Introduction

*0

Signals

Signals are **variations in energy** that carry information. The variable that carries the information (the specific energy fluctuation) depends upon the type of energy involved. For example, if the energy is electrical, the variable is a voltage or a current, while if the energy is mechanical, the variable is force or velocity (or position).

Interaction or communication with a system or between systems is done through signals. Irrespective of the type of system, its scale, or its function, we must have some way of interacting with any system of interest to us.

*0

Energy	Variables (Specific Fluctuation)	Common Measurements
Chemical	Chemical activity and/or concentration	Blood ion, O_2, CO2, pH, hormonal concentrations and other chemistry.
Mechanical	Position Force, torque or pressure	Muscle movement, Cardiovascular pressures, muscle contractility Valve and other cardiac sounds
Electrical	Voltage (potential energy of charge carriers) Current (charge carrier flow)	EEG, ECG, EMG, EOG, ERG, EGG, GSR
Thermal	Temperature	Body temperature, thermography

*0

Signal Processing

Signal processing is transformation of a signal to another representation.

Why signal processing?

1) To get more information from the signal.

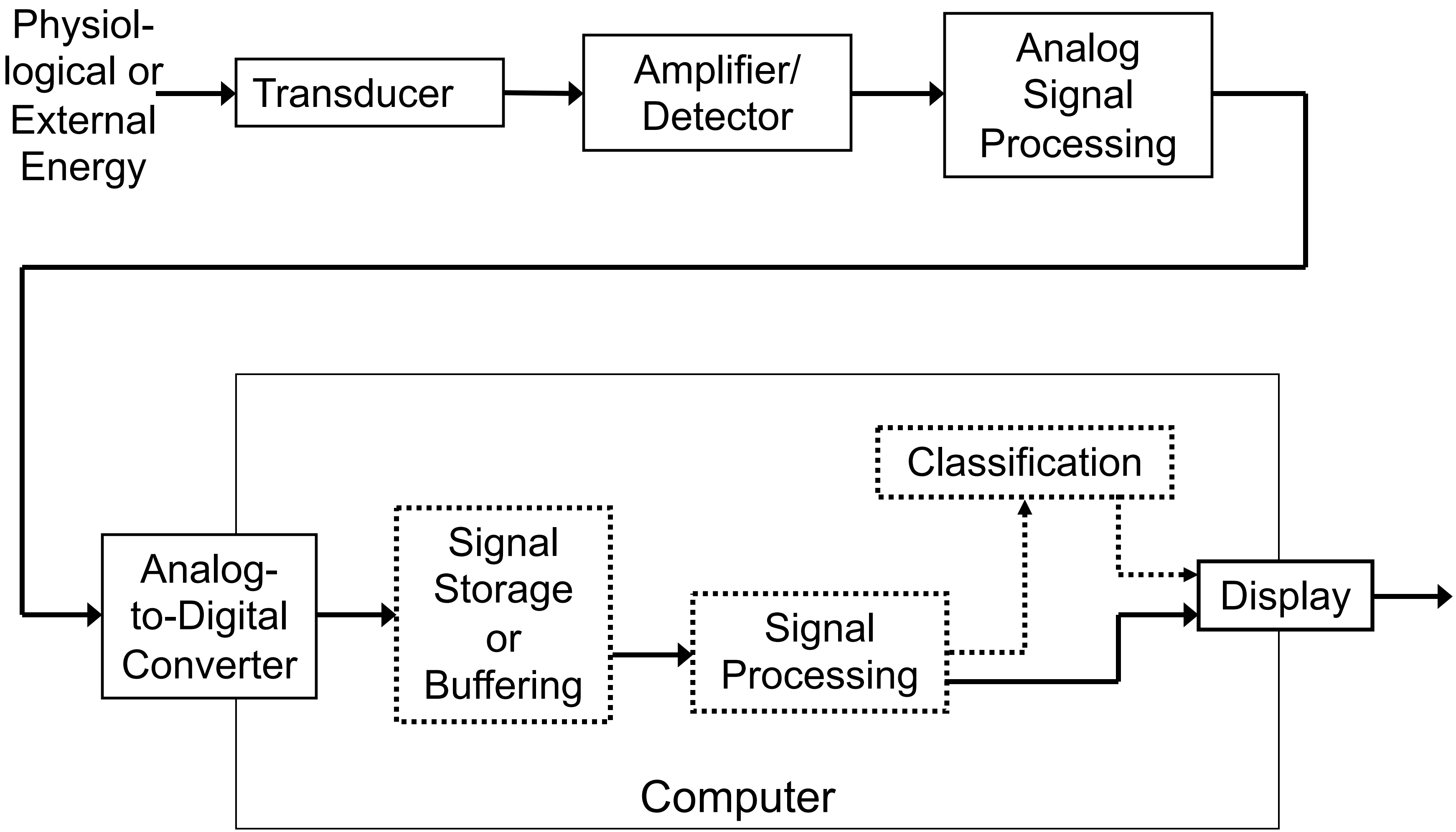
2) To reduce noise (make signal more reliable).

3) To Associate the signal with a disease (classification)

4) To reduce signal bandwidth requirements.

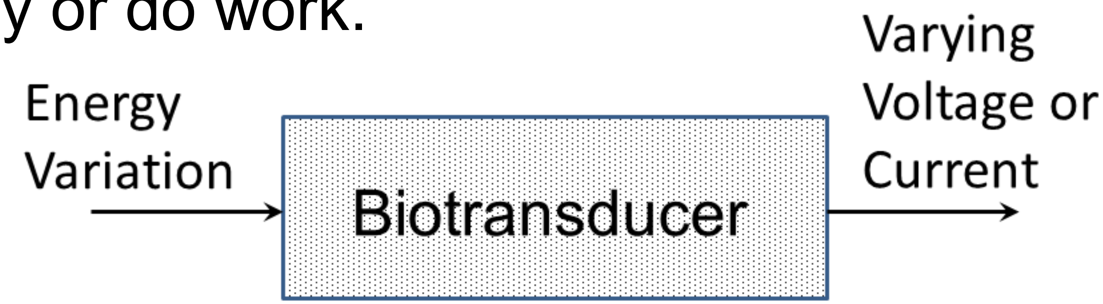
.

Biosignal Measurement Systems



Transducers

- A transducer is a device that converts energy from one form to another.
- In measurement systems, all transducers are so-called input transducers: they convert non-electrical energy into an electronic signal.
- In signal-processing applications, the purpose of energy conversion is to transfer information, not to transform energy or do work.



- **Input transducers convert signals from other energy modalities (mechanical, chemical, thermal, even other types of electrical) into electrical signals.**

*0

Energy Forms and Related Measurements

Energy	Measurement
Mechanical Length, position, and velocity Force and pressure	muscle movement, cardiovascular pressures, muscle contractility valve and other cardiac sounds
Heat	body temperature, thermography
Electrical	EEG, ECG, EMG, EOG, ERG, EGG, GSR
Chemical	ion concentrations

Amplifier/Detector

- The design of the first analog stage in a biomedical measurement system depends on the basic transducer operation.
- If the transducer is based on a variation in an electrical property, then must converted into a variation in voltage by the first stage
- Detector circuits can be single-ended or differential depending on the type of detector.
  1. If there is only one sensing element in the transducer the transducer signal is single-ended.
  2. Multiple sensors can generate a differential signal where two signals move in opposite directions.

Analog Signal Processing, Filters

- Filters are circuits specifically designed to **alter** the frequency characteristics of a signal in a well-defined manner.

- The purpose is to produce an output signal that has reduced or enhanced amplitudes within specific frequency ranges.

Filters are defined by:

1) **Basic type**: lowpass; highpass, bandpass, bandstop
2) **Bandwidth**: range of nominally unattenuated frequencies
3) **Attenuation** slope
    a) Filter order related to attenuation slope
    b) Initial sharpness

*0

Filter Types

- A filter type can be illustrated by a plot of the filter's *spectrum*, that is, a plot showing how the filter treats the signal at each frequency over the frequency range of interest.

- The filter gain, also known as the transfer function, is the ratio of output signal amplitude to input signal amplitude as a function of frequency:

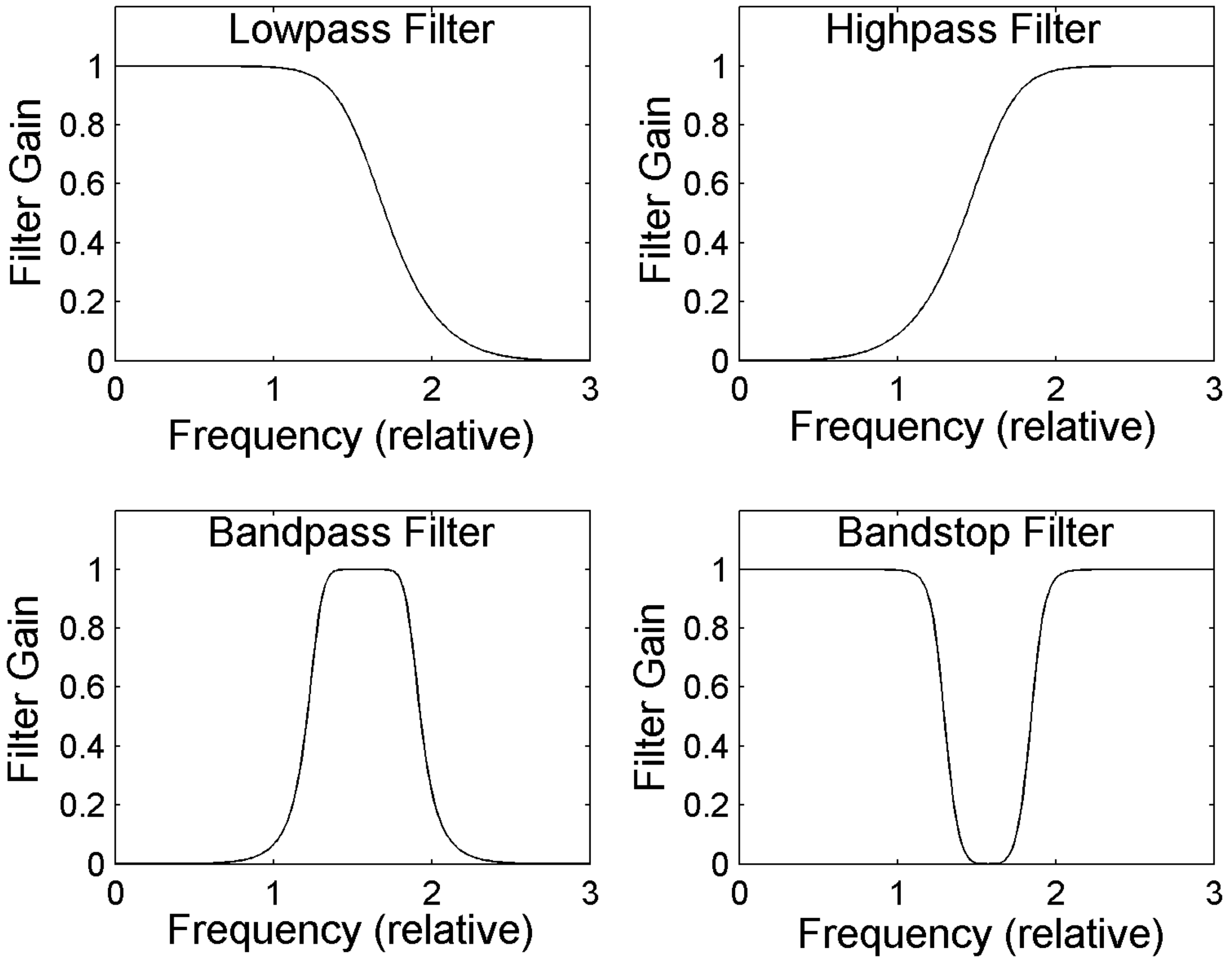
$$Gain(f) = Transfer\ \ function = \frac{Output\ Values(f)}{Input\ Values(f)}$$

- Filter types are defined in terms of the frequencies they attenuate (or do not attenuate).

The filter gain, or spectrum, of the four basic filter types is shown in the next slide.

