



ENEL 563 BIOMEDICAL SIGNAL ANALYSIS

Rangaraj M. Rangayyan



Professor

Department of Electrical and Computer Engineering
Schulich School of Engineering

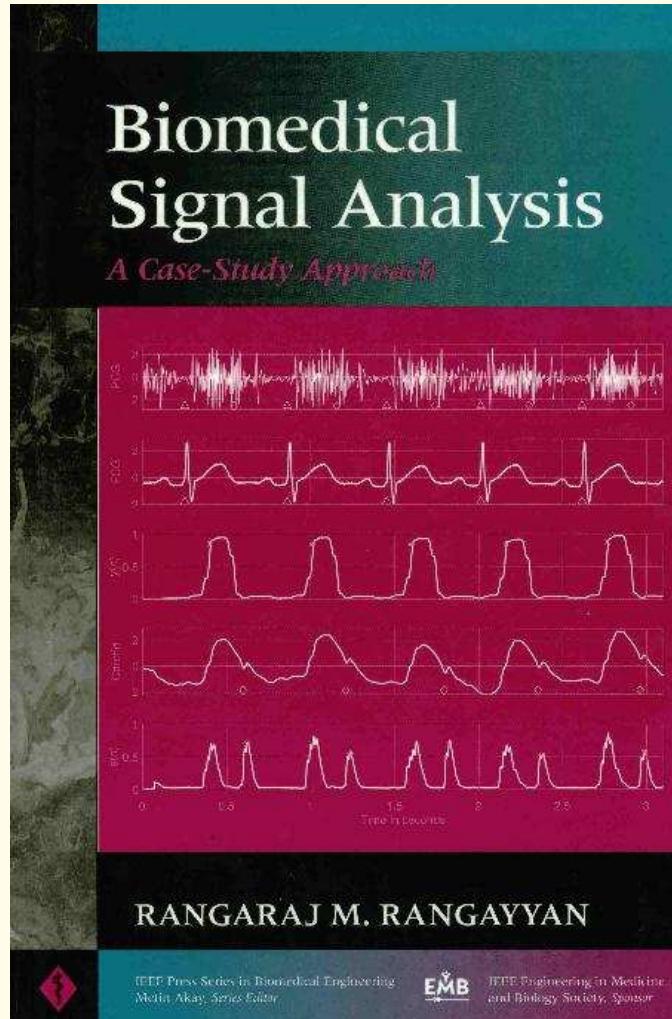
Adjunct Professor, Departments of Surgery and Radiology
University of Calgary
Calgary, Alberta, Canada T2N 1N4

Phone: +1 (403) 220-6745

e-mail: ranga@ucalgary.ca

Web: <http://www.enel.ucalgary.ca/People/Ranga/enel563>

© R.M. Rangayyan



IEEE/ Wiley, New York, NY, 2002.

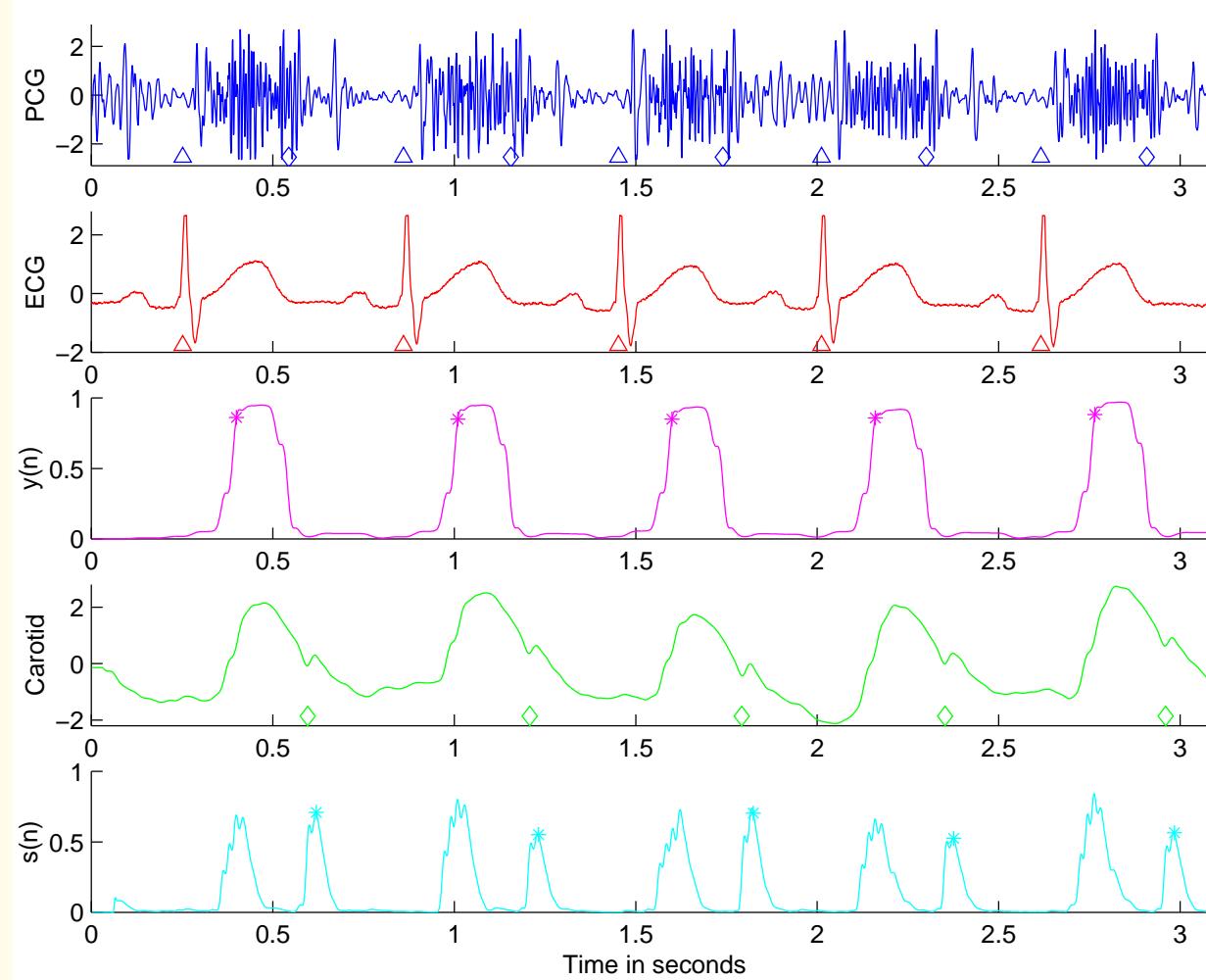
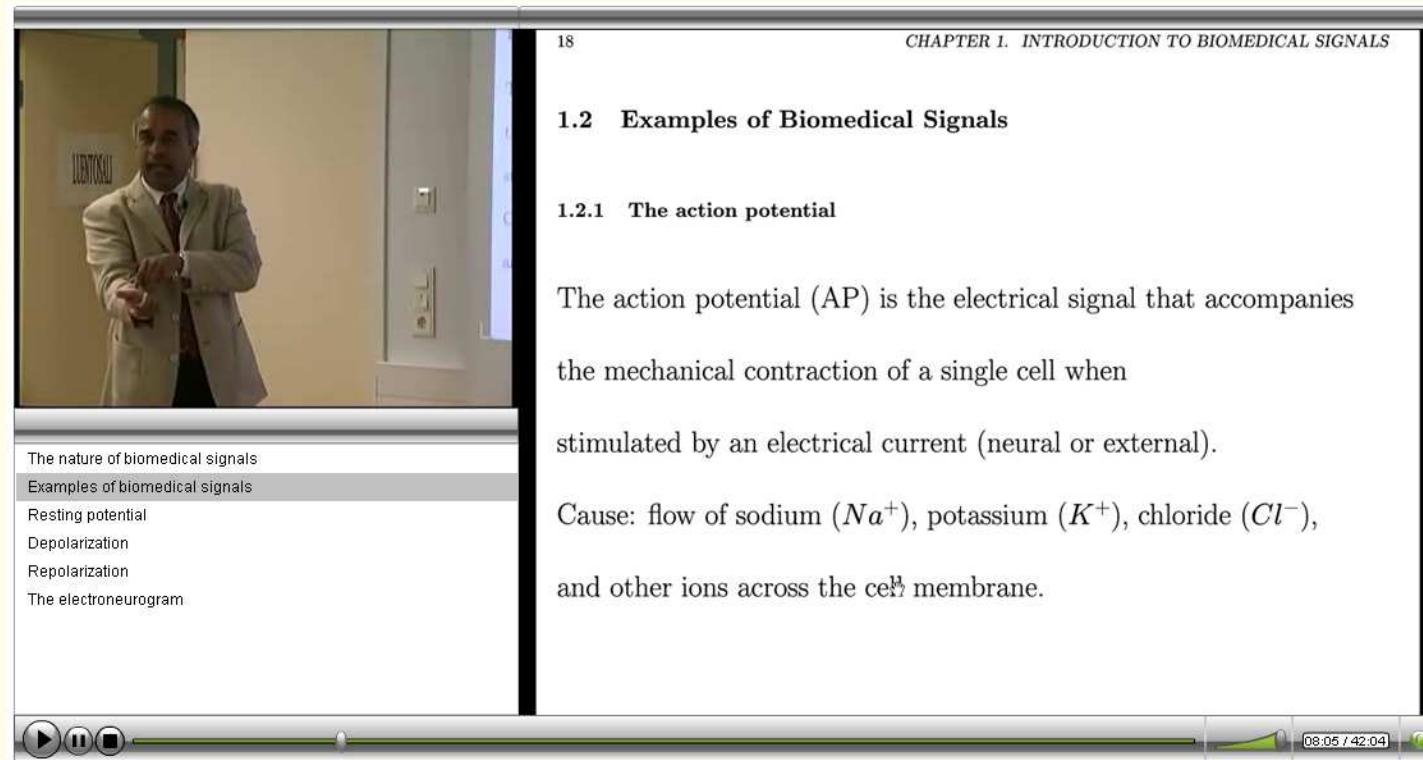


Illustration of various stages of
biomedical signal processing and analysis.



The nature of biomedical signals
Examples of biomedical signals
Resting potential
Depolarization
Repolarization
The electroneurogram

18 CHAPTER 1. INTRODUCTION TO BIOMEDICAL SIGNALS

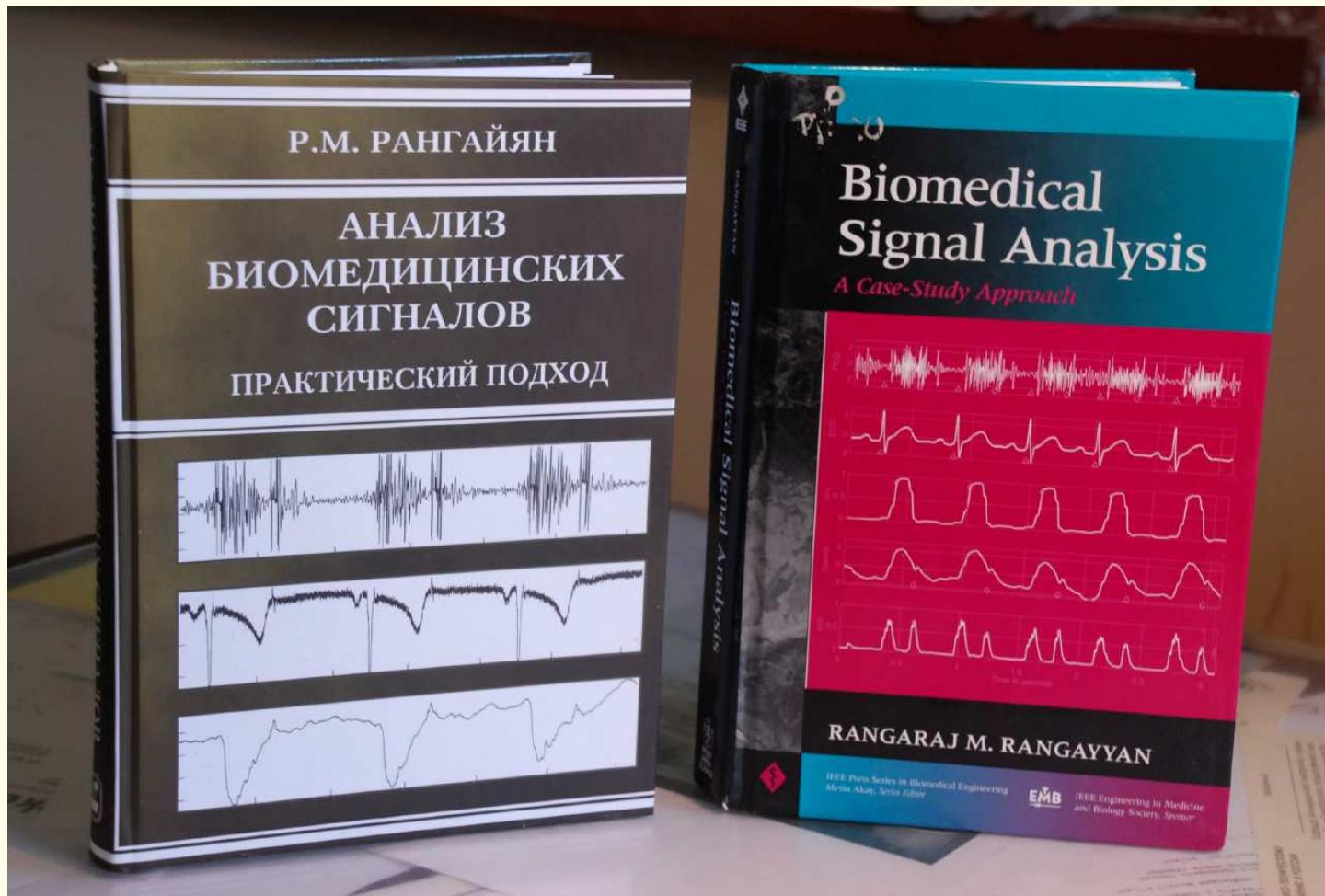
1.2 Examples of Biomedical Signals

1.2.1 The action potential

The action potential (AP) is the electrical signal that accompanies the mechanical contraction of a single cell when stimulated by an electrical current (neural or external).

Cause: flow of sodium (Na^+), potassium (K^+), chloride (Cl^-), and other ions across the cell membrane.

Video of course given at
Ragnar Granit Institute of Biomedical Engineering,
Tampere University of Technology,
Tampere, Finland.
www.evicab.eu





Important Notes: Please . . .

- Attend all lectures, tutorials, and lab sessions.
- Arrive on time.
- Do not leave during lecture or tutorial.
- Pay attention.
- Take notes.
- Ask questions.
- Cell phones, beepers off.
- No chatting: on-line or in-class.
- Use computers only to take notes: No surfing.



- Lab: Do your share of the work.
- Interpret, understand, appreciate the methods and results.
- Understand the material.

Happy learning!





1

Introduction to Biomedical Signals

1.1 The Nature of Biomedical Signals

Living organisms are made up of many component *systems*:

the human body includes several systems.



For example:

- the nervous system,
- the cardiovascular system,
- the musculo-skeletal system.



Each system is made up of several subsystems that carry on many *physiological processes*.

Cardiac system: rhythmic pumping of blood throughout the body to facilitate the delivery of nutrients, and

pumping blood through the pulmonary system for oxygenation of the blood itself.



Physiological processes are complex phenomena, including

- nervous or hormonal stimulation and control;
- inputs and outputs that could be in the form of physical material, neurotransmitters, or information; and
- action that could be mechanical, electrical, or biochemical.



Most physiological processes are accompanied by *signals* of several types that reflect their nature and activities:

- biochemical, in the form of hormones and neurotransmitters,
- electrical, in the form of potential or current, and
- physical, in the form of pressure or temperature.



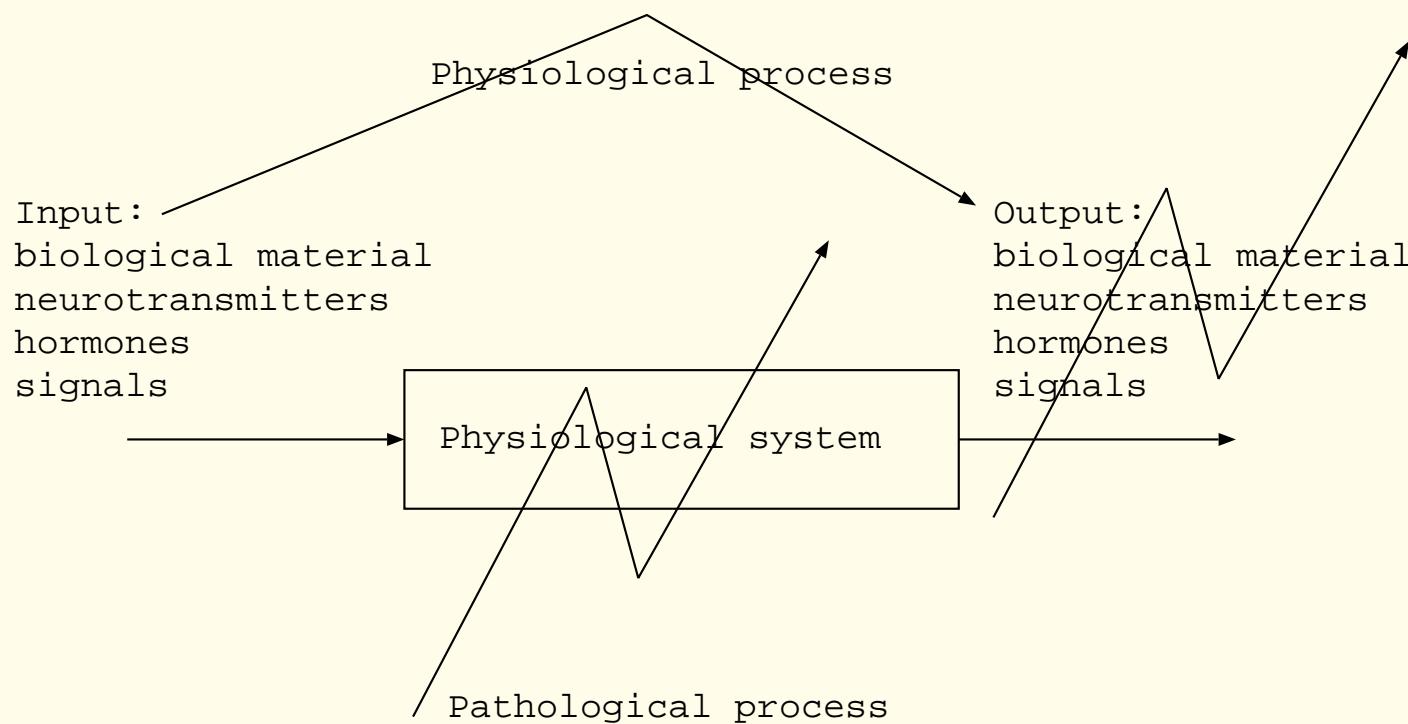
Diseases or defects in a biological system cause alterations in its normal physiological processes,

leading to *pathological processes* that affect

the performance, health, and well-being of the system.

A pathological process is typically associated with signals that are different in some respects from the corresponding normal signals.

Need a good understanding of a system of interest to observe the corresponding signals and assess the state of the system.



Schematic representation of a physiological system carrying on a physiological process.

A pathological process is indicated to represent its effects on the system and its output.



Most infections cause a rise in the temperature of the body:

sensed easily, in a relative and *qualitative* manner,

via the palm of one's hand.

Objective or *quantitative* measurement of temperature

requires an instrument, such as a thermometer.



A single measurement x of temperature is a *scalar*:

represents the thermal state of the body at a

particular or single instant of time t

and a particular position.

If we record the temperature continuously,

we obtain a *signal as a function of time*:

expressed in *continuous-time* or *analog* form as $x(t)$.



When the temperature is measured at *discrete* points of time,

it may be expressed in *discrete-time* form as $x(nT)$ or $x(n)$,

n : index or measurement sample number of the array of values,

T : uniform interval between the time instants of measurement.

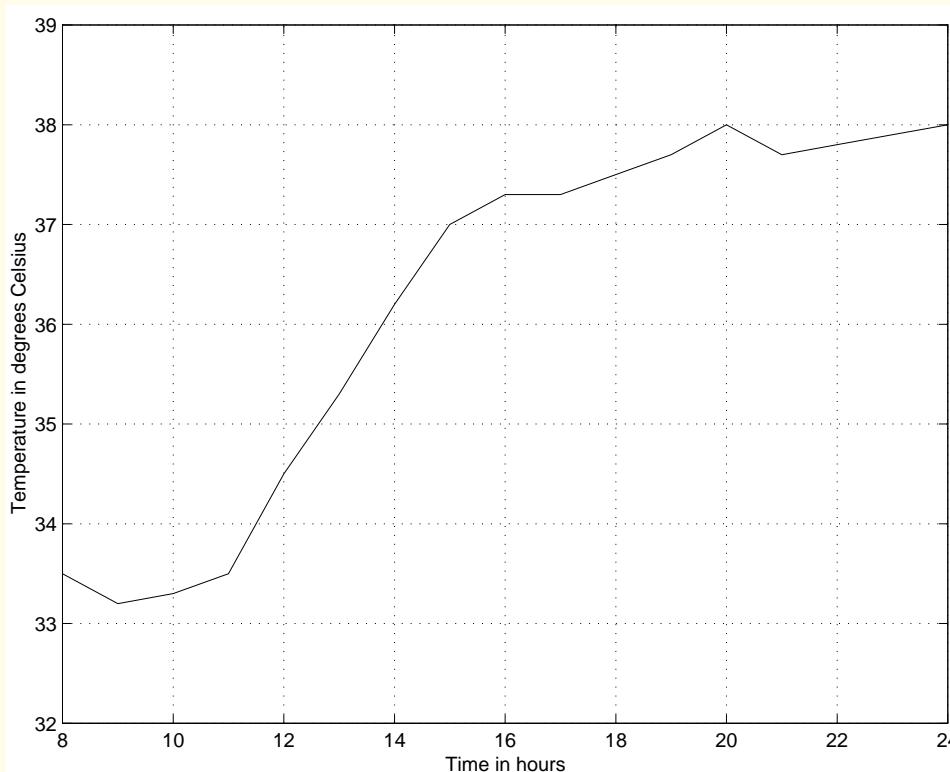
A discrete-time signal that can take amplitude values only from a limited list of *quantized* levels is called a *digital* signal.

 $33.5^{\circ} C$

(a)

Time	08:00	10:00	12:00	14:00	16:00	18:00	20:00	22:00	24:00
$^{\circ}C$	33.5	33.3	34.5	36.2	37.3	37.5	38.0	37.8	38.0

(b)



(c)

Figure 1.1: Measurements of the temperature of a patient presented as (a) a scalar with one temperature measurement x at a time instant t ; (b) an array $x(n)$ made up of several measurements at different instants of time; and (c) a signal $x(t)$ or $x(n)$. The horizontal axis of the plot represents time in **hours**; the vertical axis gives temperature in **degrees Celsius**. Data courtesy of Foothills Hospital, Calgary.



Another basic measurement in health care and monitoring:

blood pressure (BP).

Each measurement consists of two values —

the systolic pressure and the diastolic pressure.

Units: millimeters of mercury (*mm of Hg*)

in clinical practice,

although the international standard unit for pressure

is the *Pascal*.



A single BP measurement:

a *vector* $\mathbf{x} = [x_1, x_2]^T$ with two components:

x_1 indicating the systolic pressure and

x_2 indicating the diastolic pressure.

When BP is measured at a few instants of time:

an array of vectorial values $\mathbf{x}(n)$

or a function of time $\mathbf{x}(t)$.

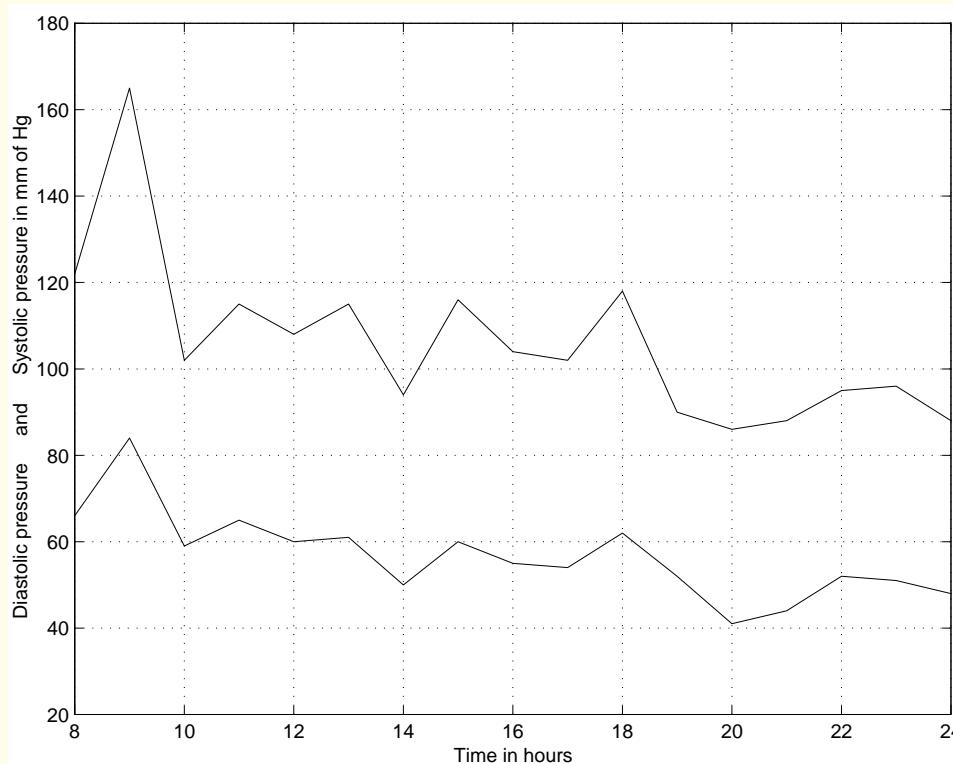


$$\begin{bmatrix} 122 \\ 66 \end{bmatrix}$$

(a)

Time	08:00	10:00	12:00	14:00	16:00	18:00	20:00	22:00	24:00
Systolic	122	102	108	94	104	118	86	95	88
Diastolic	66	59	60	50	55	62	41	52	48

(b)



(c)



Figure 1.2: Measurements of the blood pressure of a patient presented as (a) a single pair or vector of systolic and diastolic measurements \mathbf{x} in *mm of Hg* at a time instant t ; (b) an array $\mathbf{x}(n)$ made up of several measurements at different instants of time; and (c) a signal $\mathbf{x}(t)$ or $\mathbf{x}(n)$. Note the use of boldface \mathbf{x} to indicate that each measurement is a vector with two components. The horizontal axis of the plot represents time in *hours*; the vertical axis gives the systolic pressure (upper trace) and the diastolic pressure (lower trace) in *mm of Hg*. Data courtesy of Foothills Hospital, Calgary.



1.2 Examples of Biomedical Signals

1.2.1 *The action potential*

Action potential (AP): electrical signal that accompanies

the mechanical contraction of a single cell when

stimulated by an electrical current (neural or external).

Cause: flow of sodium (Na^+), potassium (K^+),

chloride (Cl^-), and other ions across the cell membrane.



Action potential:

Basic component of all bioelectrical signals.

Provides information on the nature of physiological activity at the single-cell level.

Recording an action potential requires the isolation of a single cell,

and microelectrodes with tips of the order of a few micrometers

to stimulate the cell and record the response.



Resting potential:

Nerve and muscle cells are encased in a

semi-permeable membrane:

permits selected substances to pass through; others kept out.

Body fluids surrounding cells are conductive solutions

containing charged atoms known as ions.



Resting state: membranes of excitable cells

permit entry of K^+ and Cl^- , but block Na^+ ions —

permeability for K^+ is 50–100 times that for Na^+ .

Various ions seek to establish inside vs outside balance

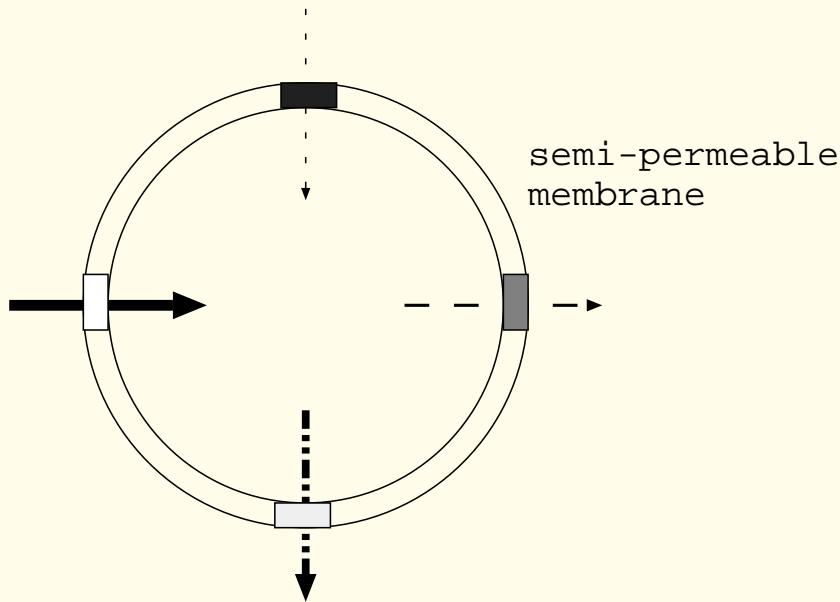
according to charge and concentration.



