notch filter with two zeros at 60 Hz; $f_s = 200$ Hz. Solid line: response of the filter with the inclusion of four poles; see the text for details.

Note that $\Theta = [\theta(0), \theta(-1), \dots, \theta(1 - M)]^T$, where

$$\theta(-k) = E[x(n - k)d(n)], \quad k = 0, 1, 2, \dots, M - 1. \quad (3.160)$$

- $E[d(n)\mathbf{x}^T(n)]$ is the transpose of $E[\mathbf{x}(n)d(n)]$; therefore,

$$\Theta^T = E[d(n)\mathbf{x}^T(n)]. \quad (3.161)$$

- $E[\mathbf{x}(n)\mathbf{x}^T(n)]$ represents the autocorrelation of the input vector $\mathbf{x}(n)$ computed as the outer product of the vector with itself, written as

$$\Phi = E[\mathbf{x}(n)\mathbf{x}^T(n)] \quad (3.162)$$

or in its full $M \times M$ matrix form as

$$\Phi = \begin{bmatrix} \phi(0) & \phi(1) & \cdots & \phi(M - 1) \\ \phi(-1) & \phi(0) & \cdots & \phi(M - 2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(-M + 1) & \phi(-M + 2) & \cdots & \phi(0) \end{bmatrix} \quad (3.163)$$

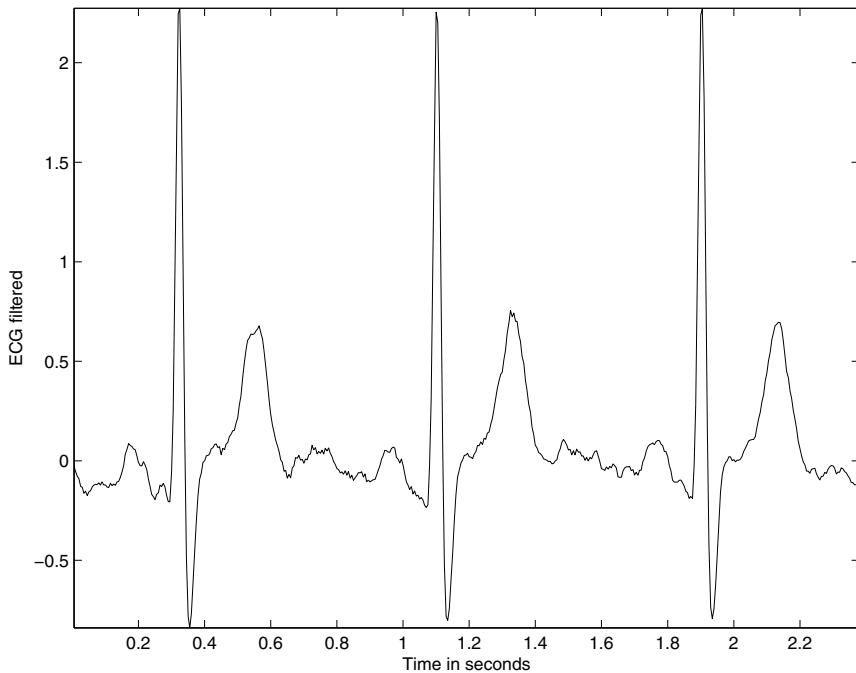


Figure 3.78 The result of filtering the noisy signal in Figure 3.76 using the notch filter with two zeros at 60 Hz.

with the element in row k and column i given by

$$\phi(i - k) = E[x(n - k)x(n - i)], \quad (3.164)$$

with the property that $\phi(i - k) = \phi(k - i)$. (Note: $\phi = \phi_{xx}$.) With the assumption of wide-sense stationarity, the $M \times M$ matrix Φ is completely specified by M values of the autocorrelation $\phi(0), \phi(1), \dots, \phi(M - 1)$ for lags $0, 1, \dots, M - 1$.

With the interpretations as listed above, the MSE expression in Equation 3.158 is simplified to

$$J(\mathbf{w}) = \sigma_d^2 - \mathbf{w}^T \boldsymbol{\Theta} - \boldsymbol{\Theta}^T \mathbf{w} + \mathbf{w}^T \boldsymbol{\Phi} \mathbf{w}. \quad (3.165)$$

This expression indicates that the MSE is a second-order function of the tap-weight vector \mathbf{w} . To determine the optimal tap-weight vector, denoted by \mathbf{w}_o , we could differentiate $J(\mathbf{w})$ with respect to \mathbf{w} , set it to zero, and solve the resulting equation. To perform this differentiation, we should note the following derivatives:

$$\frac{d}{d\mathbf{w}}(\boldsymbol{\Theta}^T \mathbf{w}) = \boldsymbol{\Theta},$$

$$\frac{d}{d\mathbf{w}}(\mathbf{w}^T \boldsymbol{\Theta}) = \boldsymbol{\Theta},$$

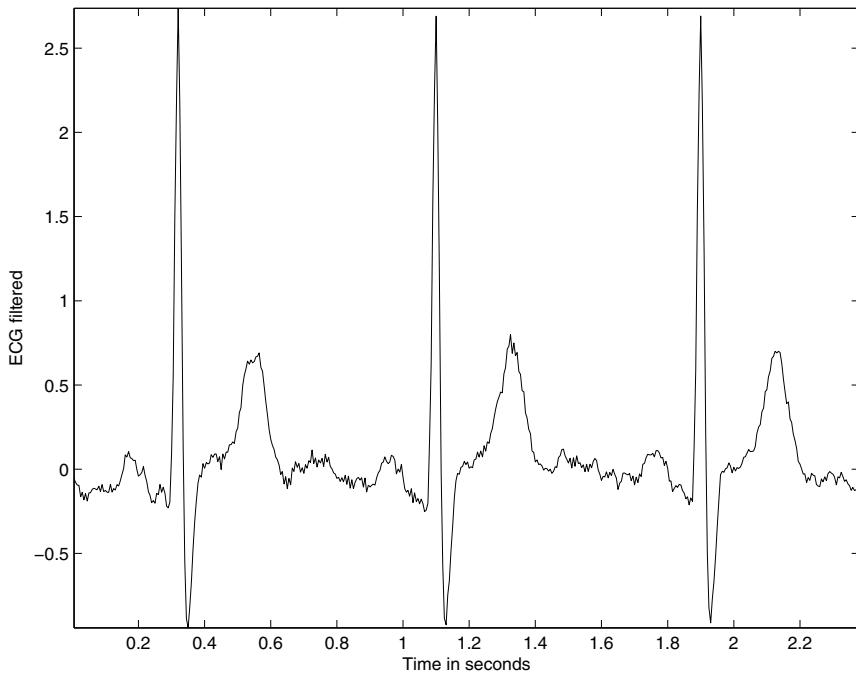


Figure 3.79 The result of filtering the noisy signal in Figure 3.76 using the notch filter with two zeros and four poles. Compare with the result in Figure 3.78.

$$\frac{d}{d\mathbf{w}}(\mathbf{w}^T \Phi \mathbf{w}) = 2\Phi\mathbf{w}.$$

Now, we obtain the derivative of $J(\mathbf{w})$ with respect to \mathbf{w} as

$$\frac{dJ(\mathbf{w})}{d\mathbf{w}} = -2\Theta + 2\Phi\mathbf{w}. \quad (3.166)$$

Setting this expression to zero, we obtain the condition for the optimal filter as

$$\Phi\mathbf{w}_o = \Theta. \quad (3.167)$$

This equation is known as the *Wiener-Hopf* equation. It is also known as the *normal equation* as it can be shown that [148], for the optimal filter, each element of the input vector $\mathbf{x}(n)$ and the estimation error $e(n)$ are mutually orthogonal and, furthermore, that the filter output $\tilde{d}(n)$ and the error $e(n)$ are mutually orthogonal (that is, the expectation of their products is zero). The optimal filter is obtained as

$$\mathbf{w}_o = \Phi^{-1} \Theta. \quad (3.168)$$

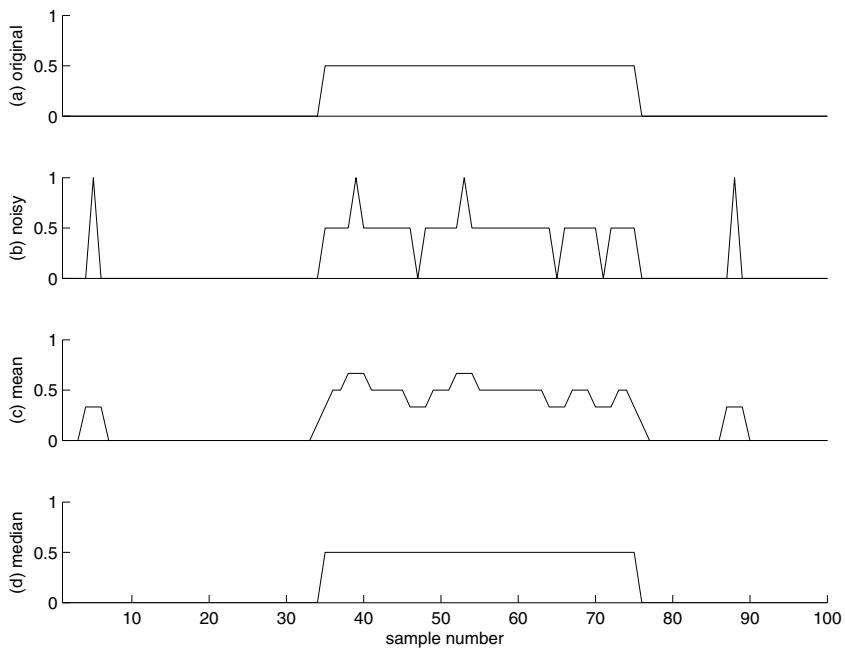


Figure 3.80 (a) A synthesized test signal with a rectangular pulse. (b) Test signal contaminated with simulated impulsive or shot noise. Result of filtering the noisy signal using (c) the mean and (d) the median with a sliding window having $M = 3$ samples.

In expanded form, we have the Wiener–Hopf equation as

$$\begin{bmatrix} \phi(0) & \phi(1) & \cdots & \phi(M-1) \\ \phi(-1) & \phi(0) & \cdots & \phi(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(-M+1) & \phi(-M+2) & \cdots & \phi(0) \end{bmatrix} \begin{bmatrix} w_{o0} \\ w_{o1} \\ \vdots \\ w_{o(M-1)} \end{bmatrix} = \quad (3.169)$$

$$\begin{bmatrix} \theta(0) \\ \theta(-1) \\ \vdots \\ \theta(1-M) \end{bmatrix}$$

or as

$$\sum_{i=0}^{M-1} w_{oi} \phi(i-k) = \theta(-k), \quad k = 0, 1, 2, \dots, M-1. \quad (3.170)$$

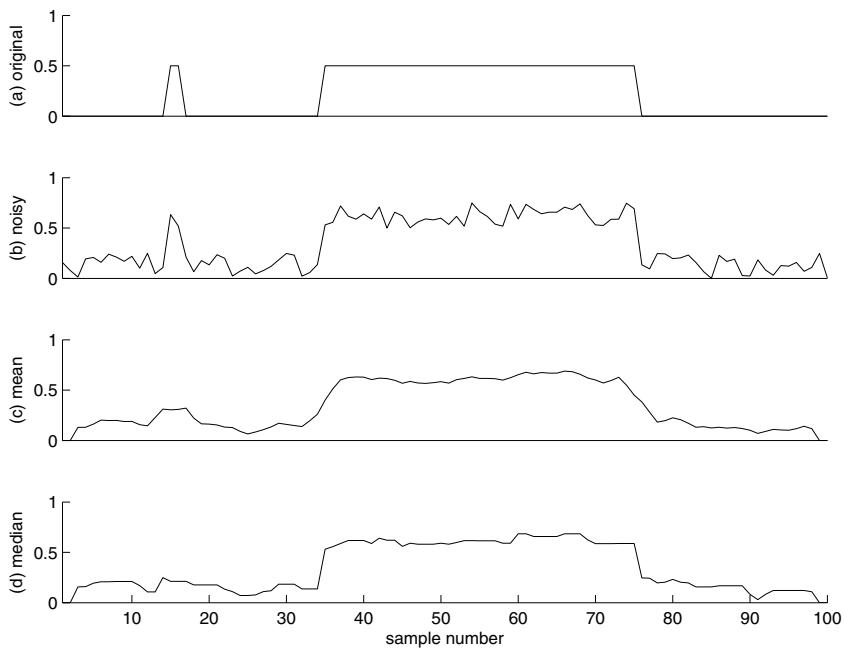


Figure 3.81 (a) A synthesized test signal with two rectangular pulses. (b) Degraded signal with uniformly distributed noise. Result of filtering the degraded signal using (c) the mean and (d) the median operation with a sliding window having $M = 5$ samples.

The minimum MSE is given by

$$J_{\min} = \sigma_d^2 - \boldsymbol{\Theta}^T \boldsymbol{\Phi}^{-1} \boldsymbol{\Theta}. \quad (3.171)$$

Given the condition that the signals involved are stationary, we have $\phi(i - k) = \phi(k - i)$ and $\theta(-k) = \theta(k)$, that is, the functions ϕ and θ are even-symmetric. Then, we may write Equation 3.170 as

$$\sum_{i=0}^{M-1} w_{oi} \phi(k - i) = \theta(k), \quad k = 0, 1, 2, \dots, M - 1. \quad (3.172)$$

Thus, we have the convolution relationship

$$w_{ok} * \phi(k) = \theta(k). \quad (3.173)$$

Applying the Fourier transform to the equation given above, we get

$$W(\omega)S_{xx}(\omega) = S_{xd}(\omega), \quad (3.174)$$

which may be modified to obtain the Wiener filter frequency response $W(\omega)$ as

$$W(\omega) = \frac{S_{xd}(\omega)}{S_{xx}(\omega)}, \quad (3.175)$$

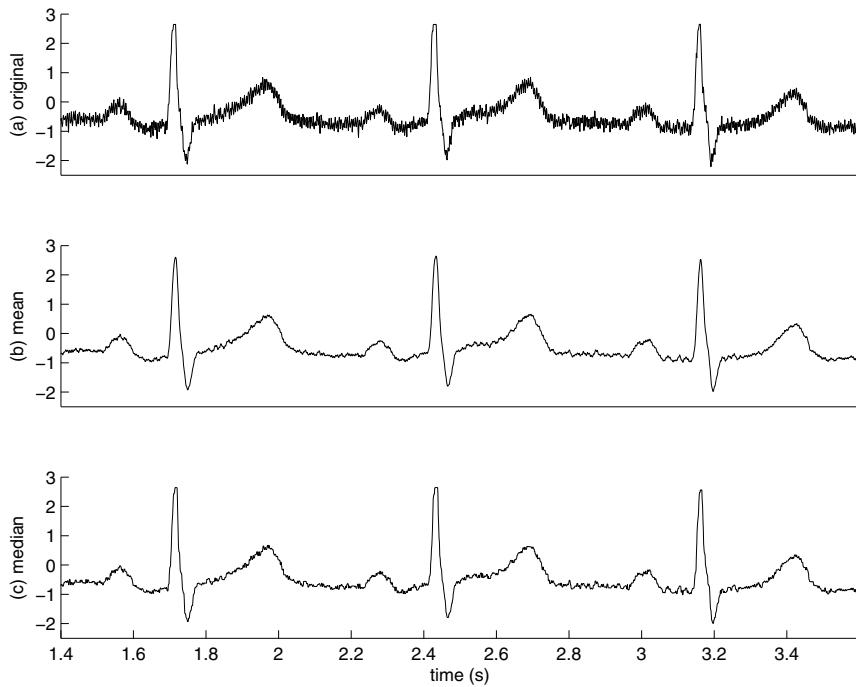


Figure 3.82 (a) An ECG signal with noise. Result of filtering the ECG signal using (b) the mean and (c) the median operation with a sliding window having $M = 9$ samples.

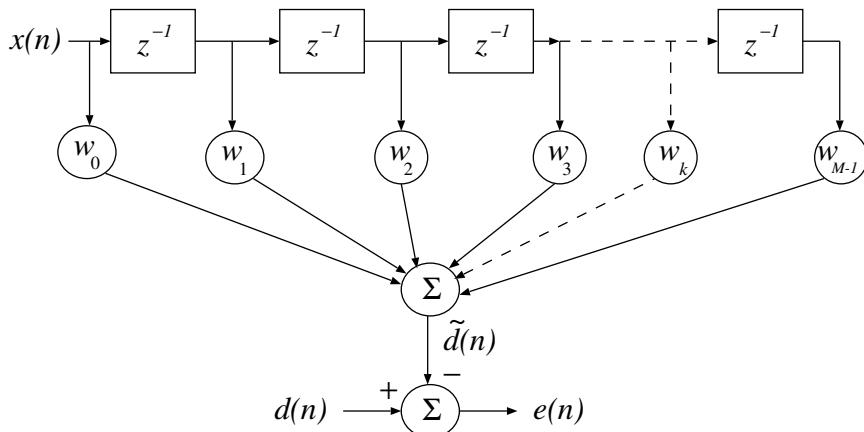


Figure 3.83 Signal-flow diagram of the Wiener filter.

where $S_{xx}(\omega)$ is the PSD of the input signal and $S_{xd}(\omega)$ is the cross-spectral density (CSD) between the input signal and the desired signal.

Note that derivation of the optimal filter requires specific knowledge about the input and the desired response in the form of the autocorrelation Φ of the input $x(n)$ and the cross-correlation Θ between the input $x(n)$ and the desired response $d(n)$. In practice, although the desired response $d(n)$ may not be known, it should be possible to obtain an estimate of its temporal or spectral statistics, which may be used to estimate Θ . Proper estimation of the statistical entities mentioned above requires a large number of samples of the corresponding signals.