*0

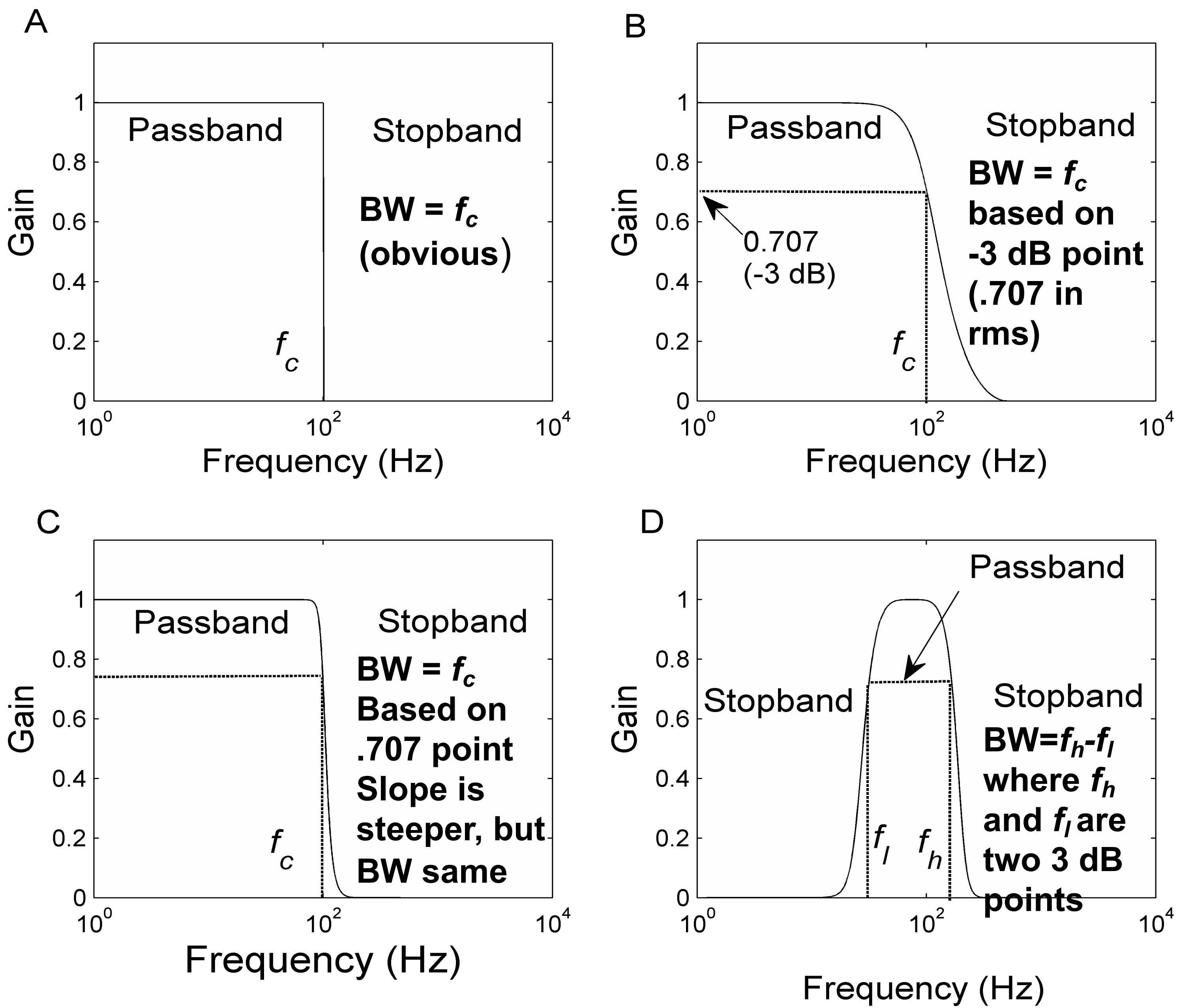
Filter Types



Except for ‘bandstop,’ filter types are named for the frequencies they “pass;” i.e., do not attenuate.

*0

Filter Bandwidth: frequency plots



The frequency plots here have linear vertical axes, but often the vertical axis is plotted in dB.

Note, the horizontal axes are in log units.

Analog-to-Digital Conversion

- It is relatively easy, and common, to convert between the analog and digital domains using electronic circuits specially designed for this purpose.

- Many medical devices acquire the data as an analog signal and convert it to a digital format

- PCs include both ADCs and *digital-to-analog converters,* DACs, as part of a sound card.

- USB compatible data transformation devices designed as general-purpose ADCs and DACs are readily available and offer greater flexibility than sound cards.

Analog-to-Digital Conversion.

- In Analog-to-Digital conversion an analog, or continuous, waveform, $x(t)$, is converted into a discrete waveform, $x[n]$, a function of real numbers that are defined only at discrete integers at discrete points in time.

- These numbers are called *samples* and the discrete points in time are usually taken at regular intervals termed the sample interval, T_s.

- The continuous signal $x(t)$ becomes just a series of numbers, $x[1]$, $x[2]$, $x[3]$,…$x[n]$ that are the signal values at times $1T_s$, $2T_s$, $3T_s$, nT_s.

Sampling Frequency

- The sample interval can also be defined by a frequency termed the sample frequency*:*

$$f_s = \frac{1}{T_s}\ \mathrm{Hz}$$

- If a waveform is sampled at T_s seconds for a total time, T_T sec, the number of values stored in the computer will be:

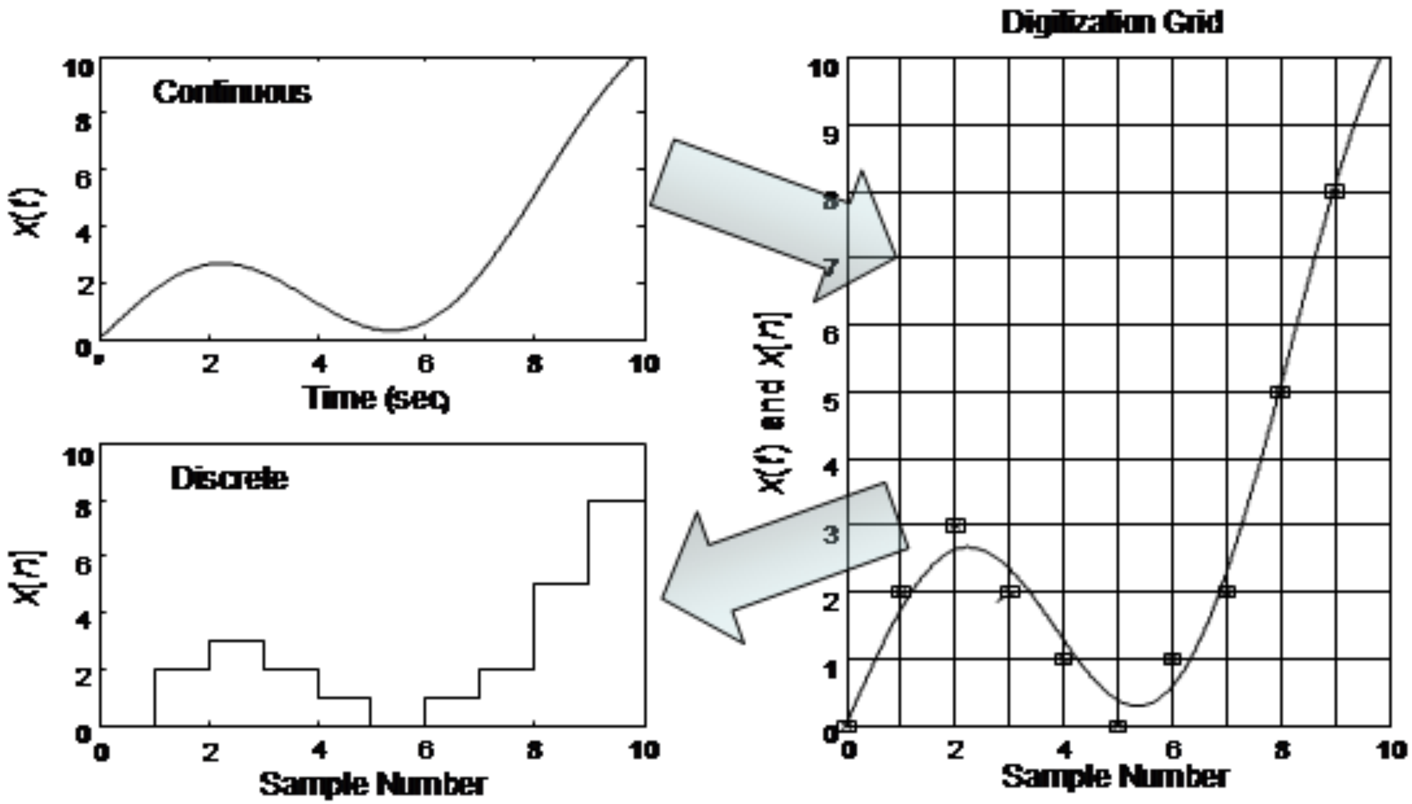
$$N = \frac{T_T}{T_s}\ \text{points}$$

- The relationship between a sample's index, *n*, and the time it was sampled is determined by:

$$t = nT_s = \frac{n}{f_s}$$

Getting the signal into a computer: Analog-to Digital Conversion

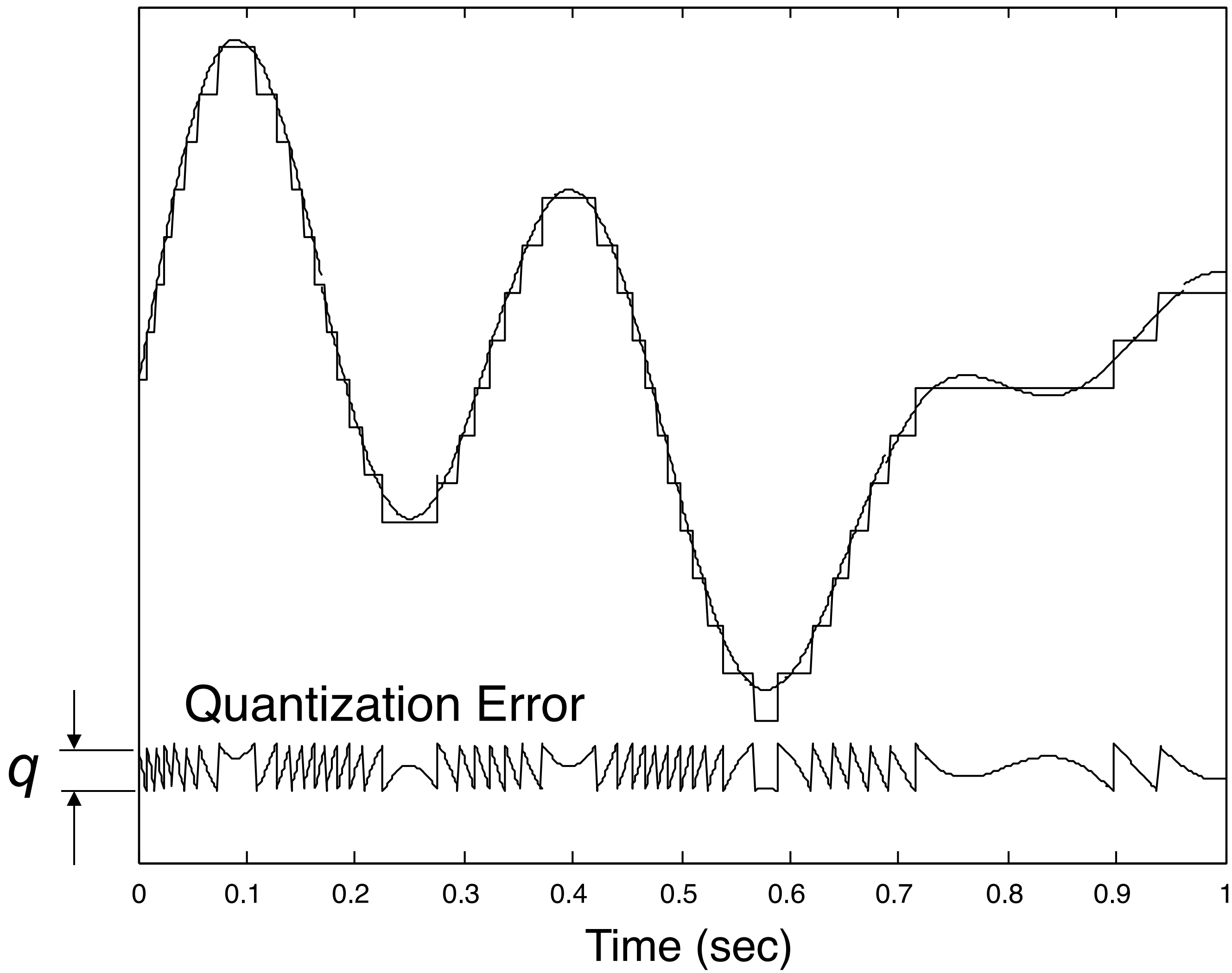
Continuous (analog) signal $\leftrightarrow$ Discrete signal $x(t) = f(t) \leftrightarrow$ Analog to digital conversion $\leftrightarrow$ $x[n] = x[1], x[2], x[3], \ldots x[n]$



Requires both time slicing and amplitude slicing.

*0

Amplitude Slicing: Quantization



The minimum voltage that can be resolved, the amplitude slice size, is known as the quantization level, *q*.

The quantization error (lower trace) looks like a type of noise.

*0

Amplitude Slicing or Quantization Errors

- The slice size in volts is the voltage range of the converter divided by the number of available discrete values, assuming all bits are converted accurately.
- If the converter produces a binary number having b accurate bits, then the number of nonzero values is:

$$(2^b-1)$$

where b is the number of bits in the binary output.

- If the voltage range of the converter varies between 0 and V_{max} volts, then the quantization step size, q, in volts, is given as:

$$q = \frac{V_{MAX}}{2^b - 1} \text{ volts}$$

where V_{max} is the maximum voltage range of the ADC and b is the number of bits converted.

*0

Quantization Level (cont)

- Typical converters feature 8-, 12-, and 16-bit outputs, although some high-end audio converters use 24 bits.

- Most biological signals do not have sufficient signal-to-noise ratio to justify a higher resolution; you are simply obtaining a more accurate conversion of the noise in the signal.

- For example, given the quantization level of a 12-bit ADC, the dynamic range is $2^{12} - 1 = 4095$: in dB this is $20 * \log(4095) = 72$ dB. Since typical biological signals have dynamic ranges rarely exceeding 40-50 dB, and are often closer to 30 dB, a 12-bit converter with the dynamic range of 72 dB is usually adequate.

*0

Quantization Error (cont)

- The variance or mean square error of the quantization noise can be determined using the expectation function from basic statistics:

$$\sigma^2 = \overline{e^2} = \int_{-\infty}^{\infty} e^2 PDF(e)\,de$$

- Assuming the noise, *e* is evenly distributed between -*q*/2 and *q*/2, the variance of this error (similar to RMS) is approximated by the equation:

$$\sigma^2 = \int_{-q/2}^{q/2} \frac{e^2}{q} de = \left.\frac{e^3}{3q}\right|_{-q/2}^{q/2} = \frac{q^2}{12}$$

Time Slicing

- Slicing the signal into discrete points in time is termed *time sampling* or simply *sampling*.

- Time slicing samples the continuous waveform, $x(t)$, at discrete prints in time, $1T_s$, $2T_s$, $3T_s$,… nT_s, where T_s is the sample interval.

- Since the purpose of sampling is to produce an acceptable (for all practical purposes) copy of the original waveform, then the critical issue is how well does this copy represent the original?