Excitable cell: enclosed in semi-permeable membrane

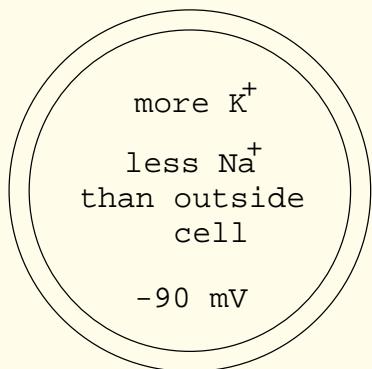
Body fluids: conductive solutions containing ions

Important ions: Na^+ , K^+ , Ca^+ , Cl^-

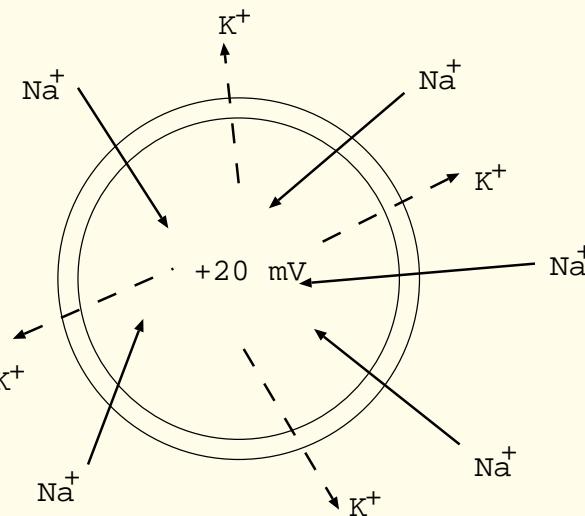


Selective permeability: some ions can move in and out of the cell easily, whereas others cannot, depending upon the state of the cell and the voltage-gated ion channels

Selective permeable membrane of an excitable cell
(nerve or muscle).



At rest: permeability for K^+
50 - 100 times that for Na^+



Depolarization: triggered by
a stimulus; fast Na^+ channels open

Resting state and depolarization of a cell.



Results of the inability of Na^+ to penetrate a cell membrane:

- Na^+ concentration inside is far less than that outside.
- The outside of the cell is more positive than the inside.
- To balance the charge, additional K^+ ions enter the cell, causing higher K^+ concentration inside than outside.
- Charge balance cannot be reached due to differences in membrane permeability for various ions.
- State of equilibrium established with a potential difference:
inside of the cell negative with respect to the outside.



A cell in its resting state is said to be *polarized*.

Most cells maintain a *resting potential* of the order of

-60 to -100 mV

until some disturbance or stimulus upsets the equilibrium.



Depolarization:

When a cell is excited by ionic currents or an external

stimulus, the membrane changes its characteristics:

begins to allow Na^+ ions to enter the cell.

This movement of Na^+ ions constitutes an ionic current,

which further reduces the membrane barrier to Na^+ ions.

Avalanche effect: Na^+ ions rush into the cell.



K^+ ions try to leave the cell

as they were in higher concentration inside the cell

in the preceding resting state,

but cannot move as fast as the Na^+ ions.

Net result: the inside of the cell becomes positive

with respect to the outside due to an imbalance of K^+ .



New state of equilibrium reached

after the rush of Na^+ ions stops.

Represents the beginning of the *action potential*,

with a peak value of about $+20\text{ mV}$ for most cells.

An excited cell displaying an action potential

is said to be *depolarized*;

the process is called *depolarization*.



Repolarization:

After a certain period of being in the depolarized state

the cell becomes polarized again and

returns to its resting potential

via a process known as *repolarization*.



Principal ions involved in repolarization: K^+ .

Voltage-dependent K^+ channels:

predominant membrane permeability for K^+ .

K^+ concentration is much higher inside the cell:

net efflux of K^+ from the cell,

the inside becomes more negative,

effecting repolarization back to the resting potential.



Nerve and muscle cells repolarize rapidly:

action potential duration of about 1 ms .

Heart muscle cells repolarize slowly:

action potential duration of $150 - 300\text{ ms}$.



The action potential is always the same for a given cell,
regardless of the method of excitation or
the intensity of the stimulus beyond a threshold:
all-or-none or *all-or-nothing* phenomenon.



After an action potential, there is a period during which

a cell cannot respond to any new stimulus:

absolute refractory period — about 1 ms in nerve cells.

This is followed by a *relative refractory period*:

another action potential may be triggered by a

much stronger stimulus than in the normal situation.

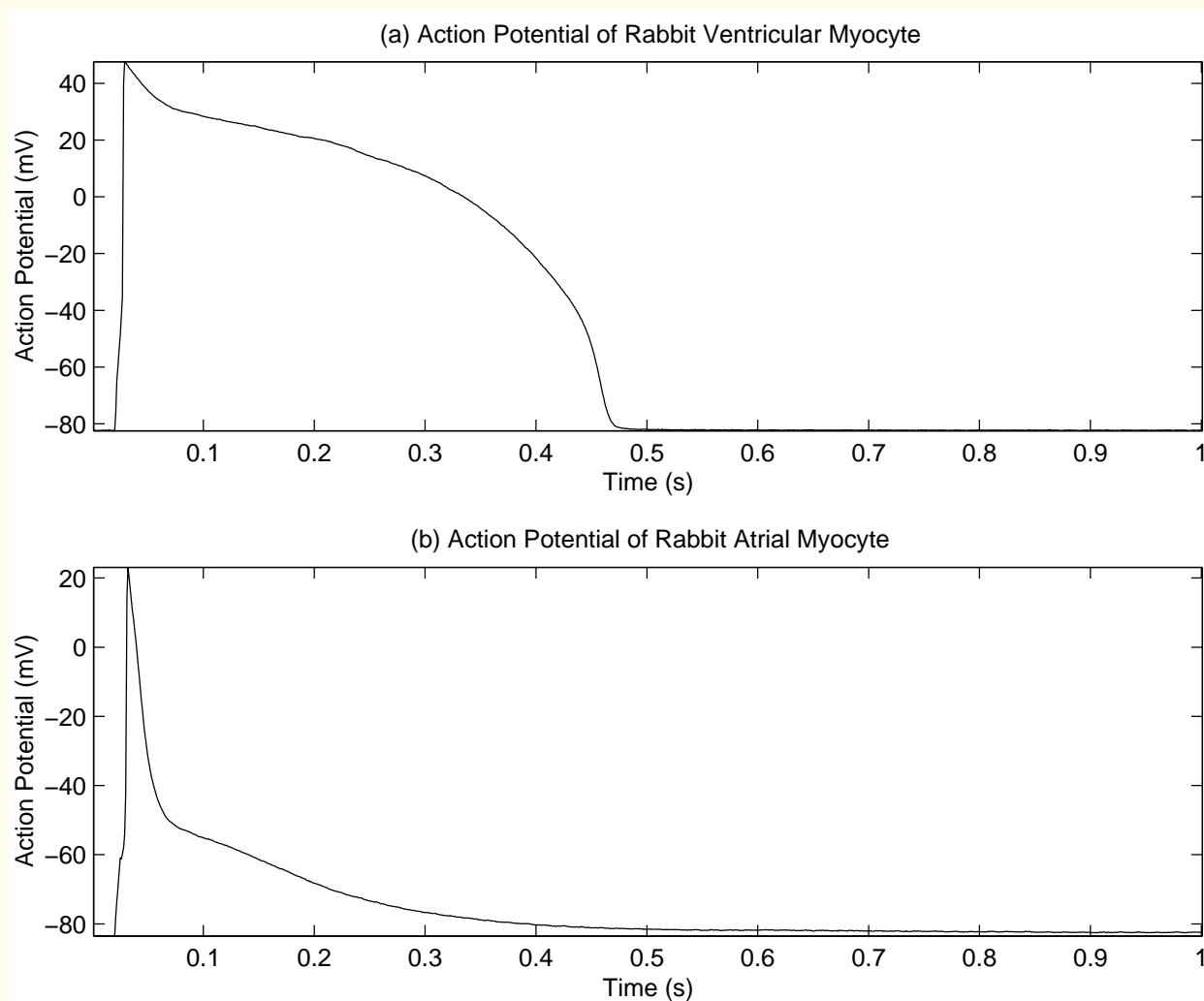
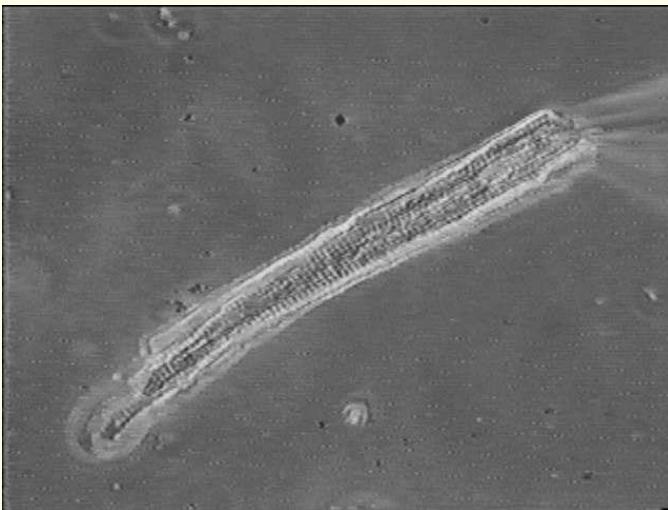
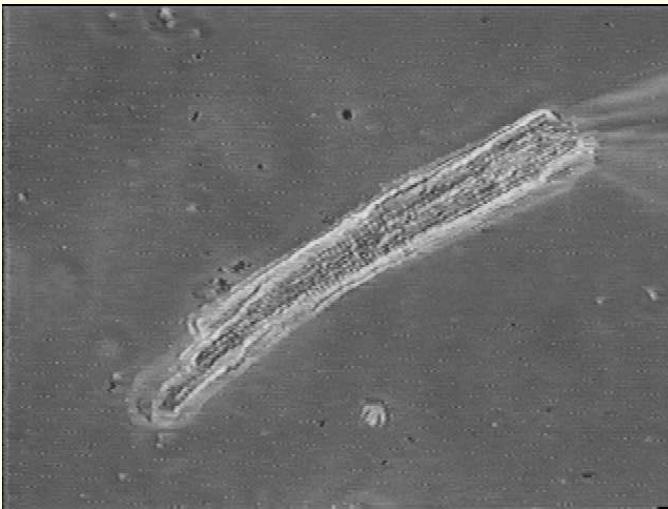


Figure 1.3: Action potentials of rabbit ventricular and atrial myocytes. Data courtesy of R. Clark, Department of Physiology and Biophysics, University of Calgary.



(a)



(b)

Figure 1.4: A single ventricular myocyte (of a rabbit) in its (a) relaxed and (b) fully contracted states. The length of the myocyte is approximately $25 \mu\text{m}$. The tip of the glass pipette, faintly visible at the upper-right end of the myocyte, is approximately $2 \mu\text{m}$ wide. A square pulse of current, 3 ms in duration and 1 nA in amplitude, was passed through the recording electrode and across the cell membrane causing the cell to depolarize rapidly. Images courtesy of R. Clark, Department of Physiology and Biophysics, University of Calgary.



An action potential propagates along a muscle fiber or

an unmyelinated nerve fiber as follows:

Once initiated by a stimulus, the action potential

propagates along the whole length of a fiber

without decrease in amplitude by

progressive depolarization of the membrane.



Current flows from a depolarized region through the intra-cellular fluid to adjacent inactive regions, thereby depolarizing them.



Current also flows through the extra-cellular fluids,

through the depolarized membrane,

and back into the intra-cellular space,

completing the local circuit.

The energy to maintain conduction is supplied by the fiber.



Myelinated nerve fibers are covered by

an insulating sheath of *myelin*,

interrupted every few millimeters by spaces known as the

nodes of Ranvier,

where the fiber is exposed to the interstitial fluid.



Sites of excitation and changes of membrane permeability

exist only at the nodes:

current flows by jumping from one node to the next

in a process known as *saltatory conduction*.



1.2.2 *The electroneurogram (ENG)*

The ENG is an electrical signal observed as a stimulus
and the associated nerve action potential
propagate over the length of a nerve.



May be used to measure the velocity of propagation

or conduction velocity of a stimulus or action potential.

ENGs may be recorded using concentric needle electrodes

or $Ag - AgCl$ electrodes at the surface of the body.



Conduction velocity in a peripheral nerve measured by

stimulating a motor nerve

and measuring the related activity at two points

at known distances along its course.

Stimulus: 100 V , $100 - 300\text{ }\mu\text{s}$.

ENG amplitude: $10\text{ }\mu\text{V}$;

Amplifier gain: $2,000$; Bandwidth $10 - 10,000\text{ Hz}$.

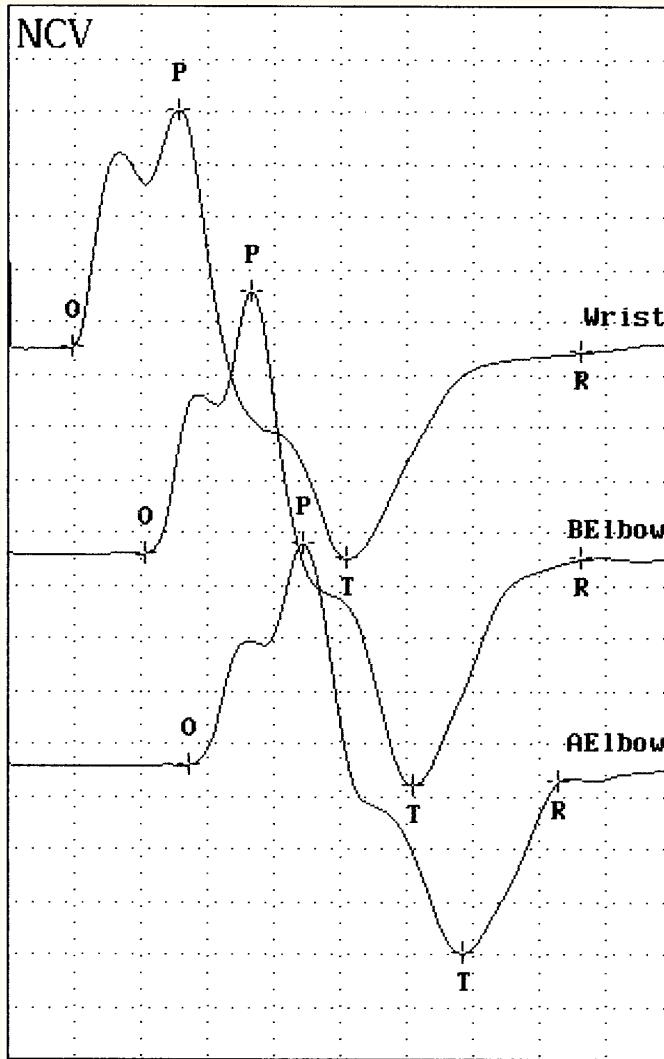


Figure 1.5: Nerve conduction velocity measurement via electrical stimulation of the ulnar nerve. The grid boxes represent 3 ms in width and 2 μ V in height. AElbow: above the elbow. BElbow: below the elbow. O: onset. P: Peak. T: trough. R: recovery of base-line. Courtesy of M. Wilson and C. Adams, Alberta Children's Hospital, Calgary.

The responses shown in the figure are normal.

BElbow – Wrist latency 3.23 ms. Nerve conduction velocity 64.9 m/s.



Typical nerve conduction velocity:

- $45 - 70 \text{ m/s}$ in nerve fibers;
- $0.2 - 0.4 \text{ m/s}$ in heart muscle;
- $0.03 - 0.05 \text{ m/s}$ in time-delay fibers between the atria and ventricles.

Neural diseases may cause a decrease in conduction velocity.



1.2.3 *The electromyogram (EMG)*

Skeletal muscle fibers are twitch fibers:

produce a mechanical twitch response for a single stimulus

and generate a propagated action potential.



Skeletal muscles made up of collections of

motor units (MUs),

each of which consists of an anterior horn cell,

or motoneuron or motor neuron,

its axon, and all muscle fibers innervated by that axon.



Motor unit: smallest muscle unit

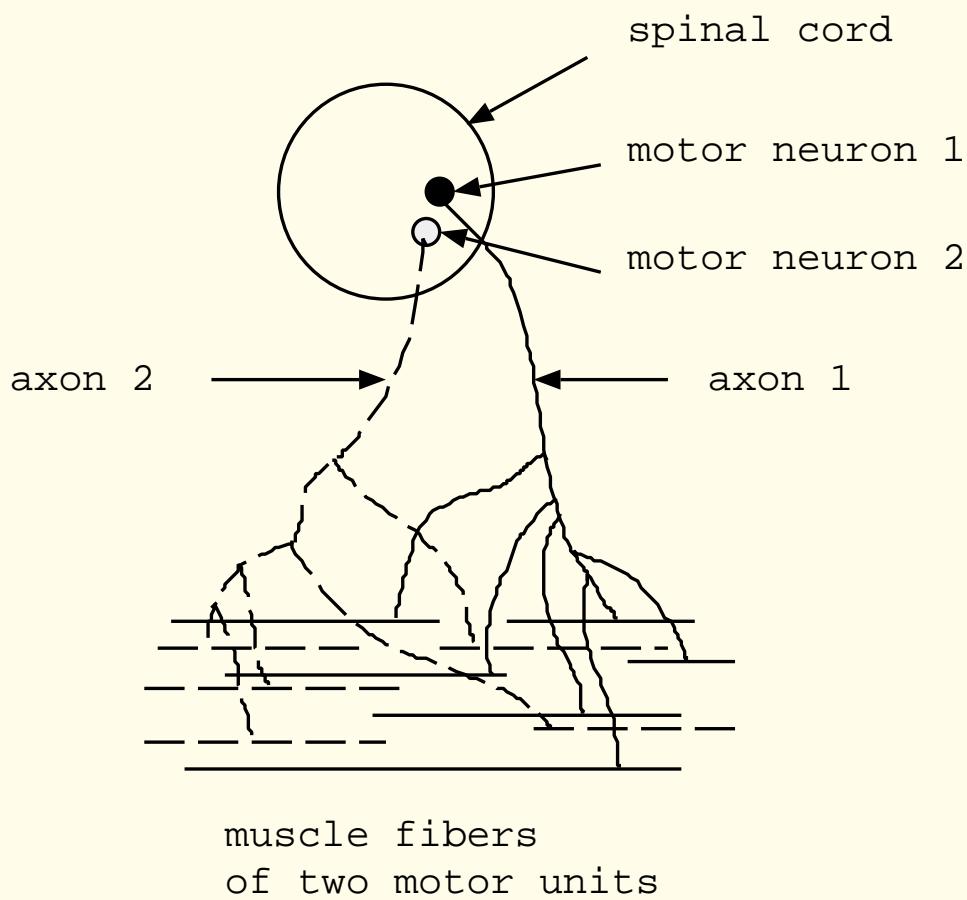
that can be activated by volitional effort.

Constituent fibers of a motor unit activated synchronously.

Component fibers of a motor unit extend lengthwise

in loose bundles along the muscle.

Fibers of an MU interspersed with the fibers of other MUs.



Schematic illustration of two motor units in a muscle.

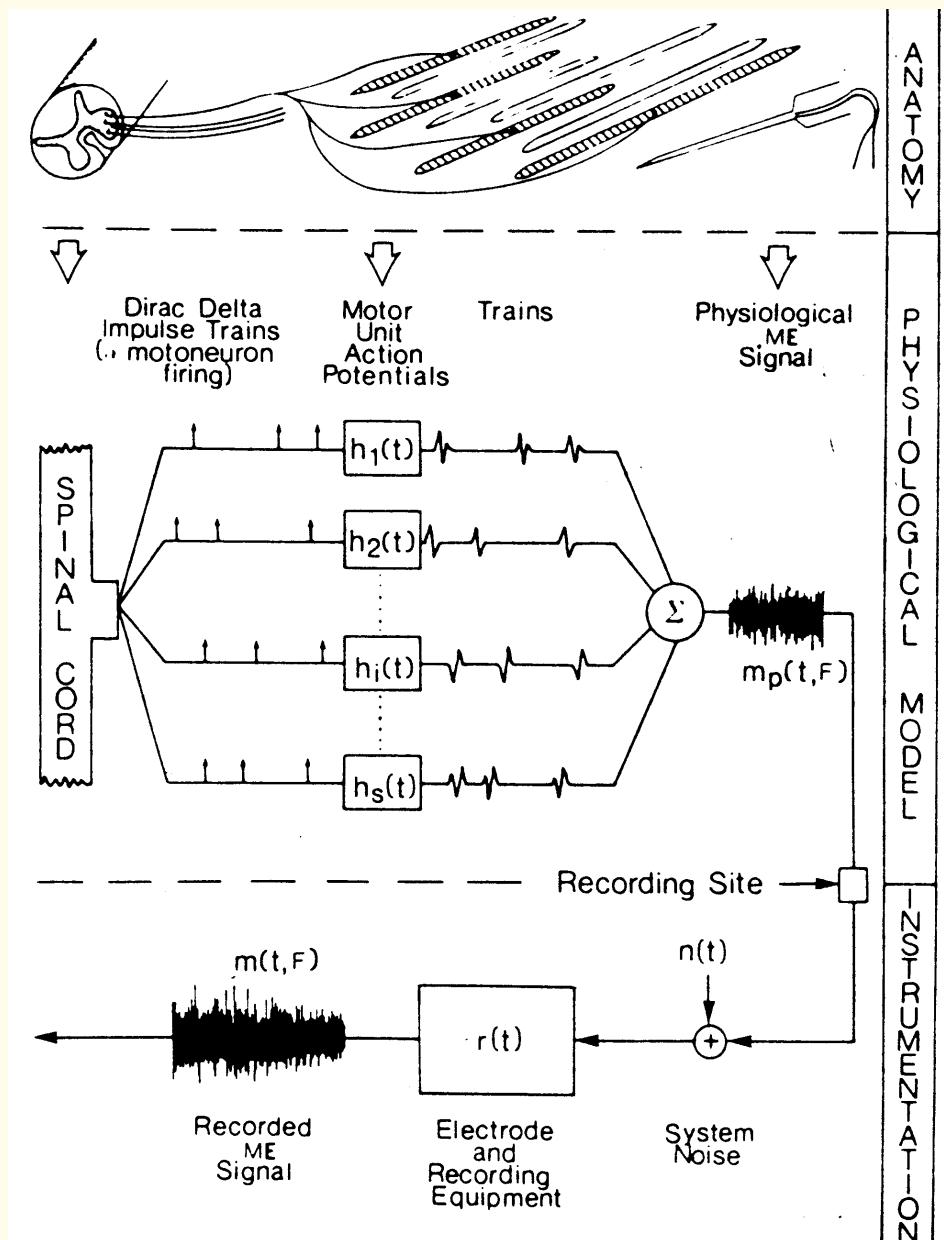




Figure 1.6: Schematic representation of a motor unit and model for the generation of EMG signals. Top panel: A motor unit includes an anterior horn cell or motor neuron (illustrated in a cross-section of the spinal cord), an axon, and several connected muscle fibers. The hatched fibers belong to one motor unit; the non-hatched fibers belong to other motor units. A needle electrode is also illustrated. Middle panel: The firing pattern of each motor neuron is represented by an impulse train. Each system $h_i(t)$ shown represents a motor unit that is activated and generates a train of SMUAPs. The net EMG is the sum of several SMUAP trains. Bottom panel: Effects of instrumentation on the EMG signal acquired. The observed EMG is a function of time t and muscular force produced F . Reproduced with permission from C.J. de Luca, Physiology and mathematics of myoelectric signals, *IEEE Transactions on Biomedical Engineering*, 26:313–325, 1979. ©IEEE.



Large muscles for gross movement have 100s of fibers/MU;
muscles for precise movement have fewer fibers per MU.

Number of muscle fibers per motor nerve fiber:

innervation ratio.

Platysma muscle of the neck: 1,826 large nerve fibers

controlling 27,100 muscle fibers with 1,096 motor units;

innervation ratio of 15.



First dorsal interosseus (finger) muscle:

199 large nerve fibers and 40,500 muscle fibers

with 119 motor units; innervation ratio of 203.

Mechanical output (contraction) of a muscle = net result of

stimulation and contraction of several of its motor units.



When stimulated by a neural signal, each MU contracts

and causes an electrical signal that is the summation

of the action potentials of all of its constituent cells:

this is known as the *single-motor-unit action potential*.

SMUAP or MUAP recorded using needle electrodes.

Normal SMUAPs usually biphasic or triphasic;

3 – 15 ms in duration, 100 – 300 μV in amplitude,

appear with frequency of 6 – 30/s.



The shape of a recorded SMUAP depends upon
the type of the needle electrode used,
its positioning with respect to the active motor unit,
and the projection of the electrical field of the activity
onto the electrodes.

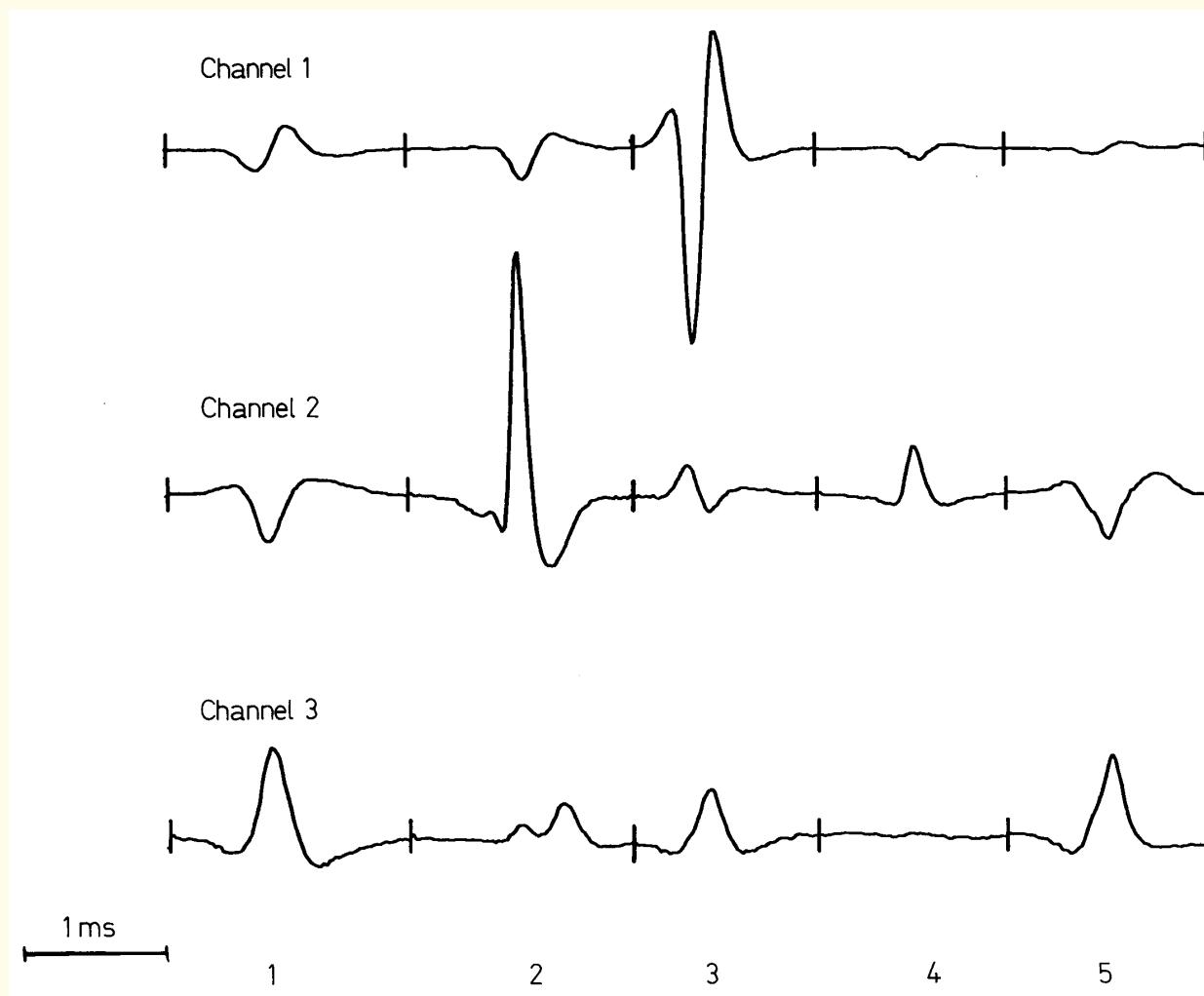


Figure 1.7: SMUAP trains recorded simultaneously from three channels of needle electrodes. Observe the different shapes of the same SMUAPs projected onto the axes of the three channels. Three different motor units are active over the duration of the signals illustrated. Reproduced with permission from B. Mambrito and C.J. de Luca, Acquisition and decomposition of the EMG signal, in *Progress in Clinical Neurophysiology*, Volume 10: Computer-aided Electromyography, Editor: J.E. Desmedt, pp 52–72, 1983. ©S. Karger AG, Basel, Switzerland.



The shape of SMUAPs is affected by disease.

Neuropathy: slow conduction,

desynchronized activation of fibers,

polyphasic SMUAP with an amplitude larger than normal.

The same MU may fire at higher rates than normal

before more MUs are recruited.



Myopathy: loss of muscle fibers in MUs,

with the neurons presumably intact.

Splintering of SMUAPs occurs due to

asynchrony in activation

as a result of patchy destruction of fibers

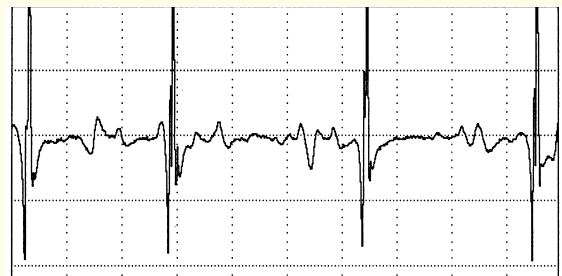
(muscular dystrophy),

leading to splintered SMUAPs.

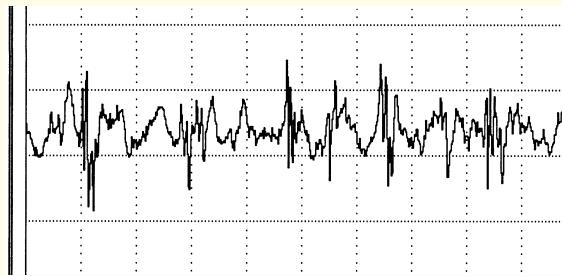
More MUs recruited at low levels of effort.