{Note: Haykin [148] allows all the entities involved to be complex. Vector transposition T is then Hermitian or complex-conjugate transposition H . Products of two entities require one to be conjugated: for example, $e^2(n)$ is obtained as $e(n)e^*(n)$, and Equation 3.153 will have w_k^* in place of w_k . Furthermore, $\frac{d}{dw}(\Theta^H w) = \mathbf{0}$ and $\frac{d}{dw}(w^H \Theta) = 2\Theta$. The final Wiener–Hopf equation, however, simplifies to the same as above in Equation 3.170.}

Let us now consider the problem of removing noise from a corrupted input signal. For this case, let the input $x(n)$ contain a mixture of the desired (original) signal $d(n)$ and noise $\eta(n)$, that is,

$$x(n) = d(n) + \eta(n). \quad (3.176)$$

Using the vector notation as before, we have

$$\mathbf{x}(n) = \mathbf{d}(n) + \boldsymbol{\eta}(n), \quad (3.177)$$

where $\boldsymbol{\eta}(n)$ is the vectorial representation of the noise function $\eta(n)$. The autocorrelation matrix of the input is given by

$$\Phi = E[\mathbf{x}(n)\mathbf{x}^T(n)] = E[\{\mathbf{d}(n) + \boldsymbol{\eta}(n)\}\{\mathbf{d}(n) + \boldsymbol{\eta}(n)\}^T]. \quad (3.178)$$

If we assume that the noise process is statistically independent of the signal process and that at least one of the processes has a mean value of zero, we have

$$E[\mathbf{d}(n)\boldsymbol{\eta}^T(n)] = E[\boldsymbol{\eta}^T(n)\mathbf{d}(n)] = \mathbf{0}. \quad (3.179)$$

(This result is valid also if the two processes are mutually orthogonal.) Then,

$$\Phi = E[\mathbf{d}(n)\mathbf{d}^T(n)] + E[\boldsymbol{\eta}(n)\boldsymbol{\eta}^T(n)] = \Phi_d + \Phi_\eta, \quad (3.180)$$

where Φ_d and Φ_η are the $M \times M$ autocorrelation matrices of the signal and noise, respectively. Furthermore,

$$\Theta = E[\mathbf{x}(n)d(n)] = E[\{\mathbf{d}(n) + \boldsymbol{\eta}(n)\}d(n)] = E[\mathbf{d}(n)d(n)] = \Phi_{1d}, \quad (3.181)$$

where Φ_{1d} is an $M \times 1$ autocorrelation vector of the desired signal. The optimal Wiener filter is then given by

$$\mathbf{w}_o = (\Phi_d + \Phi_\eta)^{-1} \Phi_{1d}. \quad (3.182)$$

The frequency response of the Wiener filter may be obtained by modifying Equation 3.175 by taking into account the spectral relationships

$$S_{xx}(\omega) = S_d(\omega) + S_\eta(\omega) \quad (3.183)$$

and

$$S_{xd}(\omega) = S_d(\omega), \quad (3.184)$$

which leads to

$$W(\omega) = \frac{S_d(\omega)}{S_d(\omega) + S_\eta(\omega)} = \frac{1}{1 + \frac{S_\eta(\omega)}{S_d(\omega)}}, \quad (3.185)$$

where $S_d(\omega)$ and $S_\eta(\omega)$ are the PSDs of the desired signal and the noise process, respectively. Note that designing the optimal filter requires knowledge of the PSDs of the desired signal and the noise process (or models thereof).

From Equation 3.185, the following important properties of the Wiener filter are evident: $W(\omega) = 0$ wherever $S_d(\omega) = 0$, $W(\omega) = 1$ wherever $S_\eta(\omega) = 0$, and $W(\omega)$ decreases as $S_\eta(\omega)$ increases. Therefore, the Wiener filter has the following characteristics:

- frequency components that are not present in the input (that is, have a value of zero) will not be modified or restored,
- the input signal's frequency components are passed with unit gain at frequencies where noise does not exist, and
- the gain of the Wiener filter decreases as the SNR , expressed as a function of frequency, decreases.

Illustrations of application: The upper trace in Figure 3.84 shows one ECG cycle extracted from the signal with noise in Figure 3.5. A piecewise linear model of the desired version of the signal was created by concatenating linear segments to provide P, QRS, and T waves with amplitudes, durations, and intervals similar to those in the given noisy signal. The baseline of the model was set to zero. The noise-free model is shown in the middle trace of Figure 3.84. The log PSDs of the given noisy signal and the noise-free model, the latter being $S_d(\omega)$ in Equation 3.185, are shown in the upper two plots of Figure 3.85.

The T-P intervals between successive cardiac cycles in an ECG (the interbeat intervals) may be taken to represent the isoelectric baseline. Then, any activity present in these intervals constitutes noise. Four T-P intervals were selected from the noisy signal in Figure 3.5, and their Fourier power spectra were averaged to derive the noise PSD $S_\eta(\omega)$ required in the Wiener filter (Equation 3.185). The estimated log PSD of the noise is shown in the third trace of Figure 3.85. Observe the relatively high levels of power in the noise PSD above 100 Hz as compared to the PSDs of the original noisy signal and the model. Observe also the peaks in the original and noise PSDs near 180 Hz, 300 Hz, and 420 Hz, representing the third, fifth, and seventh harmonics of 60 Hz, respectively; the peak at 460 Hz is an aliased version

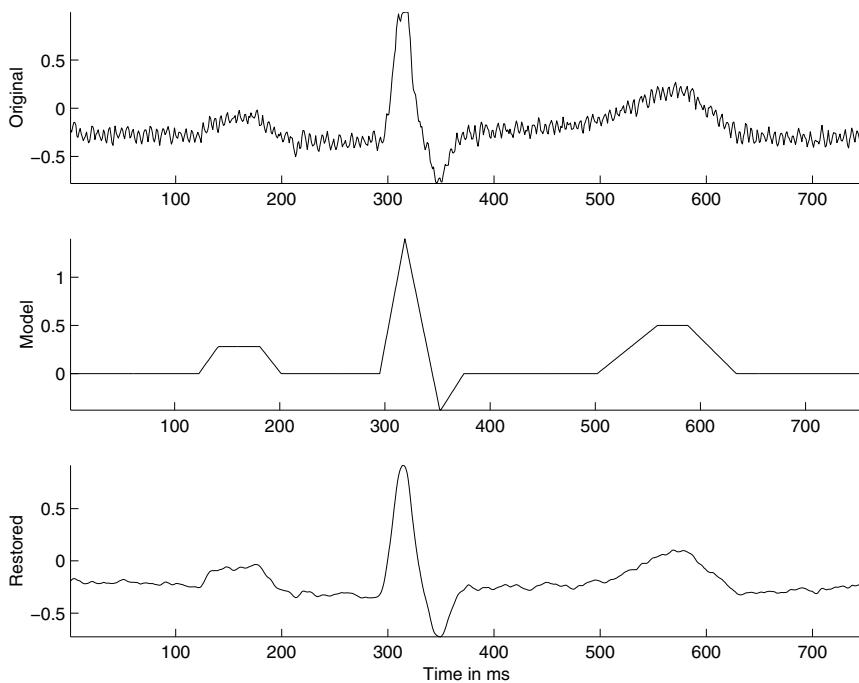


Figure 3.84 From top to bottom: one cycle of the noisy ECG signal in Figure 3.5 (labeled as Original); a piecewise linear model of the desired noise-free signal (Model); and the output of the Wiener filter (Restored).

of the ninth harmonic at 540 Hz . The 60 Hz component itself appears to have been suppressed by a notch filter in the signal acquisition system. (See Sections 3.3.4 and 3.6.3 for details.)

The Wiener filter frequency response was derived as in Equation 3.185, and is shown in the fourth plot in Figure 3.85. Observe the low gain of the filter near 180 Hz , 300 Hz , 420 Hz , and 460 Hz corresponding to the peaks in the noise spectrum. As indicated by Equation 3.185, the Wiener filter gain is inversely related to the noise PSD and directly related to the signal PSD. The result of application of the Wiener filter to the given signal is shown in the third trace of Figure 3.84. It is evident that almost all of the noise has been effectively removed by the filter. This is also confirmed by comparing the first and last plots in Figure 3.85: The spectrum of the restored signal has values below -50 dB for frequencies greater than 100 Hz .

The most important point to observe here is that the filter was derived with models of the noise and signal processes (PSDs), which were obtained from the given signal itself in the present application. No cutoff frequency was required to be specified in designing the Wiener filter, whereas the Butterworth filter requires the specification of a cutoff frequency and filter order.

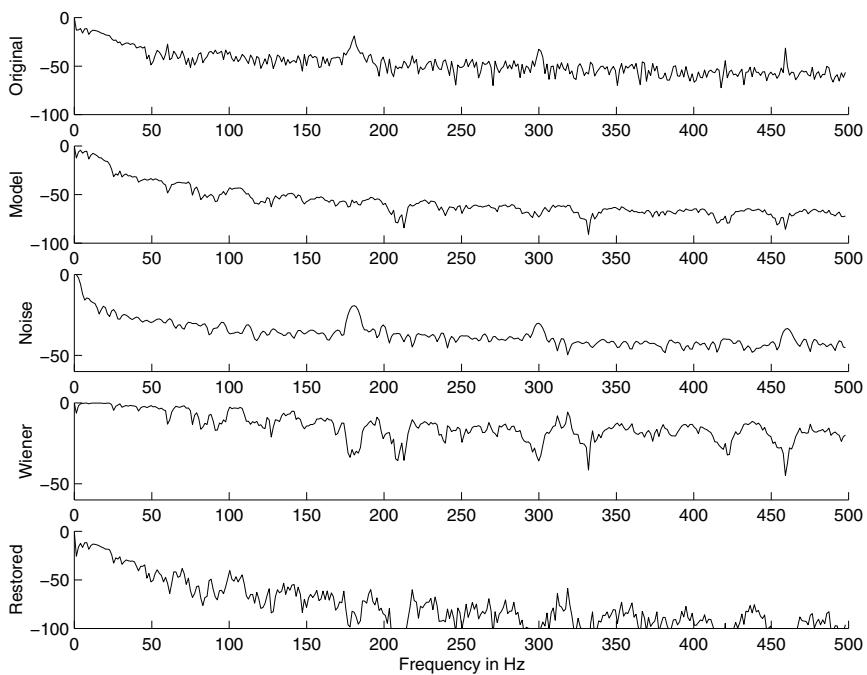


Figure 3.85 From top to bottom: log PSD (in dB) of the given noisy signal (labeled as Original); log PSD of the noise-free model (Model); estimated log PSD of the noise process (Noise); log frequency response of the Wiener filter (Wiener); and log PSD of the filter output (Restored). The scaling of the vertical axis is not the same for all PSDs in the figure.

In another experiment, a noise-free ECG signal was obtained from a subject under controlled conditions; a one-second-long segment of this signal was selected to derive the desired signal PSD (model) for use in the design of the Wiener filter. Subsequently, a noisy ECG was recorded from the same subject by including artifacts related to motion and contraction of the limbs. The sampling frequency used was 200 Hz . Ten segments of noise were selected from the noisy ECG in the T-P intervals and their average PSD was computed. The impulse response of the Wiener filter was derived by taking the inverse Fourier transform of the frequency response derived according to Equation 3.185.

Figure 3.86 shows one-second-long traces of the model ECG, the noisy ECG, and the filtered result. Also shown in the same figure are the ACFs of the model ECG and noise, as well as the impulse response of the Wiener filter. The impulse response of the Wiener filter, which possesses even symmetry due to the real-valued nature of the related frequency response, was truncated by selecting only 10 samples on either side of its maximal value (which occurs at zero time but appears at 0.5 s in Figure 3.86 due to the shift applied for causality), resulting in the filter order of $M = 21$. The coefficients obtained as above were used to filter the noisy signal in the time domain

(over a duration of about two minutes). It is evident from the last two traces of Figure 3.86 that the noise has been effectively suppressed without any noticeable degradation of the ECG wave components. Figure 3.87 shows longer traces of the noisy and filtered ECG signals. Close inspection of the results indicates that, while relatively high-frequency noise has been suppressed, noise within the bandwidth of the ECG signal remains in the output. Furthermore, low-frequency artifacts remain for the same reason, along with the fact that the short segments of noise used to model the noise PSD do not effectively incorporate the low-frequency artifacts in the model. The PSDs and frequency response of the Wiener filter shown in Figure 3.88 confirm the observations made above for the present example as well as those in the previous illustration.

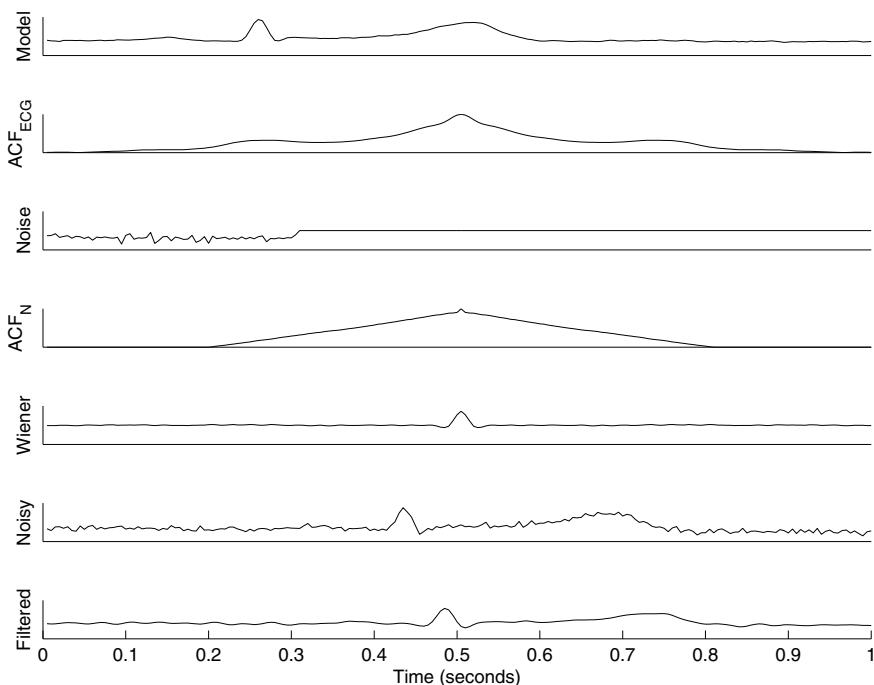


Figure 3.86 From top to bottom: one cycle of the noise-free ECG of a subject (labeled as Model); the ACF of the noise-free ECG; a sample segment of noise from a noisy ECG of the same subject (the actual noise segment has a duration of 0.3 s but has been padded with zeros to the same length as the ECG, which is 1 s); the ACF of the noise (ACF_N) obtained using 10 segments; the impulse response of the Wiener filter (shifted for causality); a segment of the noisy ECG to be filtered; and the corresponding filtered result. The amplitude labels have been suppressed to prevent clutter. ECG signal data courtesy of Emily Marasco and Matthew LaRocque, University of Calgary. See also Figure 3.87.

Most signal acquisition systems should permit the measurement of at least the variance or power level of the noise present. A uniform (white) PSD model may

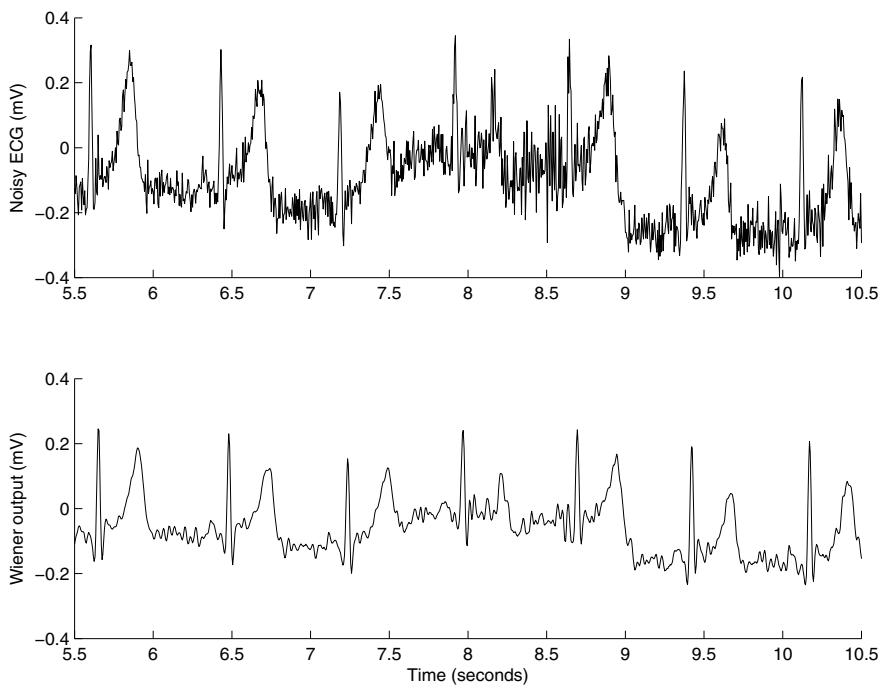


Figure 3.87 From top to bottom: a segment of the noisy ECG to be filtered and the corresponding filtered result. ECG signal data courtesy of Emily Marasco and Matthew LaRocque, University of Calgary. See also Figure 3.86.

then be easily derived. Models of the ideal signal and the noise processes may also be created using parametric Gaussian or Laplacian models either in the time domain (ACF) or directly in the frequency domain (PSD). However, it should be observed that the noise PSDs in Figures 3.85 and 3.88 are not flat or white, and that the noise ACF in Figure 3.86 is not an impulse. These characteristics indicate that the noise process is not random but has some underlying structure.