- The answer to this question depends on the frequency at which an analog waveform is sampled relative to the frequencies that it contains.

The Shannon Sampling Theorem

- The question of what sampling frequency should be used can best be addressed assuming a simple waveform, a single sinusoid. (Later we show that all finite, continuous waveforms can be represented by a series of sinusoids, so if we can determine the appropriate sampling frequency for a single sinusoid, we have also solved the more general problem.)

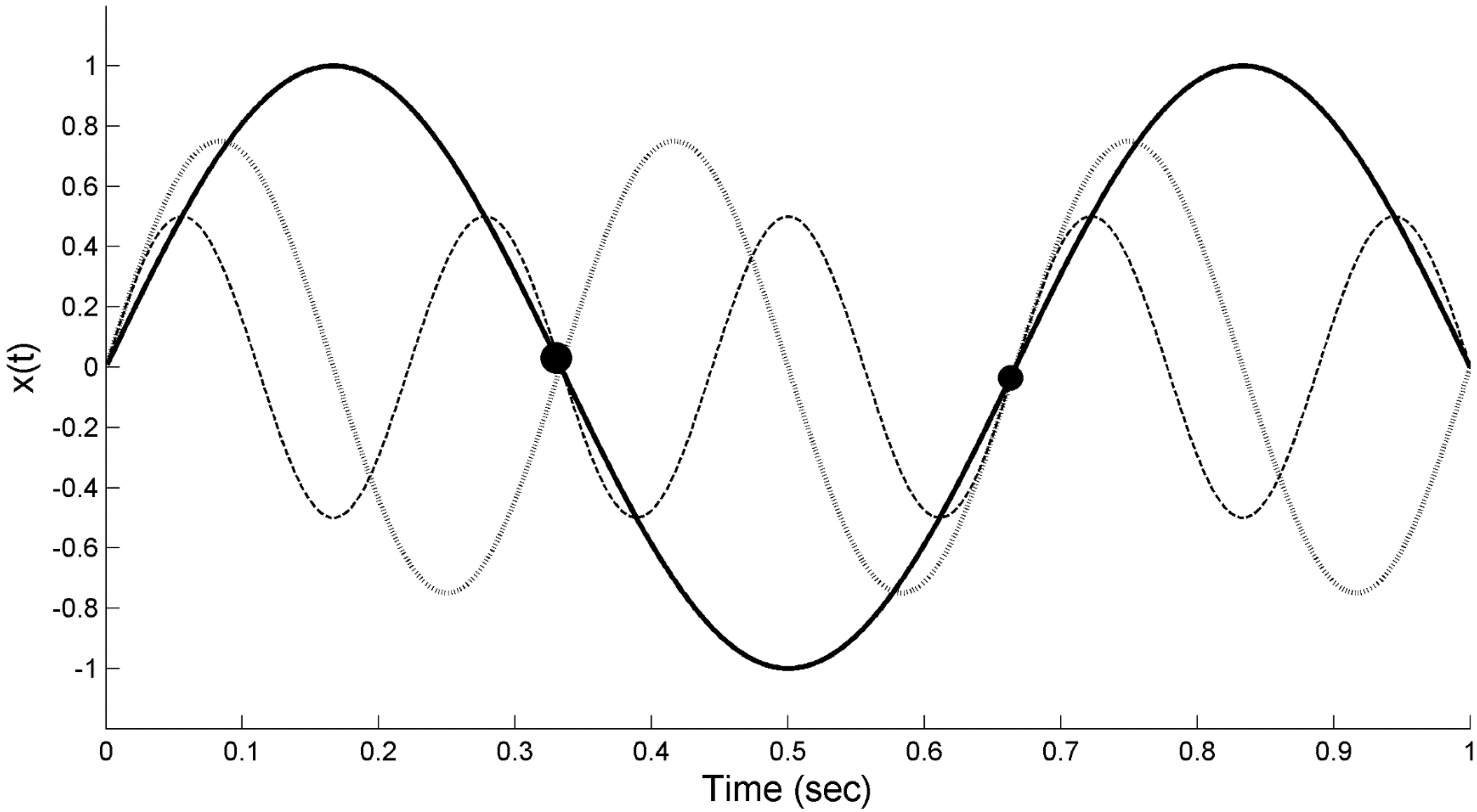
- The Shannon Sampling Theorem states that any sinusoidal waveform can be uniquely reconstructed provided it is sampled at least twice in one period. (Equally spaced samples are assumed).

- In term of frequency, the sampling frequency, f_s:

$$f_s > 2\, f_{sinusoid}.$$

*0

Two Points Define a Sine Wave



No other sine waves at a <u>lower frequency</u> can pass through the two points. The lowest-frequency sine wave is uniquely defined.

However, there is an infinite number of higher-frequency sine waves that could pass through these two points (two of which are shown here.)

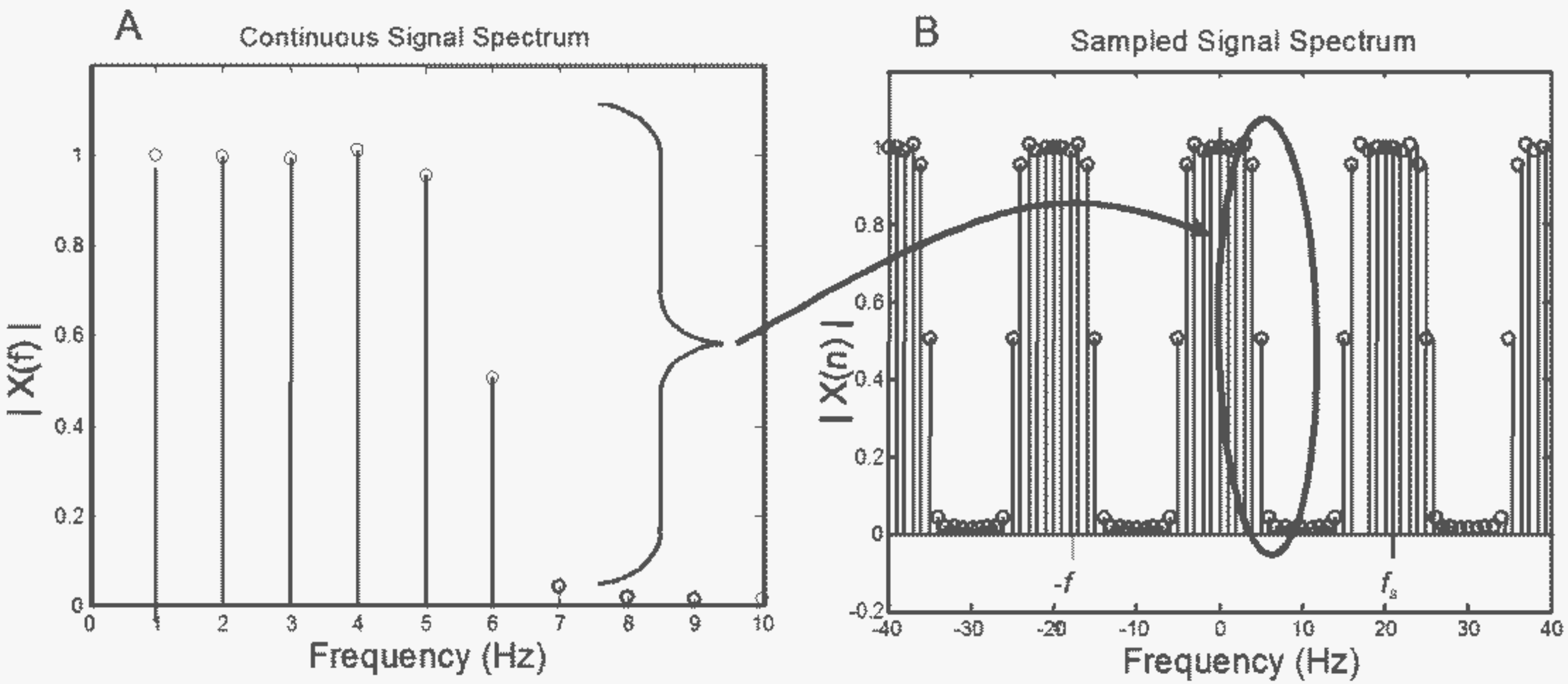
The Sampling Theorem in the Frequency Domain

- Shannon’s theorem can also be described in terms of signal frequencies: sampling has a peculiar effect on the sampled signal’s frequency characteristics or spectrum.

- The sampled signal’s spectrum contains a mirror image of the original spectrum reflected around a frequency that is half the sample frequency.

- If the sampling frequency is f_s, then the spectrum above $f_s/2$ will be the mirror image of that below $f_s/2$. The frequency $f_s/2$ is sometimes referred to as the “Nyquist frequency.”

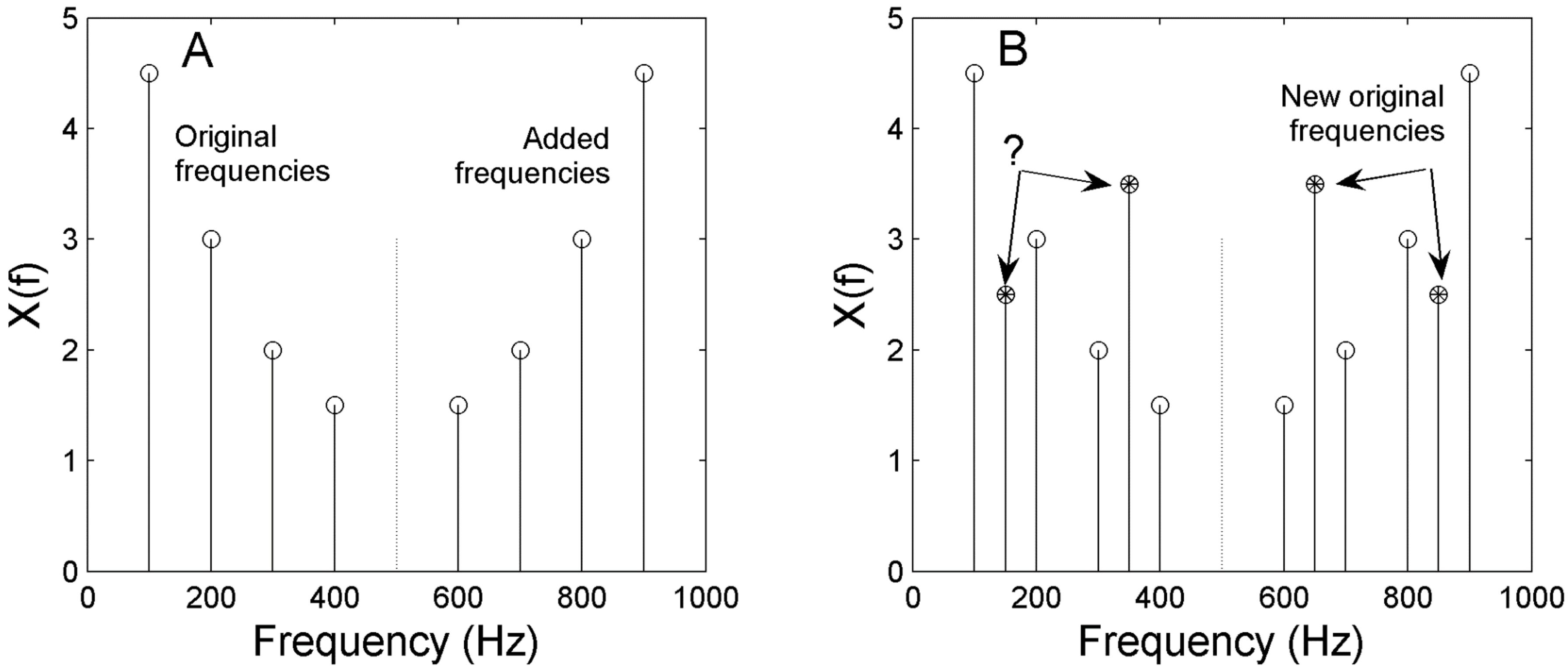
Sampled Spectrum



Sampling is a nonlinear process that produces an infinite number of additional frequencies. However, we need only be concerned with the frequencies in the range of 0 - *fs*.

*0

Sampled Spectrum: Aliasing



If the original signal contains frequencies > $f_s/2$, the added (reflected) frequencies are now less than $f_s/2$.

The generation of additional frequencies that overlap the original signal frequencies is termed "aliasing."