(a)



(b)



(c)

Figure 1.8: Examples of SMUAP trains. (a) From the right deltoid of a normal subject, male, 11 years; the SMUAPs are mostly biphasic, with duration in the range $3 - 5 \text{ ms}$. (b) From the deltoid of a six-month-old male patient with brachial plexus injury (neuropathy); the SMUAPs are polyphasic and large in amplitude ($800 \mu\text{V}$), and the same motor unit is firing at a relatively high rate at low-to-medium levels of effort. (c) From the right biceps of a 17-year-old male patient with myopathy; the SMUAPs are polyphasic and indicate early recruitment of more motor units at a low level of effort. The signals were recorded with gauge 20 needle electrodes. The width of each grid box represents a duration of 20 ms ; its height represents an amplitude of $200 \mu\text{V}$. Courtesy of M. Wilson and C. Adams, Alberta Children's Hospital, Calgary.



Gradation of muscular contraction:

Muscular contraction levels are controlled in two ways:

- *Spatial recruitment* — activating new MUs, and
- *Temporal recruitment* — increasing the frequency of discharge or firing rate of each MU,

with increasing effort.



MUs activated at different times and at different frequencies:
asynchronous contraction.

The twitches of individual MUs sum and fuse to form
tetanic contraction and increased force.

Weak volitional effort: MUs fire at about $5 - 15 \text{ pps}$.

As greater tension is developed, an *interference pattern*

EMG is obtained, with the active MUs firing at $25 - 50 \text{ pps}$.



Spatio-temporal summation of the MUAPs of all active MUs gives rise to the EMG of the muscle.

EMG signals recorded using surface electrodes:

complex signals including interference patterns

of several MUAP trains — difficult to analyze.

EMG may be used to diagnose neuromuscular diseases

such as neuropathy and myopathy.

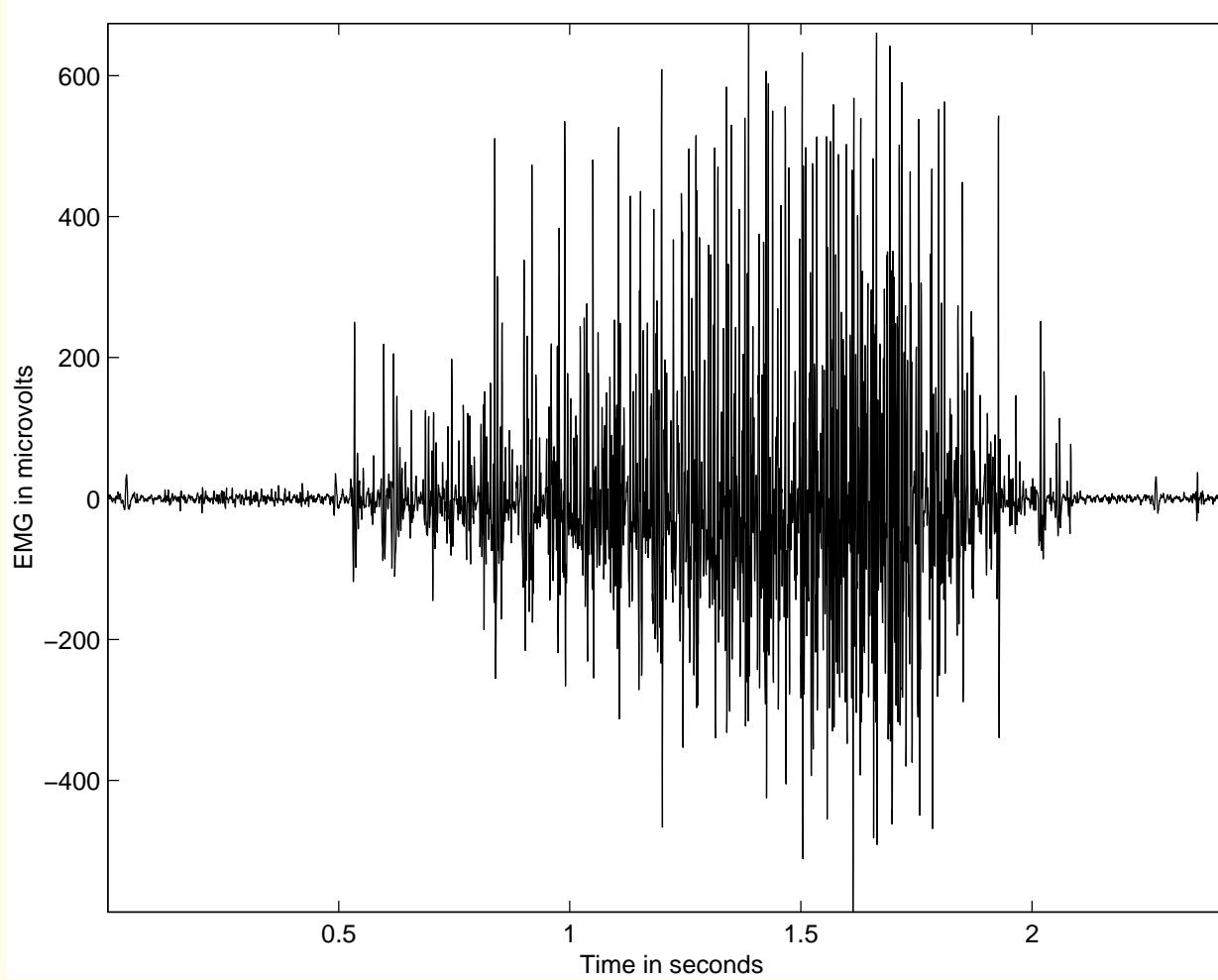


Figure 1.9: EMG signal recorded from the crural diaphragm muscle of a dog using implanted fine-wire electrodes. Data courtesy of R.S. Platt and P.A. Easton, Department of Clinical Neurosciences, University of Calgary.

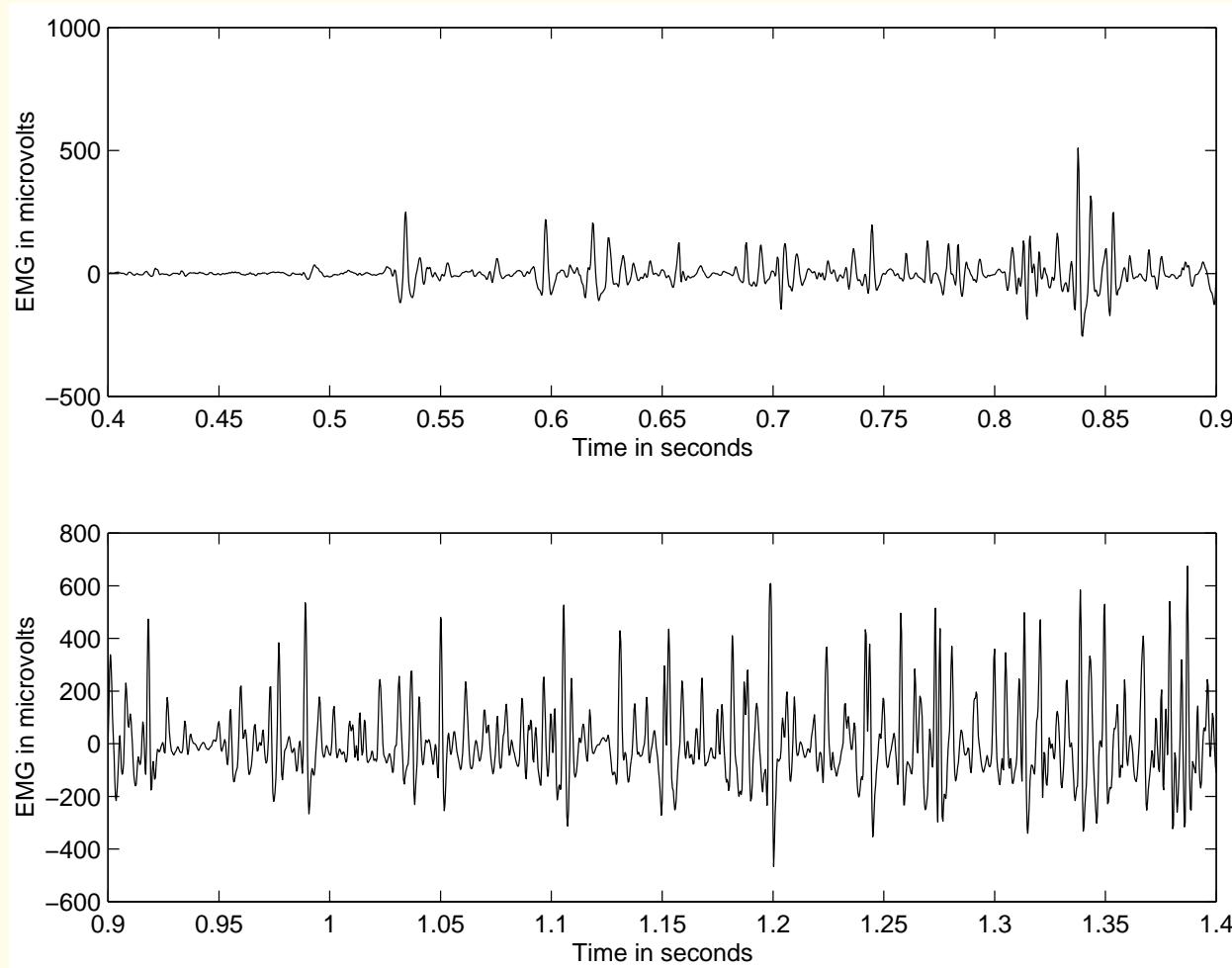


Figure 1.10: The initial part of the EMG signal in Figure 1.9 shown on an expanded time scale. Observe the SMUAPs at the initial stages of contraction, followed by increasingly complex interference patterns of several MUAPs. Data courtesy of R.S. Platt and P.A. Easton, Department of Clinical Neurosciences, University of Calgary.



1.2.4 *The electrocardiogram (ECG)*

ECG: electrical manifestation of the

contractile activity of the heart.

Recorded with surface electrodes on the limbs or chest.

ECG: most commonly known & used biomedical signal.

The rhythm of the heart in terms of beats per minute (*bpm*)

may be estimated by counting the readily identifiable waves.



ECG waveshape is altered by cardiovascular diseases and abnormalities: myocardial ischemia and infarction, ventricular hypertrophy, and conduction problems.



The heart:

A four-chambered pump with

two atria for collection of blood

and two ventricles for pumping out of blood.

Resting or filling phase of a cardiac chamber: *diastole*;

contracting or pumping phase: *systole*.



Right atrium (or auricle, RA): collects impure blood

from the superior and inferior vena cavae.

Atrial contraction: blood is passed from the right atrium

to the right ventricle (RV) through the tricuspid valve.

Ventricular systole: impure blood in the right ventricle

pumped out through the pulmonary valve

to the lungs for purification (oxygenation).

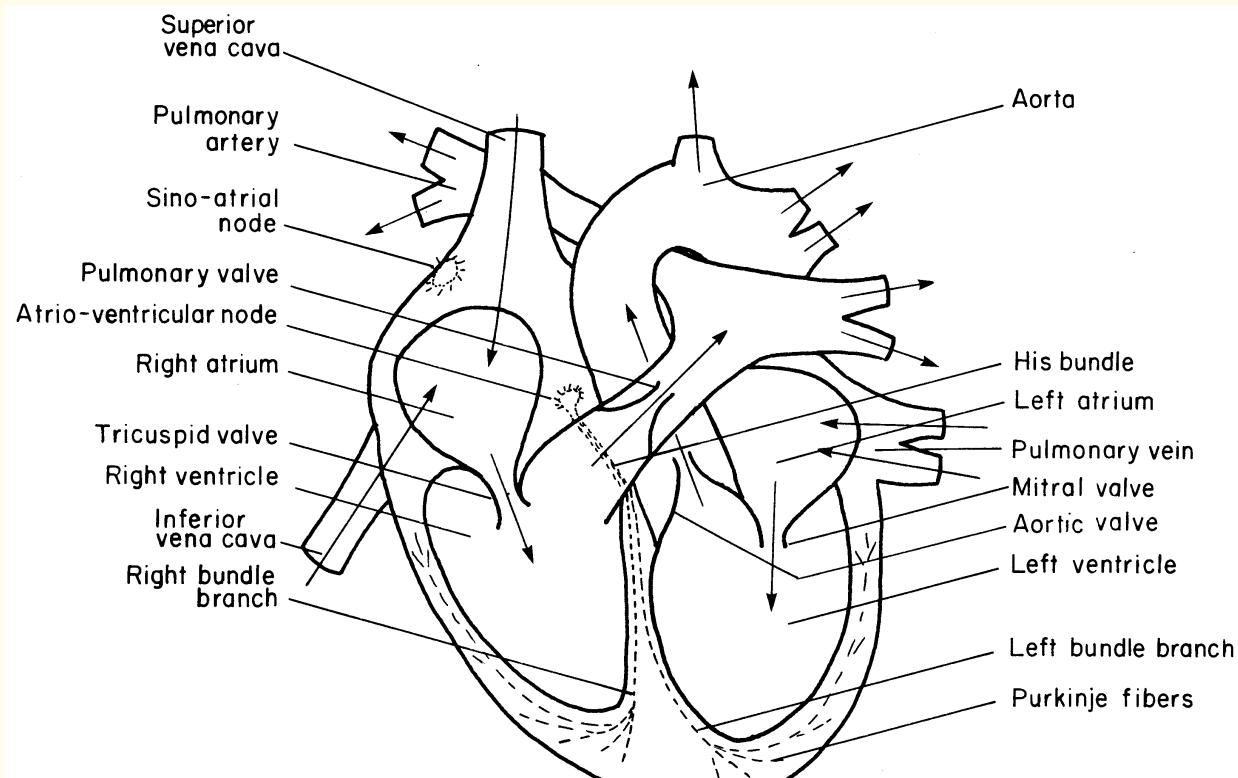


Figure 1.11: Schematic representation of the chambers, valves, vessels, and conduction system of the heart.



Left atrium (LA) receives purified blood from the lungs.

Atrial contraction: blood passed to the

left ventricle (LV) via the mitral valve.

Left ventricle: largest and most important cardiac chamber.



LV contracts the strongest among the cardiac chambers:

to pump oxygenated blood through the aortic valve

and the aorta against the pressure of the rest of the

vascular system of the body.

The terms systole and diastole are applied to the

ventricles by default.



Heart rate (HR) or cardiac rhythm controlled by specialized pacemaker cells in the sino-atrial (SA) node.

Firing rate of SA node controlled by impulses from

the autonomous and central nervous systems:

leading to the delivery of the neurotransmitters

acetylcholine for vagal stimulation — reduced HR;

epinephrine for sympathetic stimulation — increased HR.



Normal, resting heart rate: 70 bpm .

Abnormally low $HR < 60 \text{ bpm}$ during activity:

bradycardia.

High resting HR due to illness or cardiac abnormalities:

tachycardia.



The electrical system of the heart:

Co-ordinated electrical events and a specialized

conduction system intrinsic and unique to the heart:

rhythmic contractile activity.

SA node: basic, natural cardiac pacemaker —

triggers its own train of action potentials.



The action potential of the SA node
propagates through the heart,
causing a particular pattern of excitation and contraction.



Sequence of events and waves in a cardiac cycle:

1. The SA node fires.
2. Electrical activity propagated through atrial musculature at comparatively low rates, causing slow-moving depolarization or contraction of the atria:

P wave in the ECG.

Due to slow contraction and small size of the atria, the P wave is a slow, low-amplitude wave:

$0.1 - 0.2 \text{ mV}$, $60 - 80 \text{ ms}$.



3. Propagation delay at the atrio-ventricular (AV) node.

Normally iso-electric segment of $60 - 80\text{ ms}$

after the P wave in the ECG — PQ segment.

Transfer of blood from the atria to the ventricles.

4. The AV node fires.

5. The His bundle, the bundle branches, and the Purkinje system of specialized conduction fibers propagate the stimulus to the ventricles at a high rate.



6. The wave of stimulus spreads rapidly from the apex of the heart upwards, causing rapid depolarization or contraction of the ventricles:

QRS wave — sharp biphasic or triphasic wave

1 *mV* amplitude and 80 *ms* duration.



7. Ventricular muscle cells possess a relatively long action potential duration of $300 - 350 \text{ ms}$.
The plateau portion of the action potential causes a normally iso-electric segment of about $100 - 120 \text{ ms}$ after the QRS: the ST segment.
8. Repolarization or relaxation of the ventricles causes the slow T wave, with amplitude of $0.1 - 0.3 \text{ mV}$ and duration of $120 - 160 \text{ ms}$.

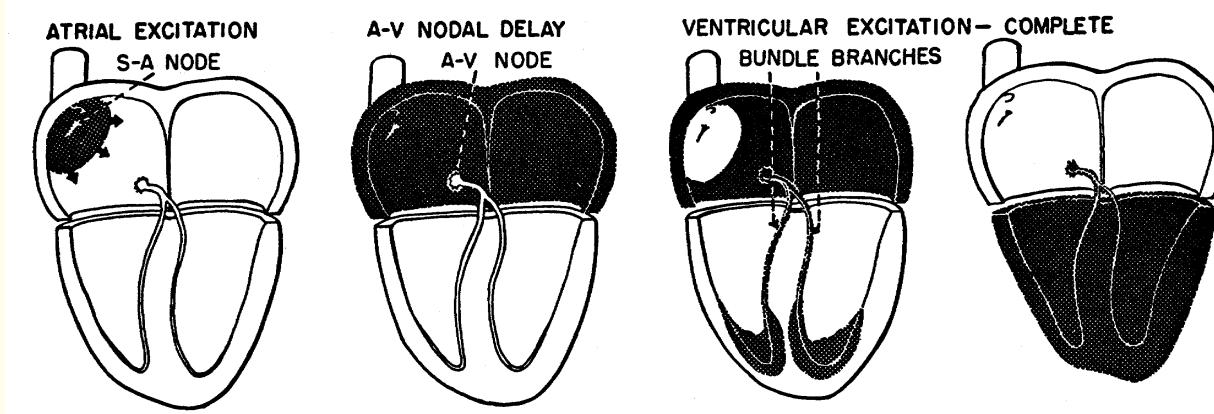


Figure 1.12: Propagation of the excitation pulse through the heart. Reproduced with permission from R.F. Rushmer, *Cardiovascular Dynamics*, 4th edition, ©W.B. Saunders, Philadelphia, PA, 1976.

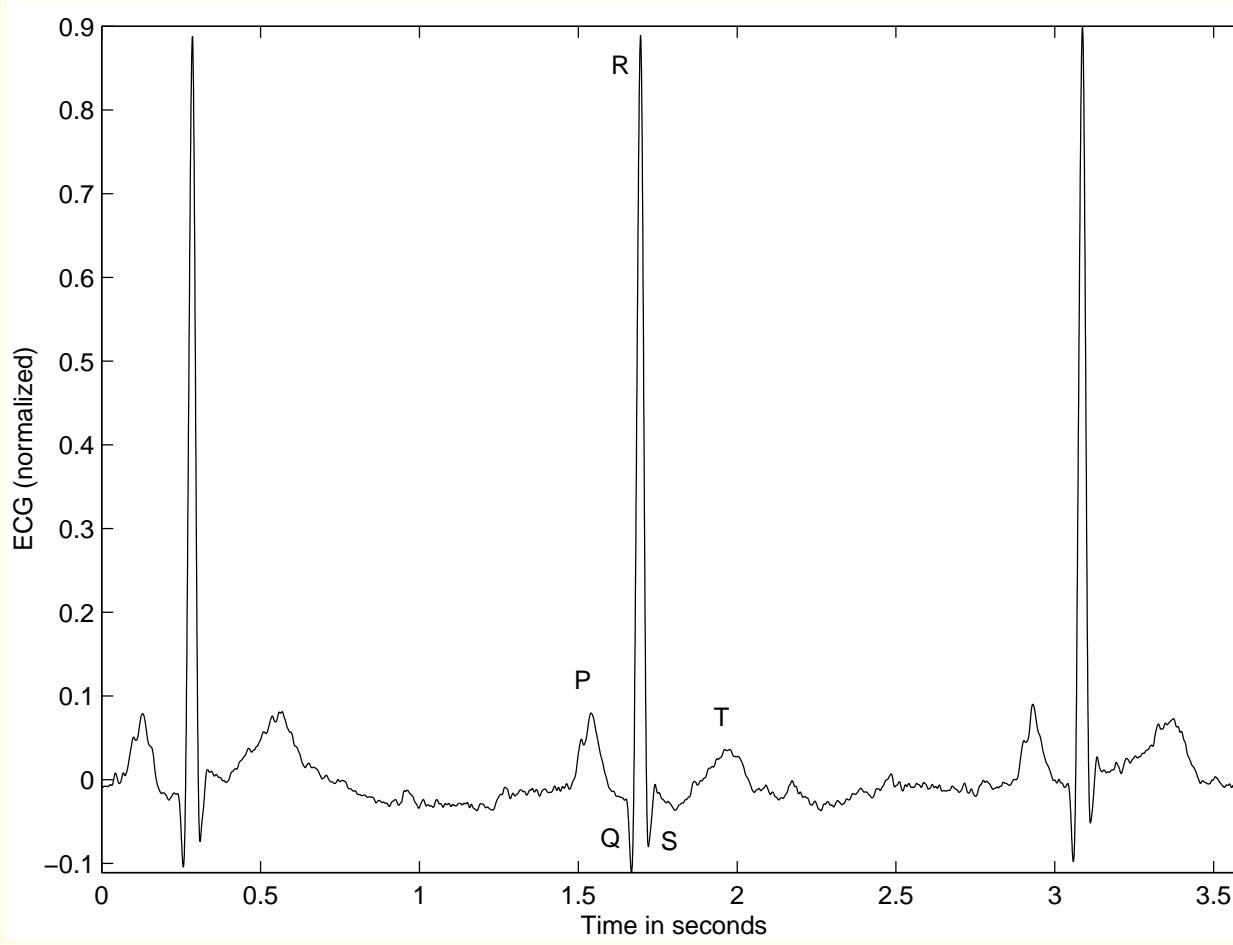


Figure 1.13: A typical ECG signal (male subject of age 24 years). (*Note:* Signal values are not calibrated, that is, specified in physical units, in many applications. As is the case in this plot, signal values in plots in this book are in arbitrary or normalized units unless specified.)



Disturbance in the regular rhythmic activity of the heart:

arrhythmia.

Cardiac arrhythmia may be caused by:

irregular firing patterns from the SA node,

abnormal and additional pacing activity

from other parts of the heart.



Many parts of the heart possess inherent rhythmicity

and pacemaker properties:

SA node, AV node, Purkinje fibers,

atrial tissue, and ventricular tissue.

If the SA node is depressed or inactive, any one of the above

may take over the role of the pacemaker

or introduce *ectopic* beats.



Different types of abnormal rhythm (arrhythmia) result from variations in the site and frequency of impulse formation.

Premature ventricular contractions (PVCs):

caused by ectopic foci on the ventricles.

May lead to ventricular dissociation and fibrillation —

a state of disorganized contraction of the

ventricles independent of the atria.

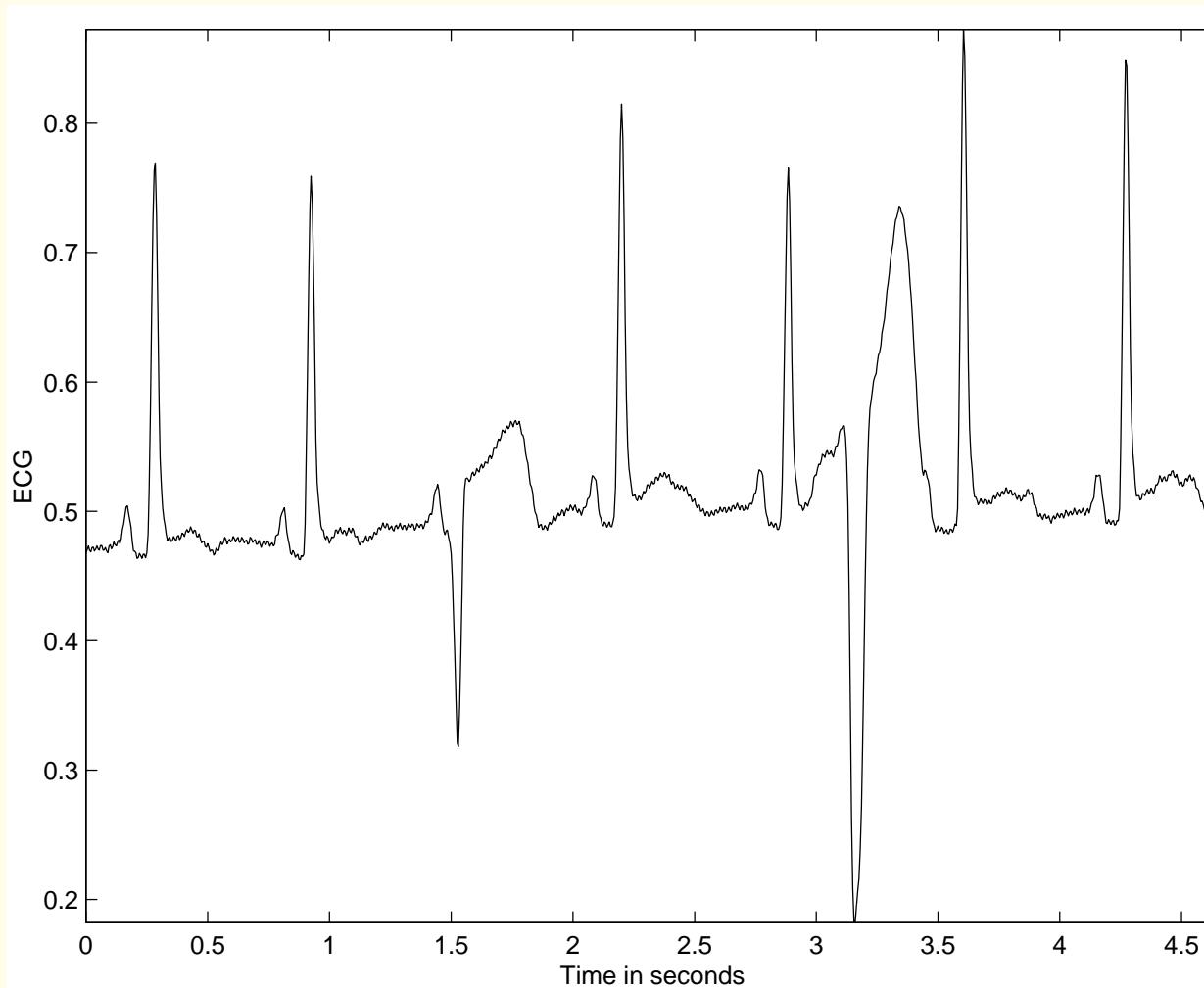


Figure 1.14: ECG signal with PVCs. The third and sixth beats are PVCs. The first PVC has blocked the normal beat that would have appeared at about the same time instant, but the second PVC has not blocked any normal beat triggered by the SA node. Data courtesy of G. Groves and J. Tyberg, Department of Physiology and Biophysics, University of Calgary.



The ECG signal of a patient (male, 65 years) with PVCs. Each strip is of duration 10 s; the signal continues from top to bottom. The second half of the seventh strip and the first half of the eighth strip illustrate an episode of bigeminy.



QRS waveshape affected by conduction disorders:

bundle-branch block causes a widened and jagged QRS.

Ventricular hypertrophy or enlargement: wide QRS.

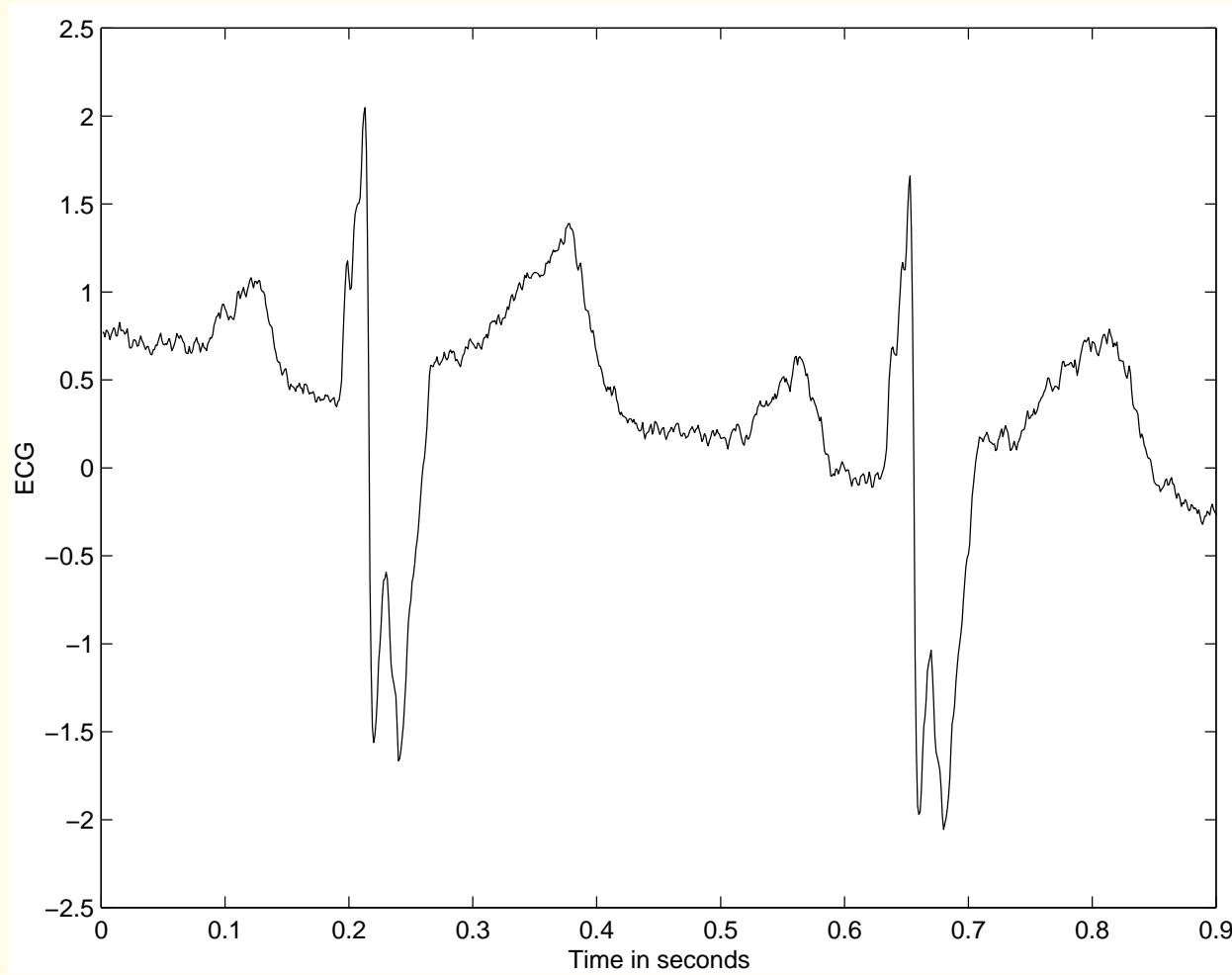


Figure 1.15: ECG signal of a patient with right bundle-branch block and hypertrophy (male patient of age 3 months). The QRS complex is wider than normal, and displays an abnormal, jagged waveform due to desynchronized contraction of the ventricles. (The signal also has a base-line drift, which has not been corrected for.)