3.9 Adaptive Filters for Removal of Interference

Filters with fixed characteristics (tap weights or coefficients), as seen in the preceding sections, are suitable when the characteristics of the signal and noise (random or structured) are stationary and known. Design of frequency-domain filters requires detailed knowledge of the spectral contents of the signal and noise. Such filters are not applicable when the characteristics of the signal and/or noise vary with time, that is, when they are nonstationary. They are also not suitable when the spectral contents of the signal and the interference overlap significantly.

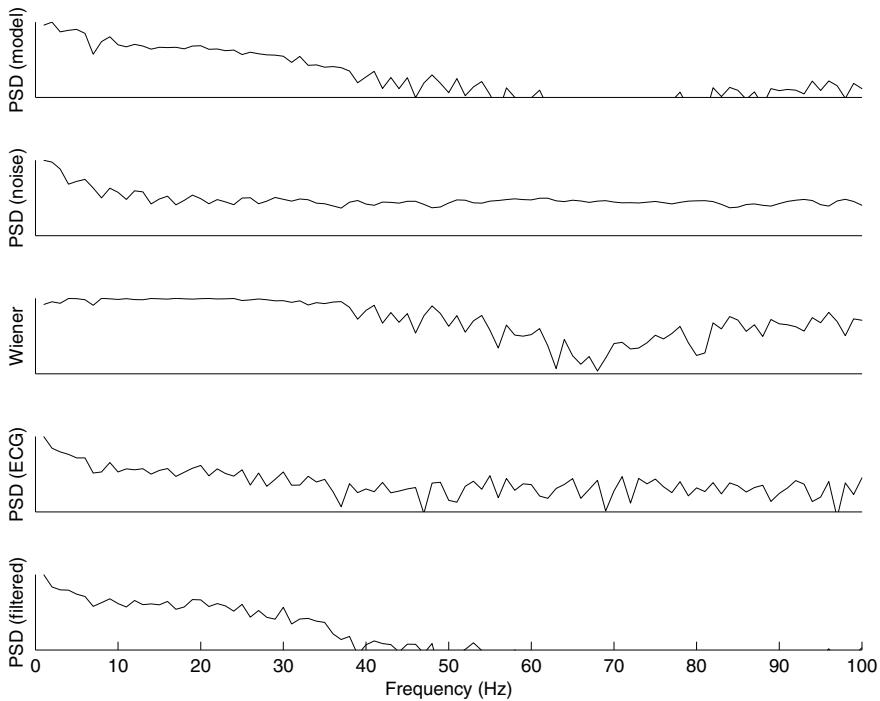


Figure 3.88 From top to bottom: log PSD (in dB) of the noise-free ECG (model); log PSD of the noise obtained using 10 segments; log frequency response of the Wiener filter; log PSD of the noisy ECG; and log PSD of the filtered result. The display of all PSDs is limited to the range $[-50, 0]$ dB . $f_s = 200$ Hz . The PSDs of ECG signals were obtained using only one cardiac cycle of each as shown in Figure 3.86.

Consider the situation when two ECG signals such as those of a fetus and the expectant mother, or two vibration signals such as the VAG and the VMG, arrive at the recording site and get added in some proportion. The spectra of the signals in the mixture span the same or similar frequency ranges, and hence fixed filtering cannot separate them. In the case of the VAG/VMG mixture, it is also possible for the spectra of the signals to vary from one point in time to another, due to changes in the characteristics of the cartilage surfaces causing the VAG signal, and due to the effect of variations in the recruitment of the motor units of a muscle on the VMG signal. Such a situation calls for the use of a filter that can learn and adapt to the characteristics of the interference, estimate the interfering signal, and remove it from the mixture to obtain the desired signal. This requires the filter to adjust automatically its impulse response (and hence its frequency response) as the characteristics of the signal and/or noise vary. (Several methods for the analysis of nonstationary and multicomponent signals are presented in Chapter 8.)

Problem: Design an optimal filter to remove a nonstationary interference from a nonstationary signal. An additional channel of information related to the interference is available for use. The filter should continuously adapt to the changing characteristics of the signal and interference.

Solution: We need to address two different concerns in this problem:

1. The filter should be *adaptive*; the tap-weight vector of the filter will then vary with time. The principles of the adaptive filter, also known as the adaptive noise canceler (ANC), are explained in Section 3.9.1.
2. The filter should be *optimal*. Two well-established methods for optimization of the adaptive filter are presented in Sections 3.9.2 and 3.9.3.

Illustrations of the application of the methods are presented at the end of Sections 3.9.2 and 3.9.3 as well as at the end of the chapter, in Sections 3.13 and 3.14.

3.9.1 The adaptive noise canceler

Figure 3.89 shows a generic block diagram of an adaptive filter or ANC [124, 185]. The “primary input” to the filter $x(n)$ is a mixture of the signal of interest $v(n)$ and the “primary noise” $m(n)$:

$$x(n) = v(n) + m(n). \quad (3.186)$$

$x(n)$ is the primary observed signal; it is desired that the interference or noise $m(n)$ be estimated and removed from $x(n)$ in order to obtain the signal of interest $v(n)$. (The notation and terminology used in the present section, while being different from those in the preceding sections, are commonly used in the literature on adaptive filters.) It is assumed that $v(n)$ and $m(n)$ are uncorrelated. Adaptive filtering requires a second input, known as the “reference input” $r(n)$, that is uncorrelated with the signal of interest $v(n)$ but closely related to or correlated with the interference or noise $m(n)$ in some manner that need not be known. The ANC filters or modifies the reference input $r(n)$ to obtain a signal $y(n)$ that is as close to the noise $m(n)$ as possible. $y(n)$ is then subtracted from the primary input to estimate the desired signal:

$$\tilde{v}(n) = e(n) = x(n) - y(n). \quad (3.187)$$

Let us now analyze the function of the filter. Let us assume that the signal of interest $v(n)$, the primary noise $m(n)$, the reference input $r(n)$, and the primary noise estimate $y(n)$ are statistically stationary and have zero means. (Note: The requirement of stationarity is removed later when the expectations are computed in moving windows.) We have already stated that $v(n)$ is uncorrelated with $m(n)$ and $r(n)$, and that $r(n)$ is correlated with $m(n)$. The output of the ANC is

$$\begin{aligned} e(n) &= x(n) - y(n) \\ &= v(n) + m(n) - y(n), \end{aligned} \quad (3.188)$$

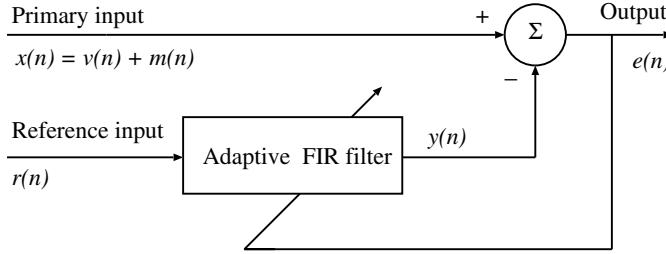


Figure 3.89 Block diagram of a generic ANC or adaptive filter.

where $y(n) = \tilde{m}(n)$ is the estimate of the primary noise obtained at the output of the adaptive FIR filter. By taking the square and expectation (statistical average) of both sides of Equation 3.188, we obtain

$$E[e^2(n)] = E[v^2(n)] + E[\{m(n) - y(n)\}^2] + 2E[v(n)\{m(n) - y(n)\}]. \quad (3.189)$$

Since $v(n)$ is uncorrelated with $m(n)$ and $y(n)$ and all of them have zero means, we have

$$E[v(n)\{m(n) - y(n)\}] = E[v(n)]E[m(n) - y(n)] = 0. \quad (3.190)$$

Equation 3.189 can be rewritten as

$$E[e^2(n)] = E[v^2(n)] + E[\{m(n) - y(n)\}^2]. \quad (3.191)$$

With reference to Figure 3.89, note that the output $e(n)$ is used (fed back) to control the adaptive FIR filter. In ANC applications, the objective is to obtain an output $e(n)$ that is a least-squares fit to the desired signal $v(n)$. This is achieved by feeding the output back to the adaptive FIR filter and adjusting the filter to minimize the total system output power. The system output serves as the error signal for the adaptive process.

The signal power $E[v^2(n)]$ will be unaffected as the filter is adjusted to minimize $E[e^2(n)]$; accordingly, the minimum output power is

$$\min E[e^2(n)] = E[v^2(n)] + \min E[\{m(n) - y(n)\}^2]. \quad (3.192)$$

As the filter is adjusted so that $E[e^2(n)]$ is minimized, $E[\{m(n) - y(n)\}^2]$ is also minimized. Thus, the filter output $y(n)$ is the MMSE estimate of the primary noise $m(n)$. Moreover, when $E[\{m(n) - y(n)\}^2]$ is minimized, $E[\{e(n) - v(n)\}^2]$ is also minimized, since from Equation 3.188, we have

$$e(n) - v(n) = m(n) - y(n). \quad (3.193)$$

Adjusting or adapting the filter to minimize the total output power is, therefore, equivalent to causing the *output e(n) to be the MMSE estimate of the signal of interest v(n)* for the given structure and adjustability of the adaptive FIR filter and for the given reference input.

The output $e(n)$ will contain the signal of interest $v(n)$ and some noise. From Equation 3.193, the output noise is given by $e(n) - v(n) = \tilde{v}(n) - v(n) = m(n) - y(n)$. Because minimizing $E[e^2(n)]$ minimizes $E[\{m(n) - y(n)\}^2]$, *minimizing the total output power minimizes the output noise power*. Since the signal component $v(n)$ in the output remains unaffected, *minimizing the total output power maximizes the output SNR*.

From Equation 3.191, we have the condition that the output power is minimum when $E[e^2(n)] = E[v^2(n)]$. When this condition is achieved, $E[\{m(n) - y(n)\}^2] = 0$. We then have $y(n) = m(n)$ and $e(n) = v(n)$; that is, the output is a perfect and noise-free estimate of the desired signal.

Optimization of the filter may be performed by expressing the error in terms of the tap-weight vector and applying the procedure of choice. The output $y(n)$ of the adaptive filter (see Figure 3.89) in response to its input $r(n)$ is given by

$$y(n) = \sum_{k=0}^{M-1} w_k r(n-k), \quad (3.194)$$

where w_k , $k = 0, 1, 2, \dots, M-1$, are the tap weights, and M is the order of the filter. The estimation error $e(n)$ or the output of the ANC system is

$$e(n) = x(n) - y(n). \quad (3.195)$$

For the sake of notational simplicity, let us define the tap-weight vector at time n as

$$\mathbf{w}(n) = [w_0(n), w_1(n), \dots, w_{M-1}(n)]^T. \quad (3.196)$$

Similarly, the input vector at each time instant n may be defined as the M -dimensional vector

$$\mathbf{r}(n) = [r(n), r(n-1), \dots, r(n-M+1)]^T. \quad (3.197)$$

Then, the estimation error $e(n)$ given in Equation 3.195 may be rewritten as

$$e(n) = x(n) - \mathbf{w}^T(n)\mathbf{r}(n). \quad (3.198)$$

It is worth noting that the derivations made above required no knowledge about the processes behind $v(n)$, $m(n)$, and $r(n)$ or their interrelationships, other than the assumptions of statistical independence between $v(n)$ and $m(n)$ and some form of correlation between $m(n)$ and $r(n)$. The arguments can be extended to situations where the primary and reference inputs contain additive random noise processes that are mutually uncorrelated and also uncorrelated with $v(n)$, $m(n)$, and $r(n)$. The procedures may also be extended to cases where $m(n)$ and $r(n)$ are deterministic or structured rather than stochastic processes, such as power-line interference, an ECG, or a VMG signal [124].

Several methods are available to maximize the output SNR ; two such methods based on the least-mean-squares (LMS) and the recursive least-squares (RLS) approaches are described in the following sections.

3.9.2 The least mean squares adaptive filter

The purpose of adaptive filtering algorithms is to adjust the tap-weight vector to minimize the MSE. By squaring the expression for the estimation error $e(n)$ given in Equation 3.198, we get

$$e^2(n) = x^2(n) - 2x(n)\mathbf{r}^T(n)\mathbf{w}(n) + \mathbf{w}^T(n)\mathbf{r}(n)\mathbf{r}^T(n)\mathbf{w}(n). \quad (3.199)$$

The squared error is a second-order (quadratic) function of the tap-weight vector (and the inputs), and may be depicted as a concave hyperparaboloidal (bowl-like) surface that is never negative. The aim of the filter optimization procedure would be to reach the bottom of the bowl-like function. Gradient-based methods may be used for this purpose.

By taking the expected values of the entities in Equation 3.199 and taking the derivative with respect to the tap-weight vector, we may derive the Wiener–Hopf equation for the present application. The LMS algorithm takes a simpler approach by assuming the square of the instantaneous error as in Equation 3.199 to stand for an estimate of the MSE [124]. The LMS algorithm is based on the method of steepest descent, where the new tap-weight vector $\mathbf{w}(n+1)$ is given by the present tap-weight vector $\mathbf{w}(n)$ plus a correction proportional to the negative of the gradient $\nabla e^2(n)$ of the squared error:

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu \nabla e^2(n). \quad (3.200)$$

The parameter μ controls the stability and rate of convergence of the algorithm: the larger the value of μ , the larger the gradient of the error that is introduced and the faster the convergence of the algorithm, and viceversa.

The LMS algorithm approximates $\nabla e^2(n)$ by the derivative of the squared error in Equation 3.199 with respect to the tap-weight vector as

$$\widetilde{\nabla e^2}(n) = -2x(n)\mathbf{r}(n) + 2\{\mathbf{w}^T(n)\mathbf{r}(n)\}\mathbf{r}(n) = -2e(n)\mathbf{r}(n). \quad (3.201)$$

Using this estimate of the gradient in Equation 3.200, we get

$$\mathbf{w}(n+1) = \mathbf{w}(n) + 2\mu e(n) \mathbf{r}(n). \quad (3.202)$$

This expression is known as the Widrow–Hoff LMS algorithm.

The advantages of the LMS algorithm lie in its simplicity and ease of implementation: Although the method is based on the MSE and gradient-based optimization, the filter expression itself is free of differentiation, squaring, or averaging. It has been shown that the expected value of the tap-weight vector provided by the LMS algorithm converges to the optimal Wiener solution when the input vectors are uncorrelated over time [124, 186]. The procedure may be started with an arbitrary tap-weight vector; it will converge in the mean and remain stable as long as μ is greater than zero but less than the reciprocal of the largest eigenvalue of the autocorrelation matrix of the reference input [124].

Illustration of application: Zhang et al. [125] used a two-stage adaptive LMS filter to cancel muscle-contraction interference from VAG signals. The first stage

was used to remove the measurement noise in the accelerometers and associated amplifiers, and the second stage was designed to cancel the muscle signal.

Zhang et al. [125] also proposed a procedure for optimization of the step size μ by using an *RMS*-error-based misadjustment factor and a time-varying estimate of the input signal power, among other entities. The LMS algorithm was implemented as

$$\mathbf{w}(n+1) = \mathbf{w}(n) + 2\mu(n) e(n) \mathbf{r}(n). \quad (3.203)$$

The step size μ was treated as a variable, its value being determined dynamically as

$$\mu(n) = \frac{\mu}{(M+1) \bar{x}^2(n) [\alpha, r(n), \bar{x}^2(n-1)]}, \quad (3.204)$$

with $0 < \mu < 1$. The notation $\bar{x}^2(n) [\alpha, r(n), \bar{x}^2(n-1)]$ indicates that the updated value of $\bar{x}^2(n)$ is a function of $\alpha, r(n)$, and $\bar{x}^2(n-1)$. A forgetting factor α was introduced in the adaptation process, with $0 \leq \alpha \ll 1$; this feature was expected to overcome problems caused by high levels of nonstationarity in the signal. $\bar{x}^2(n)$ is a time-varying estimate of the input signal power, computed as $\bar{x}^2(n) = \alpha r^2(n) + (1 - \alpha) \bar{x}^2(n-1)$.

The filtered versions of the VAG signals recorded from the midpatella and the tibial tuberosity positions, as shown in Figure 3.11 [traces (b) and (c), right-hand column], are shown in Figure 3.90. The muscle-contraction signal recorded at the distal rectus femoris position was used as the reference input [Figure 3.11, right-hand column, trace (a)]. It is seen that the low-frequency muscle-contraction artifact has been successfully removed from the VAG signals (compare the second half of each signal in Figure 3.90 with the corresponding part in Figure 3.11).

3.9.3 The RLS adaptive filter

When the input process of an adaptive system is quasistationary, the best steady-state performance results from slow adaptation. However, when the input statistics are time-variant (nonstationary), the best performance is obtained by a compromise between fast adaptation (necessary to track variations in the input process) and slow adaptation (necessary to limit the noise in the adaptive process). The LMS adaptation algorithm is a simple and efficient approach for ANC; however, it is not appropriate for fast-varying signals due to its slow convergence and due to the difficulty in selecting the correct value for the step size μ . An alternative approach based on the exact minimization of the least-squares criterion is the RLS method [148, 187]. The RLS algorithm has been widely used in real-time system identification and noise cancellation because of its fast convergence, which is about an order of magnitude higher than that of the LMS method. (The derivation of the RLS filter in this section has been adapted from Sesay [187] and Krishnan [185] with permission.)