*0

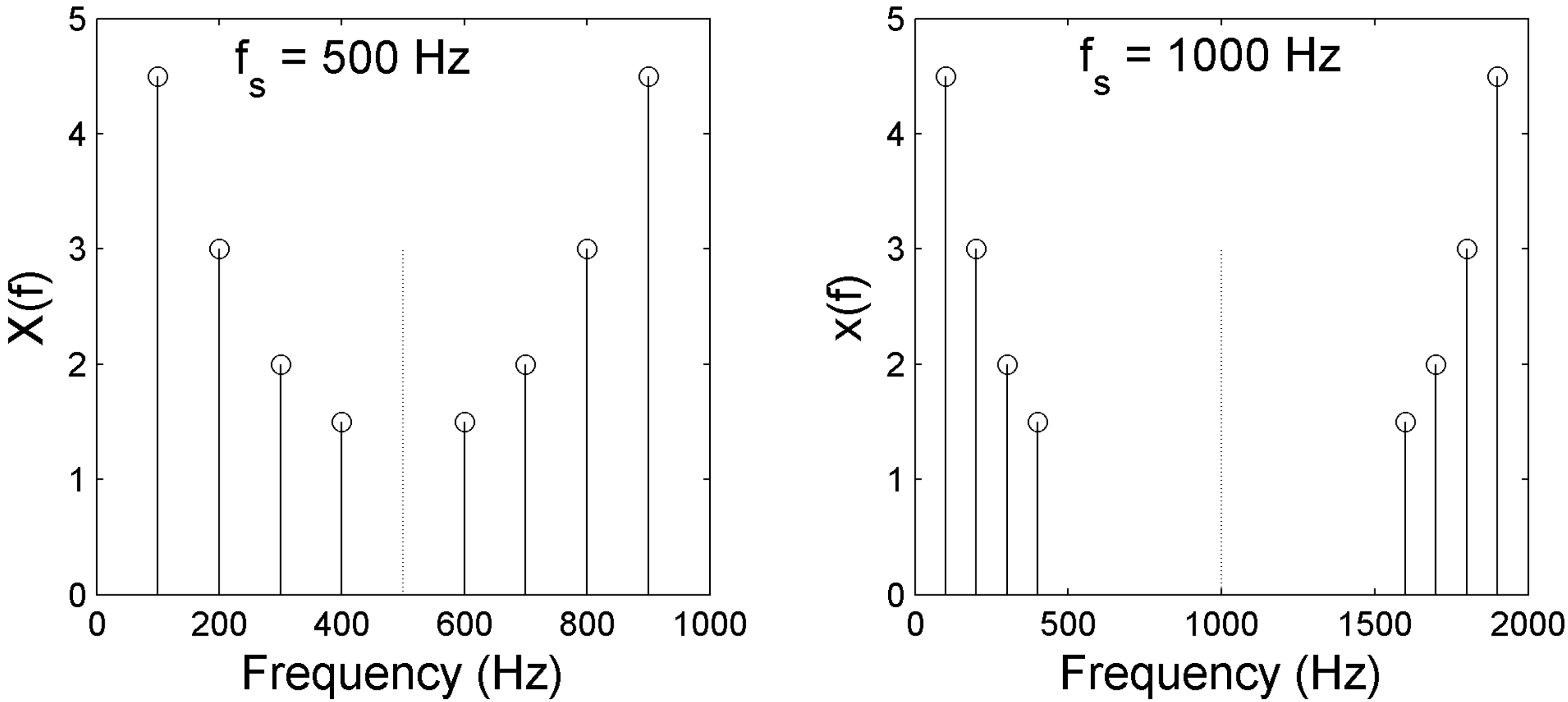
Shannon's Sample Theorem

- If the original signal contains frequencies above the Nyquist frequency, the digitized signal is corrupted.
- If the signal contains no frequencies above half the sampling frequency), the spectrum in the computer can be altered by filtering to match the original signal spectrum.
- Shannnon's Sampling Theorem states that the original signal can be recovered from a sampled version provided the sampling frequency is more than twice the <u>maximum</u> frequency contained in the original:

$$f_s > 2\, f_{max}$$

In practical situations, f_{max} is taken as the frequency above which negligible energy exists in the analog waveform.

Increasing the Sample Frequency



A higher sampling frequency (right-hand plot) provides greater separation between the original signal spectrum and the spectral components added by sampling.

Signal analysis

Biosignals

Types of Signals

- Digital signals can be classified into major subgroups: *linear* and *nonlinear*. Linear signals derive from linear processes while nonlinear signals arise from nonlinear processes such as chaotic or turbulent systems.

- These two signal classes can be further divided into *stationary* and *nonstationary* signals. Stationary signals have consistent statistical properties over the data set being analyzed.

- Applicable analytical tools depend on the signal that is being used.

*0

Analog Signal

- Analog encoding was common in consumer electronics, but most of these applications now use digital encoding. (The strange resurgence of vinyl records is a notable exception.)

- Analog encoding is important to the biomedical engineer, because most biotransducers generate analog encoded signals.

*0

Digital Signals

- A continuous analog signal after conversion to the digital domain is represented by a series of discrete samples (numbers) at specific points in time:

  $X[n] = x[1], x[2], x[3], \ldots x[n]$

- Usually this series of numbers would be stored in sequential memory locations: $x[1]$ followed by $x[2]$, , etc.

- In this case, the memory index number, n, relates to the time associated with a given sample given:

$$t = nT_s = \frac{n}{f_s}$$

where f_s is the sample frequency.

*0

Time Invariance

- If a system's response characteristics do not change over time, that is, its statistical properties are constant, it is said to be *time-invariant*.

- Time invariance is a stricter version of stationarity since a time-invariant system would also be stationary.

- Mathematically: if f is a linear function, then for time invariance:

$$y(t-T) = f\left(x(t-T)\right)$$

Note that time-invariant signals should not be confused with time-varying, that is, signals that fluctuate in amplitude. Time-invariant signals can still be time-varying.

LTI Systems

- A system that is both linear and time-invariant is referred to as a *linear time-invariant* (*LTI*) system.

- The LTI assumptions allow us to apply a powerful array of mathematical tools known collectively as linear systems analysis or linear signal analysis.

- Most living systems change over time, they are adaptive, and they are often nonlinear, but the power of linear systems analysis is sufficiently seductive that simplifying assumptions or approximations are made so that these tools can be used.

Causality

- A system that responds only to current and past inputs is termed *causal.*
- Systems that exist in the real-world (e.g., analog electronic filters) must be causal.
- Computer programs can operate on stored in the computer using values that appear to be in the future with respect to a given operation.
- Such systems are *noncausal.*

*0

Basic Signal Measurements

Mean or average value (discrete):

$$x_{avg} = \bar{x} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

Mean or average value (continuous):

$$\bar{x}(t) = \frac{1}{T}\int_0^T x(t)\, dt$$

RMS (discrete):

$$x_{rms} = \left[\frac{1}{N}\sum_{n=1}^{N} x_n^2\right]^{1/2}$$

RMS (continuous):

$$x(t)_{rms} = \left[\frac{1}{T}\int_0^T x(t)^2\, dt\right]^{1/2}$$

*0

Basic Signal Statistics

Variance, σ^2, is a measure of signal variability similar to RMS:

$$\sigma^2 = \frac{1}{N\text{-}1}\sum_{n=1}^{N}(x_n - \bar{x})^2 \qquad \sigma^2 = \frac{1}{T}\int_0^T \left(x(t) - \bar{x}\right)^2 dt$$

Standard deviation, σ, is the square root of variance: :

$$\sigma = \left[\frac{1}{N\text{-}1}\sum_{n=1}^{N}(x_n - \bar{x})^2\right]^{1/2} \qquad \sigma = \left[\frac{1}{T}\int_0^T \left(x(t) - \bar{x}\right)^2 dt\right]^{1/2}$$

Signal-to-Noise Ratio

- The Signal-to-noise ratio or SNR is simply the ratio of signal to noise, both measured in RMS (root-mean-squared) amplitude. The SNR is often expressed in dB where:

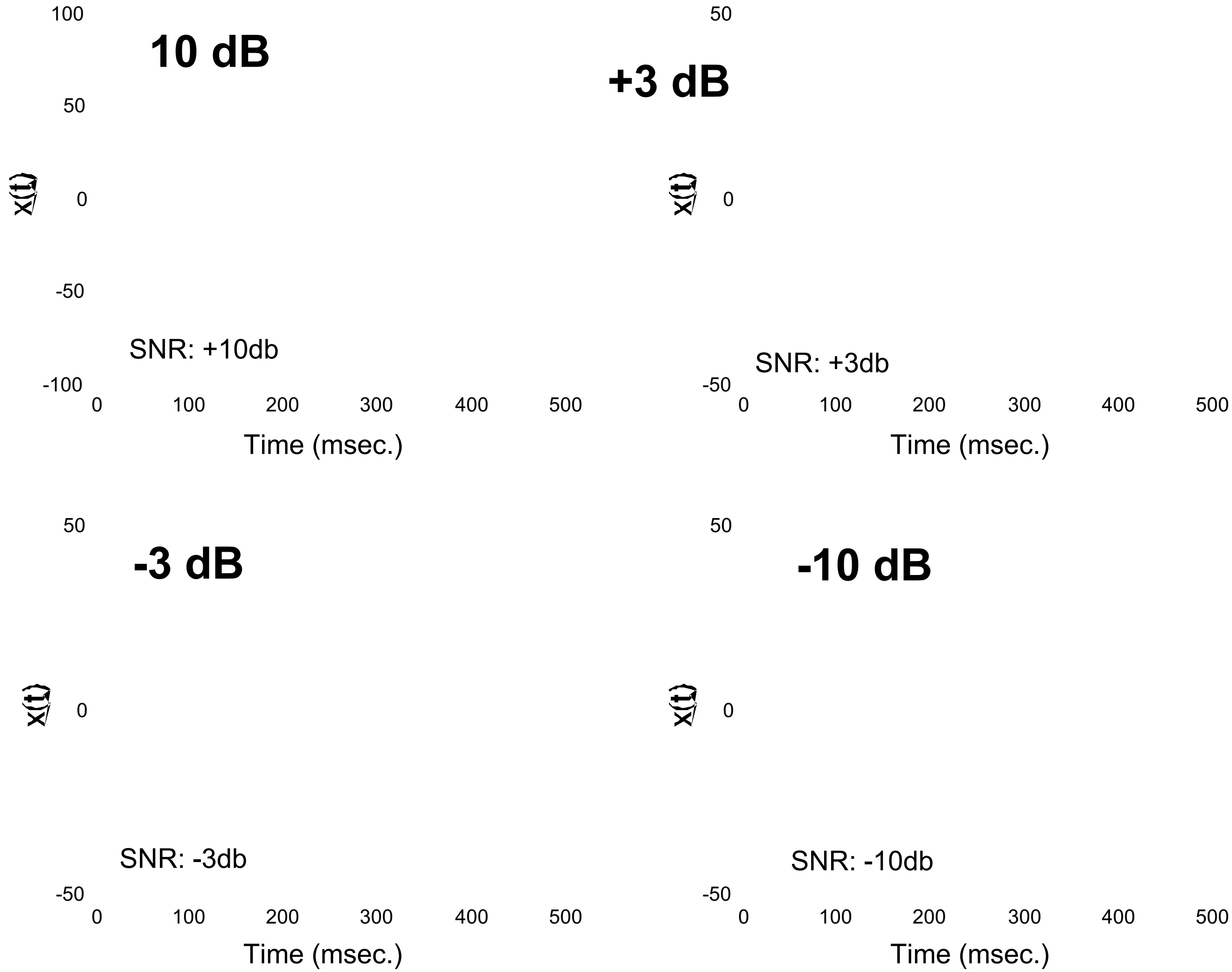
$$\mathrm{SNR} = 20 \log\left(\frac{\mathrm{signal}}{\mathrm{noise}}\right)$$

- To convert from dB scale to a linear scale:

$$\mathrm{SNR}_{\mathrm{Linear}} = 10^{\mathrm{dB}/20}$$

*0

Example of Different SNR Levels



A sinusoid with 4 levels of added noise. SNR's in dB.

*0

Noise and Variability

What is noise? Noise is what you do not want and **signal** is what you do want (noise is unwanted variability). Noise often limits the usefulness or information content of a signal

Sources of Variability

Source	Cause	Potential Remedy
Physiological variability	Measurement only indirectly related to variable of interest	Modify overall approach
Environmental (internal or external)	Other sources of similar energy form	Noise cancellation, transducer design
Artifact	Transducer responds to other energy sources	Transducer design
Electronic	Thermal or shot noise	Transducer or electronic design

Noise and Variability

Variability really is noise, but the word is often used to indicate fluctuation between, as opposed to within, measurements.

A variety of signal processing tools exist to reduce noise.

The more that is known about the characteristics of the noise, the more powerful are the signal processing methods that can be applied.

*0

Frequency analysis

*0

The Frequency Domain
Probing a Signal

- Complex signals such as the EEG signal shown previously could be analyzed by comparison with reference signals
- Crosscorrelation provides a running comparison between the original signal and a reference signal.
- Crosscorrelation can be used to find simple signals such as sinusoids embedded in a complicated signal.
- For example, crosscorrelating the EEG signal with sinusoids having frequencies we think may be embedded in the signal.
- If the crosscorrelation function shows a high value at some time shift (τ), that would suggest the presence of our sinusoid, or other reference signal, at that time shift (or, equivalently, at that phase shift).

*0

Crosscorrelation with Sinusoids

EEG signal

One reference sinusoid (2 Hz)

5
0
-5

0 1 2 3 4 5 6 7 8

Time (sec)

Crosscorrelation of EEG and 2 Hz sine

Crosscorrelation

20
10
0
-10
-20

-20 -15 -10 -5 0 5 10 15 20

Lags(n)

Peak crosscorrelation for different reference signals at 25 different frequencies.
(1-25 Hz)

Peak indicates more energy at related frequency

Maximum Correlation

200
150
100
50
0

0 5 10 15 20 25 30

Frequency (Hz)

Frequency Domain Transformations

- Many biological signals are more useful in research or diagnosis when viewed in the *frequency domain*.
- Determining the frequency content of a waveform is termed *spectral analysis.*
- Spectral analysis decomposes a time-domain waveform into its constituent frequencies just as a prism decomposes light into its spectrum of constituent colors.
- Various techniques convert signals into their spectral equivalent with different strengths and weaknesses.
- Spectral analysis methods are divided into two broad categories: the Fourier transform, and modern methods such as those based on models of the signal’s source.