ST segment: normally iso-electric —

flat and in line with the PQ segment.

May be elevated or depressed due to myocardial ischemia —

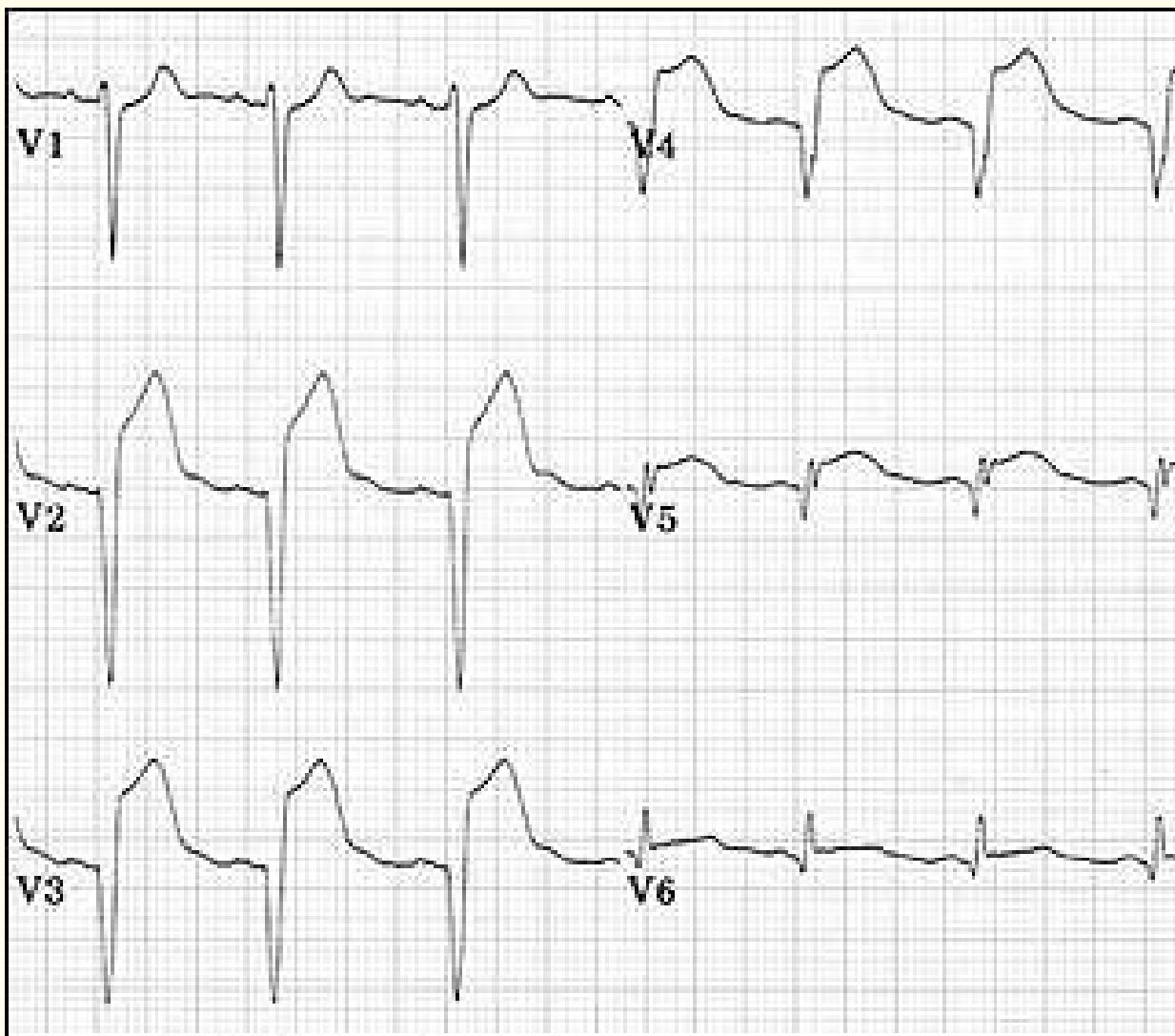
reduced blood supply to a part of the heart muscles

due to a block in the coronary arteries,

or due to myocardial infarction —

dead myocardial tissue incapable of contraction

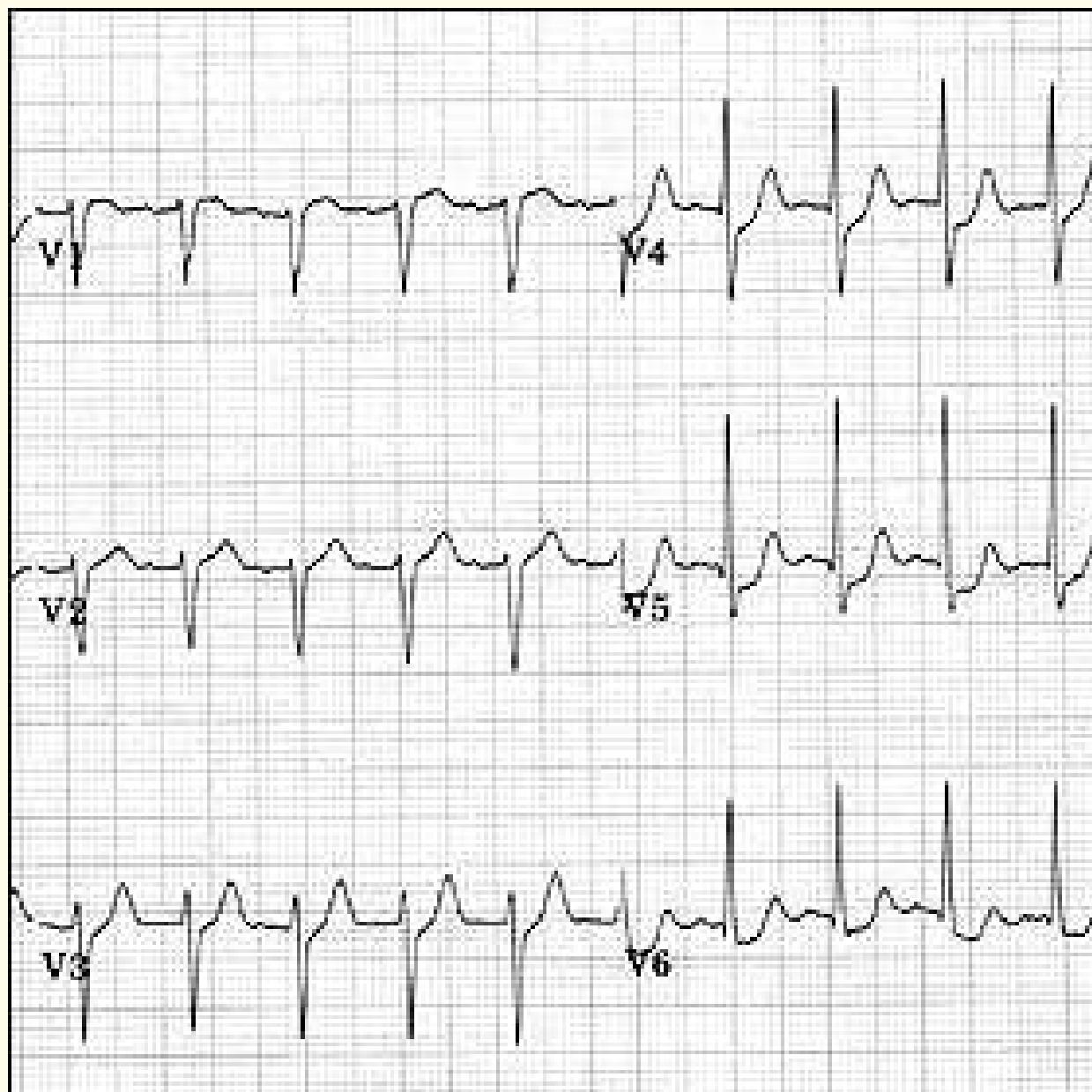
due to total lack of blood supply.



© 1997 Frank G. Yanowitz, M.D.

ST elevation: Ischemic Heart Disease: Acute transmural injury, acute anterior MI.

http://library.med.utah.edu/kw/ecg/ecg_outline/Lesson10/index.html



© 1997 Frank G. Yanowitz, M.D.

ST depression: Subendocardial ischemia: exercise induced or during angina attack.

http://library.med.utah.edu/kw/ecg/ecg_outline/Lesson10/index.html



ECG signal acquisition:

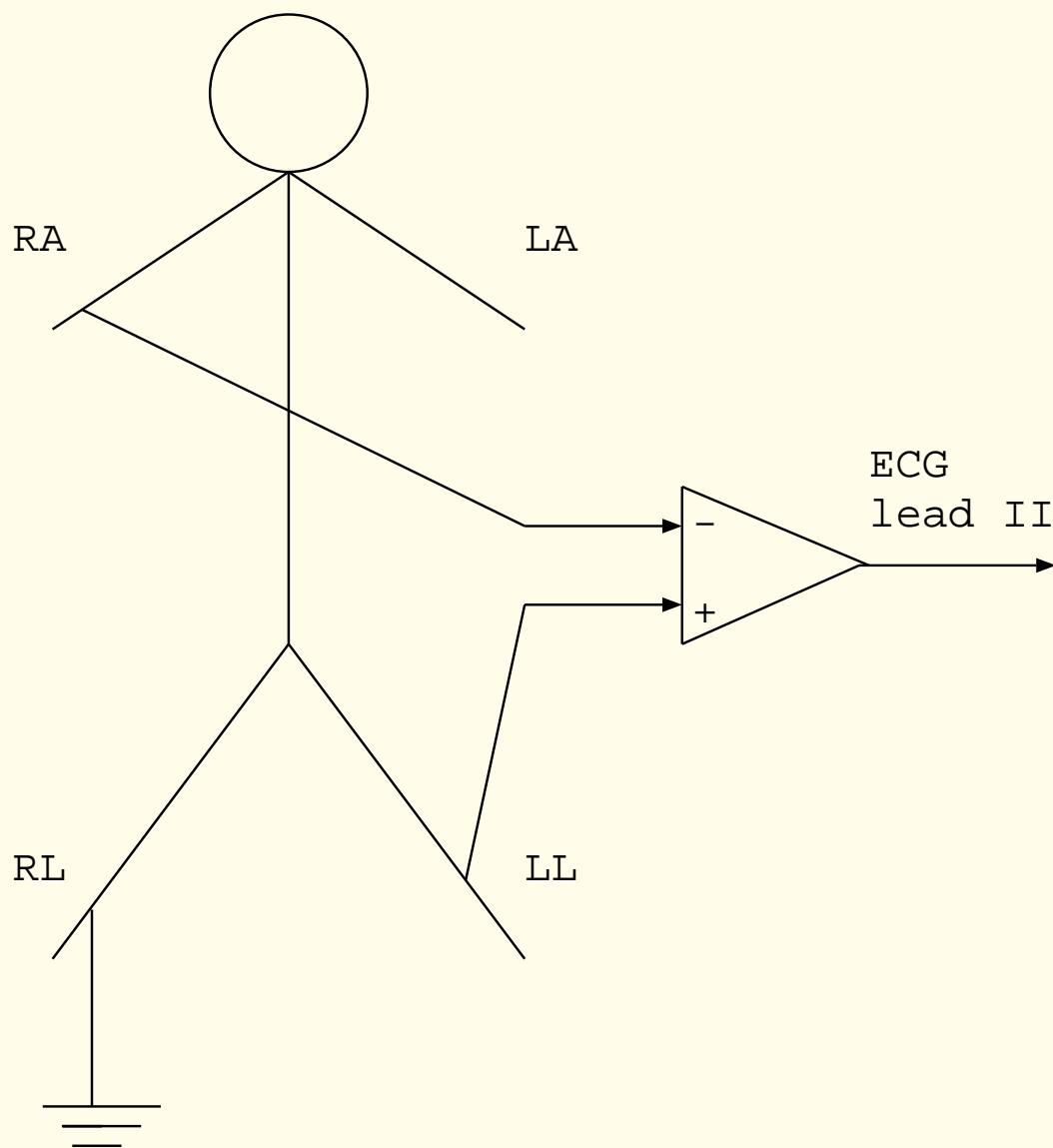
Clinical practice: standard 12-channel ECG

obtained using four limb leads

and chest leads in six positions.

Right leg: reference electrode.

Left arm, right arm, left leg: leads I, II, and III.



Lead configuration to acquire lead II ECG.



Wilson's central terminal

formed by combining left arm, right arm, and left leg leads:

used as the reference for chest leads.



The *augmented* limb leads known as aVR, aVL, and aVF —

aV for augmented lead, R for right arm,

L for left arm, and F for left foot —

obtained by using the exploring electrode on the limb

indicated by the lead name, with the reference being

Wilson's central terminal without the exploring limb lead.



Hypothetical equilateral triangle formed by

leads I, II, and III: *Einthoven's triangle*.

Center of the triangle: Wilson's central terminal.

Schematically, the heart is at the center of the triangle.



The six leads measure projections of the 3D cardiac

electrical vector onto the axes of the leads.

Six axes: sample the $0^\circ - 180^\circ$ range in steps of $\sim 30^\circ$.

Facilitate viewing and analysis of the electrical activity

of the heart from different perspectives in the frontal plane.

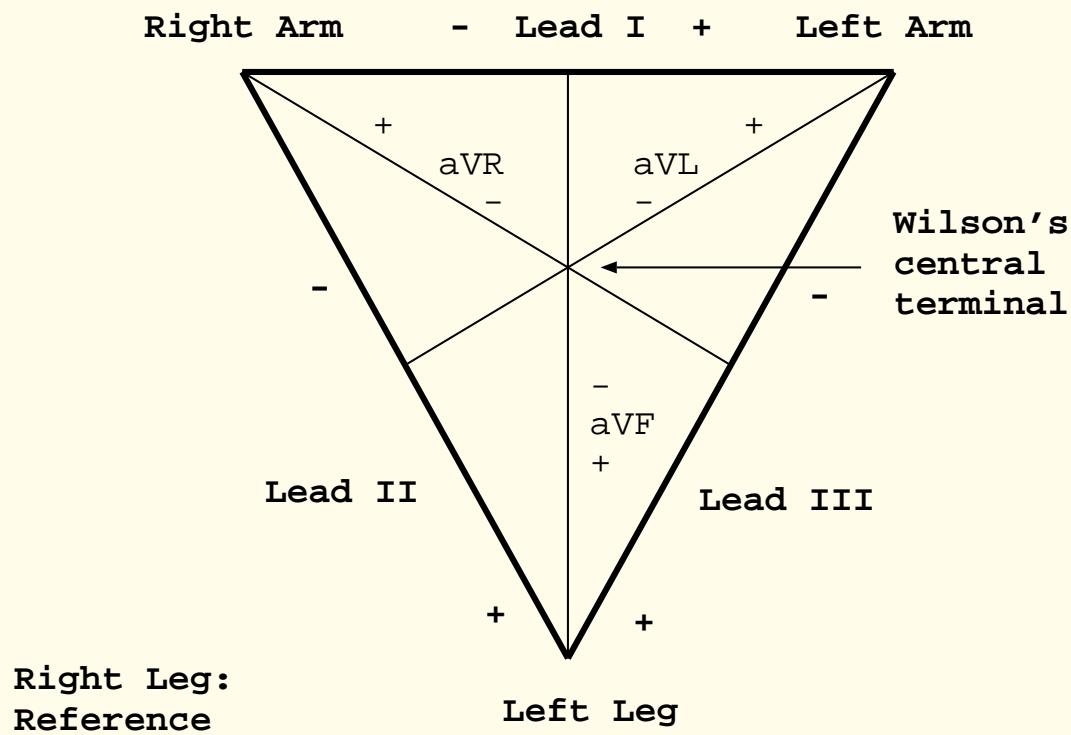


Figure 1.16: Einthoven's triangle and the axes of the six ECG leads formed by using four limb leads.



Six chest leads (V1 – V6) obtained from

six standardized positions on the chest

with Wilson's central terminal as the reference.

V1 and V2 leads placed at the fourth intercostal space

just to the right and left of the sternum, respectively.

V4: fifth intercostal space at the left midclavicular line, etc.



The six chest leads permit viewing

the cardiac electrical vector from

different orientations in a cross-sectional plane:

V5 and V6 most sensitive to left ventricular activity;

V3 and V4 depict septal activity best;

V1 and V2 reflect activity in the right-half of the heart.

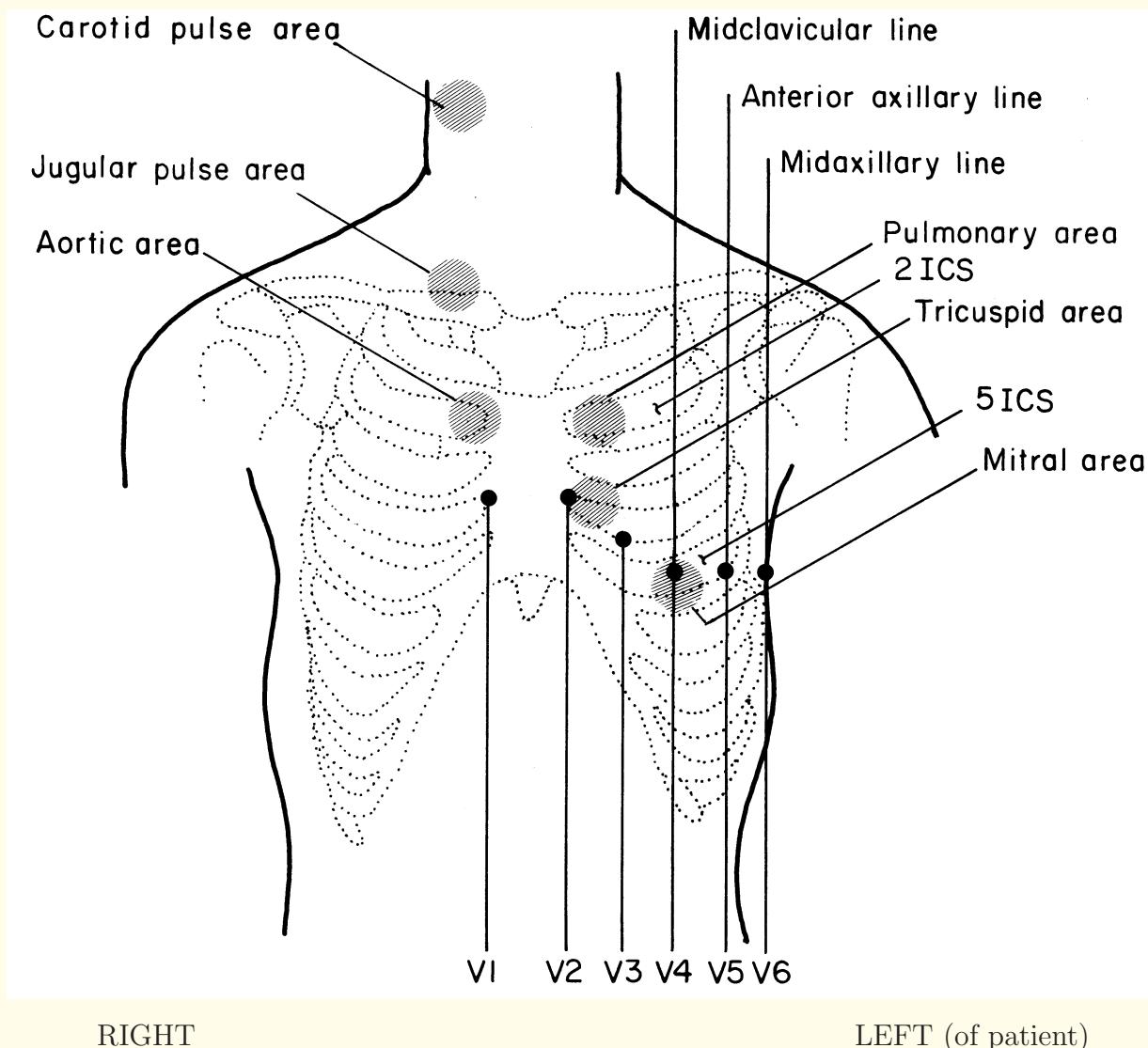


Figure 1.17: Positions for placement of the precordial (chest) leads V1 – V6 for ECG, auscultation areas for heart sounds, and pulse transducer positions for the carotid and jugular pulse signals. ICS: intercostal space.

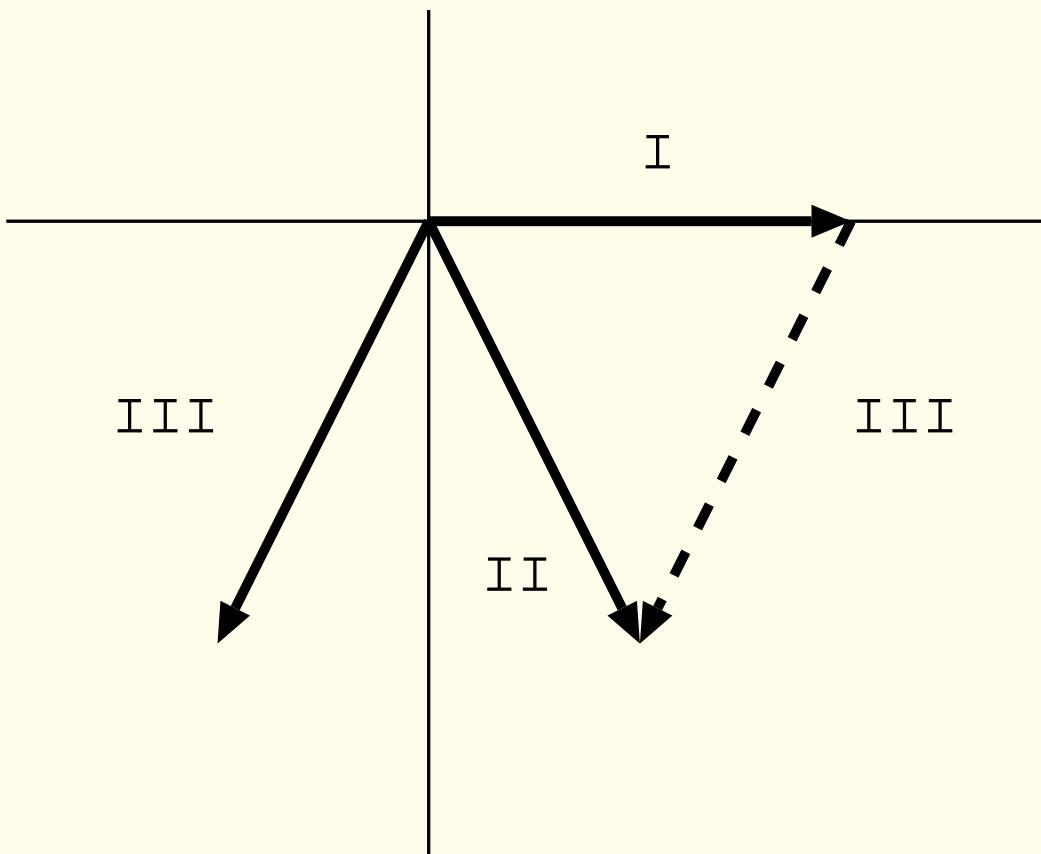


In spite of being redundant, the 12-lead system serves as the basis of the standard clinical ECG.

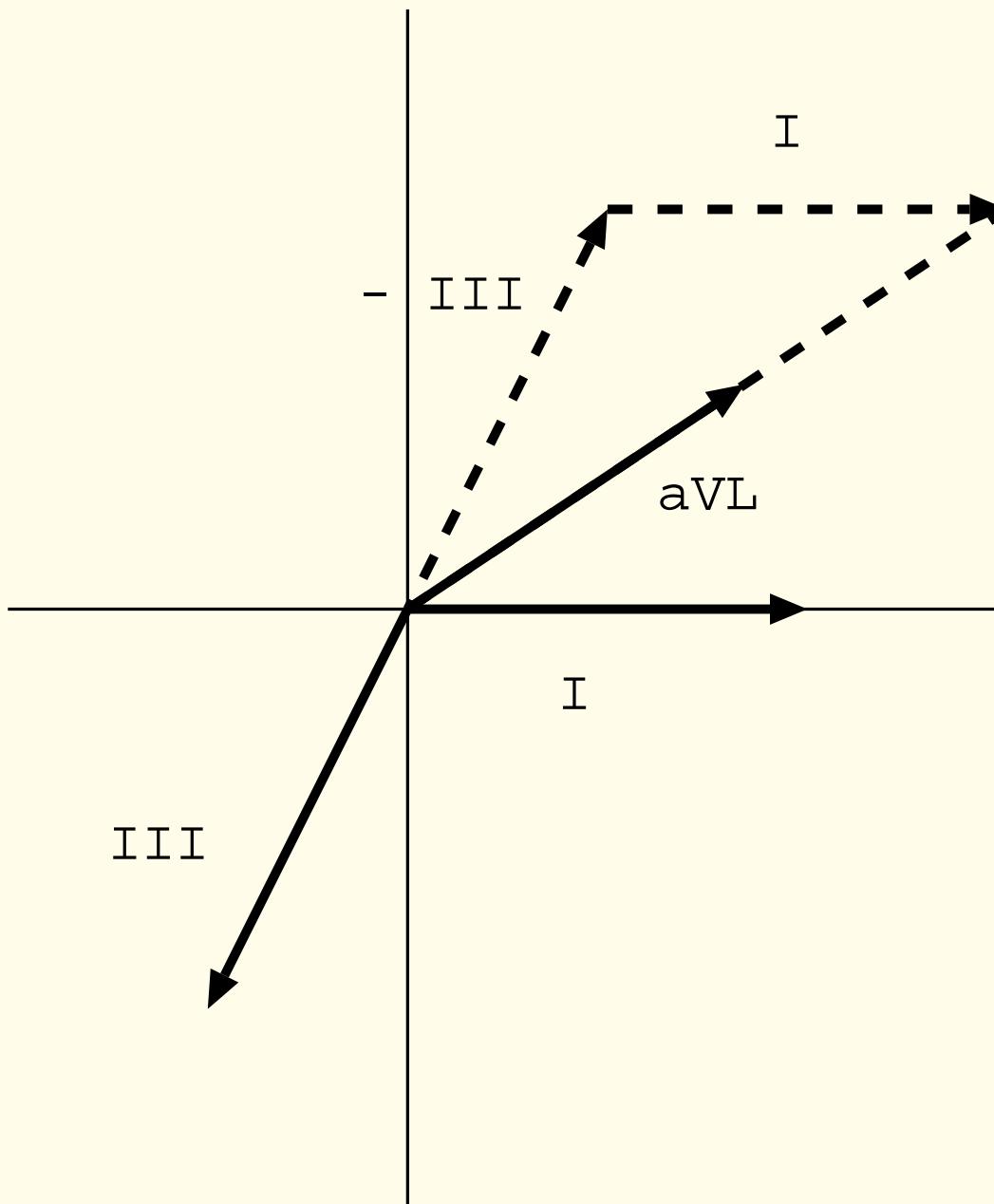
Clinical ECG interpretation is mainly empirical, based on experimental knowledge.

Some of the lead inter-relationships are:

- $\text{II} = \text{I} + \text{III}$
- $\text{aVL} = (\text{I} - \text{III}) / 2$.



Vectorial relationship between ECG leads I, II, and III.



Vectorial relationship between ECG leads I, III, and aVL.



Important features of standard clinical ECG:

- Rectangular calibration pulse, 1 mV and 200 ms :
pulse of 1 cm height on the paper plot.
- Speed 25 mm/s : 0.04 s/mm or 40 ms/mm .
Calibration pulse width: 5 mm .
- ECG signal peak value normally about 1 mV .
- Amplifier gain: 1,000.



- Clinical ECG: filtered to $0.05 - 100 \text{ Hz}$ bandwidth.

Recommended sampling rate:

500 Hz for diagnostic ECG.

Distortions in the shape of the calibration pulse may indicate improper filter settings or a poor signal acquisition system.

- ECG for heart-rate monitoring: reduced bandwidth $0.5 - 50 \text{ Hz}$.
- High-resolution ECG: greater bandwidth of $0.05 - 500 \text{ Hz}$.