An important feature of the RLS algorithm is that it utilizes information contained in the input data and extends it back to the instant of time when the algorithm was initiated [148]. Given the least-squares estimate of the tap-weight vector of the filter at time $n-1$, the updated estimate of the vector at time n is computed upon the arrival of new data.

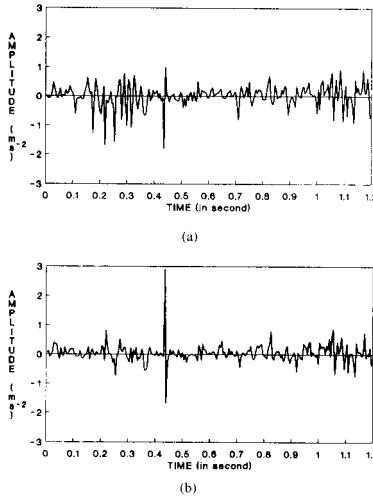


Figure 3.90 LMS-filtered versions of the VAG signals recorded from the midpatella and the tibial tuberosity positions, as shown in Figure 3.11 [traces (b) and (c), right-hand column]. The muscle-contraction signal recorded at the distal rectus femoris position was used as the reference input [Figure 3.11, right-hand column, trace (a)]. The recording setup is shown in Figure 3.10. Reproduced with permission from Y.T Zhang, R.M. Rangayyan, C.B. Frank, and G.D. Bell, Adaptive cancellation of muscle-contraction interference from knee joint vibration signals, *IEEE Transactions on Biomedical Engineering*, 41(2):181–191, 1994. ©IEEE.

In the derivation of the RLS algorithm, the *performance criterion* or *objective function* $\xi(n)$ to be minimized in the sense of least squares is defined as

$$\xi(n) = \sum_{i=1}^n \lambda^{n-i} |e(i)|^2, \quad (3.205)$$

where $1 \leq i \leq n$ is the observation interval, $e(i)$ is the estimation error as defined in Equation 3.198, and λ is a weighting factor (also known as the *forgetting factor*) with $0 < \lambda \leq 1$. The values of $\lambda^{n-i} < 1$ give more “weight” to the more recent error values. Such weighting is desired in the case of nonstationary data, where changes in the signal statistics make the inclusion of past data less appropriate. The inverse of $(1 - \lambda)$ is a measure of the memory of the algorithm.

The Wiener–Hopf equation is the necessary and sufficient condition [148] for minimizing the performance index in the least-squares sense and for obtaining the optimal values of the tap weights, and may be derived in a manner similar to that presented in Section 3.8 for the Wiener filter. The normal equation to be solved in the RLS procedure is

$$\Phi(n)\tilde{\mathbf{w}}(n) = \Theta(n), \quad (3.206)$$

where $\tilde{\mathbf{w}}(n)$ is the optimal tap-weight vector for which the performance index is at its minimum, $\Phi(n)$ is an $M \times M$ time-averaged (and weighted) autocorrelation

matrix of the reference input $\mathbf{r}(i)$, defined as

$$\Phi(n) = \sum_{i=1}^n \lambda^{n-i} \mathbf{r}(i) \mathbf{r}^T(i), \quad (3.207)$$

and $\Theta(n)$ is an $M \times 1$ time-averaged (and weighted) cross-correlation matrix between the reference input $\mathbf{r}(i)$ and the primary input $x(i)$, defined as

$$\Theta(n) = \sum_{i=1}^n \lambda^{n-i} \mathbf{r}(i) x(i). \quad (3.208)$$

The general scheme of the RLS filter is illustrated in Figure 3.91.

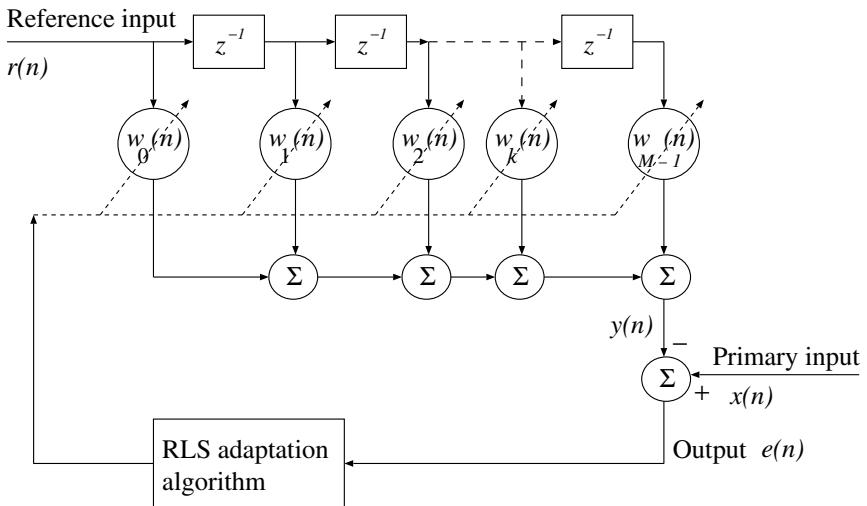


Figure 3.91 General structure of the adaptive RLS filter.

Because of the difficulty in solving the normal equation for the optimal tap-weight vector, recursive techniques need to be considered. In order to obtain a recursive solution, we could isolate the term corresponding to $i = n$ from the rest of the summation on the right-hand side (RHS) of Equation 3.207 and obtain

$$\Phi(n) = \lambda \left[\sum_{i=1}^{n-1} \lambda^{n-i-1} \mathbf{r}(i) \mathbf{r}^T(i) \right] + \mathbf{r}(n) \mathbf{r}^T(n). \quad (3.209)$$

According to the definition in Equation 3.207, the expression inside the square brackets on the RHS of Equation 3.209 equals the time-averaged and weighted autocorrelation matrix $\Phi(n-1)$. Hence, Equation 3.209 can be rewritten as a recursive expression, given by

$$\Phi(n) = \lambda \Phi(n-1) + \mathbf{r}(n) \mathbf{r}^T(n). \quad (3.210)$$

Similarly, Equation 3.208 can be written as the recursive equation

$$\Theta(n) = \lambda\Theta(n-1) + \mathbf{r}(n)x(n). \quad (3.211)$$

To compute the least-squares estimate $\tilde{\mathbf{w}}(n)$ for the tap-weight vector in accordance with Equation 3.206, we have to determine the inverse of the correlation matrix $\Phi(n)$. In practice, such an operation is time-consuming (particularly if M is large). To reduce the computational requirements, a matrix inversion lemma known as the “ABCD lemma” could be used (a similar form of the lemma can be found in Haykin [148]). According to the ABCD lemma, given matrices \mathbf{A} , \mathbf{B} , \mathbf{C} , and \mathbf{D} ,

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1}. \quad (3.212)$$

The matrices \mathbf{A} , \mathbf{C} , $(\mathbf{A} + \mathbf{BCD})$, and $(\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1})$ are assumed to be invertible. With the correlation matrix $\Phi(n)$ assumed to be positive definite and therefore nonsingular, we may apply the matrix inversion lemma to Equation 3.210 by assigning

$$\begin{aligned} \mathbf{A} &= \lambda\Phi(n-1), \\ \mathbf{B} &= \mathbf{r}(n), \\ \mathbf{C} &= 1, \\ \mathbf{D} &= \mathbf{r}^T(n). \end{aligned}$$

We then have

$$\begin{aligned} \Phi^{-1}(n) &= \lambda^{-1}\Phi^{-1}(n-1) \\ &- \lambda^{-1}\Phi^{-1}(n-1)\mathbf{r}(n) [\lambda^{-1}\mathbf{r}^T(n)\Phi^{-1}(n-1)\mathbf{r}(n) + 1]^{-1} \\ &\times \lambda^{-1}\mathbf{r}^T(n)\Phi^{-1}(n-1). \end{aligned} \quad (3.213)$$

Since the expression inside the brackets of the equation given above is a scalar, the equation can be rewritten as

$$\Phi^{-1}(n) = \lambda^{-1}\Phi^{-1}(n-1) - \frac{\lambda^{-2}\Phi^{-1}(n-1)\mathbf{r}(n)\mathbf{r}^T(n)\Phi^{-1}(n-1)}{1 + \lambda^{-1}\mathbf{r}^T(n)\Phi^{-1}(n-1)\mathbf{r}(n)}. \quad (3.214)$$

For convenience of notation, let

$$\mathbf{P}(n) = \Phi^{-1}(n), \quad (3.215)$$

with $\mathbf{P}(0) = \delta^{-1}\mathbf{I}$, where δ is a small constant and \mathbf{I} is the identity matrix. Furthermore, let

$$\mathbf{k}(n) = \frac{\lambda^{-1}\mathbf{P}(n-1)\mathbf{r}(n)}{1 + \lambda^{-1}\mathbf{r}^T(n)\mathbf{P}(n-1)\mathbf{r}(n)}. \quad (3.216)$$

$\mathbf{k}(n)$ is analogous to the *Kalman gain vector* in Kalman filter theory [148]. Equation 3.214 may then be rewritten in a simpler form as

$$\mathbf{P}(n) = \lambda^{-1}\mathbf{P}(n-1) - \lambda^{-1}\mathbf{k}(n)\mathbf{r}^T(n)\mathbf{P}(n-1). \quad (3.217)$$

By multiplying both sides of Equation 3.216 by the denominator on its RHS, we get

$$\mathbf{k}(n) [1 + \lambda^{-1} \mathbf{r}^T(n) \mathbf{P}(n-1) \mathbf{r}(n)] = \lambda^{-1} \mathbf{P}(n-1) \mathbf{r}(n), \quad (3.218)$$

or

$$\mathbf{k}(n) = [\lambda^{-1} \mathbf{P}(n-1) - \lambda^{-1} \mathbf{k}(n) \mathbf{r}^T(n) \mathbf{P}(n-1)] \mathbf{r}(n). \quad (3.219)$$

Comparing the expression inside the brackets on the RHS of the equation given above with Equation 3.217, we have

$$\mathbf{k}(n) = \mathbf{P}(n) \mathbf{r}(n). \quad (3.220)$$

$\mathbf{P}(n)$ and $\mathbf{k}(n)$ have the dimensions $M \times M$ and $M \times 1$, respectively.

By using Equations 3.206, 3.211, and 3.215, a recursive equation for updating the least-squares estimate $\tilde{\mathbf{w}}(n)$ of the tap-weight vector can be obtained as

$$\begin{aligned} \tilde{\mathbf{w}}(n) &= \Phi^{-1}(n) \Theta(n) \\ &= \mathbf{P}(n) \Theta(n) \\ &= \lambda \mathbf{P}(n) \Theta(n-1) + \mathbf{P}(n) \mathbf{r}(n) x(n). \end{aligned} \quad (3.221)$$

Substituting Equation 3.217 for $\mathbf{P}(n)$ in the first term of Equation 3.221, we get

$$\begin{aligned} \tilde{\mathbf{w}}(n) &= \mathbf{P}(n-1) \Theta(n-1) - \mathbf{k}(n) \mathbf{r}^T(n) \mathbf{P}(n-1) \Theta(n-1) \\ &\quad + \mathbf{P}(n) \mathbf{r}(n) x(n) \\ &= \Phi^{-1}(n-1) \Theta(n-1) - \mathbf{k}(n) \mathbf{r}^T(n) \Phi^{-1}(n-1) \Theta(n-1) \\ &\quad + \mathbf{P}(n) \mathbf{r}(n) x(n) \\ &= \tilde{\mathbf{w}}(n-1) - \mathbf{k}(n) \mathbf{r}^T(n) \tilde{\mathbf{w}}(n-1) + \mathbf{P}(n) \mathbf{r}(n) x(n). \end{aligned} \quad (3.222)$$

Finally, from Equation 3.220, using the fact that $\mathbf{P}(n) \mathbf{r}(n)$ equals the gain vector $\mathbf{k}(n)$, the equation given above can be rewritten as

$$\begin{aligned} \tilde{\mathbf{w}}(n) &= \tilde{\mathbf{w}}(n-1) - \mathbf{k}(n) [x(n) - \mathbf{r}^T(n) \tilde{\mathbf{w}}(n-1)] \\ &= \tilde{\mathbf{w}}(n-1) + \mathbf{k}(n) \alpha(n), \end{aligned} \quad (3.223)$$

where $\tilde{\mathbf{w}}(0) = \mathbf{0}$, and

$$\begin{aligned} \alpha(n) &= x(n) - \mathbf{r}^T(n) \tilde{\mathbf{w}}(n-1) \\ &= x(n) - \tilde{\mathbf{w}}^T(n-1) \mathbf{r}(n). \end{aligned} \quad (3.224)$$

The quantity $\alpha(n)$ is often referred to as the *a priori error*, reflecting the fact that it is the error obtained using the “old” filter (that is, the filter before being updated with the new data at the n^{th} time instant). It is evident that, in the case of ANC applications, $\alpha(n)$ will be the estimated signal of interest $\tilde{v}(n)$ after the filter has converged, that is,

$$\alpha(n) = \tilde{v}(n) = x(n) - \tilde{\mathbf{w}}^T(n-1) \mathbf{r}(n). \quad (3.225)$$

Furthermore, after convergence, the primary noise estimate, that is, the output of the adaptive filter $y(n)$, can be written as

$$y(n) = \tilde{m}(n) = \tilde{\mathbf{w}}^T(n-1)\mathbf{r}(n). \quad (3.226)$$

By substituting Equations 3.188 and 3.226 in Equation 3.225, we get

$$\begin{aligned} \tilde{v}(n) &= v(n) + m(n) - \tilde{m}(n) \\ &= v(n) + m(n) - \tilde{\mathbf{w}}^T(n-1)\mathbf{r}(n) \\ &= x(n) - \tilde{\mathbf{w}}^T(n-1)\mathbf{r}(n). \end{aligned} \quad (3.227)$$

Equation 3.223 gives a recursive relationship for obtaining the optimal values of the tap weights, which, in turn, provide the least-squares estimate $\tilde{v}(n)$ of the signal of interest $v(n)$ as in Equation 3.227.

Illustration of application: Figure 3.92 shows plots of the VAG signal of a normal subject [trace (a)] and a simultaneously recorded channel of muscle-contraction interference [labeled as MCI, trace (b)]. The characteristics of the vibration signals in this example are different from those of the signals in Figure 3.11, due to a different recording protocol in terms of speed and range of swinging motion of the leg [185]. The results of adaptive filtering of the VAG signal with the muscle-contraction interference channel as the reference are also shown in Figure 3.92: Trace (c) shows the result of LMS filtering, and trace (d) shows that of RLS filtering. A single-stage LMS filter with variable step size $\mu(n)$ as in Equation 3.204 was used; no attempt was made to remove instrumentation noise. The LMS filter used $M = 7$, $\mu = 0.05$, and a forgetting factor $\alpha = 0.98$; other values resulted in poor results. The RLS filter used $M = 7$ and $\lambda = 0.98$.

The relatively low-frequency muscle-contraction interference has been removed better by the RLS filter than by the LMS filter; the latter failed to track the nonstationarities in the interference and has caused additional artifacts in the result. The spectrograms of the primary, reference, and RLS-filtered signals are shown in Figures 3.93, 3.94, and 3.95, respectively. (The logarithmic scale is used to display better the minor differences between the spectrograms.) It is seen that the predominantly low-frequency artifact, indicated by the high energy levels at low frequencies for the entire duration in the spectrograms in Figures 3.93 and 3.94, has been removed by the RLS filter.

3.10 Selecting an Appropriate Filter

We have so far examined five approaches to remove noise and interference: (1) synchronized or ensemble averaging of multiple realizations or copies of a signal, (2) MA filtering, (3) frequency-domain filtering, (4) optimal (Wiener) filtering, and (5) adaptive filtering. The first two approaches work directly with the signal in the time domain. Frequency-domain (fixed) filtering is performed on the spectrum of the signal. Note that the impulse response of a filter designed in the frequency domain could be used to implement the filter in the time domain as an IIR or FIR filter. Furthermore,

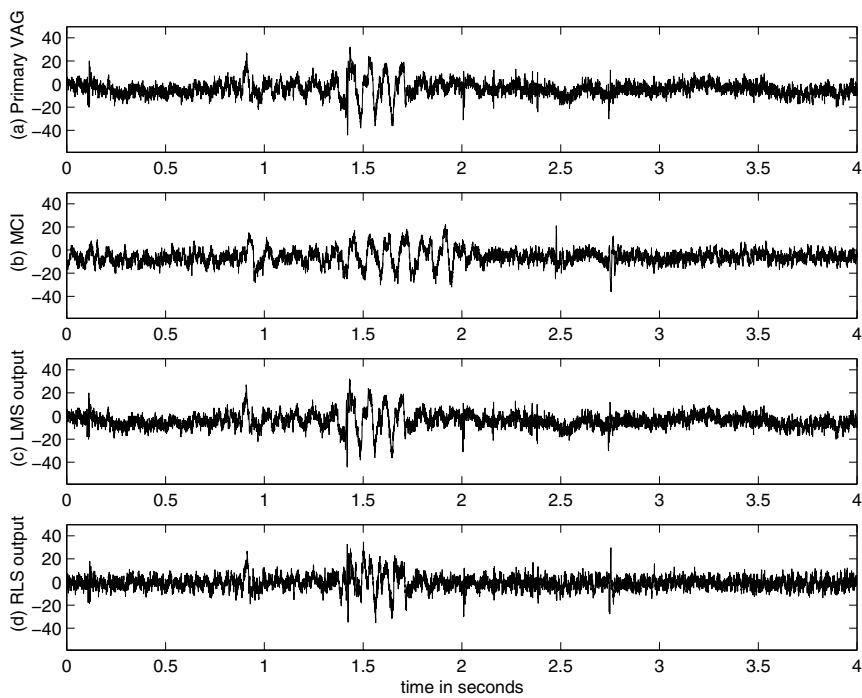


Figure 3.92 (a) VAG signal of a normal subject. (b) Muscle-contraction interference (MCI). (c) Result of LMS filtering. (d) Result of RLS filtering. The recording setup is shown in Figure 3.10.

time-domain filters may be analyzed in the frequency domain via their transfer function or frequency response to understand better their characteristics and effects on the input signal. The Wiener filter may be implemented either in the time domain as a transversal filter or in the frequency domain. Adaptive filters work directly on the signal in the time domain, but dynamically alter their characteristics in response to changes in the interference; their frequency response varies from one point in time to another.