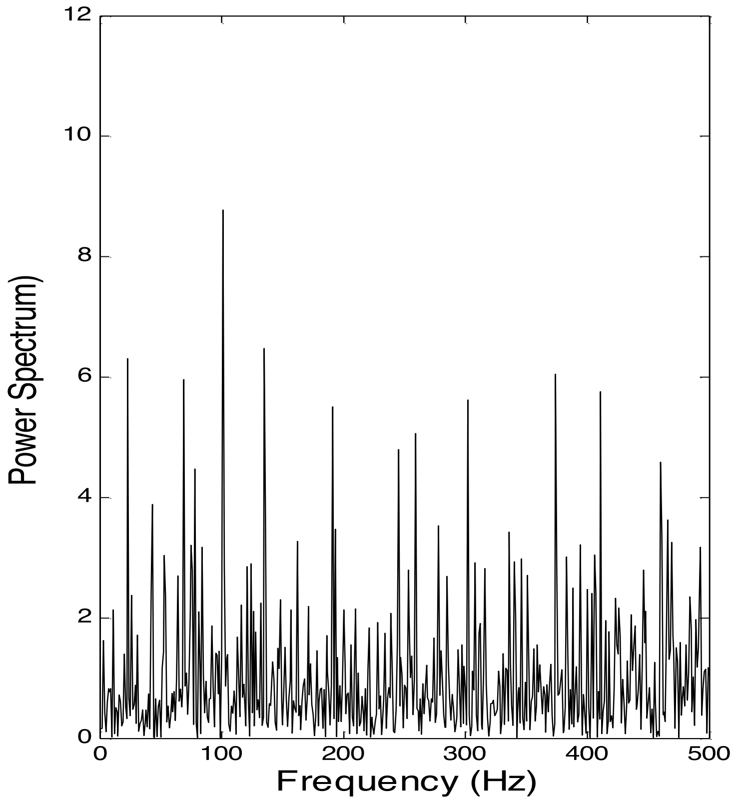
Limitations in Spectral Analysis

- Accurate spectral analysis requires the signal be periodic, or finite, and noise-free.
- Spectral analysis of real signals having short segments and noise are only <u>estimates</u> of the true spectrum.
- Intelligent spectral analysis requires an understanding of what spectral features are of interest.
- Two spectral features of potential interest are:

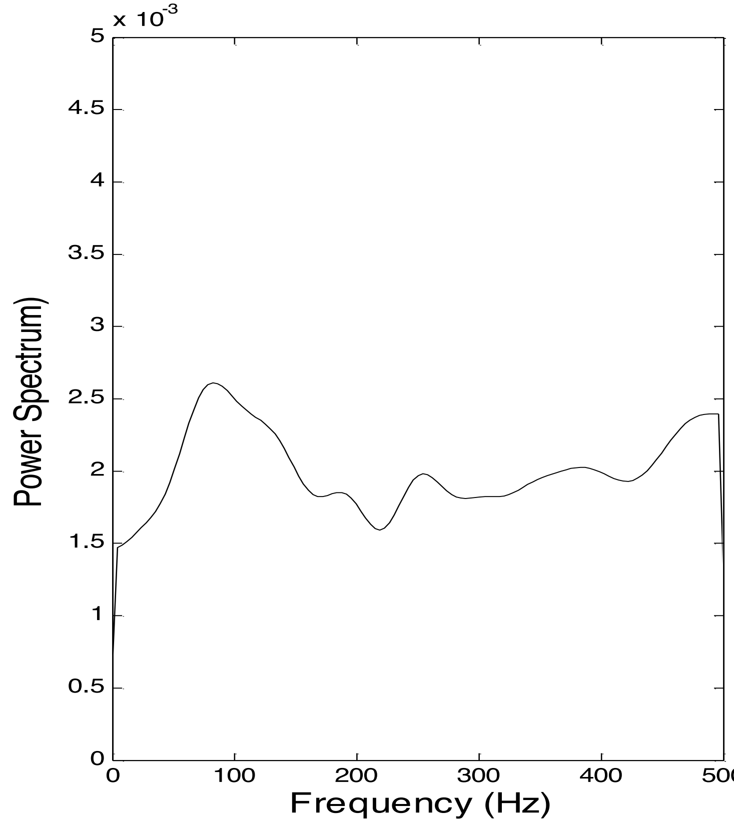
1. Overall shape of the spectrum or *broadband* features, termed the spectral estimate.
2. Local features of the spectrum or *narrowband* features, sometimes referred to as parametric estimates.

*0

Two spectra obtained from a waveform consisting of a 100-Hz sine wave and white noise using two different methods.



The spectrum produced by the Fourier transform clearly shows the signal as a spike at 100 Hz. However other spikes are present that are just noise.



An averaging technique creates a spectrum represents the white noise (which should be flat), but the 100-Hz component is missing

*0

Fourier Series Analysis
Harmonic Sinusoidal Decomposition

- A sinusoid contains energy at only one frequency.

- Sinusoidal components of a signal map directly to the frequency components of a signal.

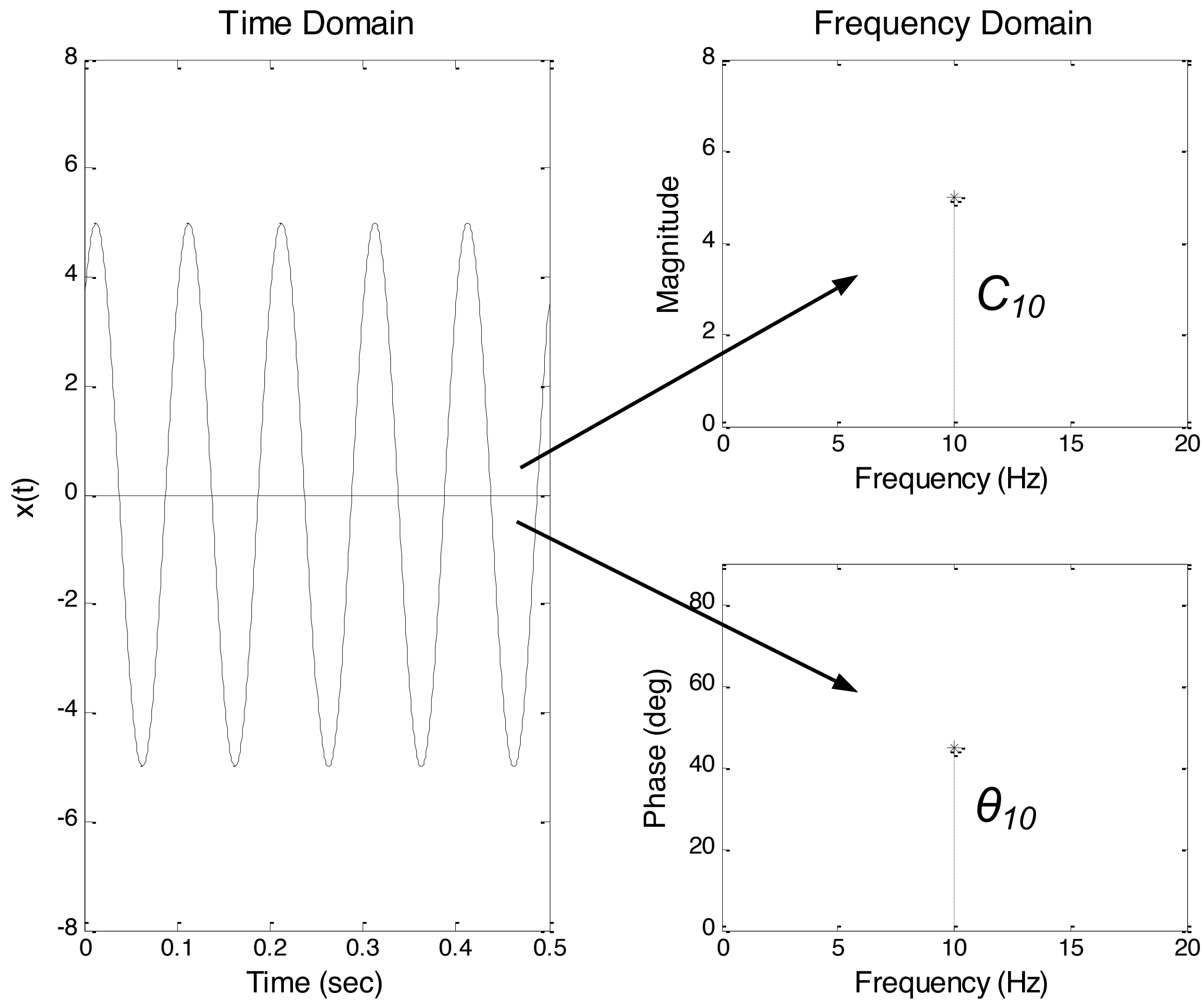
Sinusoidal Components $\Leftrightarrow$ = Frequency Components

*0

Each sinusoidal component gives us a single point on the two frequency curves: the magnitude and phase curve.

The frequency of these points is the same as the sinusoid, specifically:

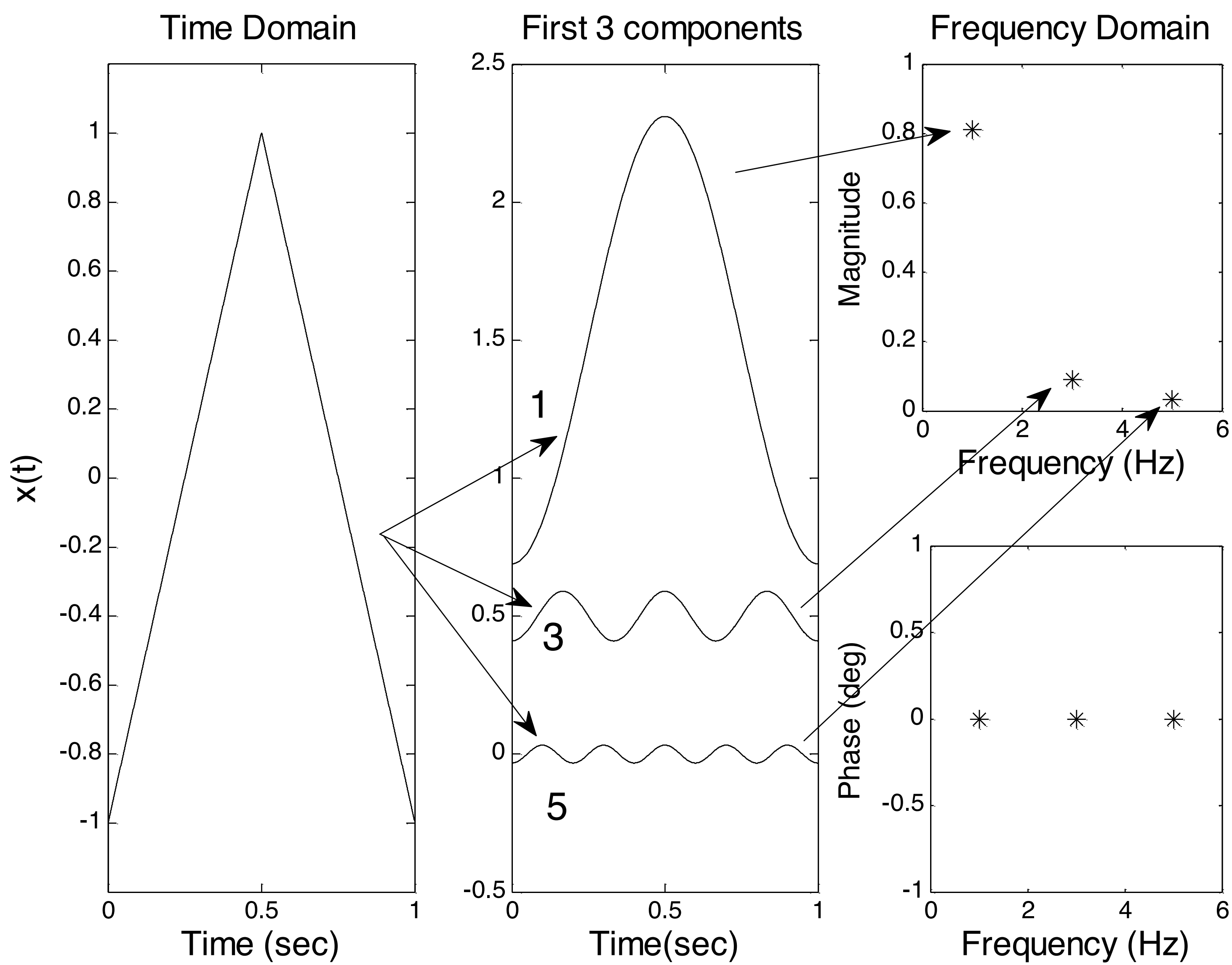
$$f_m = \frac{2\pi m}{T} = 2\pi m f_1$$



Note that the frequency is a function of *m*.

Correlation with a 10 Hz sinusoid gives the magnitude and phase characteristics of $x(t)$ at 10 Hz.

Fourier Decomposition is a Gateway to the Spectrum



*0

Power Spectrum

- The Power Spectrum is commonly defined as the Fourier Transform of the autocorrelation function. In continuous and discrete notation, the Power Spectrum equation becomes:

$$PS(f) = \int_{0}^{T} r_{xx}(t)\, e^{-j2\pi n f_1 t}\, dt$$

$$PS[n] = \sum_{n=0}^{N\text{-}1} r_{xx}[n] e^{-j2\pi n f_1 T_S}$$

where $r_{xx}(t)$ and $r_{xx}[n]$ are autocorrelation functions as described in Chapter 2.

*0

Power Spectrum (cont)

Since the autocorrelation function has even symmetry, the sine terms of the Fourier series are all zero, and the two equations can be simplified to include only real cosine terms:

$$PS(f) = \frac{1}{T}\int_0^T r_{xx}(t)\ \cos(2\pi mft)\ dt \qquad\qquad m = 0,1,2,3,\ldots$$

$$PS[m] = \sum_{n=0}^{N\text{-}1} r_{xx}[n]\cos\left(2\pi nm/N\right) \qquad\qquad m = 0,1,2,3,\ldots N$$

These equations are sometimes referred to as cosine transforms.

*0

Power Spectrum (continued)

- In the direct approach, the power spectrum is calculated as the magnitude squared of the Fourier transform of the waveform of interest:

$$PS(f) = |X(f)|^2$$

- The power spectrum does not contain phase information so the power spectrum is not a bilateral transformation -- it is not possible to reconstruct the signal from the power spectrum.
- Since the power spectrum does not contain phase information, it is applied in situations where phase is not considered useful.

Noise reduction and digital filters

Noise Reduction and Digital Filters

Noise Reduction

- The equation for this *moving average* method is:

  $y[n] = x[n]/3 + x[n-1]/3 + x[n-2]/3)$

  where $x[n]$ is the original signal and $y[n]$ is the new signal with reduced noise.