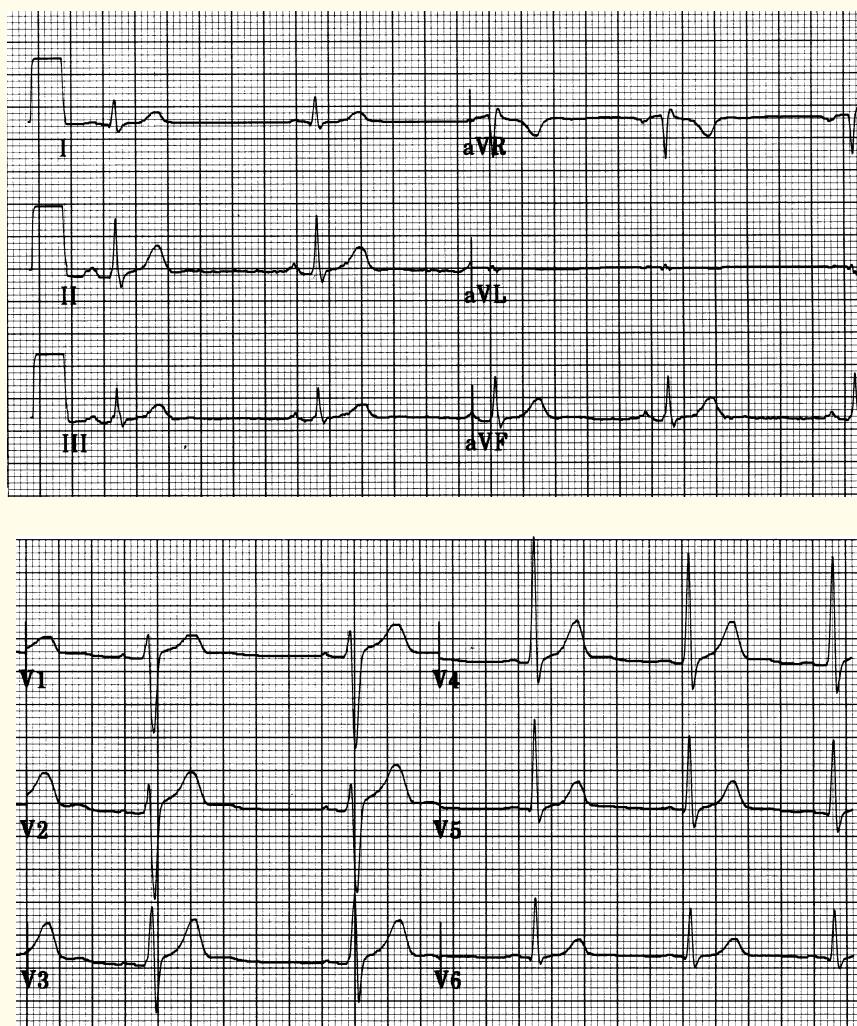


Figure 1.18: Standard 12-lead ECG of a normal male adult. Courtesy of E. Gedamu and L.B. Mitchell, Foothills Hospital, Calgary.

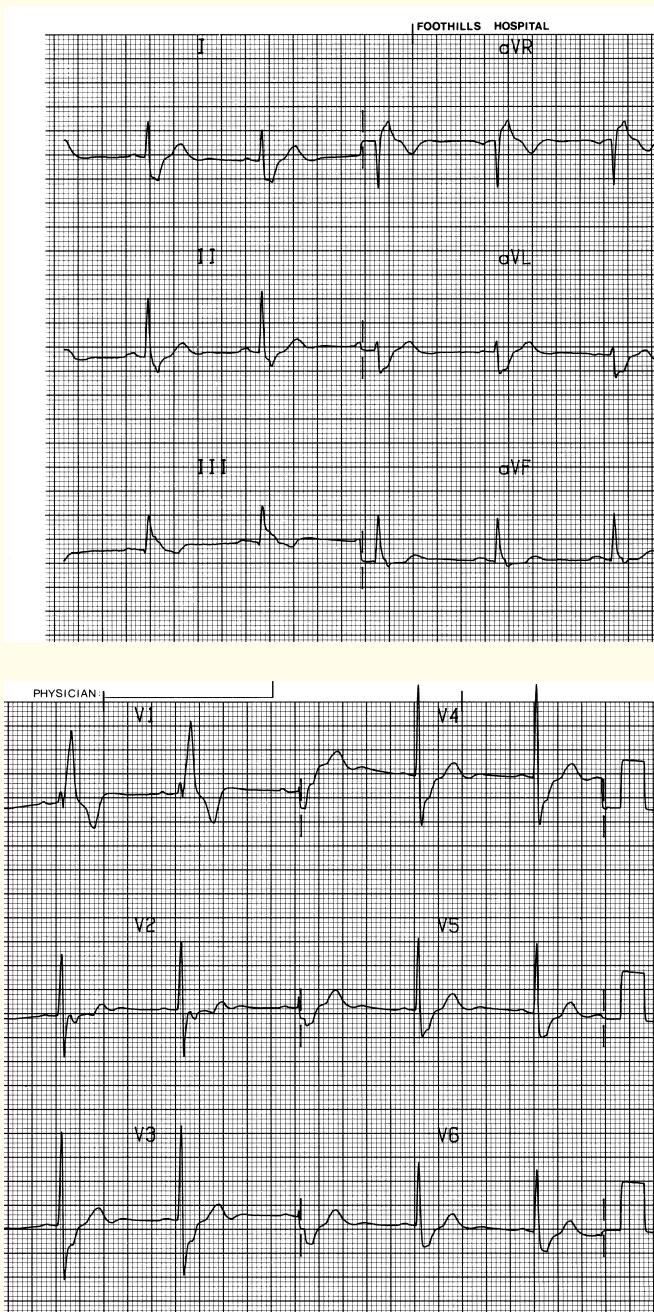


Figure 1.19: Standard 12-lead ECG of a patient with right bundle-branch block. Courtesy of L.B. Mitchell, Foothills Hospital, Calgary.



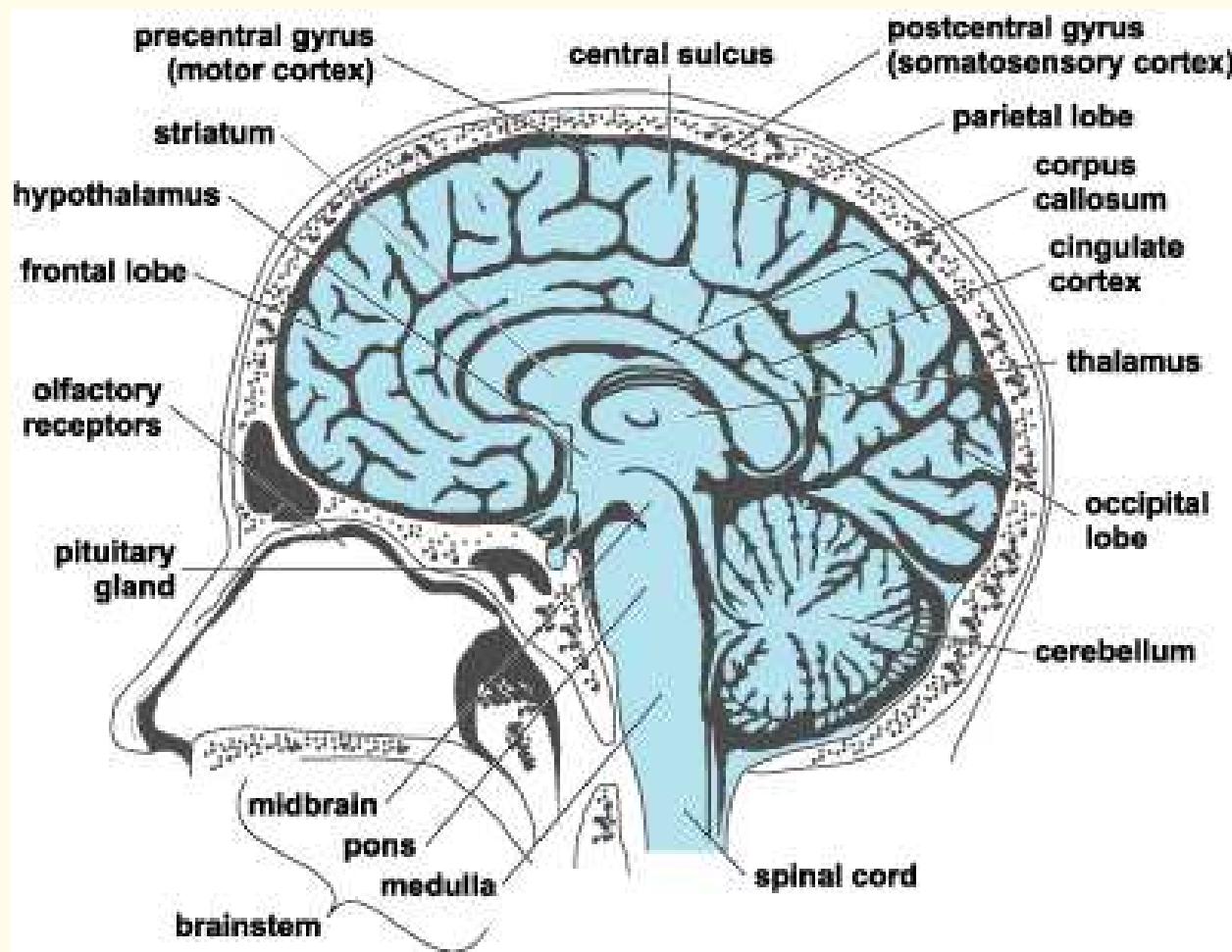
1.2.5 *The electroencephalogram (EEG)*

EEG or *brain waves*: electrical activity of the brain.

Main parts of the brain: cerebrum, cerebellum,

brain stem (midbrain, pons medulla, reticular formation),

thalamus (between the midbrain and the hemispheres).



Midsagittal section through the human brain.

From www.answers.com.



Cerebrum divided into two hemispheres,

separated by a longitudinal fissure with a large

connective band of fibers: corpus callosum.

Outer surface of the cerebral hemispheres (cerebral cortex)

composed of neurons (grey matter) in convoluted patterns,

separated into regions by fissures (sulci).

Beneath the cortex lie nerve fibers that lead to

other parts of the brain and the body (white matter).



Cortical potentials generated due to excitatory and inhibitory post-synaptic potentials developed by cell bodies and dendrites of pyramidal neurons.

Physiological control processes, thought processes, and external stimuli generate signals in the corresponding parts of the brain: recorded at the scalp using surface electrodes.



Scalp EEG: average of multifarious activities of many small zones of the cortical surface beneath the electrode.

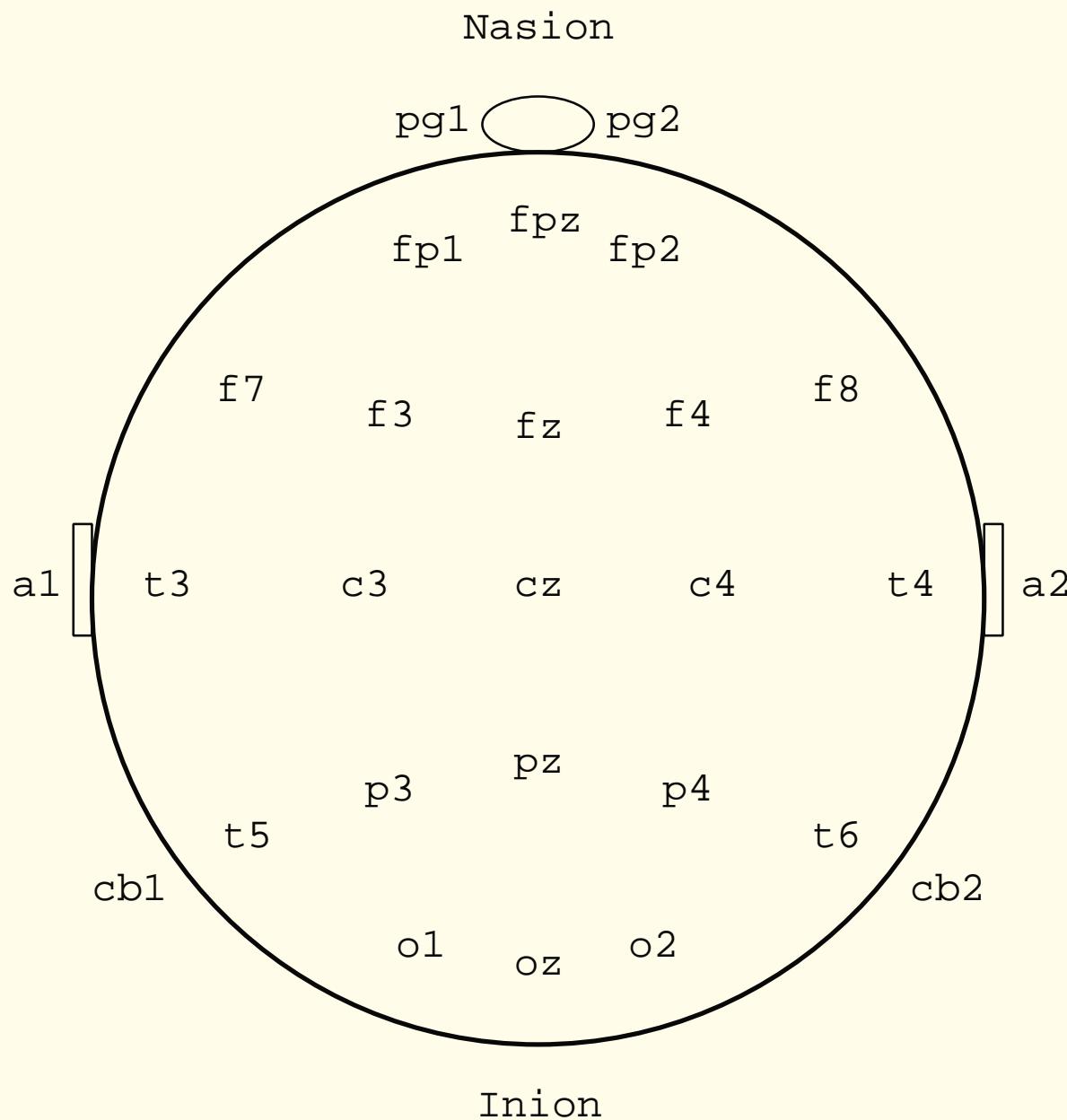


Figure 1.20: The 10 – 20 system of electrode placement for EEG recording. Notes regarding channel labels: pg– naso-pharyngeal, a– auricular (ear lobes), fp– pre-frontal, f– frontal, p– parietal, c– central, o– occipital, t– temporal, cb– cerebellar, z– midline, odd numbers on the left, even numbers on the right of the subject.



EEG instrumentation settings:

lowpass filtering at 75 Hz ,

recording at $100\text{ }\mu\text{V/cm}$ and 30 mm/s

for $10 - 20$ minutes over $8 - 16$ simultaneous channels.

Monitoring of sleep EEG and

detection of transients related to epileptic seizures:

multichannel EEG acquisition over several hours.



Special EEG techniques:

needle electrodes,

naso-pharyngeal electrodes,

electrocorticogram (ECoG) from exposed cortex,

intracerebral electrodes.



Evocative techniques for recording the EEG:

initial recording at rest (eyes open, eyes closed),

hyperventilation (after breathing at 20 respirations

per minute for 2 – 4 minutes),

photic stimulation (with 1 – 50 flashes of light per second),

auditory stimulation with loud clicks,

sleep (different stages), and pharmaceuticals or drugs.



EEG rhythms or frequency bands:

- Delta (δ): $0.5 \leq f < 4 \text{ Hz}$;
- Theta (θ): $4 \leq f < 8 \text{ Hz}$;
- Alpha (α): $8 \leq f \leq 13 \text{ Hz}$; and
- Beta (β): $f > 13 \text{ Hz}$.



EEG rhythms:

associated with physiological and mental processes.

Alpha: principal resting rhythm of the brain:

common in wakeful, resting adults,

especially in the occipital area with bilateral synchrony.

Auditory and mental arithmetic tasks with the

eyes closed lead to strong alpha waves:

suppressed when the eyes are opened.



Alpha wave replaced by

slower rhythms at various stages of sleep.

Theta waves: beginning stages of sleep.

Delta waves: deep-sleep stages.

High-frequency beta waves:

background activity in tense and anxious subjects.

Spikes and sharp waves: epileptogenic regions.

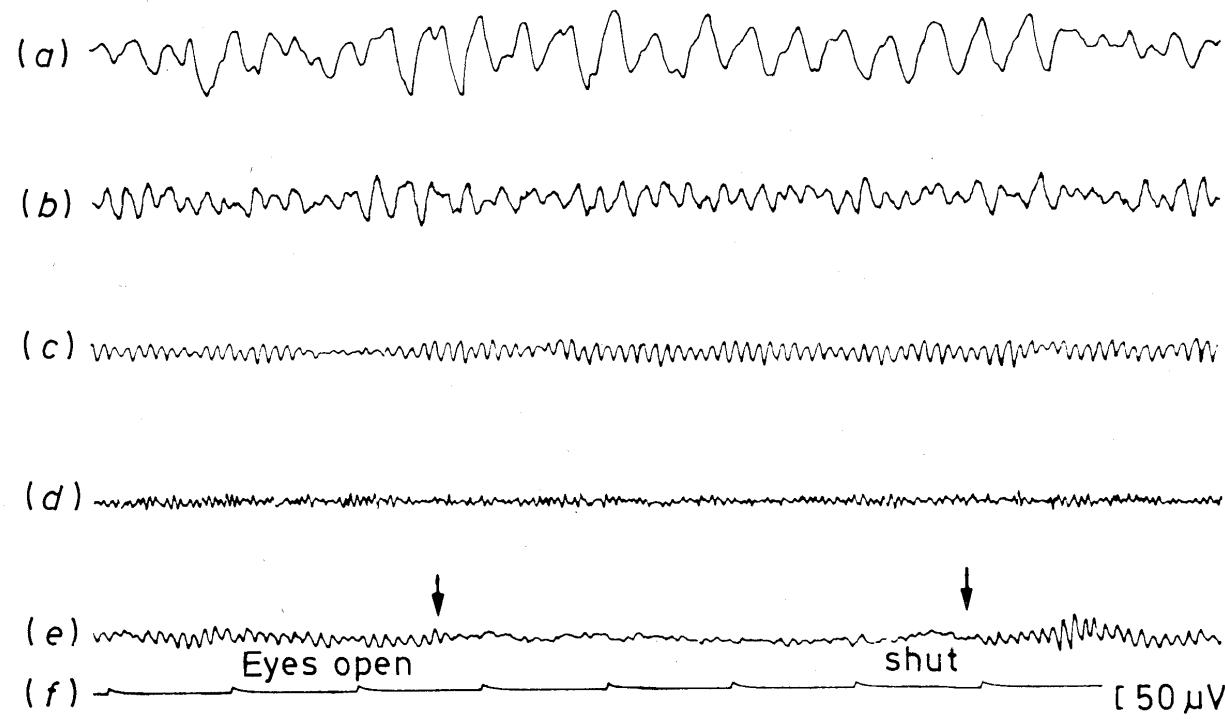


Figure 1.21: From top to bottom: (a) delta rhythm; (b) theta rhythm; (c) alpha rhythm; (d) beta rhythm; (e) blocking of the alpha rhythm by eye opening; (f) 1 s time markers and 50 μ V marker. Reproduced with permission from R. Cooper, J.W. Osselton, and J.C. Shaw, *EEG Technology*, 3rd Edition, 1980. ©Butterworth Heinemann Publishers, a division of Reed Educational & Professional Publishing Ltd., Oxford, UK.

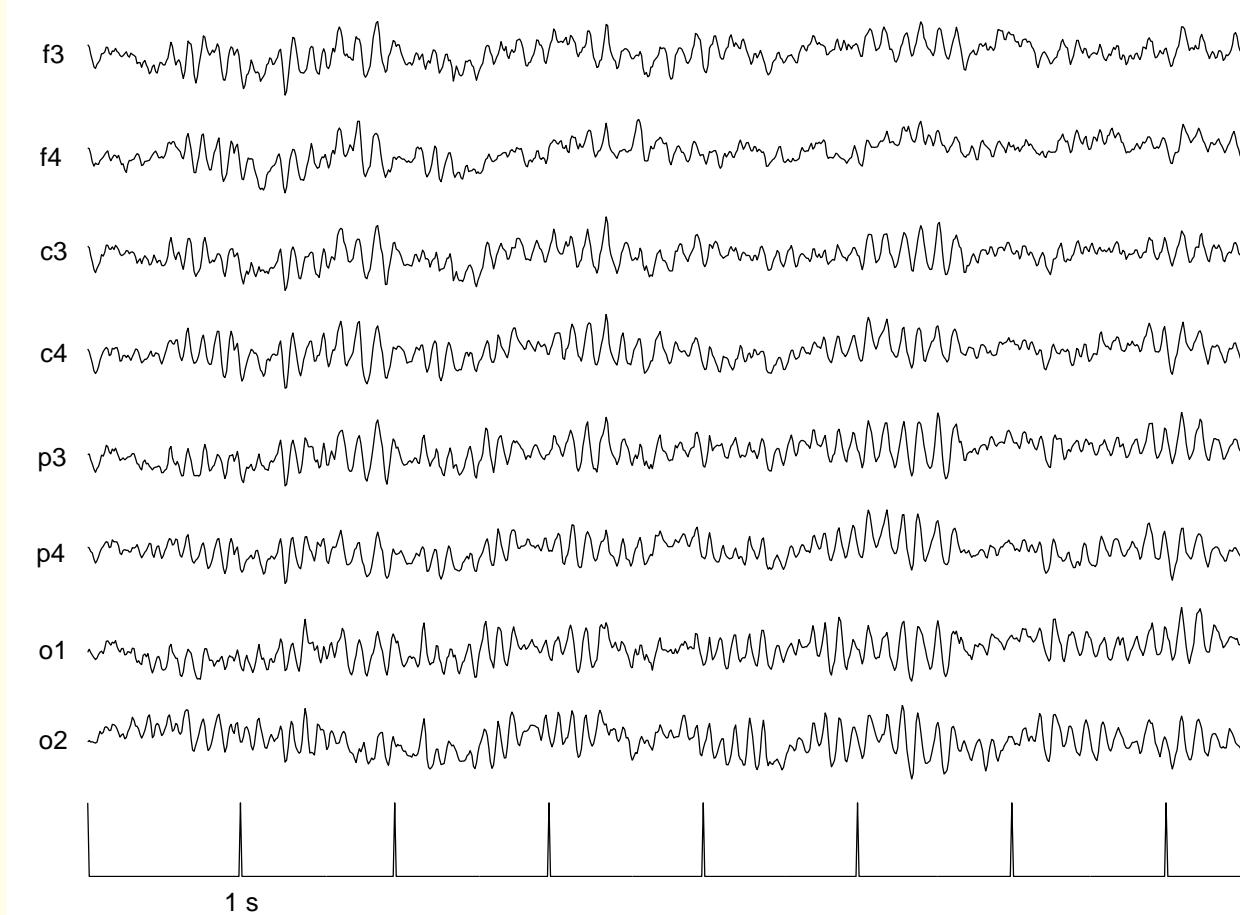


Figure 1.22: Eight channels of the EEG of a subject displaying alpha rhythm. See Figure 1.20 for details regarding channel labels. Data courtesy of Y. Mizuno-Matsumoto, Osaka University Medical School, Osaka, Japan.

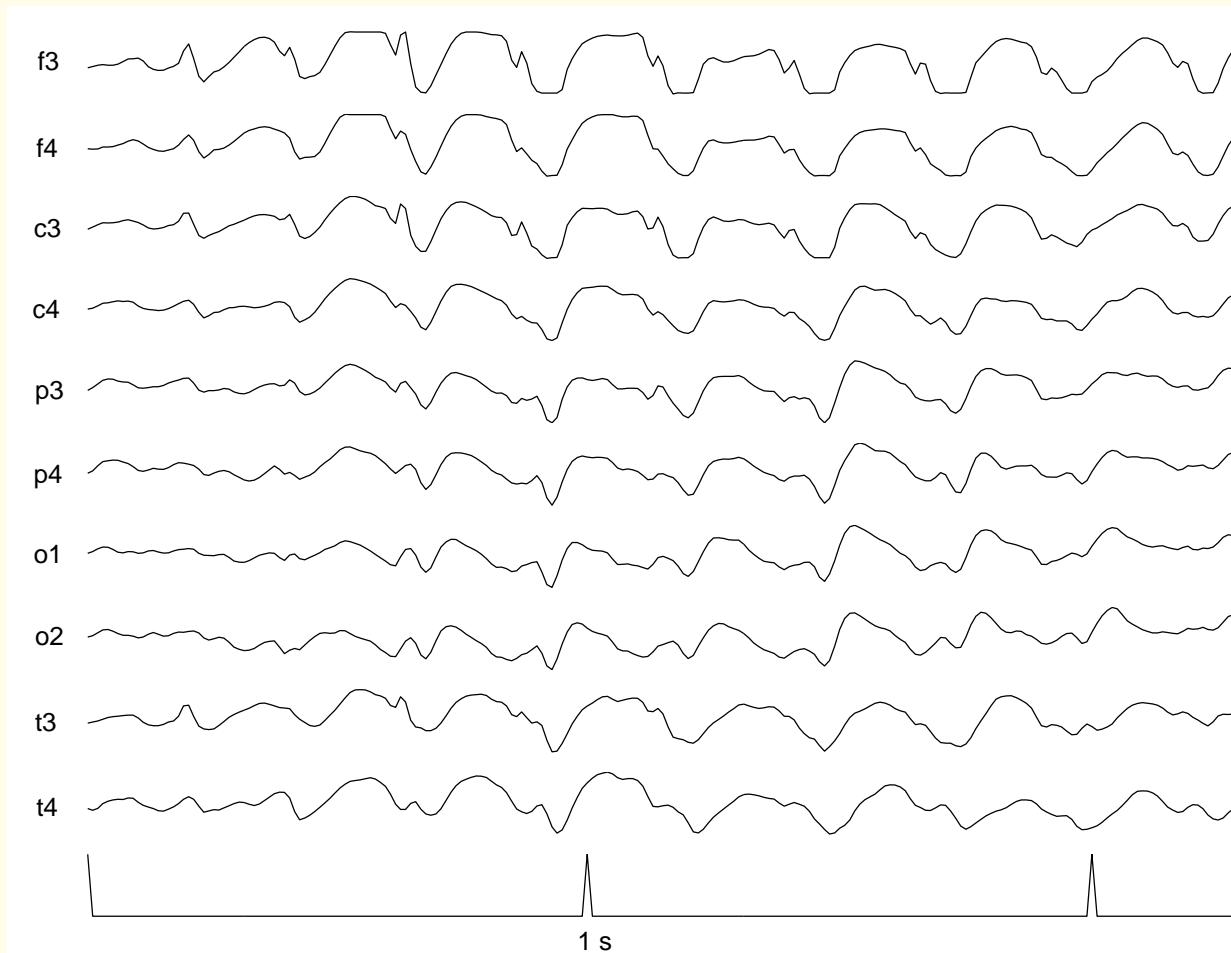


Figure 1.23: Ten channels of the EEG of a subject displaying spike-and-wave complexes. See Figure 1.20 for details regarding channel labels. Data courtesy of Y. Mizuno-Matsumoto, Osaka University Medical School, Osaka, Japan. Note that the time scale is expanded compared to that of Figure 1.22.



1.2.6 *Event-related potentials (ERPs)*

The term *event-related potential*

is more general than and preferred to

the term *evoked potential*:

includes the ENG or the EEG in response to

light, sound, electrical, or other external stimuli.



Short-latency ERPs: dependent upon the

physical characteristics of the stimulus,

Longer-latency ERPs: influenced by the

conditions of presentation of the stimuli.

Somatosensory evoked potentials:

useful for noninvasive evaluation of the nervous system

from a peripheral receptor to the cerebral cortex.



Median nerve short-latency SEPs:

obtained by placing stimulating electrodes

2 – 3 cm apart over the median nerve at the wrist

with electrical stimulation at *5 – 10 pps*,

each stimulus pulse less than *0.5 ms*, about *100 V*

(producing a visible thumb twitch).

SEPs recorded from the surface of the scalp.

Latency, duration, and amplitude of the response measured.



ERPs and SEPs are weak signals:

buried in ongoing activity of associated systems.

SNR improvement: synchronized averaging and filtering.



1.2.7 *The electrogastrogram (EGG)*

Electrical activity of the stomach:

rhythmic waves of depolarization and repolarization

of smooth muscle cells.

Surface EGG: overall electrical activity of the stomach.

Gastric dysrhythmia or arrhythmia may be detected

with the EGG.



1.2.8 *The phonocardiogram (PCG)*

PCG: vibration or sound signal related to the contractile

activity of the cardiohemic system (heart and blood).

Recording the PCG requires a transducer to convert the

vibration or sound signal into an electronic signal:

microphones, pressure transducers, or accelerometers.



Cardiovascular diseases and defects cause changes or additional sounds and murmurs: useful in diagnosis.



The genesis of heart sounds:

Heart sounds not caused by valve leaflet movements *per se*,

but by vibrations of the whole cardiovascular system

triggered by pressure gradients.

Secondary sources on the chest related to the well-known

auscultatory areas: mitral, aortic, pulmonary, tricuspid.



A normal cardiac cycle contains two major sounds —

the first heart sound (S1) and

the second heart sound (S2).

S1 occurs at the onset of ventricular contraction:

corresponds in timing to the QRS in the ECG.