What are the guiding principles to determine which of these filters is the best for a given application? The following points should assist in making this decision.

Synchronized or ensemble averaging is possible when:

- The signal is statistically stationary, (quasi)periodic, or cyclostationary.
- Multiple realizations or recordings of the signal of interest are available.
- A trigger point or time marker is available, or can be derived to extract and align the realizations of the signal.

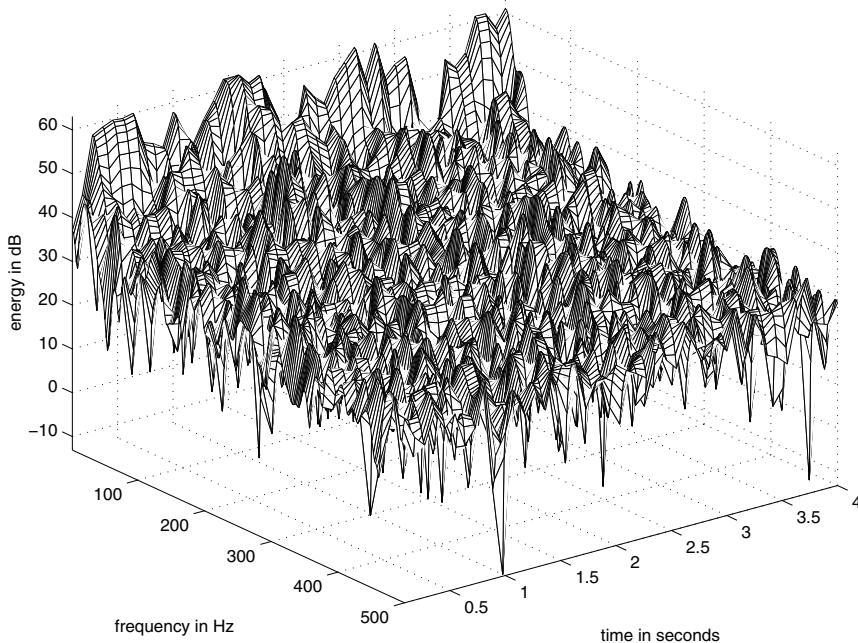


Figure 3.93 Spectrogram of the VAG signal in Figure 3.92 (a). A Hann window of length 256 samples (128 ms) was used; an overlap of 32 samples (16 ms) was allowed between adjacent segments.

- The noise is a stationary random process that is uncorrelated with the signal and has a zero mean (or a known mean).

Temporal MA filtering is suitable when:

- The signal is statistically stationary at least over the duration of the moving window.
- The noise is a zero-mean random process that is stationary at least over the duration of the moving window and is independent of the signal.
- The signal is a relatively slow (low-frequency) phenomenon.
- Fast, on-line, real-time filtering is desired.

Frequency-domain fixed filtering is applicable when:

- The signal is statistically stationary.
- The noise is a stationary random process that is statistically independent of the signal.

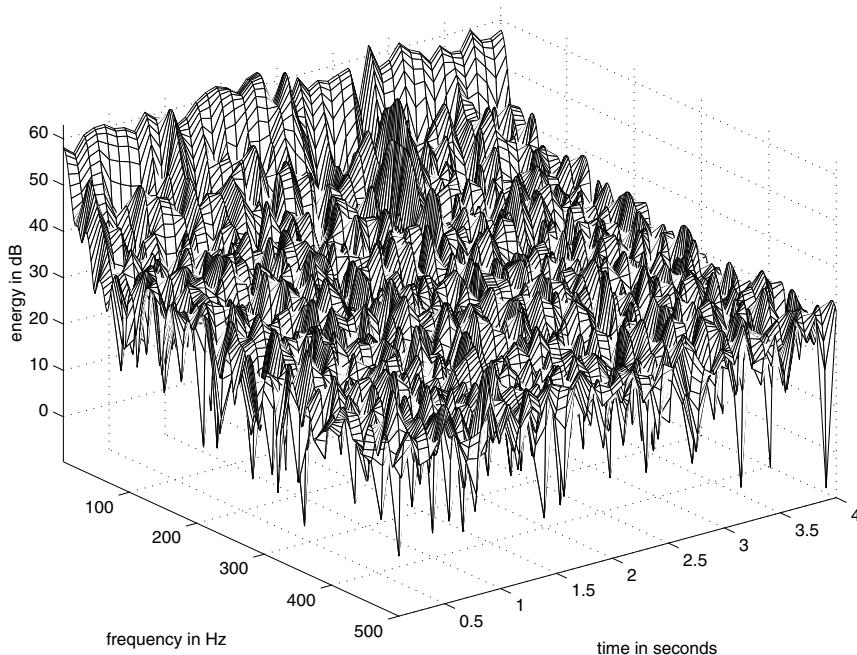


Figure 3.94 Spectrogram of the muscle-contraction interference signal in Figure 3.92 (b). A Hann window of length 256 samples (128 ms) was used; an overlap of 32 samples (16 ms) was allowed between adjacent segments.

- The signal spectrum is limited in bandwidth compared to that of the noise (or viceversa).
- Loss of information in the spectral band removed by the filter does not seriously affect the signal.
- On-line, real-time filtering is not required (if implemented in the spectral domain via the Fourier transform).

The **optimal Wiener filter** can be designed if:

- The signal is statistically stationary.
- The noise is a stationary random process that is statistically independent of the signal.
- Specific details (or models) are available regarding the ACFs or the PSDs of the signal and noise.

Adaptive filtering is called for and possible when:

- The noise or interference is not stationary and not necessarily a random process.

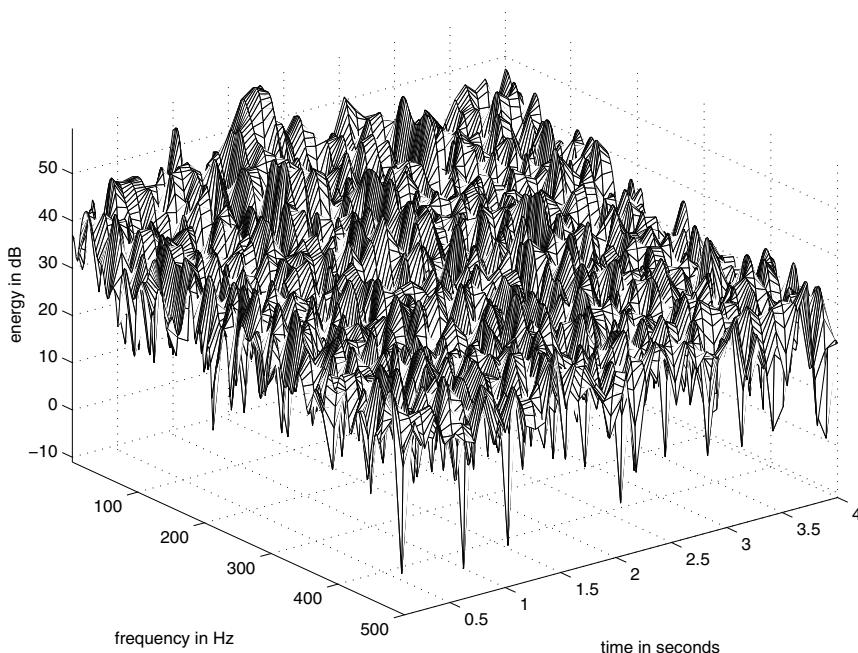


Figure 3.95 Spectrogram of the RLS-filtered VAG signal in Figure 3.92 (d). A Hann window of length 256 samples (128 ms) was used; an overlap of 32 samples (16 ms) was allowed between adjacent segments.

- The noise is uncorrelated with the signal.
- No information is available about the spectral characteristics of the signal and noise, which may also overlap significantly.
- A second source or recording site is available to obtain a reference signal that is strongly correlated with the noise but uncorrelated with the signal.

It is worth noting that an adaptive filter acts as a fixed filter when the signal and noise are stationary. An adaptive filter can also act as a notch or a comb filter when the interference is periodic. It should be noted that all of the filters mentioned above are applicable only when the noise is additive. Techniques such as homomorphic filtering (see Section 4.8) may be used as preprocessing steps if signals combined with operations other than addition need to be separated.

3.11 Application: Removal of Artifacts in ERP Signals

Problem: Propose a method to improve the SNR of ERP signals. Suggest methods to measure the improvement in the results obtained.

Solution: Kamath et al. [188] applied synchronized averaging to improve the *SNR* of cortical evoked potentials or ERPs related to electrical and mechanical stimulation of the esophagus. We have seen the details of ensemble averaging or synchronized averaging in Section 3.5.1; examples of application of synchronized averaging are shown in Figure 3.2 and 3.41. Although improvement in *SNR* was obtained in some experiments conducted by Kamath et al. [188], they also observed that habituation took place as the number of stimuli was increased beyond a certain limit, and that the use of the ERPs obtained after habituation in averaging led to a reduction in the *SNR*.

Let $y_k(n)$ represent one realization or observation of an ERP, with $k = 1, 2, \dots, M$ representing the ensemble index, and $n = 0, 1, 2, \dots, N - 1$ representing the time-sample index. Here, M is the number of realizations of the ERP available, and N is the number of the time samples in each signal. We may express the observed signal as

$$y_k(n) = x_k(n) + \eta_k(n), \quad (3.228)$$

where $x_k(n)$ represents the original uncorrupted signal and $\eta_k(n)$ represents the noise in the k^{th} realization of the observed signal. The result of ensemble averaging or synchronized averaging is obtained, for each instant of time n , by averaging the M observations of the ERP as

$$\begin{aligned} \bar{y}(n) &= \frac{1}{M} \sum_{k=1}^M y_k(n) \\ &= \frac{1}{M} \sum_{k=1}^M x_k(n) + \frac{1}{M} \sum_{k=1}^M \eta_k(n); \quad n = 0, 1, 2, \dots, N - 1. \end{aligned} \quad (3.229)$$

Figure 3.96 shows superimposed plots of all of the $M = 24$ ERPs available from the experiments conducted by Kamath et al. [188]. Whereas the overall trend of the ERPs is visible, the extent of the noise and inherent variations present in the ERPs can also be observed. Figure 3.97 shows the result of synchronized averaging of all of the 24 ERPs available. The result demonstrates the benefits of averaging in terms of reduced noise and clear depiction of the trends in the ERP signal.

To study the effects of habituation, various sets of the 24 available ERPs were averaged. Figure 3.98 shows three such results, each being the average of eight ERPs. Whereas the average of the first eight ERPs for $k = 1, 2, \dots, 8$ (solid line) appears to be a good result, the result for the next set of eight ERPs for $k = 9, 10, \dots, 16$ (dashed line) appears to be noisy. Furthermore, the average of the last set of eight ERPs for $k = 17, 18, \dots, 24$ (dotted line) is not similar to the other results of averaging at all, indicating that habituation (or fatigue) could have resulted in the last eight ERPs being substantially different from the earlier responses.

In Section 3.5.1, we saw that synchronized averaging can, theoretically, improve the *SNR* by the factor of \sqrt{M} . In a practical application, the actual improvement

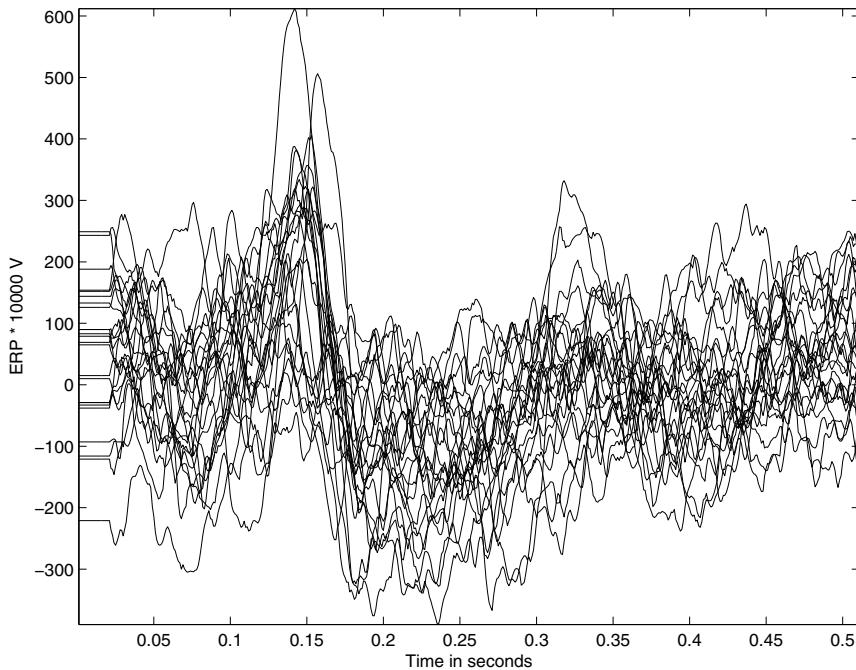


Figure 3.96 Superimposed plots of 24 cortical ERPs related to electrical stimulation of the esophagus. Data courtesy of M.V. Kamath, McMaster University, Hamilton, ON, Canada.

obtained in the *SNR* may be different. Kamath et al. [188] estimated the *SNR* by first deriving the power of the noise in the output as

$$\sigma_{\eta}^2 = \frac{1}{NT(M-1)} \sum_{k=1}^M \sum_{n=0}^{N-1} [y_k(n) - \bar{y}(n)]^2. \quad (3.230)$$

Here, $T = 0.001$ s is the sampling interval. The signal power in the output was derived as

$$\sigma_{\bar{y}}^2 = \frac{1}{NT} \left\{ \sum_{n=0}^{N-1} [\bar{y}(n)]^2 \right\} - \frac{\sigma_{\eta}^2}{M}. \quad (3.231)$$

Then, *SNR* was estimated as

$$SNR = \frac{\sigma_{\bar{y}}^2}{\sigma_{\eta}^2}. \quad (3.232)$$

Kamath et al. [188] also computed the Euclidean distance between the original ERP signals and the averaged signal obtained as

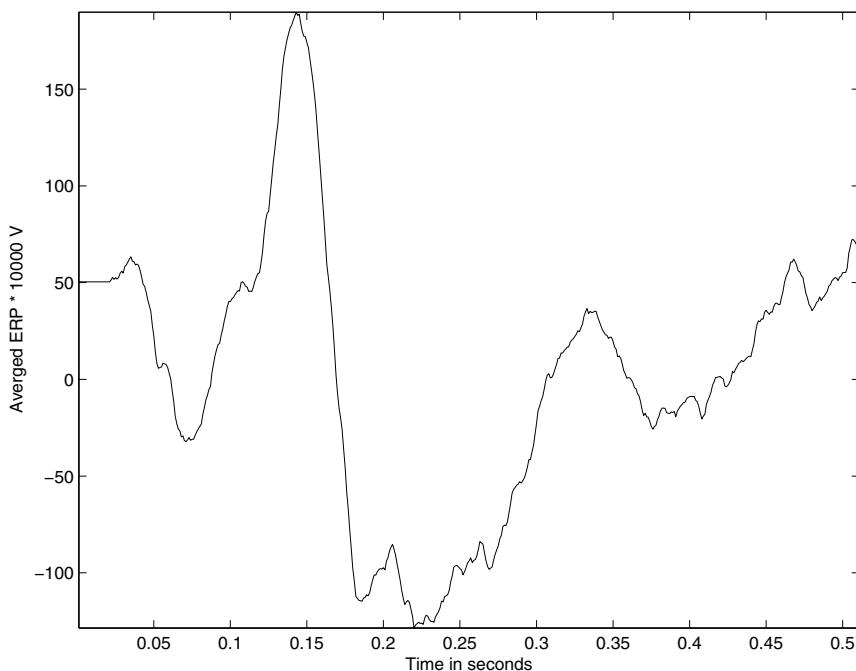


Figure 3.97 Result of synchronized averaging of all of the 24 ERPs shown in Figure 3.96.

$$D = \frac{1}{M} \sum_{k=1}^M \sqrt{\sum_{n=0}^{N-1} [y_k(n) - \bar{y}(n)]^2}. \quad (3.233)$$

The *SNR* for the average of all of the 24 ERPs, shown in Figure 3.97, was found to be 0.37. The *SNRs* for the three results of averaging eight ERPs, shown in Figure 3.98, were computed to be 0.73, 0.65, and 0.06, respectively, for the first, second, and third sets of eight ERPs. These results confirm, in a quantitative manner, the subjective observations made in the preceding paragraphs: repeated stimulation of a biological system could lead to habituation and fatigue, which could cause the response to be substantially different. Although, based on theoretical considerations, one would expect the averaging of more observations of a certain process to lead to better results with higher *SNR*, the results of the present study demonstrate otherwise. A judicious design of the experiment is essential, by taking into account various beneficial as well as adverse effects of the methods used.