- Intuitively we can see that this averaging process lessens the influence of outlying samples and smooths sharp transitions.

*0

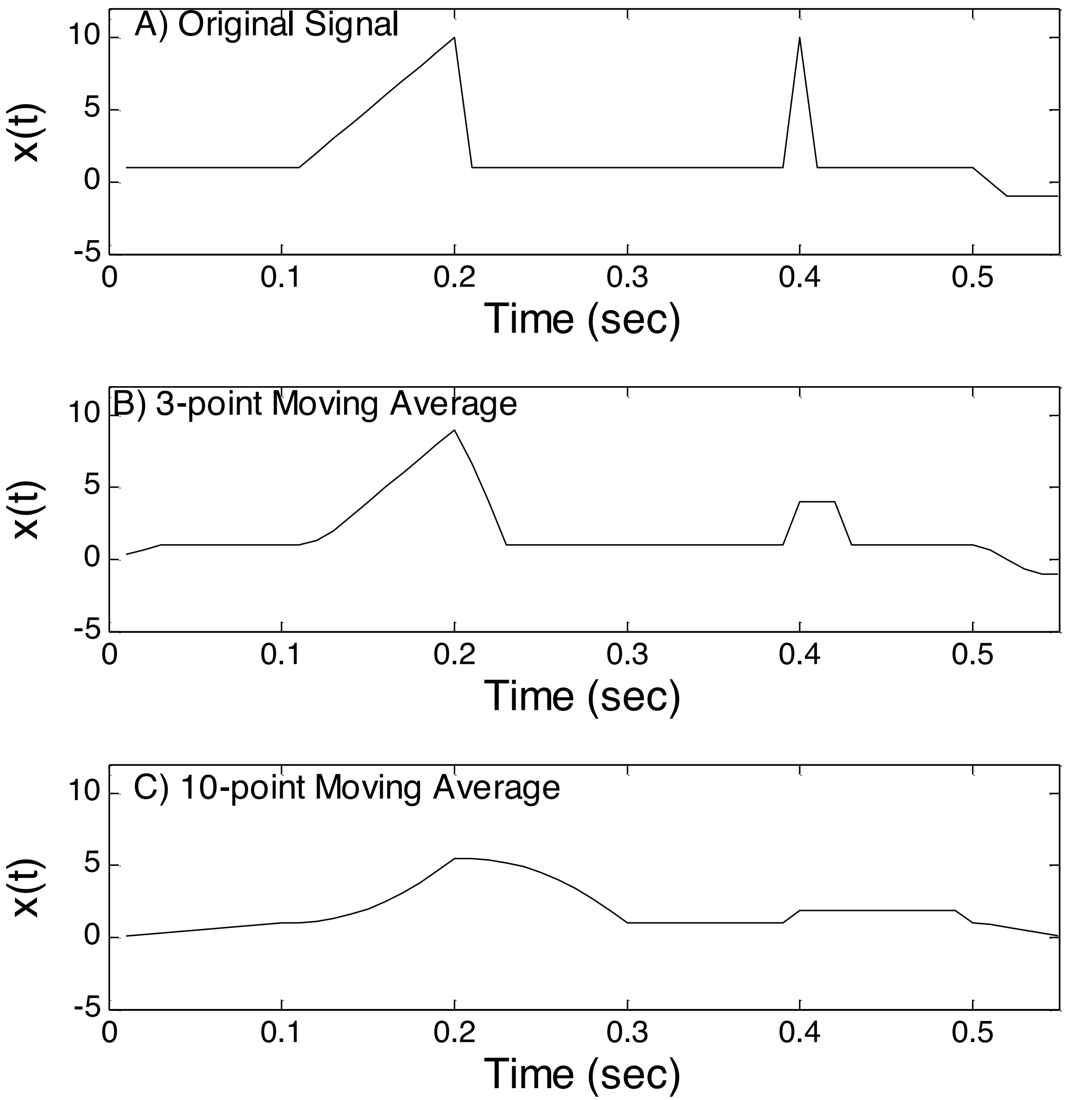
Moving Averages

Averaging more points generates greater smoothing.

Averaging is the fundamental concept behind all of the noise-reduction techniques presented in this chapter.

The weights need not all be the same.

Weights may be applied as a moving average or recursively, so that samples already averaged are gone over again.



*0

Ensemble Averaging

- In ensemble averaging, entire signals are averaged together.
- Ensemble averaging is a simple yet powerful signal processing technique for reducing noise when multiple observations of the signal are possible.
- Such multiple observations could come from multiple sensors, but in many biomedical applications the multiple observations come from repeated responses to the same stimulus.
- In ensemble averaging, a new signal with a lower noise is constructed by averaging point-by-point over all signals in the ensemble.
- There are two essential requirements for the application of ensemble averaging for noise reduction:

  1. The ability to obtain multiple observations;
  2. A reference signal closely time-linked to the response.

Noise Reduction through Ensemble Averaging

- If these measurements are combined or added together, the means add, so the combined value or signal has a mean that is the average of the individual means.
- The same is true for the variance: the variances add, and the average variance of the combined measurement is the mean of the individual variances:

$$\bar{\sigma}^2 = \frac{1}{N} \sum_{n=1}^{N} \sigma_n^2$$

- When several signals are added together, it can be shown that the noise components are reduced by a factor equal to $1/\sqrt{N}$, where N is the number of measurements that are averaged.

$$\overline{\sigma_{AVG}} = \frac{\sigma}{\sqrt{N}}$$

Digital Filters

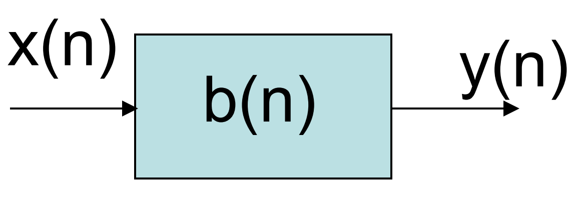
- Construction of a filter with the desired bandwidth is known as “filter design.”
- Filters are designed in the frequency domain, but implemented in the time domain (convolution).
- There are two basic filter types:

  1) FIR (Finite Impulse Response): Standard linear process: defined in the time domain by its impulse response. Frequency = fft (impulse response);

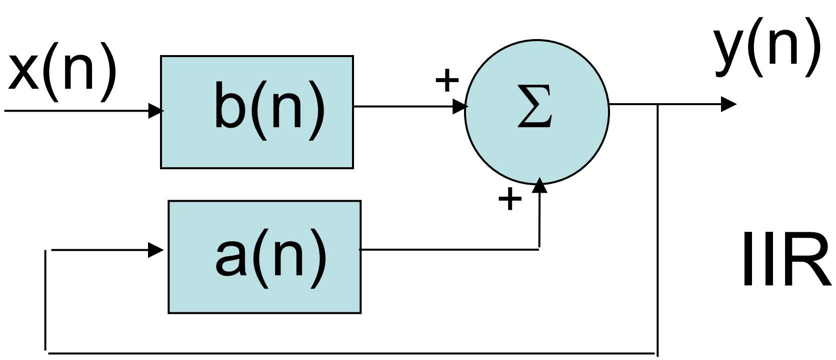
  2) IIR (Infinite Impulse Response): has additional memory-based components; defined in the time domain by a difference equation; requires additional tools for development and implementation. Frequency = fft(num)/fft(demon).

Digital Filters

- Digital filters are implemented by convolution:
  1. FIR (Finite Impulse Response) Filters: a set of filter “weights” or “coefficients,” b(n), is *convolved* with the input signal;
  2. IIR (Infinite Impulse Response) Filters: a set of filter weights, b(n), is convolved with the input signal, and another set of weights, a(n), is *convolved* with the output signal (a past version of the signal).



FIR



IIR

Digital Transfer Function

From discrete time to frequency

The Digital transfer function is analogous to, but not the same as, the analog transfer function that uses the complex variable *s* except that the z complex variable is used:

$$H[z] = \frac{Y[z]}{x[z]}$$

Typical Digital Transfer Functions will have numerators and denominators that are powers of *z*.

$$H[z] = \frac{b[0] + b[1]z^{-1} + b[2]z^{-2} + \cdots + b[K]z^{-K}}{1 + a[1]z^{-1} + a[2]z^{-2} + \cdots + a[L]z^{-L}}$$

Note *a*(1) always equals 1.0

Frequency Domain Implementation

From the digital Transfer Function, H(*z*), it is possible to determine the output given any input.

$$H(z) = \frac{\sum_{k=0}^{K} b[k] z^{-k}}{\sum_{\ell=0}^{L} a[\ell] z^{-\ell}}$$

It is easy to find the frequency spectrum of *H*(*z*)

Simply substituting $z = e^{-j\omega}$ into $H(z) = Y(z)/X(z)$:

$$H(m) = \frac{Y(\mathrm{m})}{X(m)} = \frac{\sum_{n=0}^{N\text{-}1} b(n)\,\mathrm{e}^{(-j2\pi mn/N)}}{\sum_{n=0}^{D\text{-}1} a(n)\,\mathrm{e}^{(-j2\pi mn/N)}} = \frac{\mathrm{fft}(b_n)}{\mathrm{fft}(a_n)}$$

Time Domain Implementation

$$Y[z]=H[z]x[z]=\frac{\sum_{k=0}^{K} b[k]z^{-k}}{\sum_{\ell=0}^{L} a[\ell]z^{-\ell}}$$

Multiplying by the denominator and noting $a[0] = 1$

$$Y[z]\sum_{\ell=1}^{L} a[\ell]z^{-n} = X[z]\sum_{k=0}^{K} b[k]\,z^{-n}$$

The input-output or *difference equation* analogous to the time domain equation of a linear system can be obtained by applying the time-shift interpretation to z^{-n} ,i.e.:

$$Z[\,x[n-k]\,] = Z^{-k}\,Z[x[n]\,]$$

to the equation above.

$$y[n]=\sum_{k=0}^{K-1} b[k]x[n-k]-\sum_{\ell=1}^{L} a[\ell]y[n-\ell]$$

Fundamental Filter Equation

Components

$$y[n] = \sum_{k=0}^{K-1} b[k]x[n-k] - \sum_{\ell=1}^{L} a[\ell]y[n-\ell]$$

Convolution with input
(standard linear system *b*(*k*) is impulse response)

Convolution with past output

(recursive component. Recall *a*(0) = 1)

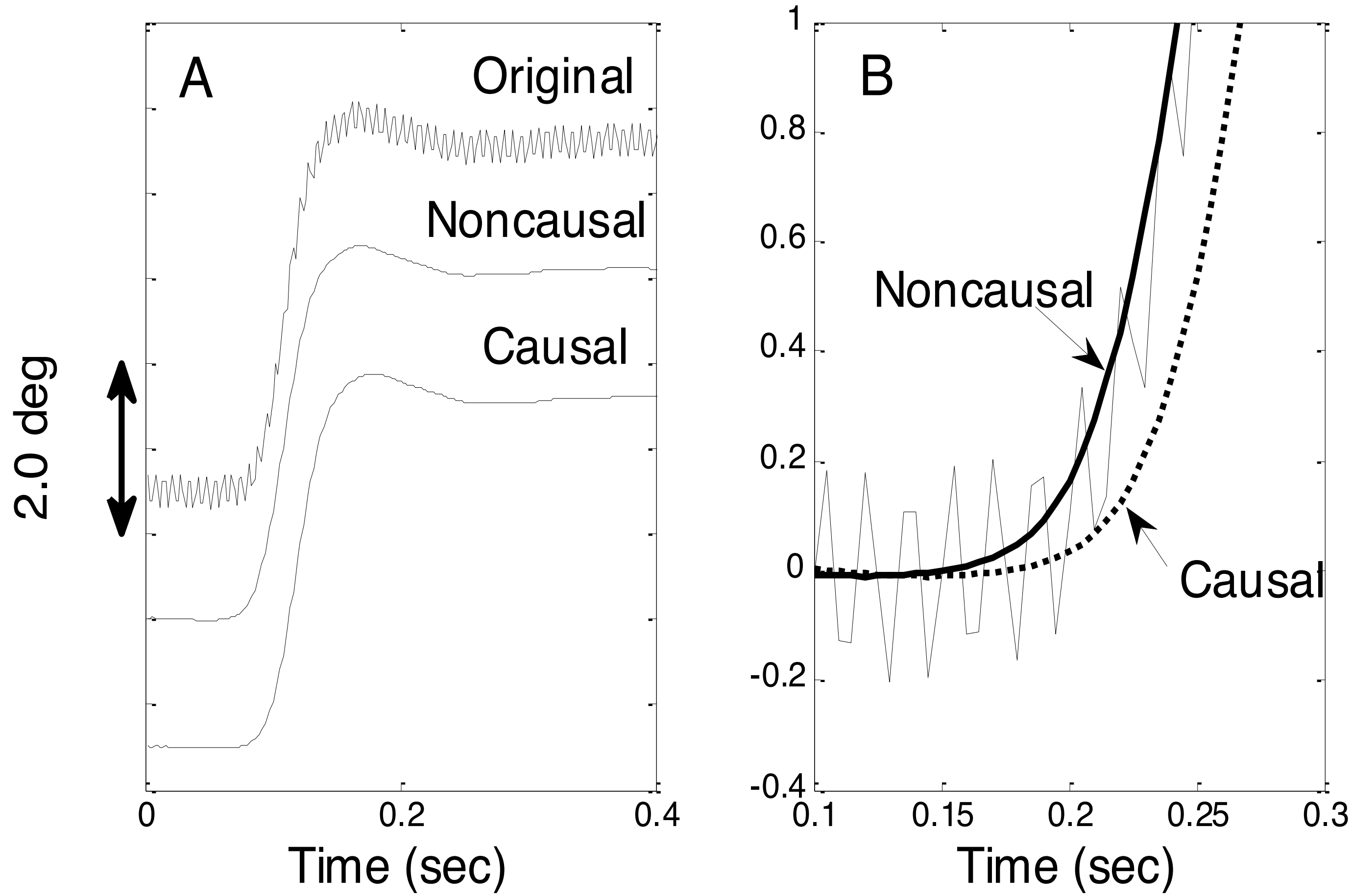
Causal and Non-Causal Filters

- A filter that uses only current and past data is a causal filter.
- All real-time systems (that operate on incoming data) must be causal since they do not have access to the future; they have no choice but to operate on current and past values of a signal.
- If the data are already stored in a computer, it is possible to use future signal values along with current and past values to compute an output signal; that is, future with respect to a given sample in the output signal.
- Filters (or systems) that use future values of a signal in their computation are noncausal.