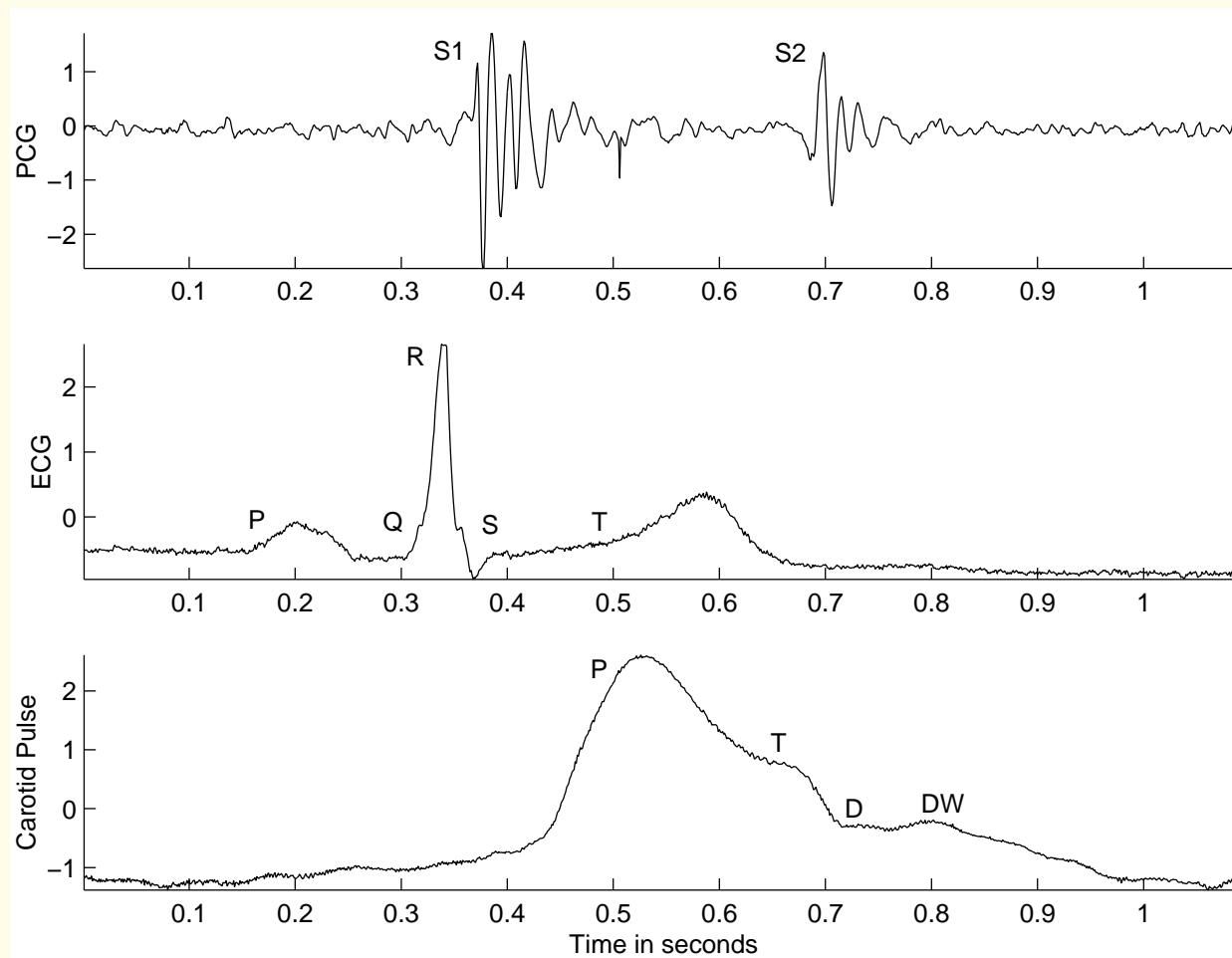


Figure 1.24: Three-channel simultaneous record of the PCG, ECG, and carotid pulse signals of a normal male adult.



Initial vibrations in S1: first myocardial contractions

in the ventricles move blood toward the atria,

sealing the AV (mitral and tricuspid) valves.

Second component of S1: abrupt tension of the

closed AV valves, decelerating the blood.

Next, the semilunar (aortic and pulmonary) valves open:

blood is ejected out of the ventricles.



Third component of S1: caused by oscillation of blood

between the root of the aorta and the ventricular walls.

Fourth component of S1: vibrations caused by turbulence

in the ejected blood flowing rapidly

through the ascending aorta and the pulmonary artery.



A. COMPONENTS OF FIRST HEART SOUND

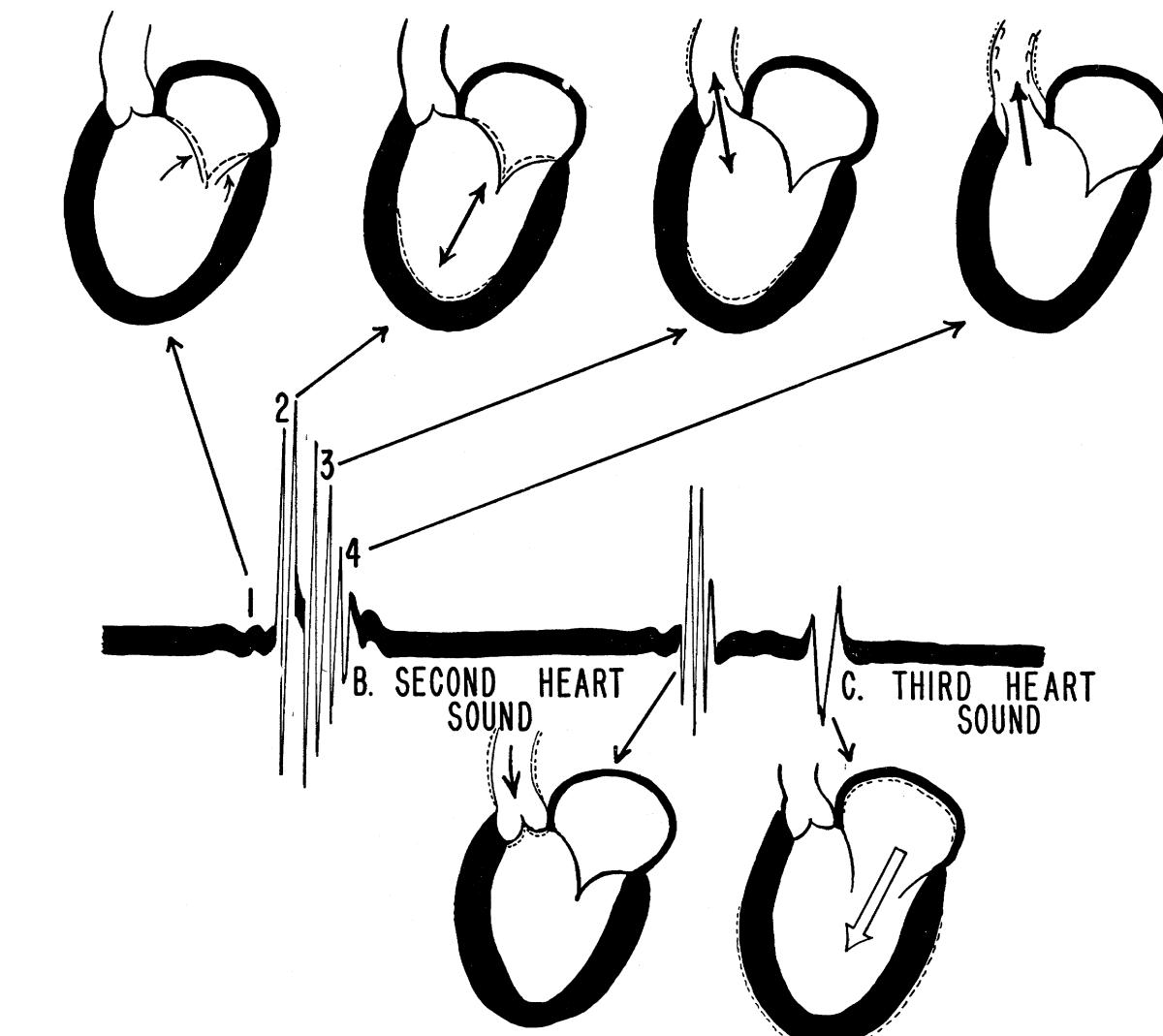


Figure 1.25: Schematic representation of the genesis of heart sounds. Only the left portion of the heart is illustrated as it is the major source of the heart sounds. The corresponding events in the right portion also contribute to the sounds. The atria do not contribute much to the heart sounds. Reproduced with permission from R.F. Rushmer, *Cardiovascular Dynamics*, 4th edition, ©W.B. Saunders, Philadelphia, PA, 1976.



Following the systolic pause in a normal cardiac cycle,
second sound S2 caused by the
closure of the semilunar valves.

Primary vibrations occur in the arteries due to
deceleration of blood;
the ventricles and atria also vibrate, due to transmission
of vibrations through the blood, valves, and the valve rings.



S2 has two components:

one due to closure of the aortic valve (A2)

another due to closure of the pulmonary valve (P2).

The aortic valve normally closes before the

pulmonary valve; A2 precedes P2 by a few milliseconds.



Pathologic conditions could cause this gap to widen,
or may also reverse the order of occurrence of A2 and P2.
A2 – P2 gap is also widened in normal subjects
during inspiration.



Other sounds:

S3: sudden termination of the ventricular rapid-filling phase.

S4: atrial contractions displacing blood into the distended ventricles.

Valvular clicks and snaps.

Murmurs.



Heart murmurs:

S1 – S2 and S2 – S1 intervals normally silent:

corresponding to ventricular systole and diastole.

Murmurs caused by cardiovascular defects and diseases

may occur in these intervals.

Murmurs are high-frequency, noise-like sounds:

arise when the velocity of blood becomes high

as it flows through an irregularity (constriction, baffle).



Conditions that cause turbulence in blood flow:

valvular stenosis and insufficiency.

Valve stenosed due to the deposition of calcium:

valve leaflets stiffened and do not open completely —

obstruction or baffle in the path of the blood being ejected.

Valve insufficient when it cannot close effectively:

reverse leakage or regurgitation of blood through
a narrow opening.



Systolic murmurs (SM) caused by ventricular septal defect (VSD) — hole in the wall between the left and right ventricles; aortic stenosis (AS), pulmonary stenosis (PS), mitral insufficiency (MI), and tricuspid insufficiency (TI).



Semilunar valvular stenosis (AS, PS):

obstruction in the path of blood being ejected during systole.

AV valvular insufficiency (MI, TI):

regurgitation of blood to the atria during

ventricular contraction.



Diastolic murmurs (DM) caused by

aortic or pulmonary insufficiency (AI, PI),

mitral or tricuspid stenosis (MS, TS),

atrial septal defect (ASD).



Features of heart sounds and murmurs:

intensity, frequency content, and timing

affected by physical and physiological factors such as

recording site on thorax, intervening thoracic structures,

left ventricular contractility,

position of the cardiac valves at the onset of systole,

the degree of the defect present,

the heart rate, and blood velocity.



- S1 is loud and delayed in mitral stenosis;
- right bundle-branch block causes wide splitting of S2;
- left bundle-branch block results in reversed splitting of S2;
- acute myocardial infarction causes a pathologic S3;
- severe mitral regurgitation (MR) leads to an increased S4.



Although murmurs are noise-like events, their features aid in distinguishing between different causes.

Aortic stenosis causes a diamond-shaped

midsystolic murmur.

Mitral stenosis causes a decrescendo – crescendo type

diastolic – presystolic murmur.



Recording PCG signals:

Piezoelectric contact sensors sensitive to displacement

or acceleration at the skin surface.

Hewlett Packard HP21050A transducer:

nominal bandwidth of $0.05 - 1,000 \text{ Hz}$.



PCG recording performed in a quiet room;

patient in supine position, head resting on a pillow.

PCG transducer placed firmly at the desired position

on the chest using a suction ring and/or a rubber strap.

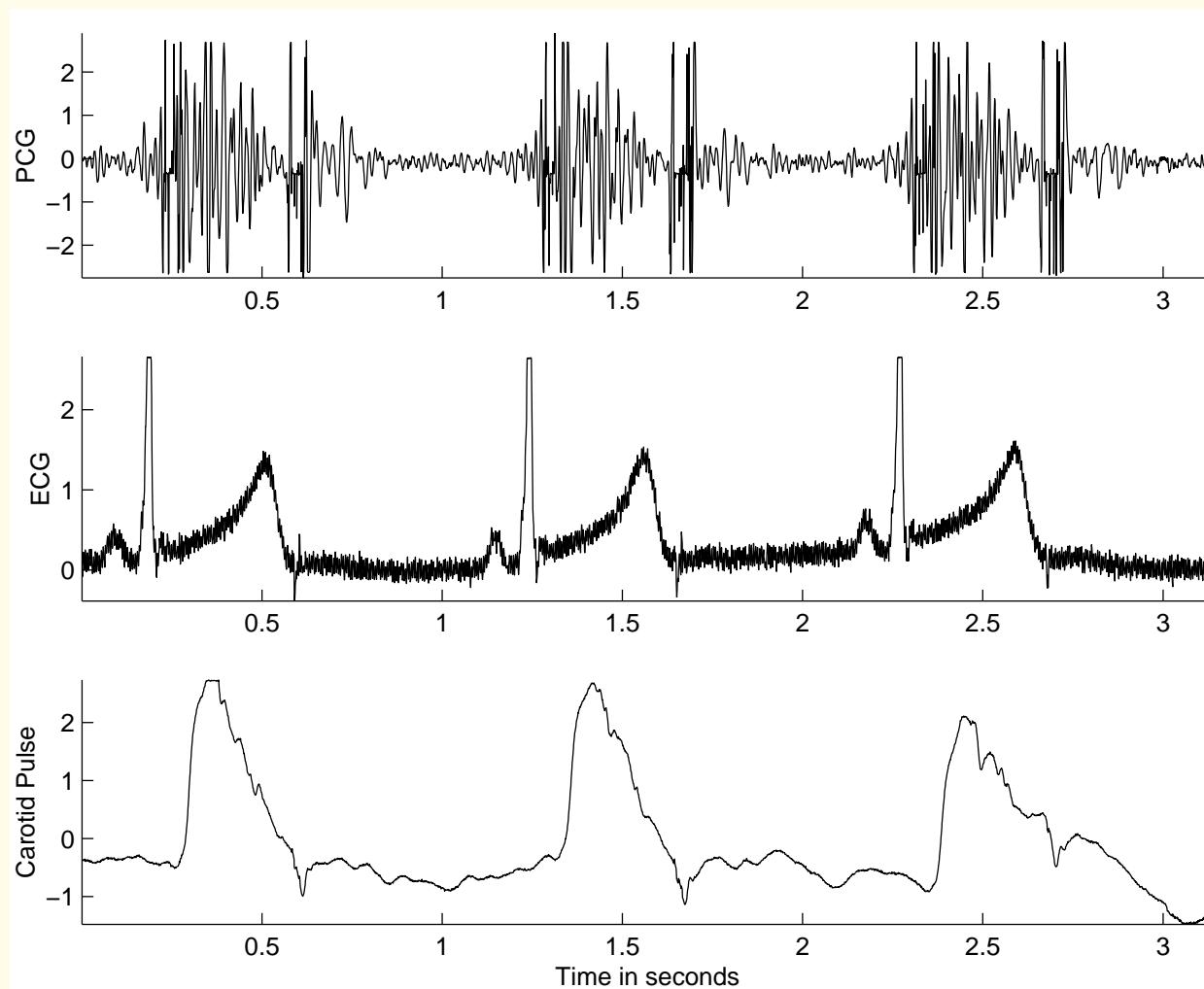


Figure 1.26: Three-channel simultaneous record of the PCG, ECG, and carotid pulse signals of a patient (female, 11 years) with aortic stenosis. Note the presence of the typical diamond-shaped systolic murmur and the split nature of S2 in the PCG.



1.2.9 *The carotid pulse (CP)*

Carotid pulse:

pressure signal recorded over the carotid artery

as it passes near the surface of the body at the neck.

Pulse signal indicating the variations in arterial

blood pressure and volume with each heart beat.

Resembles the pressure signal at the root of the aorta.

HP21281A pulse transducer: bandwidth of $0 - 100 \text{ Hz}$.



Carotid pulse: rises abruptly with the ejection of blood from the left ventricle to the aorta.

Peak: percussion wave (P).

Plateau or secondary wave: tidal wave (T):

caused by reflected pulse returning from the upper body.

Closure of the aortic valve: dicrotic notch.

Dicrotic wave (DW): reflected pulse from the lower body.



1.2.10 *Signals from catheter-tip sensors*

Sensors placed on catheter tips inserted into the

cardiac chambers:

left ventricular pressure, right atrial pressure,

aortic (AO) pressure, and intracardiac sounds

Procedures invasive and associated with certain risks.

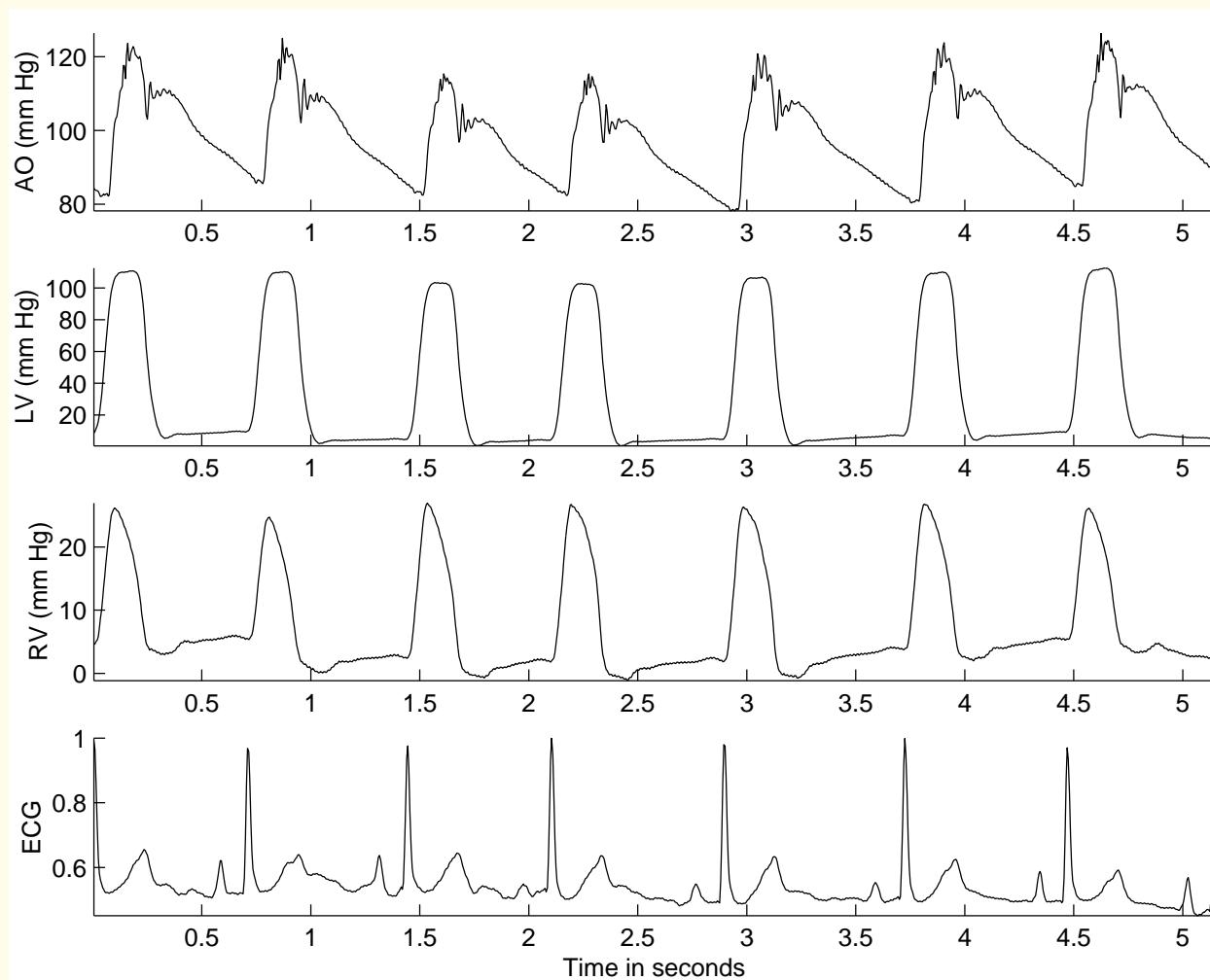


Figure 1.27: Normal ECG and intracardiac pressure signals from a dog. AO represents aortic pressure near the aortic valve. Data courtesy of R. Sas and J. Tyberg, Department of Physiology and Biophysics, University of Calgary.

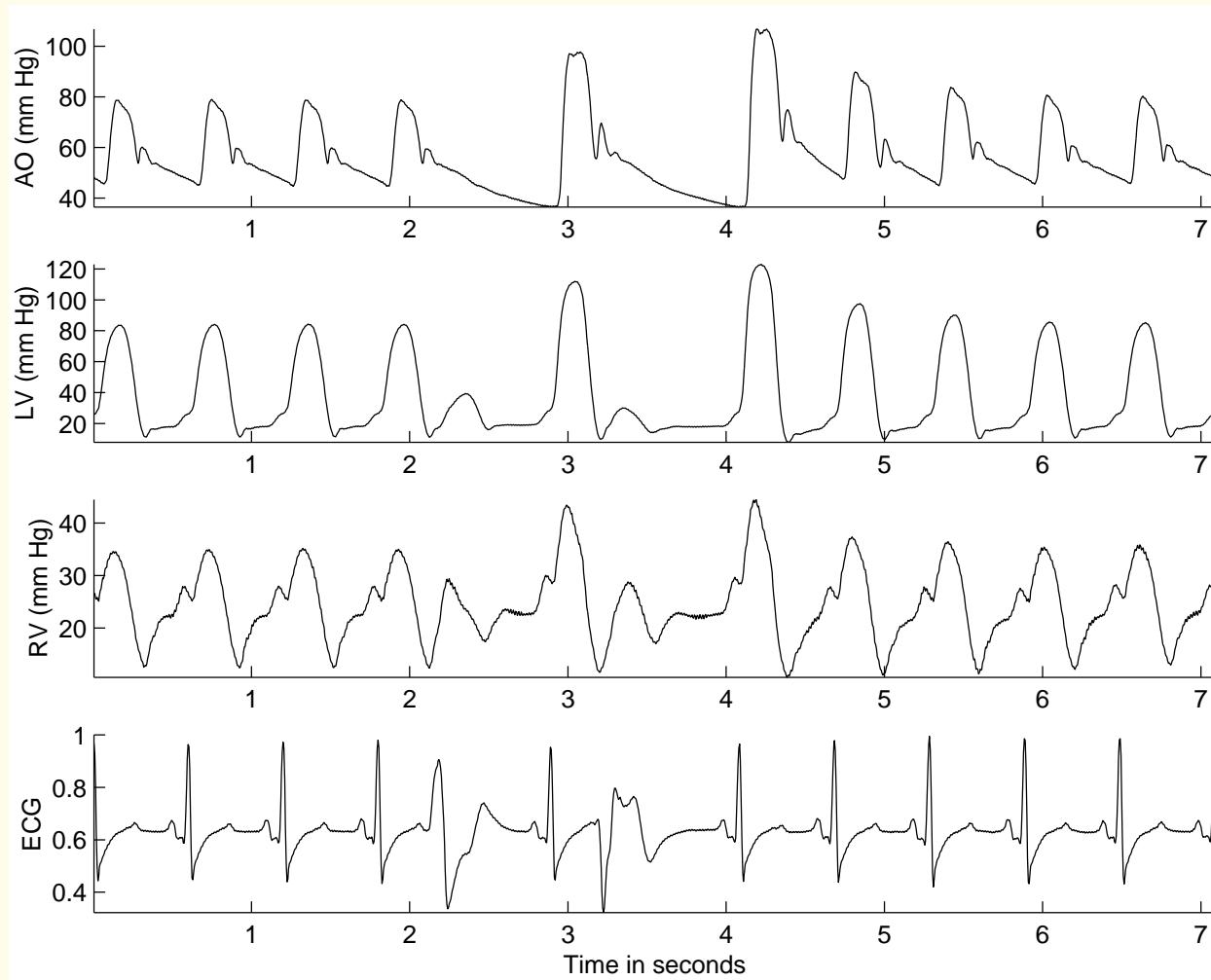


Figure 1.28: ECG and intracardiac pressure signals from a dog with PVCs. Data courtesy of R. Sas and J. Tyberg, Department of Physiology and Biophysics, University of Calgary.



1.2.11 *The speech signal*

Speech produced by transmitting puffs of air from the lungs

through the vocal tract as well as the nasal tract.

Vocal tract: starts at the vocal cords or glottis in the throat

and ends at the lips and the nostrils.

Shape of vocal tract varied to produce different types of

sound units or *phonemes* which form speech.



The vocal tract acts as a filter that modulates the spectral characteristics of the input puffs of air.

The system is dynamic: the filter and the speech signal have time-varying characteristics — they are nonstationary.



Speech sounds classified mainly as

voiced, unvoiced, and plosive sounds.

Voiced sounds involve participation of the glottis:

air forced through vocal cords held at a certain tension.

The result is a series of quasi-periodic pulses of air

which is passed through and filtered by the vocal tract.



The input to the vocal tract may be treated as an

impulse train that is almost periodic.

Upon convolution with the impulse response of the

vocal tract, held at a certain configuration for the

duration of the voiced sound desired,

a quasi-periodic signal is produced

with a characteristic waveshape that is repeated.



Vowels are voiced sounds.

Features of interest in voiced signals are the pitch and

resonance or formant frequencies of the vocal-tract system.

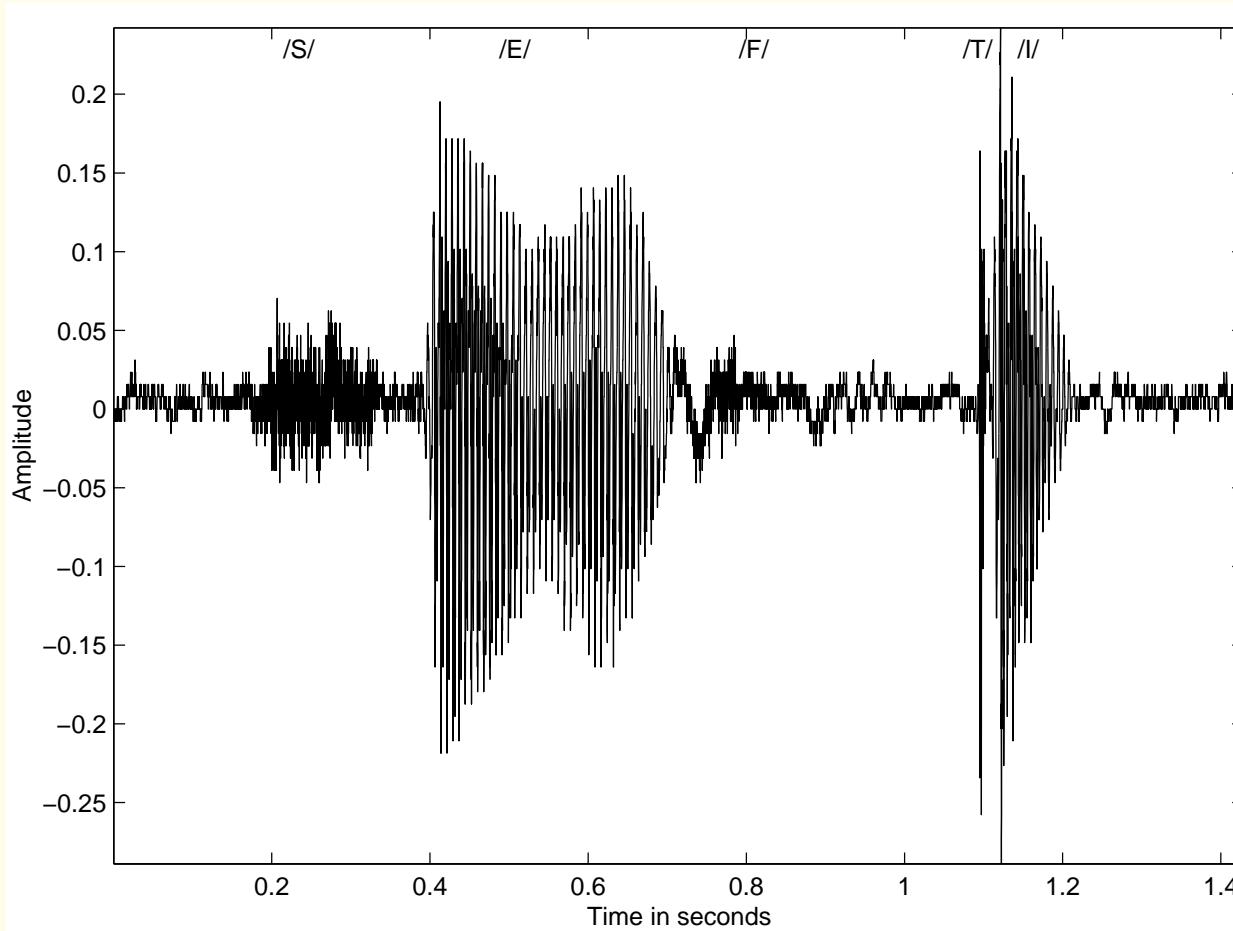


Figure 1.29: Speech signal of the word “safety” uttered by a male speaker. Approximate time intervals of the various phonemes in the word are /S/: 0.2 – 0.35 s; /E/: 0.4 – 0.7 s; /F/: 0.75 – 0.95 s; /T/: transient at 1.1 s; /I/: 1.1 – 1.2 s. Background noise is also seen in the signal before the beginning and after the termination of the speech, as well as during the stop interval before the plosive /T/.