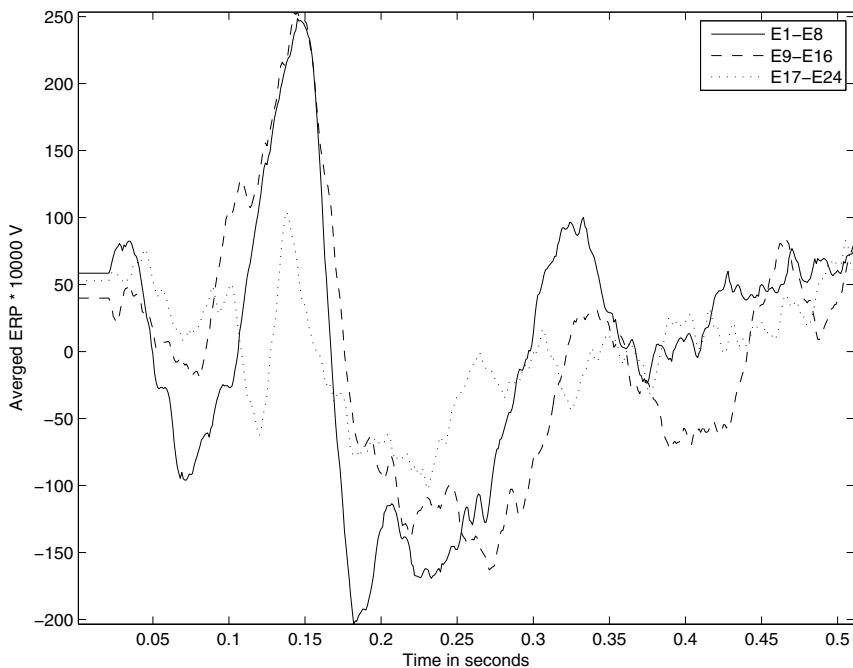


Figure 3.98 Results of synchronized averaging of selected sets of the ERPs shown in Figure 3.96. The three plots shown were obtained by averaging ERPs in groups of eight, for $k = 1, 2, \dots, 8$ (solid line); $k = 9, 10, \dots, 16$ (dashed line); and $k = 17, 18, \dots, 24$ (dotted line).

3.12 Application: Removal of Artifacts in the ECG

Problem: Figure 3.99 (top trace) shows an ECG signal with a combination of baseline drift, high-frequency noise, and power-line interference. Design filters to remove the artifacts.

Solution: The power spectrum of the given signal is shown in the topmost plot in Figure 3.100. Observe the relatively high amount of spectral energy present near DC, from 100 Hz to 500 Hz, and at the power-line frequency and its harmonics located at 60 Hz, 180 Hz, 300 Hz, and 420 Hz. The fundamental component at 60 Hz is lower than the third, fifth, and seventh harmonics, perhaps due to a notch filter included in the signal acquisition system, which has not been effective.

A Butterworth lowpass filter with order $N = 8$ and $f_c = 70$ Hz (see Section 3.6.1 and Equation 3.145), a Butterworth highpass filter of order $N = 8$ and $f_c = 2$ Hz (see Section 3.6.2 and Equation 3.148), and a comb filter with zeros at 60 Hz, 180 Hz, 300 Hz, and 420 Hz (see Section 3.6.3 and Equation 3.151) were applied in series to the signal. The signal spectrum displays the presence of further harmonics (ninth and eleventh) of the power-line interference at 540 Hz and 660 Hz

that have been aliased to the peaks apparent at 460 Hz and 340 Hz, respectively. However, the comb filter in the present example was not designed to remove these components. The lowpass and highpass filters were applied in the frequency domain to the Fourier transform of the signal using the form indicated by Equations 3.145 and 3.148. The comb filter was applied in the time domain using the coefficients in Equation 3.151.

The combined frequency response of the filters is shown in the middle plot in Figure 3.100. The spectrum of the ECG signal after the application of all three filters is shown in the bottom plot in Figure 3.100. The filtered signal spectrum has no appreciable energy beyond about 100 Hz, and it displays significant attenuation at 60 Hz.

The outputs after the lowpass filter, the highpass filter, and the comb filter are shown in Figure 3.99. Observe that the baseline drift is present in the output of the lowpass filter, and that the power-line interference is present in the outputs of the lowpass and highpass filters. The final trace is free of all three types of interference. Note, however, that the highpass filter has introduced a noticeable distortion (undershoot) in the P and T waves.

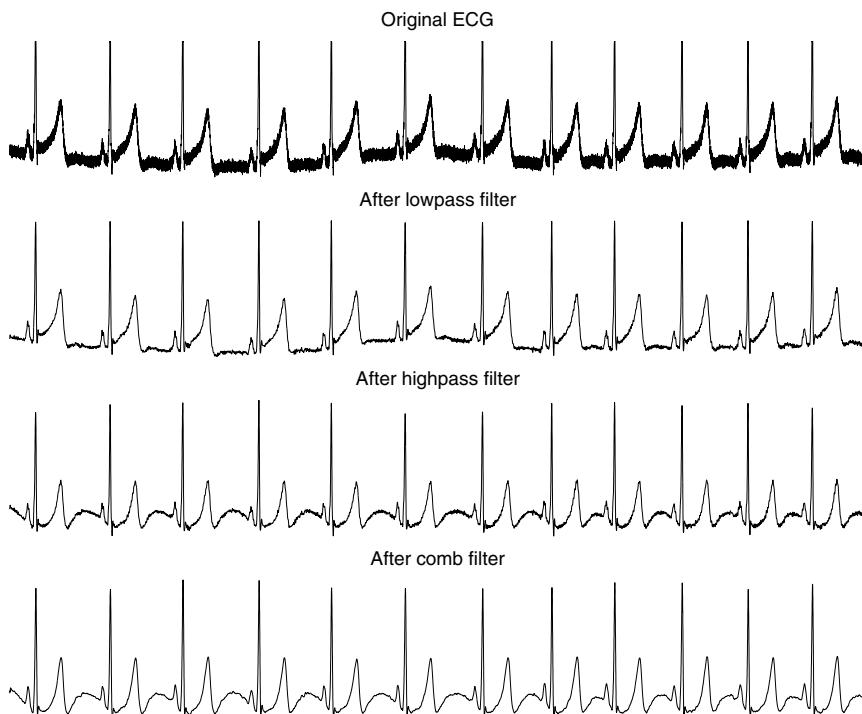


Figure 3.99 ECG signal with a combination of artifacts and its filtered versions. The duration of the signal is 10.7 s.

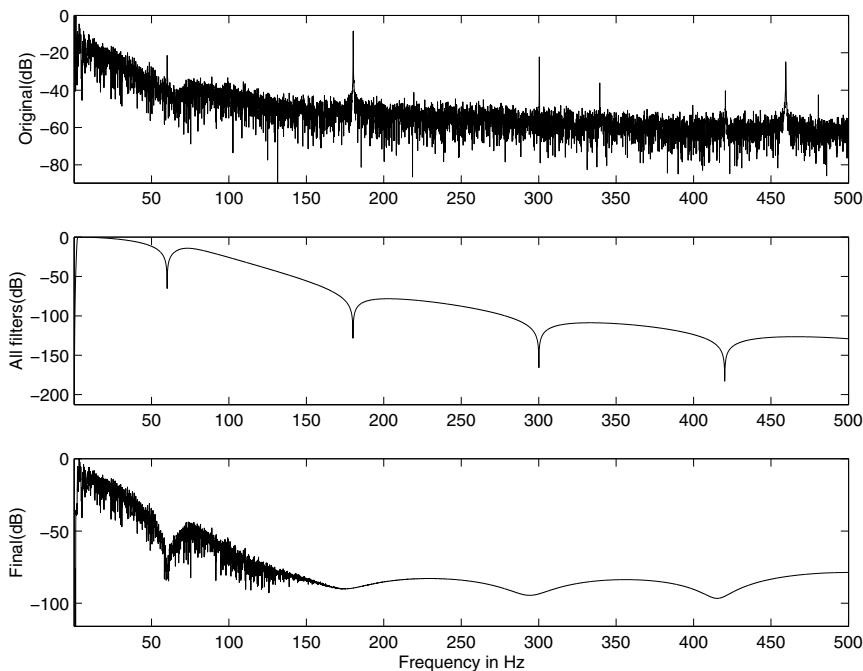


Figure 3.100 Top and bottom plots: Power spectra of the ECG signals in the top and bottom traces of Figure 3.99. Middle plot: Frequency response of the combination of lowpass, highpass, and comb filters. The cutoff frequency of the highpass filter is 2 Hz; the highpass portion of the frequency response is not clearly seen in the plot.

3.13 Application: Adaptive Cancellation of the Maternal ECG to Obtain the Fetal ECG

Problem: Propose an adaptive noise cancellation filter to remove the maternal ECG signal from the abdominal-lead ECG shown in Figure 3.9 to obtain the fetal ECG. Chest-lead ECG signals of the expectant mother may be used for reference.

Solution: Widrow et al. [124] described a multiple-reference ANC for removal of the maternal ECG in order to obtain the fetal ECG. The combined ECG was obtained from a single abdominal lead, whereas the maternal ECG was obtained via four chest leads. The model was designed to permit the treatment of not only multiple sources of interference, but also of components of the desired signal present in the reference inputs, and further to consider the presence of uncorrelated noise components in the reference inputs. It should be noted that the maternal cardiac vector is projected on to the axes of different ECG leads in different ways, and hence the characteristics of the maternal ECG in the abdominal lead would be different from those of the chest-lead ECG signals used as reference inputs.

Each filter channel used by Widrow et al. [124] had 32 taps and a delay of 129 ms. The signals were prefiltered to the bandwidth 3 – 35 Hz and a sampling rate of 256 Hz was used. The optimal Wiener filter used (see Section 3.8) included transfer functions and cross-spectral vectors between the input source and each reference input. Further extension of the method to more general multiple-source, multiple-reference noise cancelling problems was also discussed by Widrow et al.

The result of cancellation of the maternal ECG from the abdominal lead ECG signal in Figure 3.9 is shown in Figure 3.101. Comparing the two figures, it is seen that the filter output has successfully extracted the fetal ECG and suppressed the maternal ECG. See Widrow et al. [124] for details; see also Ferrara and Widrow [189] and Zarzoso and Nandi [190] for additional related discussion.

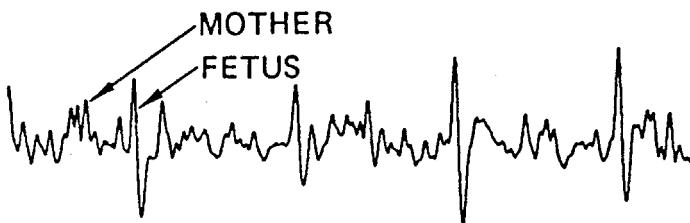


Figure 3.101 Result of adaptive cancellation of the maternal chest ECG from the abdominal ECG in Figure 3.9. The QRS complexes extracted correspond to the fetal ECG. Reproduced with permission from B. Widrow, J.R. Glover, Jr., J.M. McCool, J. Kaunitz, C.S. Williams, R.H. Hearn, J.R. Zeidler, E. Dong, Jr., and R.C. Goodlin, Adaptive noise cancelling: Principles and applications, *Proceedings of the IEEE*, 63(12):1692–1716, 1975. ©IEEE.

3.14 Application: Adaptive Cancellation of Muscle contraction Interference in VAG Signals

Problem: Study the applicability of adaptive noise cancellation filters to remove the muscle-contraction interference caused by the rectus femoris in the VAG signal recorded at the patella.

Solution: Rangayyan et al. [191] conducted a study on the impact of cancellation of muscle-contraction interference on modeling and classification of VAG signals and further classification of the filtered signals as normal or abnormal. Both the LMS (see Section 3.9.2) and the RLS (see Section 3.9.3) methods were investigated, and the RLS method was chosen for its more efficient tracking of nonstationarities in the input and reference signals.

Figure 3.102 shows plots of the VAG signal of a subject with chondromalacia patella of grade II [trace (a)] and a simultaneously recorded channel of muscle-contraction interference [labeled as MCI, trace (b)]. The results of adaptive filtering of the VAG signal with the muscle-contraction interference channel as the reference are also shown in Figure 3.102: Trace (c) shows the result of LMS filtering and trace

(d) shows that of RLS filtering. A single-stage LMS filter with variable step size $\mu(n)$ as in Equation 3.204 was used, with $M = 7$, $\mu = 0.05$, and $\alpha = 0.98$. The RLS filter used $M = 7$ and $\lambda = 0.98$.

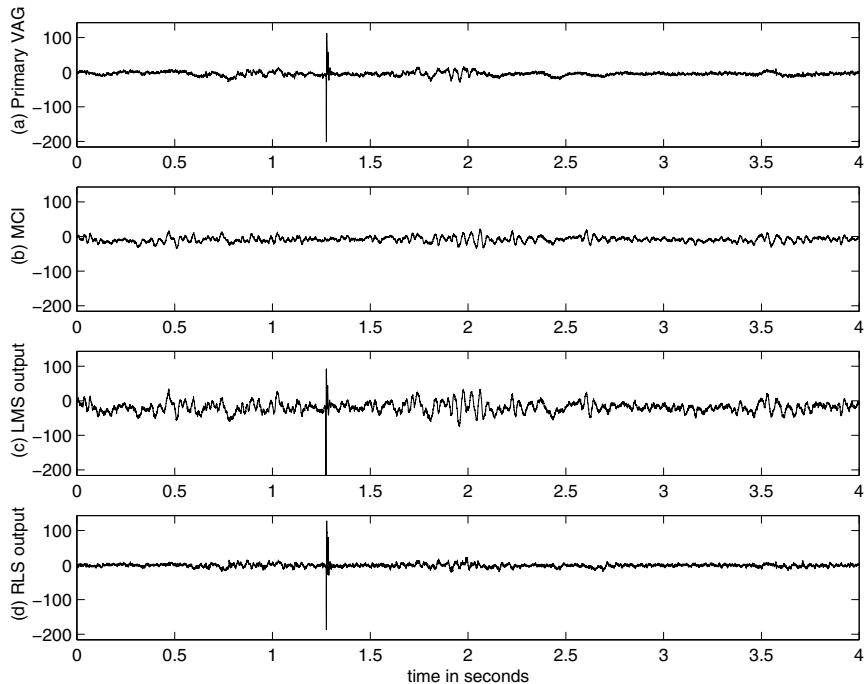


Figure 3.102 Top to bottom: (a) VAG signal of a subject with chondromalacia patella of grade II; (b) muscle-contraction interference (MCI); (c) result of LMS filtering; and (d) result of RLS filtering. The recording setup is shown in Figure 3.10.

As in the example in Figure 3.92, it is seen that the muscle-contraction interference has been removed by the RLS filter; however, the LMS filter failed to perform well, due to its limited capabilities in tracking the nonstationarities in the interference. The spectrograms of the primary, reference, and RLS-filtered signals are shown in Figures 3.103, 3.104, and 3.105, respectively. (The logarithmic scale is used to display better the minor differences between the spectrograms.) It is seen that the frequency components of the muscle-contraction interference have been suppressed by the RLS filter.

The primary (original) and filtered VAG signals of 53 subjects were adaptively segmented and modeled in the study of Rangayyan et al. [191] (see Chapter 8). The segment boundaries were observed to be markedly different for the primary and the filtered VAG signals. Parameters extracted from the filtered VAG signals were expected to provide higher discriminant power in pattern classification when compared to the same parameters of the unfiltered or primary VAG signals. However, classification experiments indicated otherwise: The filtered signals gave a lower classification

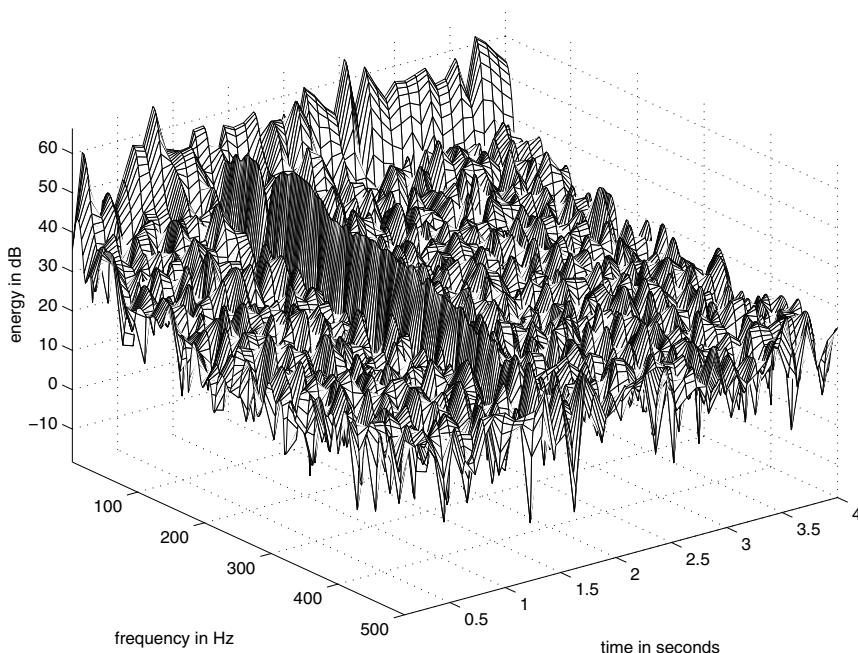


Figure 3.103 Spectrogram of the original VAG signal in Figure 3.102 (a). A Hann window of length 256 samples (128 ms) was used; an overlap of 32 samples (16 ms) was allowed between adjacent segments.

accuracy by almost 10%. It was reasoned that, after removal of the predominantly low-frequency muscle-contraction interference, the transient VAG signals of clinical interest were not modeled well by the prediction-based methods. It was concluded that the adaptive filtering procedure used was not an appropriate preprocessing step before signal modeling for pattern classification. However, it was noted that cancellation of muscle-contraction interference may be a desirable step before auditory or spectral analysis of VAG signals.