Causal and Non-Causal Filters

This figure illustrates the advantage of noncausal filters.

A noisy biological signal is filtered with a 10 moving average filter using future and past samples (noncausal) and only past samples (causal).



The causal filter introduces a slight delay as shown in the blow-up right-hand image. The noncausal filter has no added delay.

Finite Impulse Response (FIR) Filters

The general equation for an FIR filter is a simplification of the difference equation and, after changing the limits to conform with MATLAB notation, becomes:

$$y[n] = \sum_{k=1}^{K} b[k]x[n-k]$$

The frequency response of this filter can be determined as:

$$X[m] = \sum_{k=1}^{K} b[k]e^{-j2\pi mn/N} = FT\big(b[k]\big)$$

Other FIR Filters

Rectangular window filters can highpass, bandpass, or bandstop by modifying the basic equation.

Highpass

$$b[k] = \begin{cases} \dfrac{-\sin(2\pi k f_c)}{\pi k} & k \neq 0 \\ 1 - 2 f_c & k = 0 \end{cases}$$

Bandpass

$$b[k] = \begin{cases} \dfrac{\sin(2\pi k f_h)}{\pi k} - \dfrac{\sin(2\pi k f_l)}{\pi k} & k \neq 0 \\ 2(f_h - f_l) & k = 0 \end{cases}$$

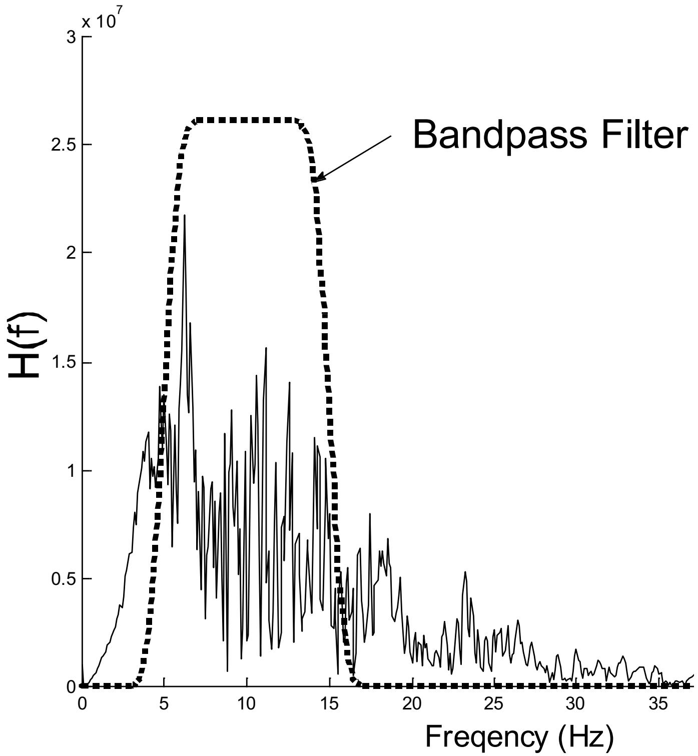
Bandstop

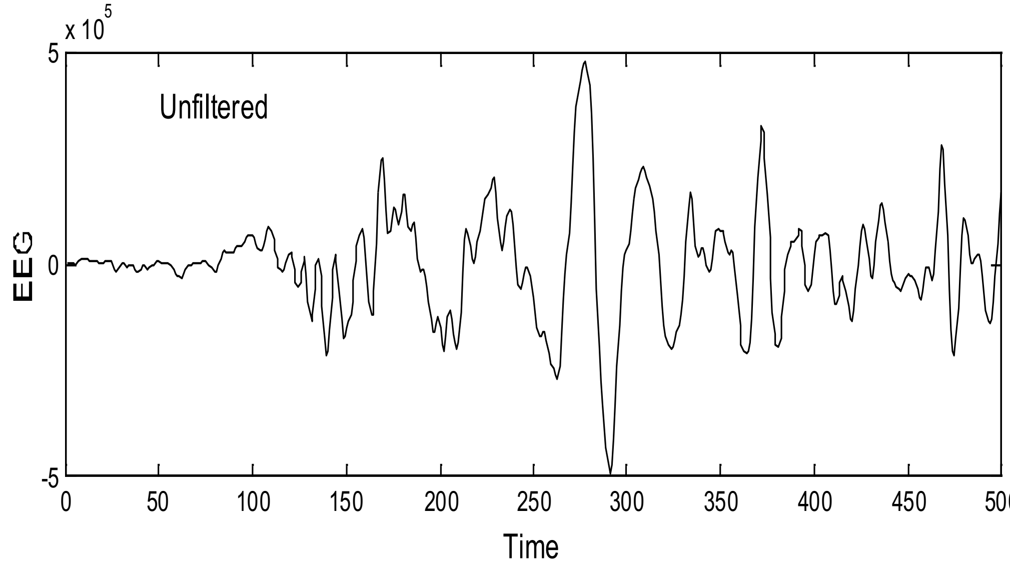
$$b[k] = \begin{cases} \dfrac{\sin(2\pi k f_l)}{\pi k} - \dfrac{\sin(2\pi k f_h)}{\pi k} & k \neq 0 \\ 1 - 2(f_h - f_l) & k = 0 \end{cases}$$

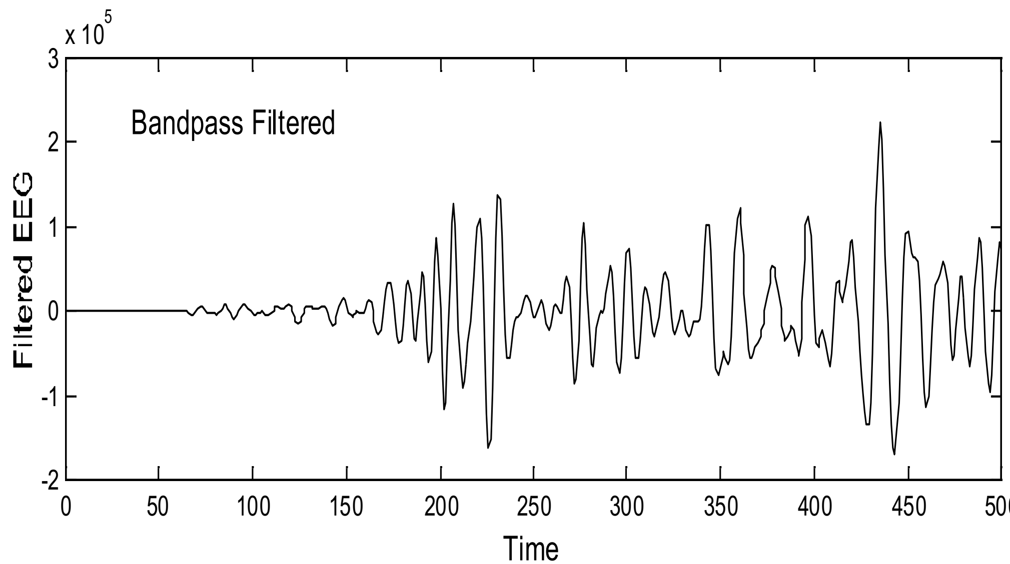
The order of highpass and bandstop filters should always be even, so the number of coefficients in these filters should be odd.

*0

Bandpass Filtering of the EEG Signal



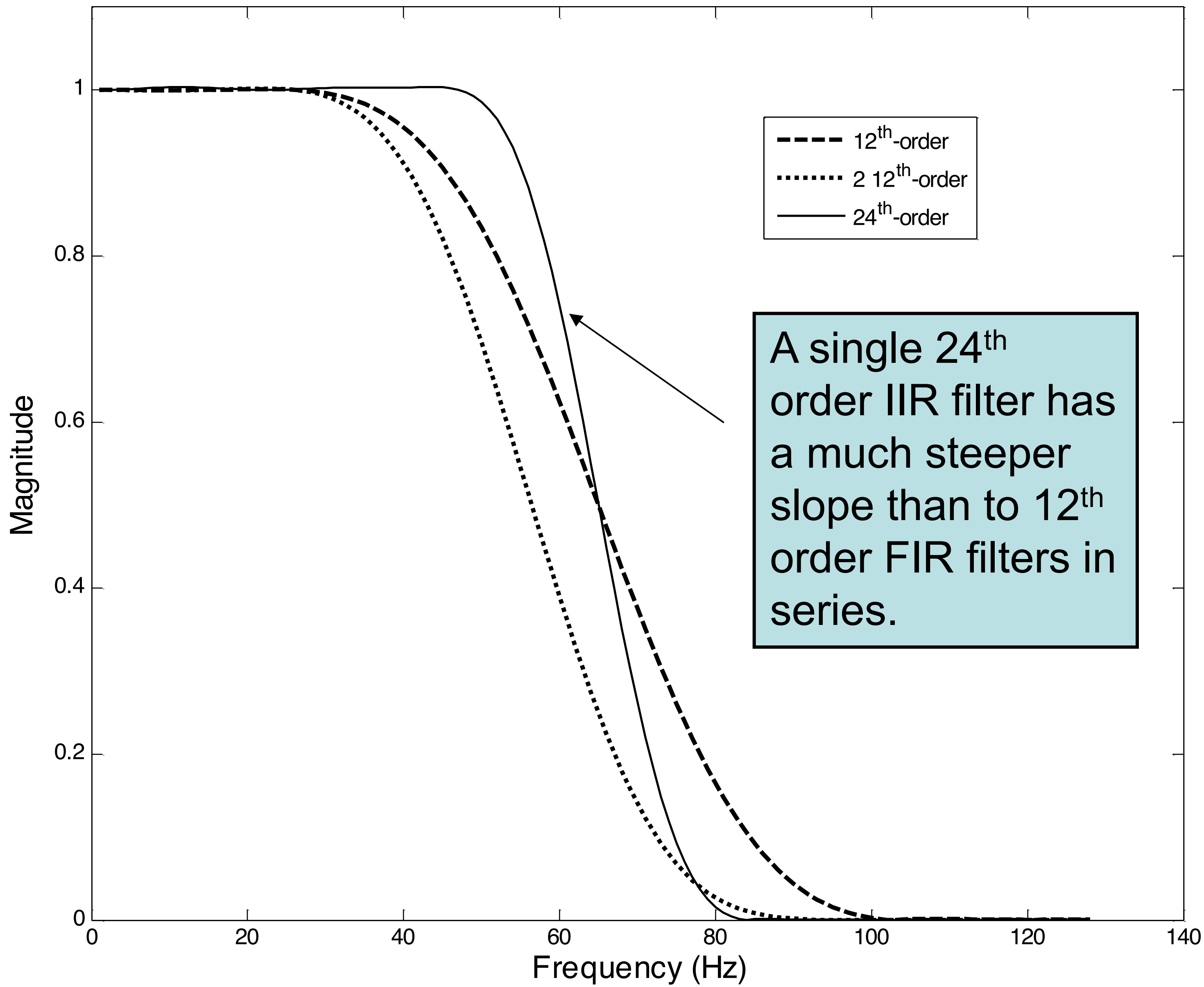




Infinite Impulse Response Filters

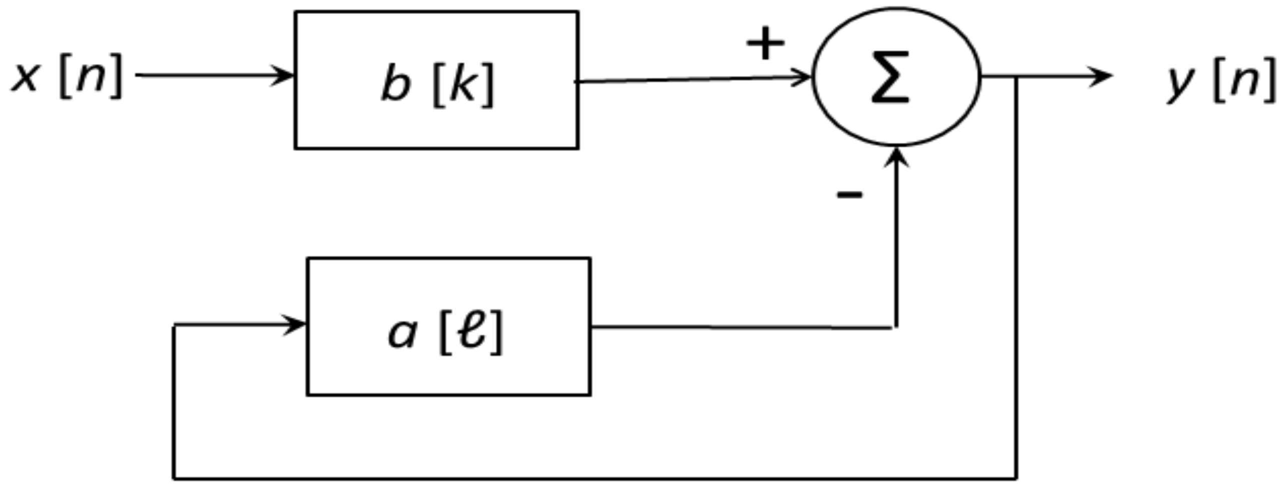
- To increase the attenuation slope of a filter, we can apply two or more filters in sequence.
- In the frequency domain, the frequency spectra multiply. This leads to a steeper attenuation slope in the spectrum.
- The next slide shows the magnitude spectrum of a 12th-order FIR rectangular window filter (dashed line) and the spectra of two such filters in series (dotted line).
- The next slide also shows that for the same computational effort you can use a single 24th-order filter and get an even steeper attenuation slope.