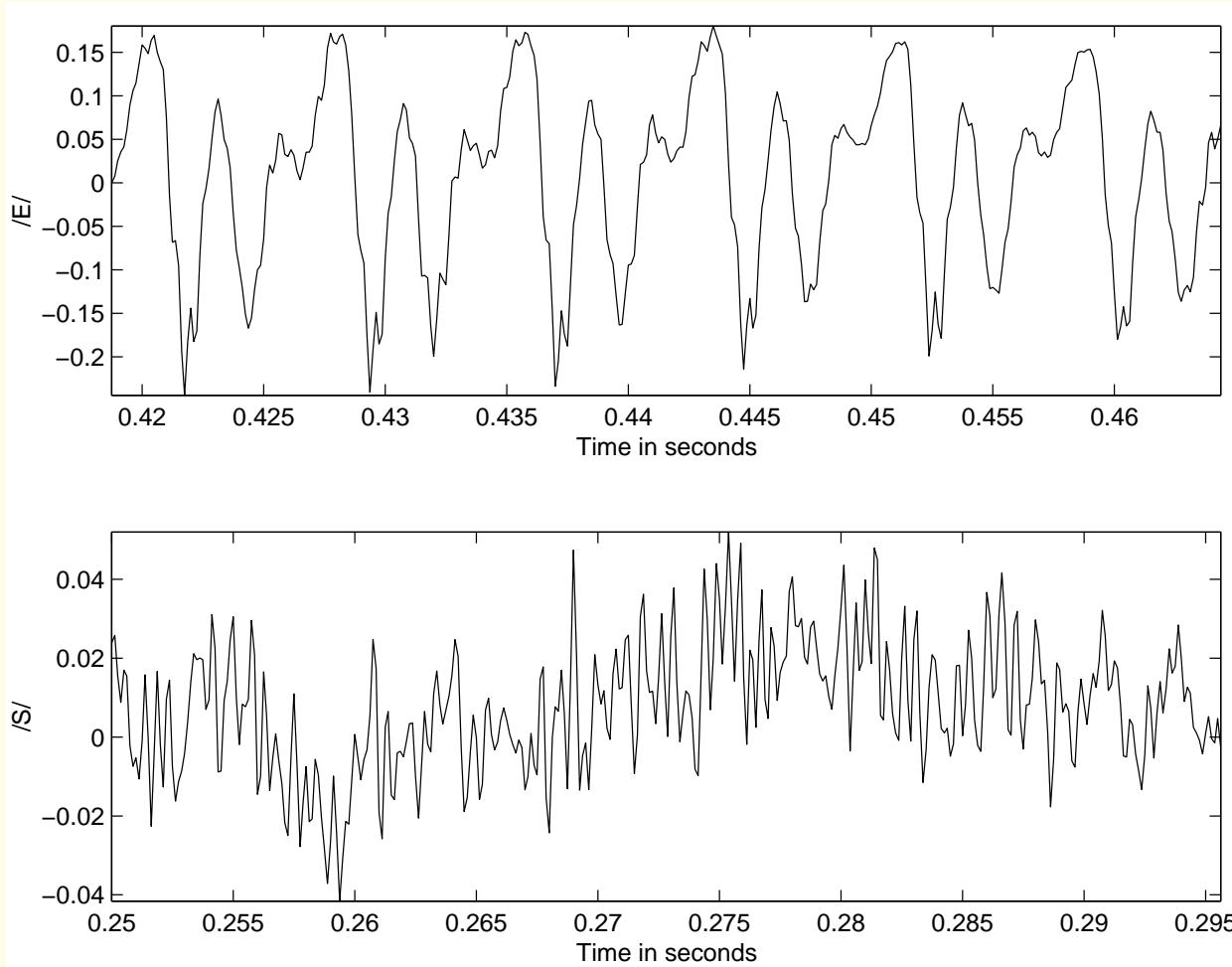


Figure 1.30: Segments of the signal in Figure 1.29 on an expanded scale to illustrate the quasi-periodic nature of the voiced sound $/E/$ in the upper trace, and the almost-random nature of the fricative $/S/$ in the lower trace.



Unvoiced sound (fricative) produced by

forcing a steady stream of air through

a narrow opening or constriction formed

at a specific position along the vocal tract:

Turbulent signal that appears like random noise.

The input to the vocal tract is a broadband random signal:

filtered by the vocal tract to yield the desired sound.



Fricatives are unvoiced sounds:

do not involve any activity (vibration) of the vocal cords.

Fricatives: /S/, /SH/, /Z/, /F/.

Plosives (stops): complete closure of the vocal tract,

followed by an abrupt release of built-up pressure.

Plosives: /P/, /T/, /K/, /D/.



1.2.12 *The vibromyogram (VMG)*

VMG: mechanical manifestation of contraction of

skeletal muscle;

vibration signal that accompanies the EMG.

Muscle sounds or vibrations related to the change in

dimensions (contraction) of the constituent muscle fibers.

Recorded using contact microphones or accelerometers.



VMG frequency and intensity vary in proportion to contraction level.

VMG and EMG useful in studies on neuromuscular control, muscle contraction, athletic training, and biofeedback.



1.2.13 *The vibroarthrogram (VAG)*

The knee joint: the largest articulation in the human body

formed between the femur, the patella, and the tibia.

0° extension to 135° flexion;

20° to 30° rotation of the flexed leg on the femoral condyles.



The joint has four important features:

- (1) a joint cavity,
- (2) articular cartilage,
- (3) a synovial membrane, and
- (4) a fibrous capsule.



Knee joint is a synovial joint: contains a

lubricating substance called the synovial fluid.

The patella (knee cap), a sesamoid bone, protects the joint:

precisely aligned to slide in the groove (trochlea)

of the femur during leg movement.



Knee joint is made up of three compartments:

- (1) the patello-femoral,
- (2) the lateral tibio-femoral, and
- (3) the medial tibio-femoral compartments.



Patello-femoral compartment: synovial gliding joint;

tibio-femoral: synovial hinge joint.

The anterior and posterior cruciate ligaments

as well as the lateral and medial ligaments

bind the femur and tibia together,

give support to the knee joint, and

limit movement of the joint.

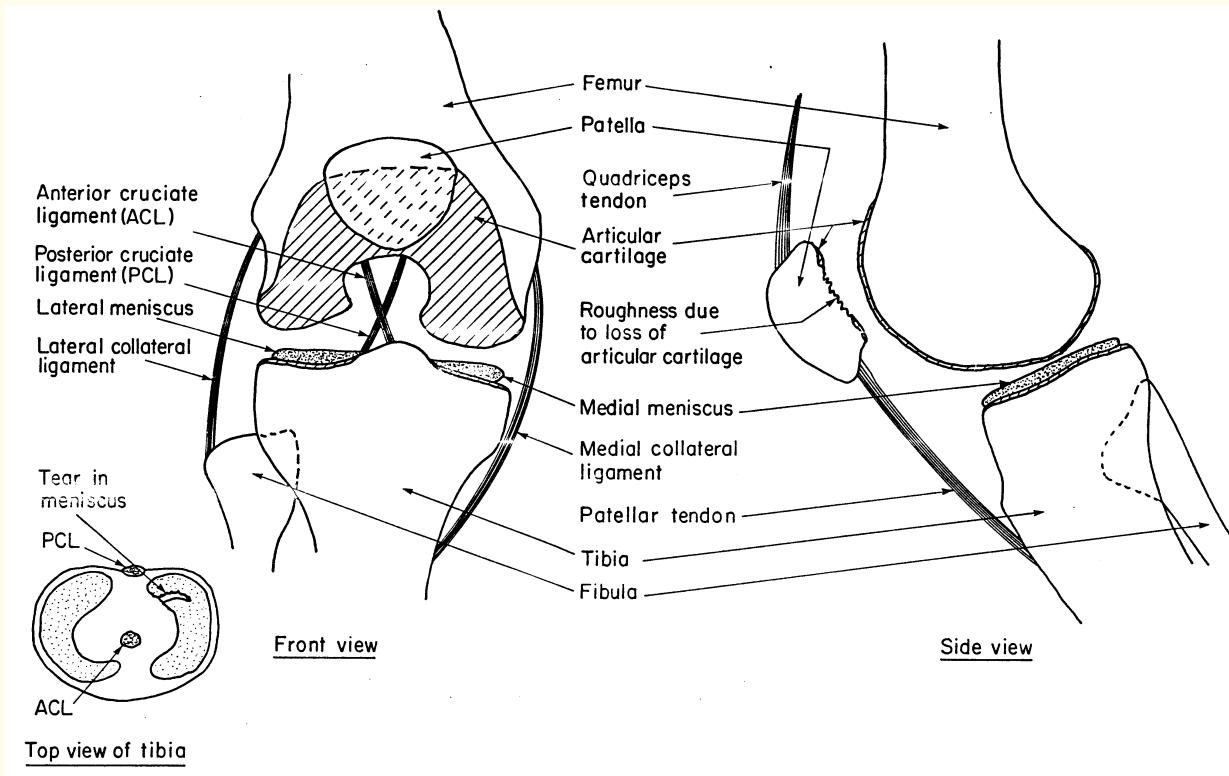


Figure 1.31: Front and side views of the knee joint (the two views are not mutually orthogonal). The inset shows the top view of the tibia with the menisci.



Two types of cartilage in the knee joint:

the *articular cartilage* covers the ends of bones;

the wedge-shaped fibrocartilaginous structure called the

menisci, located between the femur and the tibia.

Cartilage is vital to joint function:

protects the underlying bone during movement.



Loss of cartilage function leads to pain, decreased mobility, deformity and instability.

Chondromalacia patella: articular cartilage softens,

fibrillates, and sheds off the undersurface of the patella.

Meniscal fibrocartilage can soften: degenerative tears.



Knee-joint sounds:

Noise associated with degeneration of knee-joint surfaces.

VAG: vibration during movement or articulation of the joint.

Normal joint surfaces: smooth and

produce little or no sound.

Joints affected by osteoarthritis and degenerative diseases:

cartilage loss leads to grinding sounds.



Detection of knee-joint problems

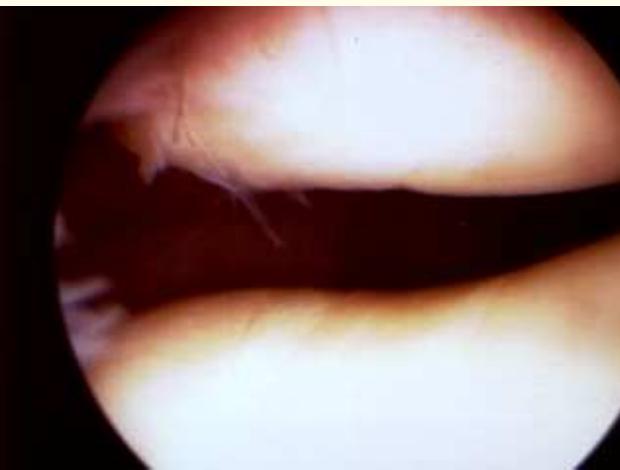
via the analysis of VAG signals

could help avoid unnecessary exploratory surgery,

and aid in more accurate diagnosis.



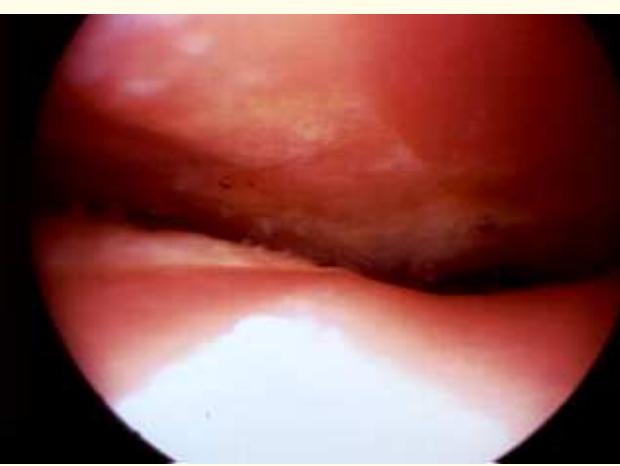
(a)



(b)



(c)



(d)

Arthroscopic views of the patello-femoral joint. (a) Normal cartilage surfaces. (b) Chondromalacia Grade II at the patella. (c) Chondromalacia Grade III at the patella. (d) Chondromalacia Grade IV at the patella and the femur; the bones are exposed. The under-surface of patella is at the top and the femoral condyle is at the bottom.

Figure courtesy: G.D. Bell, Sport Medicine Centre, University of Calgary.



1.2.14 *Oto-acoustic emission (OAE) signals*

OAE: acoustic energy emitted by the cochlea

either spontaneously or in response to an acoustic stimulus.

May assist in screening of hearing function and

diagnosis of hearing impairment.



1.3 Objectives of Biomedical Signal Analysis

- *Information gathering* — measurement of phenomena to interpret a system.
- *Diagnosis* — detection of malfunction, pathology, or abnormality.
- *Monitoring* — obtaining continuous or periodic information about a system.



- *Therapy and control* — modification of the behavior of a system based upon the outcome of the activities listed above to ensure a specific result.
- *Evaluation* — objective analysis to determine the ability to meet functional requirements, obtain proof of performance, perform quality control, or quantify the effect of treatment.

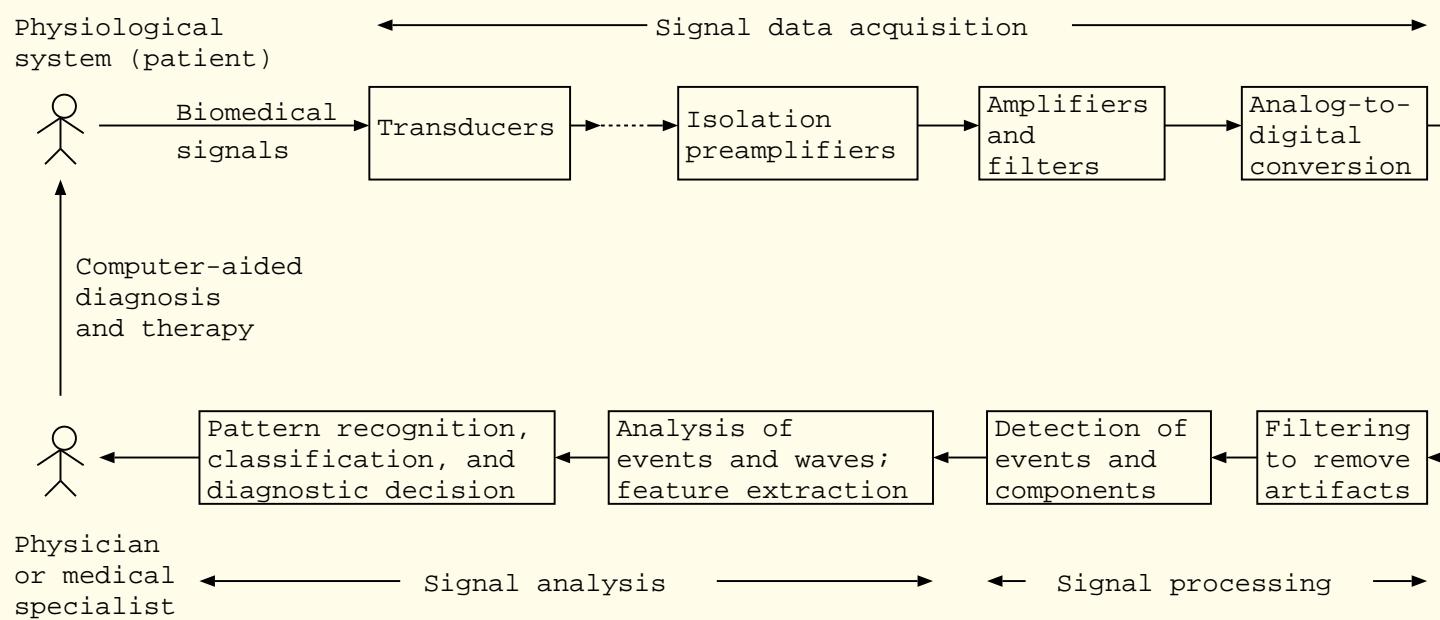


Figure 1.32: Computer-aided diagnosis and therapy based upon biomedical signal analysis.



Signal acquisition procedures may be categorized as

invasive or noninvasive, active or passive.

Risk – benefit analysis.

Be prepared to manage adverse reactions.



Ethical approval by specialized committees required

for experimental procedures involving

human or animal subjects:

minimize the risk and discomfort to the subject,

maximize benefits to the subjects and the investigator.



The human – instrument system:

- *The subject or patient.*
- *Stimulus or procedure of activity.*
- *Transducers.*
- *Signal-conditioning equipment.*
- *Display equipment.*
- *Recording, data processing, and transmission equipment.*
- *Control devices.*



1.4 Difficulties Encountered in Biomedical Signal Acquisition and Analysis

Accessibility of the variables to measurement.

Variability of the signal source.

Inter-relationships and interactions among

physiological systems.

Effects of the instrumentation or procedure on the system.



Physiological artifacts and interference.

Energy limitations.

Patient safety.



1.5 Computer-aided Diagnosis: CAD

- Humans highly skilled and fast in the analysis of visual patterns but slow in arithmetic operations.
- Humans could be affected by fatigue, boredom, and environmental factors: susceptible to committing errors.
- Computers are inanimate but accurate and consistent machines: can be designed to perform specific and repetitive tasks.
- Analysis by humans is usually subjective and qualitative.



- Analysis by humans is subject to inter-observer as well as intra-observer variations over time.
- *On-line, real-time* analysis of biomedical signals is feasible with computers.
- *Quantitative analysis* becomes possible by the application of computers to biomedical signals.
- The logic of clinical diagnosis via signal analysis *objectively* encoded and *consistently* applied in routine or repetitive tasks using computers.



- End-goal of biomedical signal analysis: computer-*aided* diagnosis and not automated diagnosis.
- Results of signal analysis need to be integrated with clinical signs, symptoms, and information by a physician.
- The *intuition* of the specialist plays an important role in arriving at the final diagnosis.
- Quantitative and objective analysis facilitated by the application of computers to biomedical signal analysis: more accurate diagnostic decision by the physician.



On the importance of quantitative analysis:

“When you can measure what you are speaking about,
and express it in numbers,
you know something about it;
but when you cannot measure it,
when you cannot express it in numbers,
your knowledge is of a meager and
unsatisfactory kind:
it may be the beginning of knowledge,
but you have scarcely, in your thoughts,
advanced to the stage of *science*. ”

— *Lord Kelvin (William Thomson, 1824 – 1907)*



On assumptions made in quantitative analysis:

“Things do not in general run around
with their measure stamped on them
like the capacity of a freight car;
it requires a certain amount of investigation
to discover what their measures are ...

What most experimenters take for granted
before they begin their experiments
is infinitely more interesting
than any results to which their experiments lead.”

— *Norbert Wiener (1894 – 1964)*







2

Analysis of Concurrent, Coupled, and Correlated Processes

The human body is a complex integration of
a number of biological systems with several ongoing
physiological, functional, and pathological processes.



Most biological processes within a body are
not independent of one another;
they are mutually correlated and bound together
by physical or physiological control and
communication phenomena.



2.1 Illustration of the Problem with Case-studies

2.1.1 *The electrocardiogram and the phonocardiogram*

Problem: *Identify the beginning of S1 in a PCG signal and extract the heart sound signal over one cardiac cycle.*

Solution: Use the QRS wave in the ECG as

reference or trigger.



2.1.2 *The phonocardiogram and the carotid pulse*

Problem: *Identify the beginning of S2 in a PCG signal.*

Solution: Use the dicrotic notch in the carotid pulse.



2.1.3 *The ECG and the atrial electrogram*

Problem: *Obtain an indicator of atrial contraction to measure the PR interval.*

Solution: Jenkins et al. developed a pill electrode

to obtain a strong and clear signal of atrial activity.

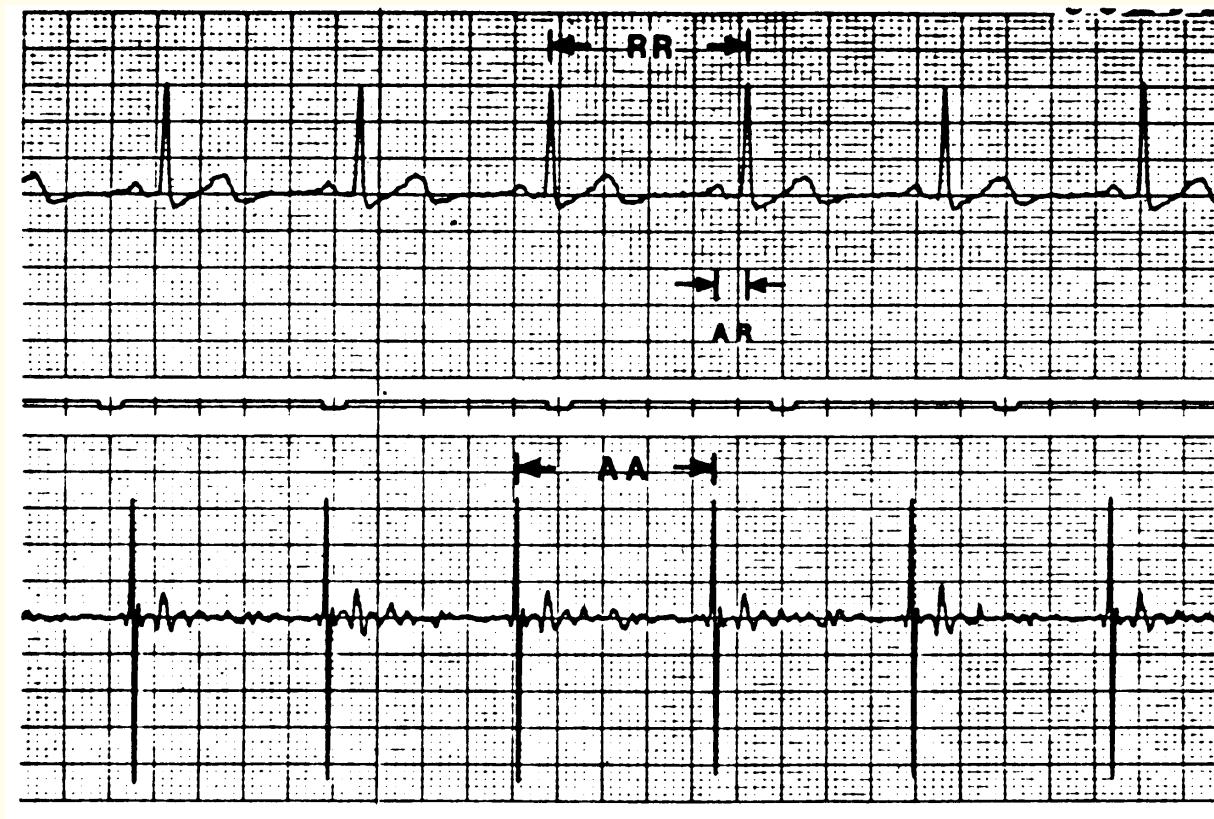


Figure 2.1: Pill-electrode recording of the atrial electrogram (lower tracing) and the external ECG (upper tracing) of a normal subject. The pulse train between the two signals indicates intervals of 1 s. Reproduced with permission from J.M. Jenkins, D. Wu, and R. Arzbaecher, Computer diagnosis of abnormal cardiac rhythms employing a new P-wave detector for interval measurement, *Computers and Biomedical Research*, 11:17–33, 1978. ©Academic Press.



Jenkins et al. developed a four-digit code for each beat.

The first digit was coded as

0: abnormal waveshape, or

1: normal waveshape,

determined by a correlation coefficient between

the beat being processed and a normal template.



The remaining three digits encoded the nature of the

RR, AR, and AA intervals.

0: short,

1: normal, or

2: long.

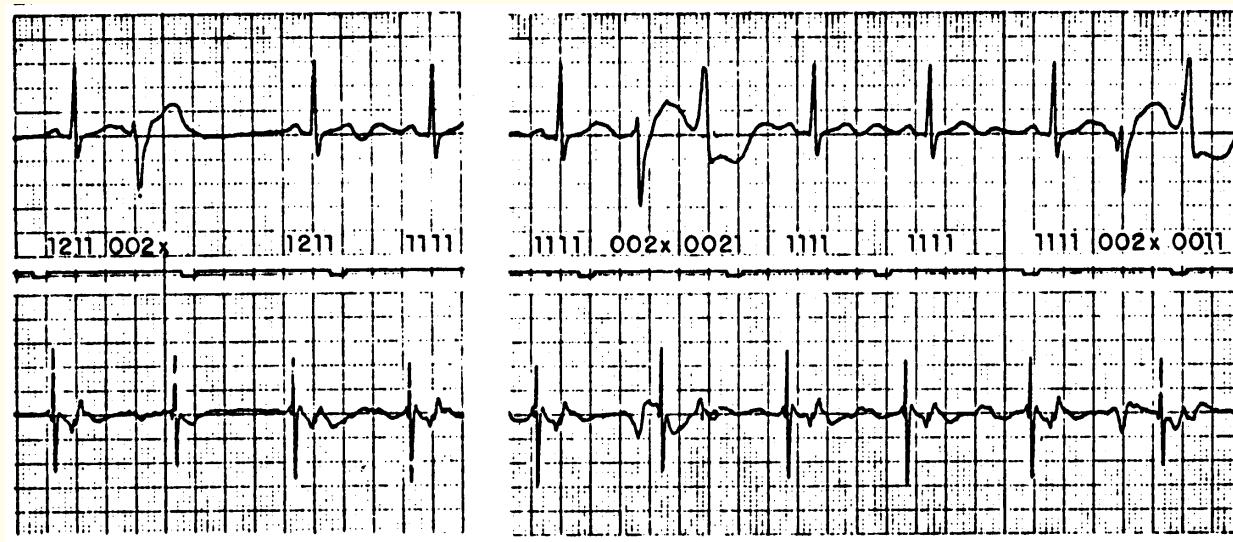


Figure 2.2: Atrial electrogram (lower tracing) and the external ECG (upper tracing) of a subject with ectopic beats. The pulse train between the two signals indicates intervals of 1 s. Reproduced with permission from J.M. Jenkins, D. Wu, and R. Arzbaecher, Computer diagnosis of abnormal cardiac rhythms employing a new P-wave detector for interval measurement, *Computers and Biomedical Research*, 11:17–33, 1978. ©Academic Press.



2.1.4 *Cardio-respiratory interaction*

The heart rate is affected by normal breathing due to the coupling and interaction between the cardiac and respiratory systems.

Baroreceptors exist in the aorta and carotid artery.

Vagus nerve activity impeded when we inhale:

HR increases.



Breathing also affects the transmission of the heart sounds from the cardiac chambers to the chest surface.



ECG signal with subject breathing.



ECG signal with subject holding breath.



2.1.5 *The electromyogram and the vibromyogram*

Problem: *Obtain a mechanical signal that is a direct indicator of muscle-fiber or motor unit activity to study muscle contraction and force development.*

Solution: Use the VMG —

direct manifestation of the contraction of muscle fibers.

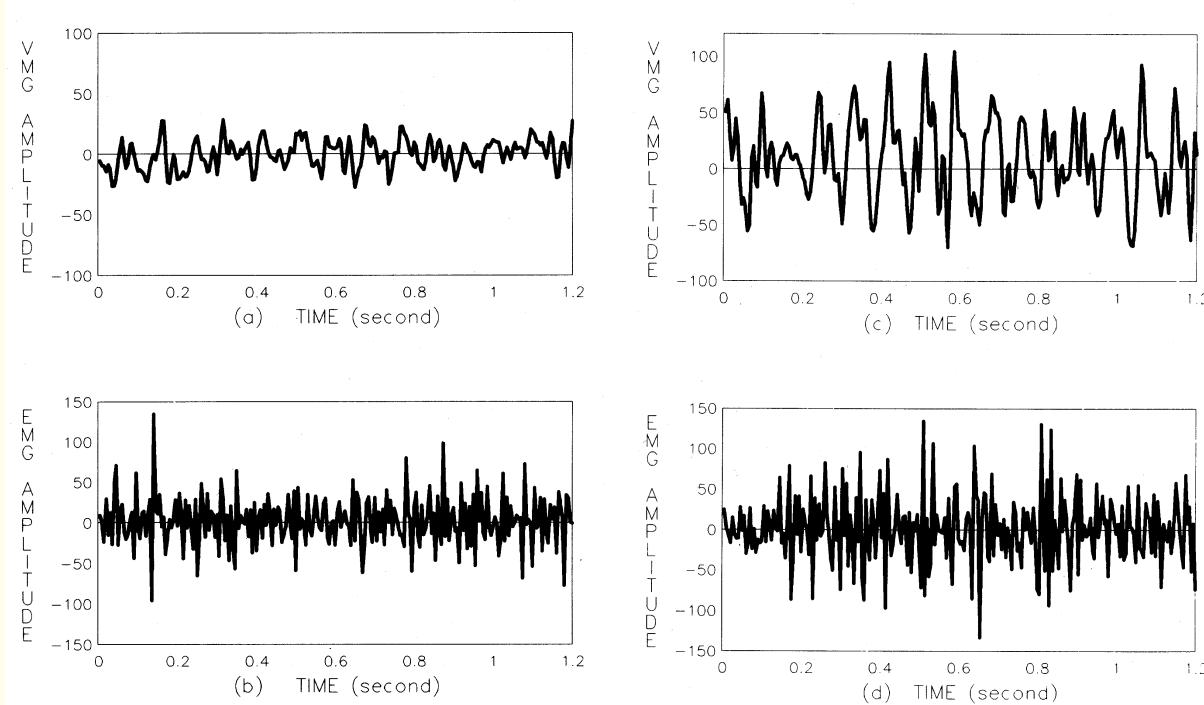


Figure 2.3: Simultaneous EMG – VMG records at two levels of contraction of the rectus femoris muscle. (a) VMG at 40% of the maximal voluntary contraction (MVC) level. (b) EMG at 40% MVC. (c) VMG at 60% MVC. (d) EMG at 60% MVC. Reproduced with permission from Y.T. Zhang, C.B. Frank, R.M. Rangayyan, and G.D. Bell, Relationships of the vibromyogram to the surface electromyogram of the human rectus femoris muscle during voluntary isometric contraction, *Journal of Rehabilitation Research and Development*, 33(4): 395–403, 1996. ©Department of Veterans Affairs.



2.1.6 *The knee-joint and muscle vibration signals*

VMG associated with the rectus femoris muscle

that must necessarily be active during extension of the leg

appears as an interference and corrupts the VAG signal.

Problem: *Suggest an approach to remove the*

muscle-contraction interference from the

knee-joint vibration signal.



Solution:

The rectus femoris muscle and the knee-joint systems are

coupled dynamic systems with vibration characteristics

that vary with activity level and time: *nonstationary*.

Record the VMG signal at the rectus femoris at the

same time as the VAG signal is acquired from the patella.

Apply adaptive filtering and noise cancellation techniques.



2.2 Application: Segmentation of the PCG into Systolic and Diastolic Parts

Problem: *Show how the ECG and carotid pulse signals may be used to break a PCG signal into its systolic and diastolic parts.*

Solution: QRS in ECG \longleftrightarrow S1 in PCG.

Dicrotic notch in carotid pulse \longleftrightarrow S2 in PCG.