3.15 Remarks

We have investigated problems posed by artifact, noise, and interference of various forms in the acquisition and analysis of several biomedical signals. Random noise, structured interference, and physiological interference have been identified and analyzed separately. Attention has been drawn to the different characteristics of various types of noise, such as frequency content and nonstationarity. Fixed, optimal, and adaptive filters were developed in the time and frequency domains for several applications, and guidelines were drawn to assist in choosing the appropriate filter for various types of artifacts. Advanced methods for adaptive denoising based on

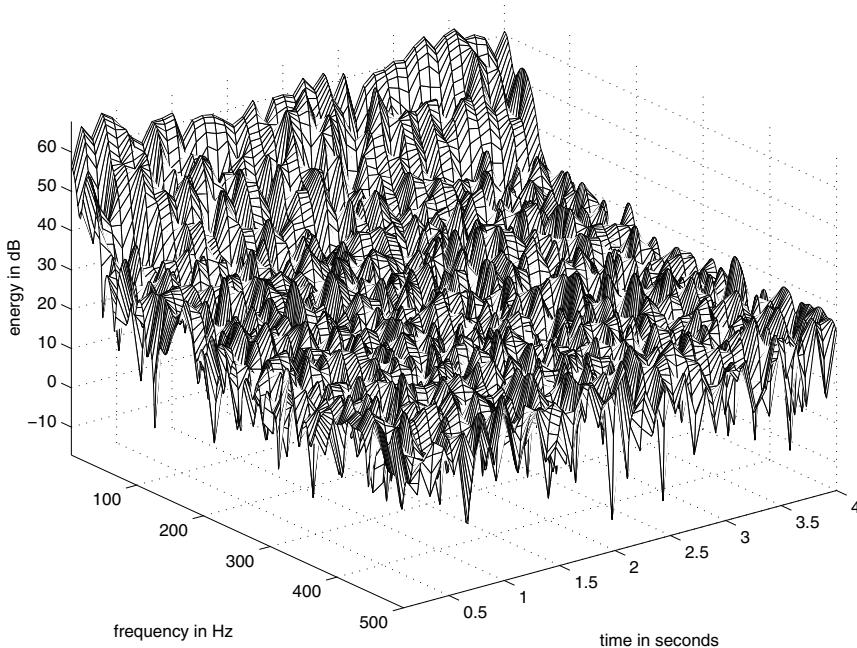


Figure 3.104 Spectrogram of the muscle-contraction interference signal in Figure 3.102 (b). A Hann window of length 256 samples (128 ms) was used; an overlap of 32 samples (16 ms) was allowed between adjacent segments.

wavelet and time-frequency decomposition methods have not been discussed in this chapter, but are described by Krishnan and Rangayyan [192] for filtering VAG signals; see Chapter 8 for related material. Another category of filters that has not been considered in this chapter is that of morphological filters [193, 194], which include nonlinear statistics-based operations and could be formulated under certain conditions to include linear filter operations as well.

It is important to observe that each practical problem needs to be studied carefully to determine the type and characteristics of the artifact present; the nature of the signal and its relationship to, or interaction with, the artifact; and the effect of the filter being considered on the desired signal or features computed from the filtered result. Different filters may be suitable for different subsequent steps of signal analysis. It is unlikely that a single filter will address all of the problems and the requirements in a wide variety of practical situations and applications. Regardless of one's expertise in filters, it should be remembered that *prevention is better than cure*: most filters, while removing an artifact, may introduce another. Attempts should be made at the outset to acquire artifact-free signals to the extent possible.

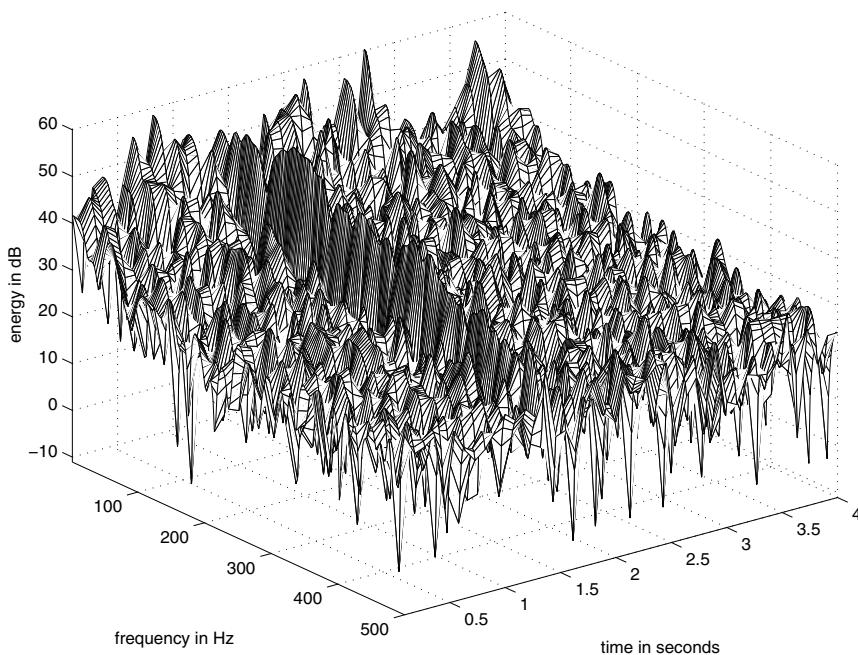


Figure 3.105 Spectrogram of the RLS-filtered VAG signal in Figure 3.102 (d). A Hann window of length 256 samples (128 ms) was used; an overlap of 32 samples (16 ms) was allowed between adjacent segments.

3.16 Study Questions and Problems

Note: Some of the questions deal with the fundamentals of signals and systems, and may require background preparation with other sources such as Lathi [1,2] or Oppenheim et al. [3]. Such problems are included for the sake of recollection of the related concepts.

1. What are the potential sources of instrumentation and physiological artifacts in recording the PCG signal? Propose nonelectronic methods to prevent or suppress the latter type of artifacts.
2. List four potential sources of instrumentation and physiological artifacts in recording the ECG signal. Describe methods to prevent or remove each artifact. Identify the possible undesired effects of your procedures on the ECG signal.
3. Identify at least three potential sources of physiological artifacts in recording the EEG signal.
4. In recording the EEG in a clinical laboratory, some channels were found to contain the ECG as an artifact. Will simple lowpass or bandpass filtering help in removing the artifact? Why (not)? Propose a scheme to remove the artifact.
5. A biomedical signal is bandpass filtered to the range $0 - 150 \text{ Hz}$. Assume the filter to be ideal, and assume any distribution of spectral energy over the bandwidth of the signal.

- (a) What is the minimum frequency at which the signal should be sampled in order to avoid aliasing errors? (b) A researcher samples the signal at 500 Hz. Draw a schematic representation of the spectrum of the sampled signal. (c) Another researcher samples the signal at 200 Hz. Draw a schematic representation of the spectrum of the sampled signal. Explain the differences between case (b) and case (c).
6. Distinguish between ensemble averages and temporal (time) averages. Identify applications of first-order and second-order averages of both types in EEG analysis.
7. Explain how one may apply ensemble averaging and temporal (time) averaging procedures to process ECG signals. Identify applications of first-order and second-order averages of both types in ECG analysis.
8. Explain how you would apply synchronized averaging to remove noise in (a) ECG signals, (b) event-related (or evoked) potentials, (c) heart sound (PCG) signals, and (d) EMG signals. In each case, explain (i) how you will obtain the information required for synchronization of the signals epochs or episodes; (ii) sources of artifacts and how you will deal with them; (iii) limitations and practical difficulties; and (iv) potential for success of the method.
9. Draw a typical ECG waveform over one cardiac cycle indicating the important component waves. How is the waveform affected by passage through the following: (a) A lowpass filter with a cutoff frequency of 40 Hz? (b) A highpass filter with a cutoff frequency of 5 Hz? Draw schematic representations of the expected outputs and explain their characteristics.
10. What is the z -transform of a signal whose samples are given in the series $\{4, 3, 2, 1, 0, -1, 0, 1, 0\}$? (The first sample represents zero time in all the signal sample arrays given in the problems, unless stated otherwise.)
11. A digital filter is used to process a signal at a sampling rate of 2,000 Hz. (a) Draw the unit circle in the complex z -plane and identify the frequencies corresponding to the points $z = (1 + j0)$, $z = (0 + j1)$, $z = (-1 + j0)$, and $z = (0 - j1)$, as well as the point $z = (1 + j0)$ again as approached in the counterclockwise direction. (b) What are the frequencies corresponding to these same points if the sampling rate is 500 Hz?
12. What is the transfer function of an LSI system whose impulse response is given by the series $\{2, 1, 0, 0, -1, 0, 1, 0\}$ for $n = 0, 1, 2, \dots, 7$?
13. The impulse response of a digital filter is $\{1, -2, 1\}$. What will be the response of the filter to the unit step?
14. The impulse response of a filter is $\{3, -2, 2\}$. What will be the response of the filter to the input $\{6, 4, 2, 1\}$?
15. The transfer function of a filter is $H(z) = z^{-1} - 3z^{-2} + 2z^{-4} - z^{-6}$. What is the difference equation relating the output to the input? What is the impulse response of the filter?
16. The impulse response of a filter is given by the series of values $\{3, 2, 1, 0, -1, 0, 0, 1\}$. What is its transfer function?
17. The impulse response of a filter is specified by the series of sample values $\{3, 1, -1, 1\}$.
(a) What will be the response of the filter to the input whose sample values are $\{4, 4, 2, 1\}$?
(b) Is the filter response obtained by linear convolution or circular convolution of the input with the impulse response? (c) What will be the response with the type of convolution other than the one you indicated as the answer to the question above? (d) How

would you implement convolution of the two signals listed above using the DFT? Which type of convolution will this procedure provide? How would you get the other type of convolution for the signals in this problem via the DFT-based procedure?

18. A biomedical signal is expected to be band-limited to 100 Hz , with significant components of interest up to 80 Hz . However, the signal is contaminated with a periodic artifact with a fundamental frequency of 60 Hz and significant third and fifth harmonics. A researcher samples the signal at 200 Hz without prefiltering the signal. Draw a schematic representation of the spectrum of the signal and indicate the artifactual components. Label the frequency axis clearly in Hz . What kind of a filter would you recommend to remove the artifact?
19. A biomedical signal sampled at 500 Hz was found to have a significant amount of 60 Hz interference. (a) Design a notch filter with two zeros to remove the interference. (b) What is the effect of the filter if a signal sampled at 100 Hz is applied as the input?
20. Two filters with transfer functions $H_1(z) = \frac{1}{3}(1 + z^{-1} + z^{-2})$ and $H_2(z) = 1 - z^{-1}$ are cascaded. (a) What is the transfer function of the complete system? (b) What is its impulse response? (c) What is its gain at DC and at the folding frequency (that is, $f_s/2$)?
21. A filter has the transfer function $H(z) = (1 + 2z^{-1} + z^{-2})/(1 - z^{-2})$. (a) Write the difference equation relating the output to the input. (b) Draw the signal-flow diagram of a realization of the filter. (c) Draw its pole-zero diagram.
22. A digital filter has zeros at $0.5 \pm j0.5$ and poles at $-0.6 \pm j0.3$. (a) Derive the transfer function of the filter. (b) Derive the time-domain difference equation (input-output relationship) of the filter. (c) If the filter is used at the sampling frequency of 1,000 Hz , what are the frequencies at which the gain of the filter is at its maximum and minimum?
23. Two filters with the transfer functions $H_1(z) = \frac{1}{2T}(1 - z^{-2})$ and $H_2(z) = \frac{1}{1 - \frac{1}{2}z^{-1}}$ are cascaded. (a) What is the transfer function of the complete system? (b) Draw its pole-zero diagram. (c) Write the difference equation relating the output to the input. (d) Draw the signal-flow diagram of a realization of the filter. (e) Compute the first six values of the impulse response of the filter. (f) The filter is used to process a signal sampled at 1,000 Hz . What is its gain at 0, 250, and 500 Hz ?
24. A filter is described by the difference equation $y(n) = y(n-1) + \frac{1}{4}x(n) - \frac{1}{4}x(n-4)$. (a) What is its transfer function? (b) Draw the signal-flow diagram of a realization of the filter. (c) Draw its pole-zero diagram.
25. Under what conditions will synchronized averaging fail to reduce noise?
26. A signal sampled at the rate of 100 Hz has the samples $\{0, 10, 0, -5, 0\}$ in mV . The signal is passed through a filter described by the transfer function $H(z) = \frac{1}{T}(1 - z^{-1})$. What will be the output sequence? Plot the output and indicate the amplitude and time scales in detail with appropriate units.
27. A signal sampled at the rate of 100 Hz has the samples $\{0, 10, 0, -5, 0\}$ in mV . It is supposed to be processed by a differentiator with the difference equation $y(n) = \frac{1}{T}[x(n) - x(n-1)]$ and then squared. By mistake the squaring operation is performed before the differentiation. What will be the output sequence? Plot the outputs for both cases and indicate the amplitude and time scales in detail with appropriate units. Explain the differences between the two results.
28. A certain signal analysis technique requires the following operations in order: (a) differentiation, (b) squaring, and (c) lowpass filtering with a filter $H(\omega)$. Considering a

generic signal $x(t)$ as the input, write the time-domain and frequency-domain expressions for the output of each stage. Will changing the order of the operations change the final result? Why (not)?

29. A signal sampled at the rate of 100 Hz has the samples $\{0, 10, 0, -5, 0\}$ in mV. The signal is processed by a differentiator with the difference equation $y(n) = \frac{1}{T}[x(n) - x(n-1)]$, and then filtered with a 4-point MA filter. (a) Derive the transfer function and frequency response of each filter and the combined system. (b) Derive the values of the signal samples at each stage. (c) Does it matter which filter is placed first? Why (not)? (d) Plot the output and indicate the amplitude and time scales in detail with appropriate units.
30. Distinguish between ensemble averages and temporal (time) averages. Identify potential applications of first-order and second-order averages of both types in heart sound (PCG) analysis. Explain how you would obtain a trigger for synchronization.
31. Is the heart sound signal (PCG) a stationary signal or not? Provide your answer in the context of one full cardiac cycle and give reasons. If you say that the PCG signal is nonstationary, identify parts (segments) that could possibly be stationary, considering the possibility of murmurs in both systole and diastole.
32. A signal $x(t)$ is transmitted through a channel. The received signal $y(t)$ is a scaled, shifted, and noisy version of $x(t)$ given as $y(t) = \alpha x(t - t_0) + \eta(t)$ where α is a scale factor, t_0 is the time delay, and $\eta(t)$ is noise. Assume that the noise process has zero mean and is statistically independent of the signal process, and that all processes are stationary. Derive expressions for the PSDs of $y(t)$ in terms of the PSDs of x and η .
33. A signal $x(n)$ that is observed in an experiment is modeled as a noisy version of a desired signal $d(n)$ as $x(n) = d(n) + \eta(n)$. The noise process η is a zero-mean, unit-variance random process with uncorrelated samples [“white” noise, with ACF $\phi_\eta(\tau) = \delta(\tau)$] that is statistically independent of the signal process d . The ACF $\phi_d(\tau)$ of d is given by the sequence $\{1.0, 0.6, 0.2\}$, for $\tau = 0, 1, 2$, respectively. Prepare the Wiener–Hopf equation and derive the coefficients of the optimal Wiener filter.
34. Draw the waveform of a typical normal ECG over one cardiac cycle. Label the names of the component waves and give their typical durations and intervals. Draw a version of the ECG waveform including power-line artifact at 60 Hz. Draw another version of the ECG waveform including high-frequency noise. Describe potential causes of the two types of artifact and methods to prevent or remove each artifact.
35. A researcher is designing a Wiener filter and is in need of help in obtaining the autocorrelation matrix, defined as $\Phi = E[\mathbf{x} \mathbf{x}^T]$. Help the researcher by computing the autocorrelation matrix for the sample vector $\mathbf{x} = [1 \ 2 \ 2 \ 1]^T$. Explain the meaning of the operator $E[\cdot]$ and describe how you would implement the operation in practice.
36. A researcher uses a combination of two digital filters in cascade (series). The output of the first filter is the first derivative or difference of the input. The output of the second filter is the average of the current input sample and the preceding input sample. (a) Give the input–output relationship in the time domain (difference equation) for each filter. (b) Derive the transfer function for each filter. (c) Derive the impulse response of the complete system. (d) Derive the transfer function of the complete system. (e) Does it matter which filter is placed first? Explain. (f) Compute the gain of the complete system at 0, $f_s/4$, and $f_s/2$, where f_s is the sampling frequency.
37. A noisy signal $x(n)$ is expressed in vector notation as $\mathbf{x}(n) = \mathbf{d}(n) + \boldsymbol{\eta}(n)$, where $\mathbf{d}(n)$ is the original (ideal) signal and $\boldsymbol{\eta}(n)$ is random noise that is statistically independent of