*0



Magnitude
Frequency (Hz)
0
0.2
0.4
0.6
0.8
1
0
20
40
60
80
100
120
140
12th-order
2 12th-order
24th-order
A single 24th order IIR filter has a much steeper slope than to 12th order FIR filters in series.

IIR Filter Configuration



- An IIR filter adds a second set of filter coefficients that get the input from the output signal.
- To avoid stability problems this second loop operates on past vales of the output signal

*0

Infinite Impulse Response (IIR) Filters

- Definition:

$$y[n] = \sum_{k=0}^{K} \underbrace{b[k]x[n-k]}_{\text{Upper path}} - \sum_{\ell=1}^{L} \underbrace{a[\ell]y[n-\ell]}_{\text{Lower path}}$$

where *b*[*n*] are the numerator coefficients also found in FIR filters, *a*[*n*] *are the denominator coefficients, x*[*n*] *is the input, and* y[*n*] the output.

While the *b*[*n*] coefficients operate only on values of the input, *x*[*n*], the *a*[*n*] coefficients operate on past values of the output and are sometimes referred to as “recursive” coefficients.

*0

FIR versus IIR Filters: Features and Applications

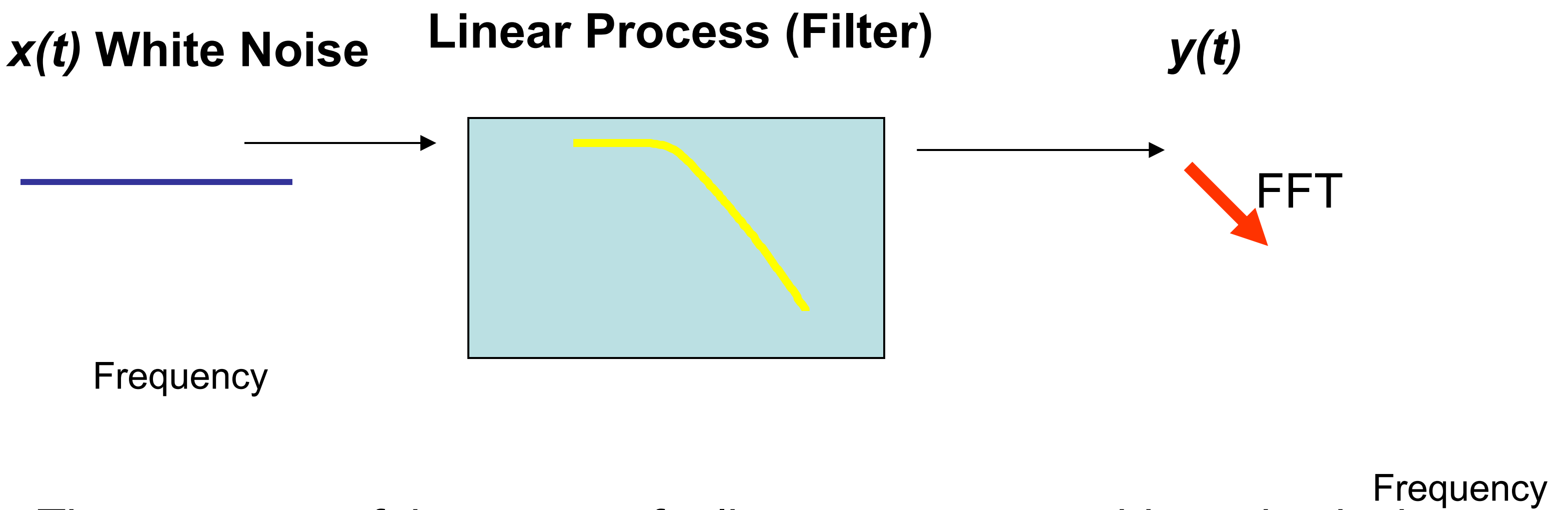
Filter type	Features	Applications
FIR	Easy to design Stable Applicable to two-dimensional data (i.e., images)	Fixed, one-dimensional filters Adaptive filters Image filtering
IIR	Require fewer coefficients for the same attenuation slope, but can become unstable. Particularly good for low cutoff frequencies. Mimic analog (circuit) filters	Fixed, one-dimensional filters, particularly at low cutoff frequencies. Real-time applications where speed is important

Modern spectral analysis

Alternative Spectral Analysis

- The FT provides a complete and accurate representation of a waveform.
- But most waveforms contain both signal and noise and the FT will include both
- The FT cannot be altered to give a simplified spectrum that may better represent the signal.
- Modern spectral methods can simplify the spectrum that might improve the signal spectrum over the noise

Recall the response of a linear process to white noise



• The spectrum of the output of a linear process to white noise is the frequency characteristics of the linear process itself.

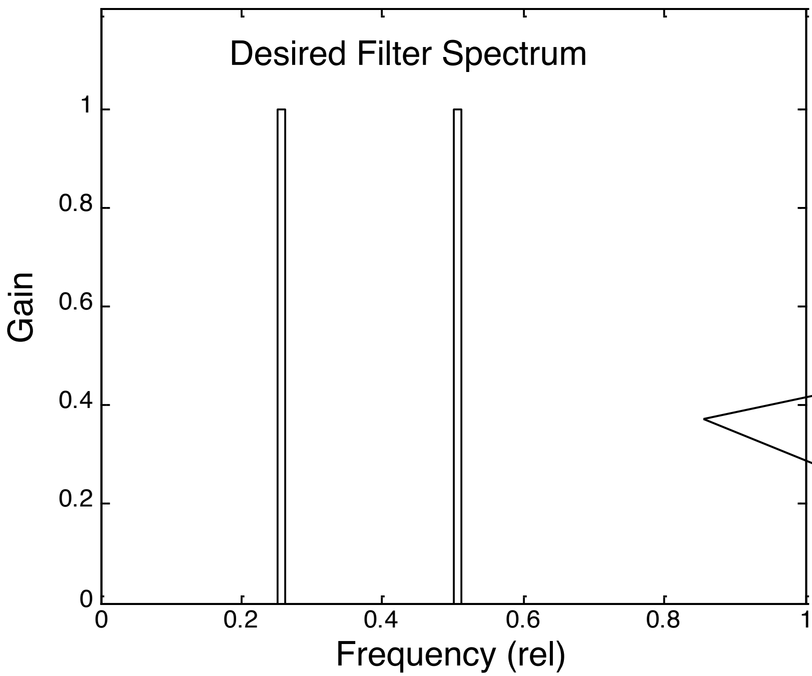
$$Y(f) = X(f)H(f) = H(f) \quad \text{if } X(f) = 1$$

• The coefficients of a linear process such as a filter are directly related to the frequency characteristics of the process through the Fourier transform.

$$H[m] = \frac{Y[\mathrm{m}]}{X[m]} = \frac{\sum_{n=0}^{N\text{-}1} b[n]\, \mathrm{e}^{(-j2\pi mn/N)}}{\sum_{n=0}^{D\text{-}1} a[n]\, \mathrm{e}^{(-j2\pi mn/N)}} = \frac{\mathrm{fft}(b_n)}{\mathrm{fft}(a_n)}$$

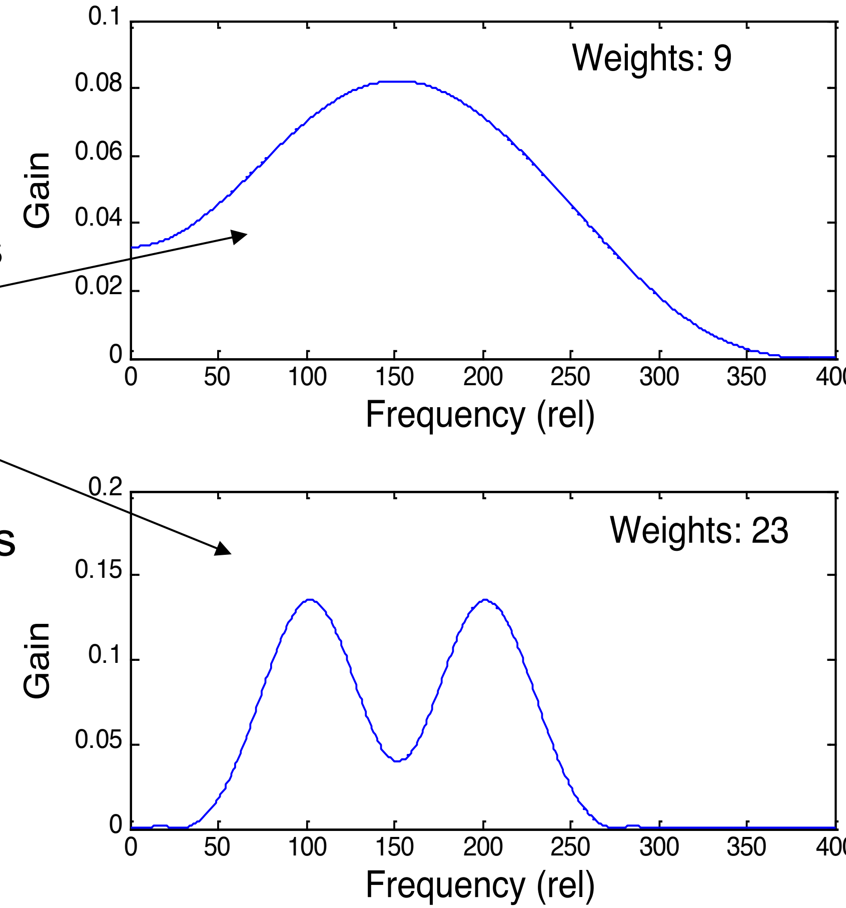
The spectral characteristics a filter is capable of producing depend on the complexity of filter (i.e., the type of filter and the number of weights.

FIR Filter spectrum created with fir2.m



9 weights

23 weights



Adjusting the filter complexity allows you to *control* spectral complexity.

*0

Approach to Model-Based Spectral Analysis

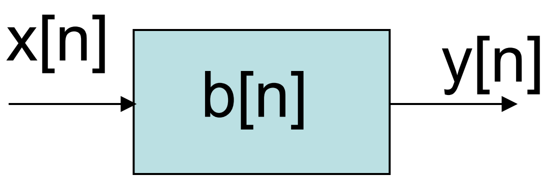
- If a set of filter parameters can be found to achieve a match between the filtered white noise and the an unknown signal, the filter parameters can be used to determine the spectrum of the unknown signal.

Filter or linear model

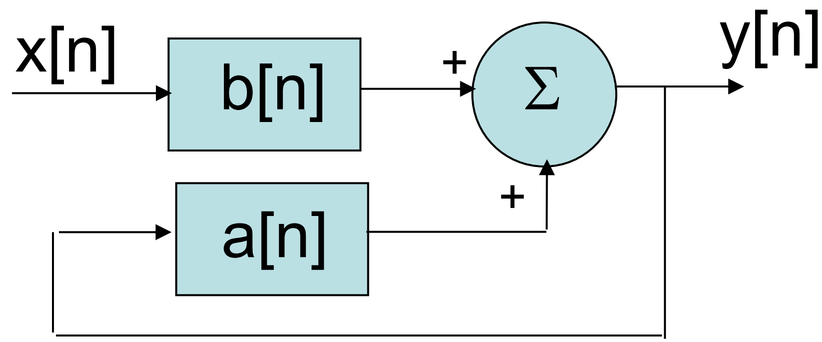
Filter complexity determines spectral complexity

(using parameters)

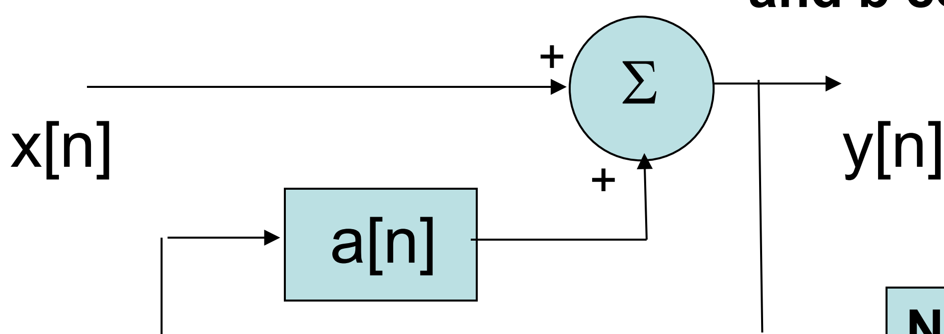
Possible Filters (Models)



(FIR)
Moving Average (MA) b coefficients only



(IIR) **Autoregressive Moving Average (ARMA) a and b coefficients**



Autoregressive (AR) a coefficients only

Not used for filtering, but could be used as a spectral model.

Filter or Model Equations

MA $$y[k]=\sum_{n=1}^{q} b[n]x[k-n]$$

ARMA $$y[k]=\sum_{n=1}^{q} b[n]x[k-n]-\sum_{n=1}^{p} a[n]x[k-n]$$

AR $$y[k]=x[n]-\sum_{n=1}^{p} a[n]x[k-n]$$ $b[0] = 1$

where $x[n]$ is the input or noise function and *p and q* is the model order.

By selecting a specific model type and model order, the spectrum that you get can be controlled.

The AR Model

- While many techniques exist for evaluating the parameters of an AR model, algorithms for MA and ARMA are less plentiful.
- In general, these algorithms involve significant computation and are not guaranteed to converge, or may converge to the wrong solution.
- The MA approach cannot model narrowband spectra well: it is not a high-resolution spectral estimator limiting its usefulness in biosignal analysis.
- Because of the computational challenges of the ARMA model and the spectral limitations of the MA model, we restrict textbook coverage to AR spectral estimation.