- the signal. All signals are assumed to be second-order stationary. Derive an expression for the ACF matrix of \mathbf{x} in terms of the ACF matrices of \mathbf{d} and $\boldsymbol{\eta}$. Explain each step of your derivation.
38. A random signal $x(t)$ is characterized by its PDF $p_x(x)$. Write the basic definition of the mean-squared value of x based upon its PDF. Given an observation of the signal in terms of its samples, expressed as $x(n), n = 0, 1, 2, \dots, N - 1$, give a formula to compute the mean-squared value of x (without the PDF). Give a formula to obtain the *RMS* value of x . If x represents an electrical signal, what do the mean-squared and *RMS* values represent?
 39. For a discrete-time signal $x(n)$, write the basic definition of the z -transform. The signal $x(n)$ is processed by an LSI system with the impulse response $h(n)$. Write the complete mathematical expression for the output $y(n)$. Showing all steps, derive the relationship between the z -transforms of $x(n)$, $h(n)$, and $y(n)$.
 40. A digital filter is specified by the difference equation $y(n) = x(n) - x(n - 2)$, where $x(n)$ is the input and $y(n)$ is the output. Derive the transfer function of the filter. Derive the magnitude and phase parts of the frequency response of the filter. Let the sampling frequency be normalized to unity. Plot the magnitude and phase responses. Indicate the values of the functions at normalized frequency values of 0, 0.25, and 0.5. Explain the nature and effects of the filter.
 41. A researcher is designing an experiment to record visual ERPs. Provide advice to the researcher on the following: (a) Identify a potential source of artifact in the form of random noise. Propose a strategy or method to prevent or remove the artifact. (b) Identify a potential source of structured noise. Propose a strategy or method to prevent or remove the artifact. (c) Identify a potential source of physiological interference. Propose a strategy or method to prevent or remove the artifact. No equation is required for your answer to this question.
 42. Two LSI filters are specified in terms of their impulse responses as $\delta(n) - \delta(n - 1)$ and $\delta(n) + 2\delta(n - 1) + \delta(n - 2)$. A researcher prepares a new filter by connecting the two filters described above in series. (a) Derive the transfer function of each filter. (b) Derive the transfer function of the combined filter. (c) Derive the impulse response of the combined filter. (d) Does it matter which filter is placed first? Explain and justify your answer. (e) Draw the signal-flow diagram of the combined filter. (f) What is the gain of the combined filter at DC, $f_s/4$, and $f_s/2$, where f_s is the sampling frequency? From these values, give an interpretation of the nature of the combined filter.
 43. A signal of interest $x(t)$ is affected by additive noise $\eta(t)$ and is observed as $y(t) = x(t) + \eta(t)$. It is assumed that the processes are random, and that the noise process is statistically independent of the signal process. It is also assumed that the mean of η is zero. Write the basic definition of the expected value (mean) of the random process x based upon its PDF. Derive the relationship between the mean of y and the statistics of x and η . Derive the relationship between the variance of y and the statistics of x and η . Interpret and explain your results.
 44. A researcher obtains a set of ERP signals $x_k(n)$, $k = 1, 2, \dots, M$, and $n = 0, 1, 2, \dots, N - 1$, where $x_k(n)$ is the k^{th} signal, M is the number of signals recorded, and N is the number of samples in each signal. Help the researcher with the following: (a) Give a step-by-step procedure (algorithm) to obtain the synchronized average of the ERPs. (b) State the conditions and requirements for synchronized averaging to be applicable and yield good results. (c) Give an equation to define the synchronized or ensemble

- average of the ERPs. (d) Give an equation to define the temporal average of one of the ERPs over a specified window of time. (e) Give an equation to obtain the average Euclidean distance between the result of synchronized averaging and the M ERPs available. Explain the meaning of the Euclidean distance in this context.
45. A researcher is designing an experiment to record ENGs. Provide advice to the researcher on the following: (a) Identify a potential source of artifact in the form of random noise. Propose a strategy or method to prevent or remove the artifact. (b) Identify a potential source of structured noise. Propose a strategy or method to prevent or remove the artifact. (c) Identify a potential source of physiological interference. Propose a strategy or method to prevent or remove the artifact. No equation is required for your answer to this question.
46. You are given two signals, $x(n)$ and $y(n)$, each with N samples, $n = 0, 1, 2, \dots, N - 1$. No information is available regarding the PDFs of the related processes. Give an equation to compute the normalized correlation coefficient between the two signals. Explain the meaning and purpose of each part of your equation.
47. A filter is specified with the transfer function $H(z) = \frac{1}{T} \left[\frac{1-z^{-1}}{1-0.95 z^{-1}} \right]$. (a) Derive an expression for the input-output relationship of the filter in the time domain. (b) Draw a signal-flow diagram of the filter using only one delay element. (c) Draw the pole-zero plot of the system. (d) What is the gain of the system at zero frequency and at one-half of the sampling frequency?
48. You are given a set of M signals, $x_k(n)$, $k = 1, 2, \dots, M$, each with N samples, $n = 0, 1, 2, \dots, N - 1$. The index n represents sampled time and the index k represents the k^{th} signal in the set. (a) Explain the differences between ensemble averages and temporal averages. Give equations to define the following: (b) the ensemble mean at an instant of time $n = n_1$; (c) the temporal mean of the k^{th} signal computed over the period $n = n_1$ to $n = n_2$; and (d) the average signal or template, $\bar{x}(n)$, $n = 0, 1, 2, \dots, N - 1$, computed over the set of signals.
49. You are given two processes that generate two signals, x and y , with the PDFs $p_x(x)$ and $p_y(y)$, respectively. The joint PDF between the two processes is $p_{x,y}(x, y)$. (a) Write an expression to define the mean of the signal x . (b) Write an expression to define the variance of the signal x . (c) Write an expression to define the covariance between the two signals x and y . (d) Write an expression to define the normalized correlation coefficient between the two signals. (e) Give interpretations of the variance and normalized correlation coefficient and indicate practical use or applications of these measures.
50. You are given an ensemble of M signals, $y_k(n)$, $k = 1, 2, \dots, M$, with each signal having N samples, indexed as $n = 0, 1, 2, \dots, N - 1$. Write mathematical equations or expressions for the following: (a) The ensemble or synchronized average, $s(n)$, of the M signals. (b) The total power of the difference, error, or noise between the result of averaging, $s(n)$, and all of the given signals, $y_k(n)$, $k = 1, 2, \dots, M$. (c) The total power of the result, $s(n)$. (d) The SNR of the result.
51. You are given two signals, $x(n)$ and $y(n)$, $n = 0, 1, 2, \dots, N$. Write a mathematical equation or expression to derive the normalized correlation coefficient between the two signals. Suppose that $x(n) = u(n) - u(n - 4)$, where $u(n)$ is the unit step function, and $y(n) = x(n - 8)$. Sketch the signals $x(n)$ and $y(n)$. Compute the correlation coefficient (without normalization) between the two signals and plot the result. Explain the relationships between the signals and the result.

52. A digital filter is specified by the difference equation $y(n) = \frac{1}{4} \sum_{k=0}^{k=3} x(n - k)$, where $x(n)$ is the input and $y(n)$ is the output. (a) Give an equation for and draw a sketch of the impulse response of the filter. (b) Draw the signal-flow diagram of the filter. (c) Derive the transfer function of the filter. (d) Derive the magnitude and phase parts of the frequency response of the filter. (e) What is the gain of the filter at 0 Hz and $f_s/2$ Hz, where f_s is the sampling frequency? (f) Explain the nature and effects of the filter.
53. Considering the acquisition of the ECG signal, identify and describe one potential source or cause of each of the following types of artifact: (a) high-frequency noise, (b) periodic artifact, and (c) a physiological artifact. In each case, explain how the artifact is caused, how the artifact gets combined with the ECG, and how you would remove or prevent the artifact.
54. Write the equations to define (a) the convolution of two signals, and (b) the Fourier transform of a signal. Prove that the Fourier transform of the convolution of two signals is equal to the product of the Fourier transforms of the two individual signals. Show and explain all steps. You may use continuous-time or discrete-time notation.
55. Write an equation to define the cross-correlation between two signals. Explain the computational procedures required to obtain the cross-correlation. Derive an expression for the Fourier transform of the cross-correlation of two signals in terms of the Fourier transforms of the individual signals. Show all steps. If you use any property of the Fourier transform, give its proof.
56. A biomedical engineer working in a neurophysiology laboratory is frustrated by the appearance of the ECG as an artifact in the EEG of a patient. You are hired to help the engineer. (a) Compare the typical amplitude ranges and frequency bandwidths of the of the two signals. (b) Would you recommend the use of a fixed lowpass, highpass, or bandpass filter to remove the artifact? If yes, give the essential characteristics of the filter that you recommend. If not, explain your reasons. (c) The engineer has heard about adaptive filters (ANCs). Draw a schematic (block) diagram of an ANC. In the context of the problem mentioned above, explain what the primary input should be; how and from where on the patient you would obtain an appropriate reference input; and how the various inputs and outputs of the system relate to one another. Explain the basic assumptions made in the design and application of the filter. You do not have to derive any equation in your answer to this problem.
57. A discrete-time signal $x(n)$ is passed through an LSI filter with the impulse response $h(n)$. Write the full expression that gives the output $y(n)$ in terms of $x(n)$ and $h(n)$. Starting with the basic definition of the z -transform, derive the relationship between the z -transforms of $y(n)$, $x(n)$, and $h(n)$.
58. (a) A filter is specified to have a zero in the z domain at $z = 1$. (i) Derive the transfer function $H_1(z)$ of the filter. (ii) Derive the impulse response $h_1(n)$ of the filter. (iii) Is this a lowpass, highpass, bandpass, or band-reject filter?
 (b) Another filter is specified to have a double-zero (or two zeros) at $z = -1$. (i) Derive the transfer function $H_2(z)$ of the filter. (ii) Derive the impulse response $h_2(n)$ of the filter. (iii) Is this a lowpass, highpass, bandpass, or band-reject filter?
 (c) A researcher uses the two filters $H_1(z)$ and $H_2(z)$ as above in cascade (series). (i) Derive the transfer function $H(z)$ of the combined filter. (ii) Derive the impulse response $h(n)$ of the combined filter. (iii) Draw the pole-zero plot for the combined

- filter. (iv) Draw the signal-flow diagram for the combined filter. (v) Is this a lowpass, highpass, bandpass, or band-reject filter?
59. A biomedical signal is sampled at 400 Hz with no aliasing error. The signal is known to contain power-line artifact at the fundamental frequency of 50 Hz and all of its harmonics up to and including the maximum frequency present in the signal. Draw the unit circle in the z -domain and show the positions of the zeros of a comb filter to remove the artifact. Mark the angle and frequency of each zero. (You do not have to derive the transfer function of the filter.)
60. Two discrete-time filters are specified in terms of their impulse responses as $h_1(n) = \delta(n) + \delta(n - 1) + \delta(n - 2) + \delta(n - 3)$ and $h_2(n) = \delta(n) - \delta(n - 1)$. The two filters are used in series to filter a signal. Derive and plot the impulse response of the combined filter. Derive the transfer function and frequency response of the combined filter.
61. In the optimization procedure for the derivation of the Wiener filter, the coefficients of the filter are expressed as the vector $\mathbf{w} = [w_0, w_1, w_2, \dots, w_{M-1}]^T$, where M is the order of the filter and T indicates the transpose. The current input sample $x(n)$ and $M - 1$ previous input samples are expressed in another vector as $\mathbf{x}(n) = [x(n), x(n - 1), \dots, x(n - M + 1)]^T$. (a) Write the full expression of convolution to define the output of the filter in terms of the input signal and the impulse response of the filter. Write the equivalent expression using the vectors as defined above and explain how the two methods lead to the same result. (b) Explain the difference between $\mathbf{x}(n) \mathbf{x}^T(n)$ and $\mathbf{x}^T(n) \mathbf{x}(n)$. (c) What does $E[\mathbf{x}(n) \mathbf{x}^T(n)]$ represent? Write a mathematical expression to give the detailed relationship between an element in the result and the input sample values.
62. A signal $x(n)$ is given in terms of its samples as $\{4, 3, 2, 1, 2, 4, 2, 1\}$, for $n = 0, 1, 2, \dots, 7$. The signal is processed using an LSI filter with the impulse response, $h(n)$, having the sampled values $\{1, 2, 1\}$, for $n = 0, 1, 2$. Give the procedure to compute the output of the filter. Derive the output of the filter showing all steps.
63. Draw a block diagram of the ANC (adaptive filter). State the expected relationships between the inputs to the filter. State the conditions that must be met for optimal functioning of the adaptive filter. Give equations for the output of the filter using both the summation form and the vectorial form for convolution of signals. Explain the composition of the vectors in the vectorial form of the equation.
64. A filter is specified in terms of its pole-zero plot as follows: a zero at $z = 1$ and a zero at $z = -1$. (a) Derive the transfer function of the filter. (b) Derive the difference equation and draw a signal-flow diagram of the filter. (c) Derive and plot the impulse response of the filter. (d) Derive the magnitude and phase of the frequency response of the filter. (e) Draw a sketch of the magnitude of the frequency response of the filter. Assuming the sampling rate to be 200 Hz , label the frequency axis in Hz . (f) Derive the gain at 0 Hz and 100 Hz and explain the nature of the filter.
65. Draw a schematic sketch of a speech signal including segments of voiced and unvoiced speech. Explain their characteristics. Explain the test for randomness. Indicate the result you would expect if you were to apply the test for randomness to your speech signal example.
66. A biomedical signal sampled at 240 Hz contains power-line interference at 60 Hz . Design a notch filter to remove the artifact. Draw the unit circle in the complex z domain. Indicate the values of z and the frequency in Hz at the intersections of the circle with

the axes. Mark the locations of the poles and/or zeros in your filter design. Derive the transfer function and the impulse response of the filter. Show all steps.

67. A researcher is interested in recording vibration signals from the knee joint during swinging movement of the leg. However, muscle vibration signals from the thigh muscle were observed to contaminate the knee-joint vibration signal. Provide recommendations to the researcher on how a filter for ANC may be designed to reduce the muscle artifact. Give a schematic block diagram of the ANC filter. You do not need to derive the mathematical procedures for the filter. Indicate where and which signals need to be provided as input to the ANC filter, and where the filtered output is to be obtained. Which part of the ANC filter has time-varying characteristics? Write an equation to describe the input–output relationship of this part.
68. Compare the schematic representations of a cell in Figure 1.9 and an LSI system in Figures 3.14 and 3.17. Draw analogies between their inputs, outputs, and characteristics.

3.17 Laboratory Exercises and Projects

Note: Data files related to the exercises are available at the site

<http://people.ucalgary.ca/~ranga/enel563>

1. The data file `ecg2x60.dat` contains an ECG signal, sampled at 200 Hz , with a significant amount of 60 Hz power-line artifact. (See also the file `ecg2x60.m`.) (a) Design a notch filter with two zeros to remove the artifact and implement it in MATLAB.[®] (b) Add two poles at the same frequencies as those of the zeros, but with a radius that is less than unity. Study the effect of the poles on the output of the filter as their radius is varied.
2. A noisy ECG signal is provided in the file `ecg_hfn.dat`. (See also the file `ecg_hfn.m`.) The sampling rate of this signal is $1,000\text{ Hz}$.

Develop a MATLAB[®] program to perform synchronized averaging as described in Section 3.5.1. Select a QRS complex from the signal for use as the template and use a suitable threshold on the cross-correlation function in Equation 3.96 for beat detection. Plot the resulting averaged QRS complex. Ensure that the averaged result covers one full cardiac cycle. Plot a sample ECG cycle from the noisy signal for comparison.

Select the QRS complex from a different beat for use as the template and repeat the experiment. Observe the results when the threshold on the cross-correlation function is low (for example, 0.4) or high (for example, 0.95) and comment.

3. Filter the noisy ECG signal in the file `ecg_hfn.dat` (see also the file `ecg_hfn.m`; $f_s = 1,000\text{ Hz}$) using four different Butterworth lowpass filters (individually) realized with MATLAB[®] with the following characteristics: (a) Order 2, cutoff frequency 10 Hz . (b) Order 8, cutoff frequency 20 Hz . (c) Order 8, cutoff frequency 40 Hz . (d) Order 8, cutoff frequency 70 Hz . Use “help butter” and “help filter” in MATLAB[®] to get details about the Butterworth filter.

Compare the results obtained using each of the four Butterworth filters (individually) with those obtained by synchronized averaging, and comment upon the improvement or distortion in the outputs. Relate your discussions to specific characteristics observed in plots of the signals.

4. The ECG signal in the file `ecg_lfn.dat`, with $f_s = 1,000 \text{ Hz}$, has a wandering baseline (low-frequency artifact). (See also the file `ecg_lfn.m`.) Filter the signal with the derivative-based filters described in Section 3.5.3 and study the results. Study the effect of variation of the position of the pole in the filter in Equation 3.131 on the signal.
5. Filter the signal in the file `ecg_lfn.dat` ($f_s = 1,000 \text{ Hz}$) using Butterworth highpass filters with orders 2 to 8 and cutoff frequencies 0.5 to 5 Hz . (See also the file `ecg_lfn.m`.) Study the efficacy of the filters in removing the baseline artifact and the effect on the ECG waveform itself. Determine the best compromise acceptable.
6. Design a Wiener filter to remove the artifacts in the ECG signal in the file `ecg_hfn.dat`. (See also the file `ecg_hfn.m`.) The equation of the desired filter is given in Equation 3.185. The required model PSDs may be obtained as follows:

Create a piecewise linear model of the desired version of the signal by concatenating linear segments to provide P, QRS, and T waves with amplitudes, durations, and intervals similar to those in the given noisy ECG signal. Compute the PSD of the model signal.

Select a few segments from the given ECG signal that are expected to be isoelectric (for example, the T-P intervals). Compute their PSDs and obtain their average. The selected noise segments should have zero mean or have the mean subtracted out.

Compare the results of the Wiener filter with those obtained by synchronized averaging and lowpass filtering.