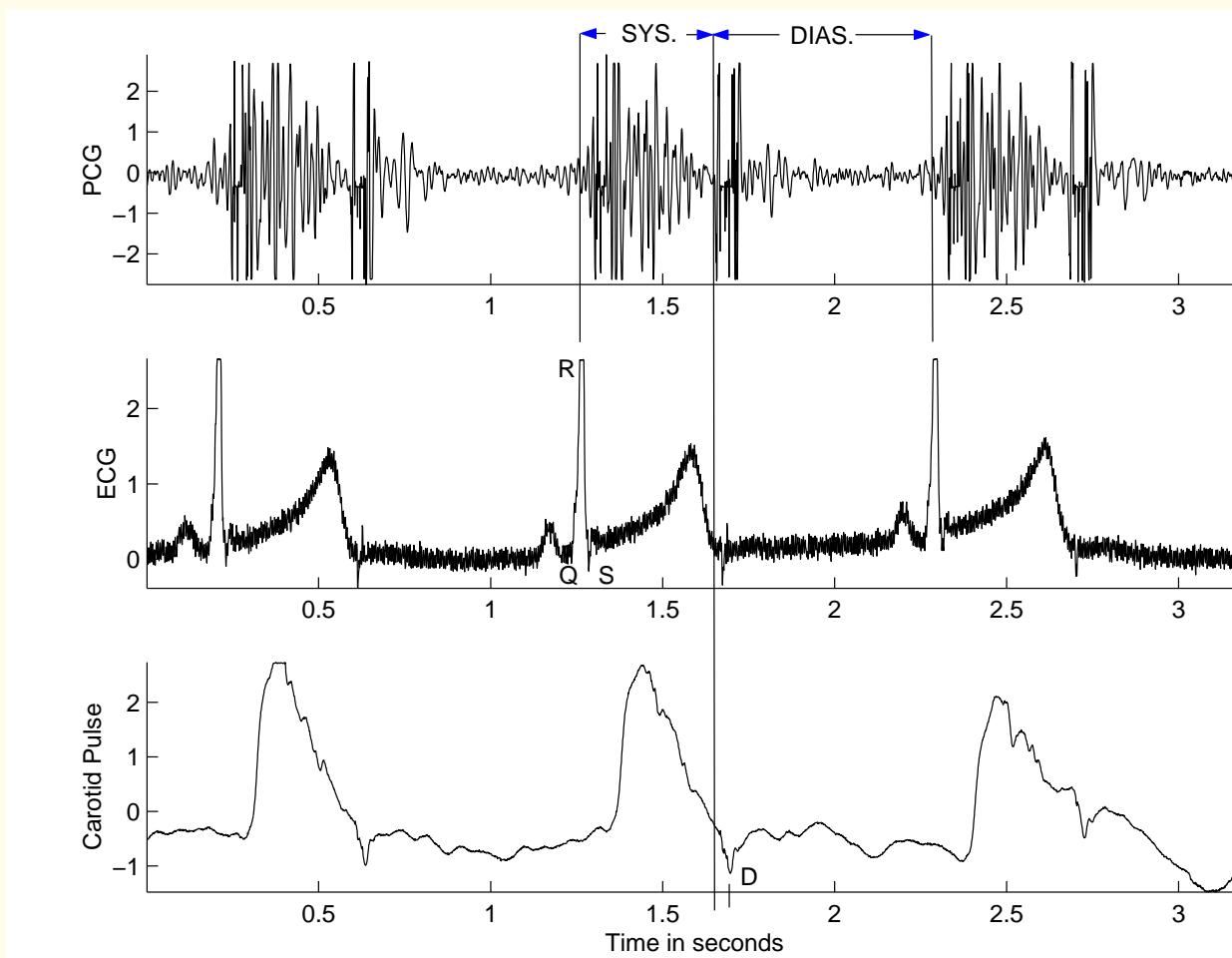


Figure 2.4: Demarcation of the systolic (SYS.) and diastolic (DIAS.) parts of the PCG signal in Figure 1.26 by using the ECG and carotid pulse as reference signals. The QRS complex and the dicrotic notch D are marked on the ECG and carotid pulse signals, respectively.







3

Filtering for Removal of Artifacts

Biomedical signals are weak signals

in an environment that is teeming

with many other signals of various origins.



Any signal other than that of interest is

interference, artifact, or *noise*.

Sources of noise: physiological, instrumentation,

or the environment of the experiment.



3.1 Problem Statement

Noise is omnipresent!

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



3.1.1 *Random noise, structured noise, and physiological interference*

Deterministic signal:

value at a given instant of time may be computed

using a closed-form mathematical function of time,

or predicted from a few past values of the signal.

A signal that does not meet this condition:

nondeterministic or random signal.



Test for randomness:

Based upon the number of peaks or troughs in the signal:

collectively known as *turning points*.

Test for a turning point: sign of the first-order

difference or derivative at the current sample

not equal to that at the preceding sample.

Given a signal of N samples, signal is random

if the number of turning points $> \frac{2}{3}(N - 2)$.



Nonstationary signal:

conduct the test using a running window of N samples.

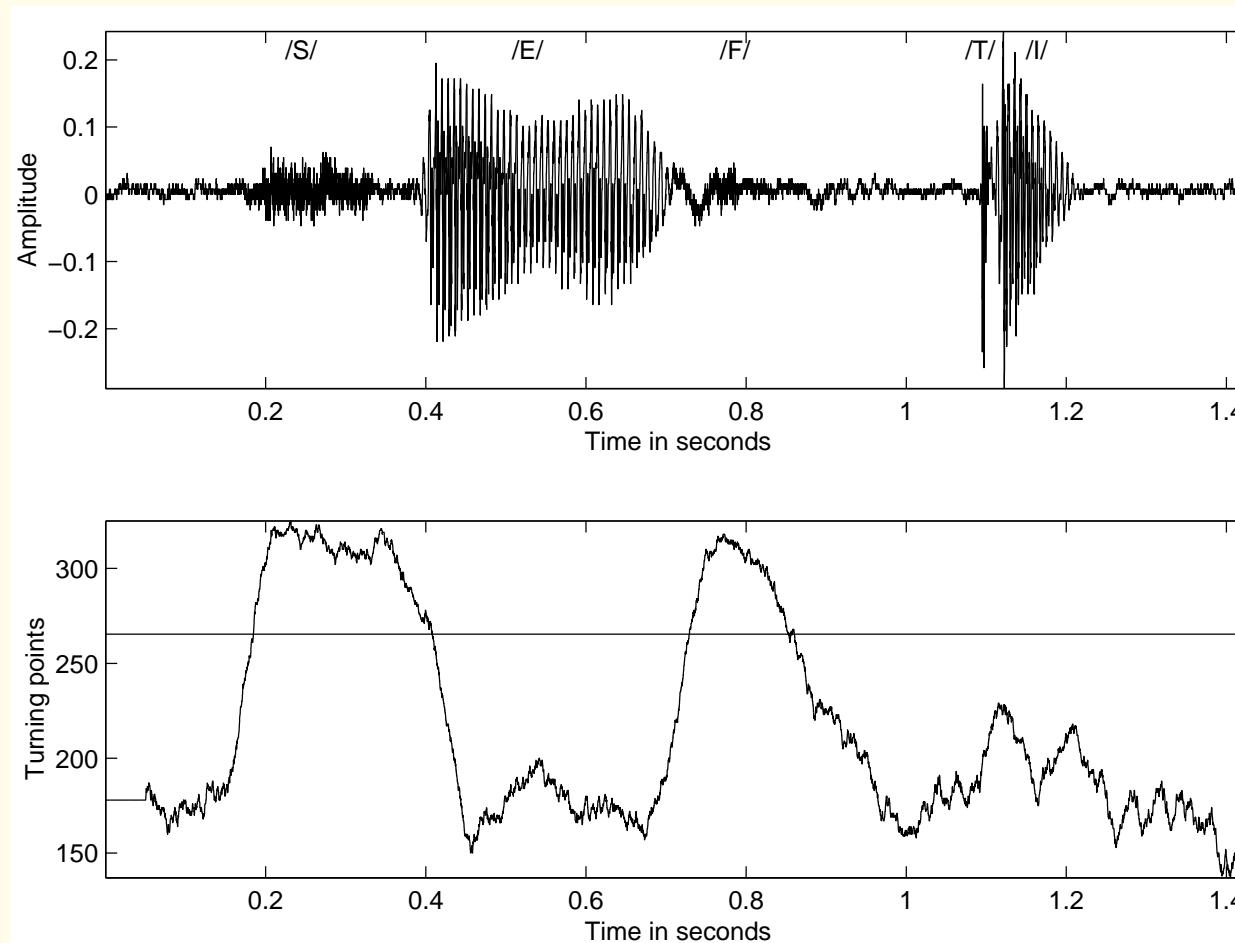


Figure 3.1: Top: Speech signal of the word “safety” uttered by a male speaker. Bottom: Count of turning points in a moving window of 50 ms (400 samples with $f_s = 8\text{ kHz}$). The threshold for randomness for $N = 400$ is 265.



Random noise:

Interference that arises from a random process

such as thermal noise in electronic devices.

A random process is characterized by the

probability density function (PDF)

representing the probabilities of occurrence

of all possible values of a random variable.



Random process η characterized by PDF $p_\eta(\eta)$.

Mean μ_η = first-order moment of the PDF:

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

where $E[]$ represents the *statistical expectation operator*.

Common to assume mean of a random noise process = 0.



Mean-squared (MS) value = second-order moment:

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



Variance σ_{η}^2 = second central moment:

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

Square root of variance = standard deviation (SD) = σ_{η} .

$$\sigma_{\eta}^2 = E[\eta^2] - \mu_{\eta}^2.$$

If $\mu_{\eta} = 0$, then $\sigma_{\eta}^2 = E[\eta^2]$:

variance = MS.



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 $\eta(t)$.

Then, the statistical measures have physical meanings:

mean = DC component;

MS = average power;

square root of MS = root mean-squared or RMS value

= average noise magnitude.



Signal-to-noise ratio:

$$SNR = \frac{\text{peak-to-peak amplitude range of signal}}{\text{RMS value of noise}}$$

or

$$SNR = \frac{\text{average power of signal}}{\text{average power of noise}}.$$



Any biomedical signal of interest $x(t)$ may also,

for the sake of generality, be considered to be

a realization of a random process x .



When a signal $x(t)$ is observed 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:

$$y(t) = x(t) + \eta(t). \quad (3.4)$$

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

may be assumed to be *statistically independent processes*.

Two random processes x and η are said to be

statistically independent if their joint PDF

$$p_{x,\eta}(x, \eta) = p_x(x) p_\eta(\eta).$$



$$E[y] = \mu_y = \mu_x + \mu_\eta. \quad (3.5)$$

If $\mu_\eta = 0$, then $\mu_y = \mu_x$.

$$E[(y - \mu_y)^2] = \sigma_y^2 = \sigma_x^2 + \sigma_\eta^2. \quad (3.6)$$



Ensemble averages:

When the PDFs of the random processes of concern are not known, approximate the statistical expectation operation by averages of a collection or *ensemble* of sample observations of the random process: *ensemble averages*.



Suppose we have M observations of the random process

x as functions of time: $x_1(t), x_2(t), \dots, x_M(t)$.

Estimate of the mean of the process at a particular

instant of time t_1 :

$$\mu_x(t_1) = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{k=1}^M x_k(t_1). \quad (3.7)$$

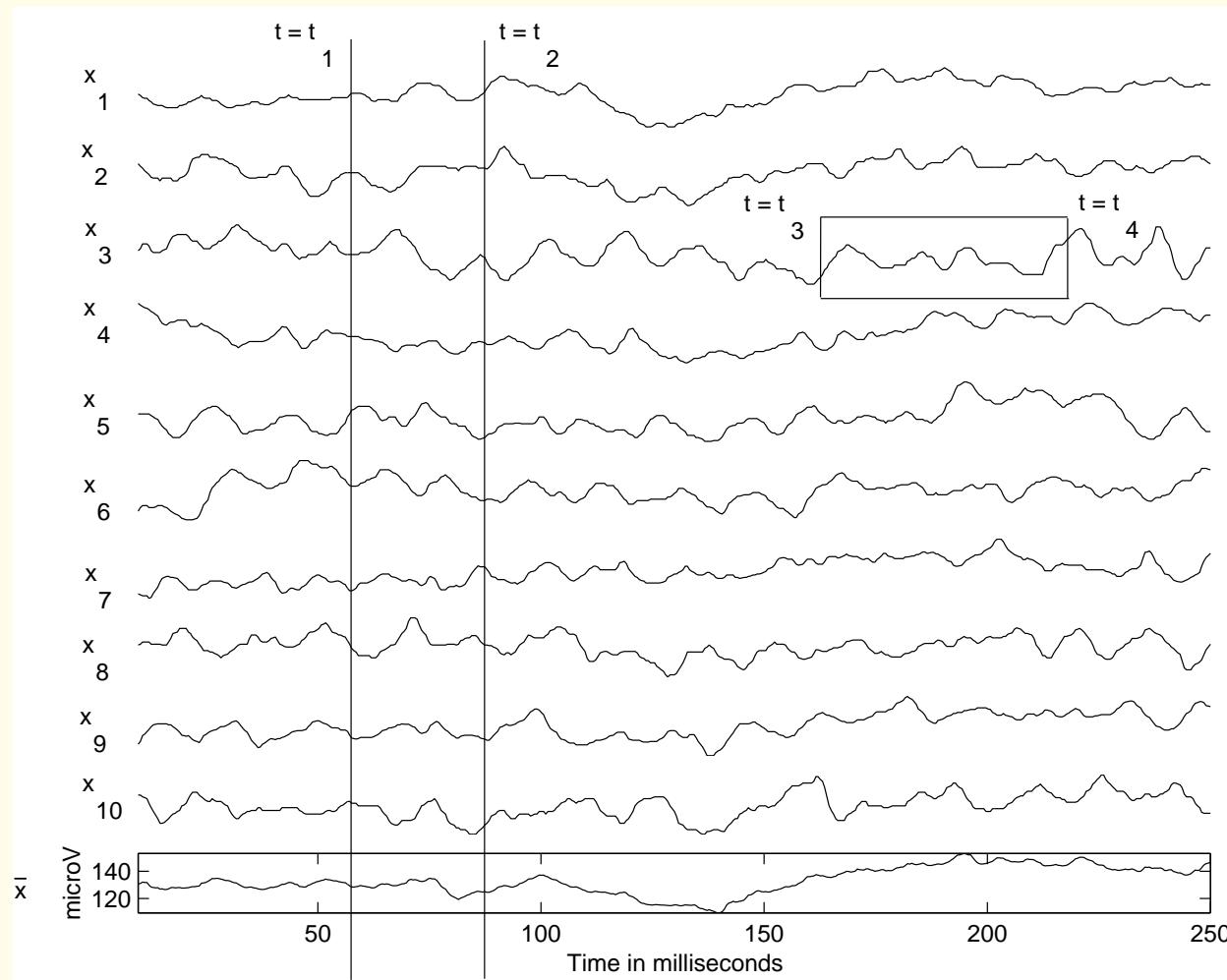


Figure 3.2: Ten sample acquisitions (x_1 to x_{10}) of individual flash visual ERPs from the occipital midline (oz) position of a normal adult male (the author of this book!). The ear lobes were used to form the reference lead (a1a2), and the left forehead was used as the reference (see Figure 1.20). 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 3.12. Data courtesy of L. Alfaro and H. Darwish, Alberta Children's Hospital, Calgary.



Autocorrelation function (ACF) $\phi_{xx}(t_1, t_1 + \tau)$:

$$\phi_{xx}(t_1, t_1 + \tau) = E[x(t_1)x(t_1 + \tau)]$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(t_1) x(t_1 + \tau) p_{x_1, x_2}(x_1, x_2) dx_1 dx_2, \quad (3.8)$$

x_1 and x_2 : random variables corresponding to

$x(t_1)$ and $x(t_1 + \tau)$;

$p_{x_1, x_2}(x_1, x_2)$: joint PDF of the two processes.



The ACF may be estimated as

$$\phi_{xx}(t_1, t_1 + \tau) = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{k=1}^M x_k(t_1) x_k(t_1 + \tau), \quad (3.9)$$

where τ is the delay parameter.



If the signals are complex: take the conjugate of one.

The two vertical lines at $t = t_1$ and $t = t_2 = t_1 + \tau$

in Figure 3.2: ensemble averaging process

to compute $\phi_{xx}(t_1, t_2)$.

ACF indicates how the values of a signal

at a particular instant of time

are statistically related to values at another instant of time.



Random processes as functions of time

(*stochastic processes*):

compute ensemble averages at every point of time.

Averaged function of time $\bar{x}(t)$:

$$\bar{x}(t) = \mu_x(t) = \frac{1}{M} \sum_{k=1}^M x_k(t) \quad (3.10)$$

for all time t .

Signal $\bar{x}(t)$ may be used to represent

the random process x as a *prototype*.



Time averages:

Sample observation of a random process $x_k(t)$:

compute *time averages* or *temporal statistics*

by integrating over time:

$$\mu_x(k) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x_k(t) dt. \quad (3.11)$$

Integral replaced by summation for sampled signals.



Time-averaged ACF $\phi_{xx}(\tau, k)$:

$$\phi_{xx}(\tau, k) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x_k(t) x_k(t + \tau) dt. \quad (3.12)$$

The horizontal box superimposed on the third trace in

Figure 3.2: 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.



Random noise may thus be characterized in terms of

ensemble and/or *temporal* statistics.

Mean: not important; assumed to be zero, or subtracted out.

ACF: plays an important role in the

characterization of random processes.

Fourier transform of ACF = power spectral density (PSD).



Covariance and cross-correlation:

When two random processes x and y need to be compared:

$$\begin{aligned} C_{xy} &= E[(x - \mu_x)(y - \mu_y)] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) p_{x,y}(x, y) dx dy. \end{aligned} \tag{3.13}$$

$p_{x,y}(x, y)$: joint PDF of the two processes.



Covariance normalized to get the correlation coefficient:

$$\rho_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y}, \quad (3.14)$$

with $-1 \leq \rho_{xy} \leq +1$.



High covariance: the two processes have similar statistical variability or behavior.

The processes x and y are uncorrelated if $\rho_{xy} = 0$.

Two statistically independent processes are also

uncorrelated; the converse of this property is,

in general, not true.



Cross-correlation function (CCF) of random processes

x and y that are functions of time:

$$\theta_{xy}(t_1, t_1 + \tau) = E[x(t_1)y(t_1 + \tau)] \quad (3.15)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(t_1) y(t_1 + \tau) p_{x,y}(x, y) dx dy.$$

CCF useful in analyzing the nature of variability,

spectral bandwidth of signals, and

detection of events by template matching.



Structured noise:

Power-line interference at 50 Hz or 60 Hz :

typical waveform of the interference known in advance.

Phase of the interfering waveform not known.

Interfering waveform may not be an exact sinusoid:

presence of harmonics of the fundamental 50 Hz or 60 Hz .



Physiological interference:

Human body: complex conglomeration of several systems.

Several physiological processes active at a given time,

each one producing many signals of different types.

Patient or subject not able to exercise control on all

physiological processes and systems.

Appearance of signals from systems or processes other than

those of interest: *physiological interference*.



- EMG related to coughing, breathing, squirming in ECG.
- EGG interfering with precordial ECG.
- Maternal ECG getting added to fetal ECG.
- ECG interfering with the EEG.
- Ongoing EEG in ERPs and SEPs.



- Breath, lung, bowel sounds in heart sounds (PCG).
- Heart sounds in breath or lung sounds.
- Muscle sound (VMG) in joint sounds (VAG).
- Needle-insertion activity at beginning of EMG.



Physiological interference not characterized by any

specific waveform or spectral content —

typically dynamic and nonstationary:

linear bandpass filters will not be applicable.

Need adaptive filters with reference inputs.



3.1.2 *Stationary versus nonstationary processes*

A random process is *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 used.

A random process is *weakly stationary*

or *stationary in the wide sense*

if its mean is a constant and

ACF depends only upon the difference (or shift) in time.



Then, we have

$$\mu_x(t_1) = \mu_x \text{ and}$$

$$\phi_{xx}(t_1, t_1 + \tau) = \phi_{xx}(\tau).$$

ACF function of the delay parameter τ only;

the PSD of the process does not vary with time.



A stationary process is *ergodic*

if the temporal statistics are

independent of the sample observed;

same result obtained for any sample observation $x_k(t)$.

Time averages are then independent of k :

$$\mu_x(k) = \mu_x \text{ and } \phi_{xx}(\tau, k) = \phi_{xx}(\tau).$$



All ensemble statistics may be replaced by temporal statistics when analyzing ergodic processes.

Ergodic processes:

important type of stationary random processes

because their statistics may be computed from

a single observation as a function of time.



Signals or processes that do not meet the conditions

described above: *nonstationary processes*.

A nonstationary process possesses

statistics that vary with time.

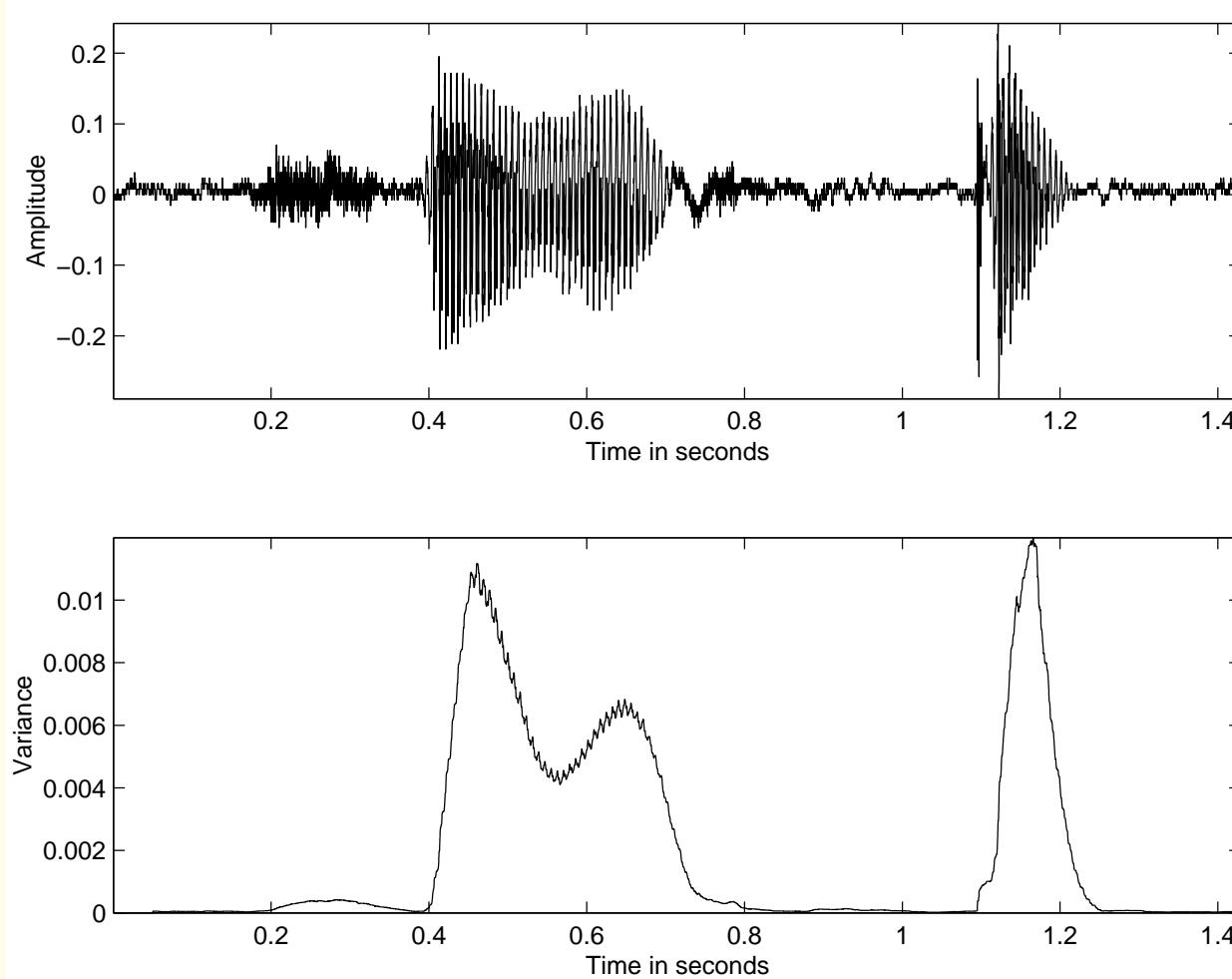


Figure 3.3: Top: Speech signal of the word “safety” uttered by a male speaker. Bottom: Variance computed in a moving window of 50 ms (400 samples with $f_s = 8 \text{ kHz}$).



Most biomedical systems are dynamic:

produce nonstationary signals

Examples: EMG, EEG, VMG, PCG, VAG, speech.

However, a physical or physiological system has

limitations in the rate of change of its characteristics.



This facilitates breaking the signal into

segments of short duration

over which the statistics of interest are not varying,

or may be assumed to remain the same.

Quasi-stationary process: short-time analysis.

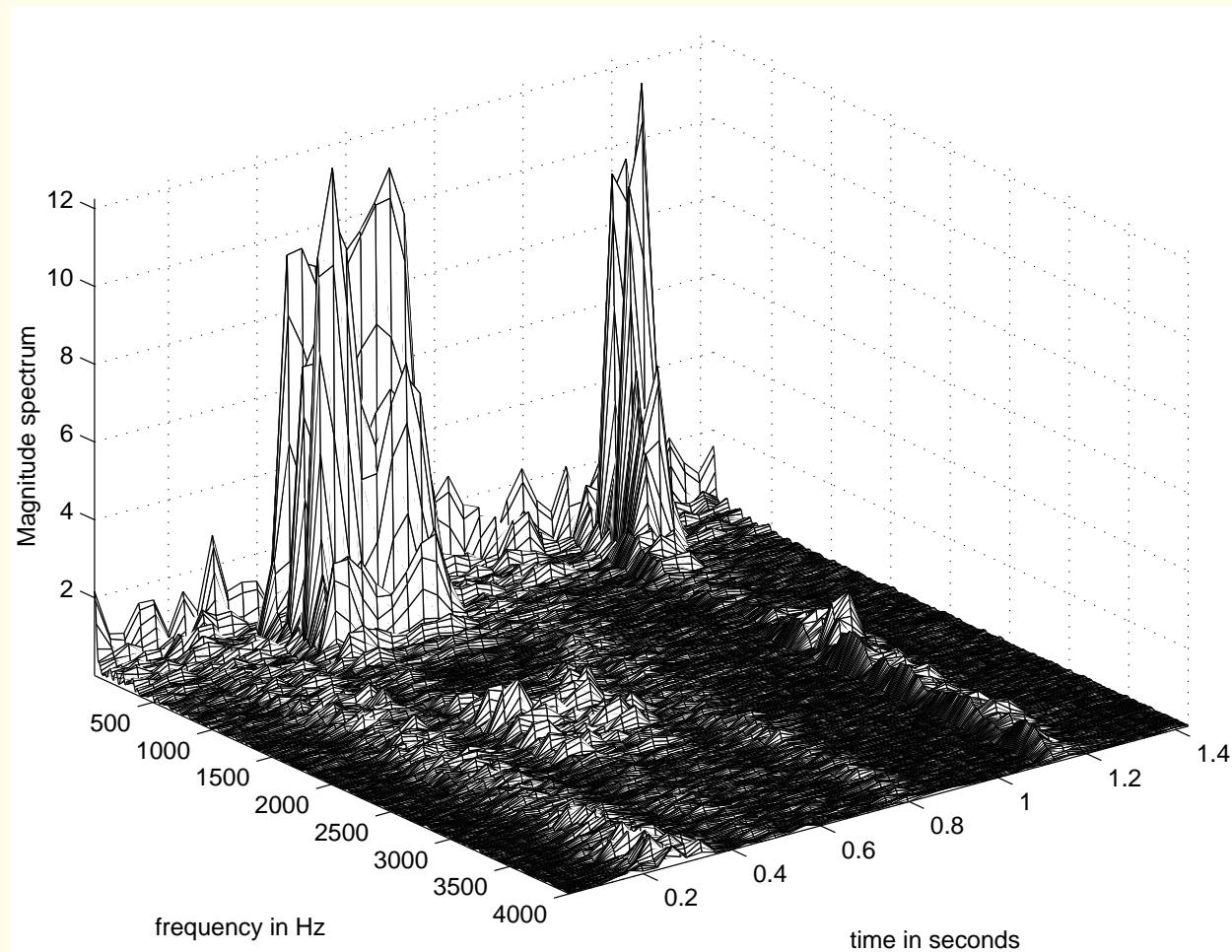


Figure 3.4: Spectrogram of the speech signal of the word “safety” uttered by a male speaker. (The signal is also illustrated in Figures 1.29, 3.1, and 3.3.) Each curve represents the magnitude spectrum of the signal in a moving window of duration 64 ms (512 samples with $f_s = 8 \text{ kHz}$), with the 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.

ECG, PCG, carotid pulse: almost periodic —

cyclo-stationary signals.

Statistics of PCG vary within the duration of a cardiac cycle,

especially when murmurs are present,

but repeat themselves at regular intervals.



Cyclic repetition facilitates ensemble averaging,
using epochs or events extracted from
an observation of the signal over many cycles
(strictly speaking, a single function of time).



3.2 Illustration of the Problem with Case-studies

3.2.1 *Noise in event-related potentials*

ERP: signal obtained in response to a stimulus.

Response of small amplitude ($\sim 10 \mu V$),

buried in ambient EEG activity and noise.

A single response may not be recognizable.



Figure 3.2: ten individual flash visual ERP signals.

Recorded at occipital midline (oz) position

against left and right ear lobes combined (a1 – a2).

Left forehead used as the reference.

ERP buried in ongoing EEG and

power-line (60 Hz) interference.



3.2.2 *High-frequency noise in the ECG*

Figure 3.5: ECG with high-frequency noise.

Noise could be due to instrumentation amplifiers,

the recording system,

pickup of ambient EM signals by cables, etc.

Also power-line interference at 60 Hz and harmonics.

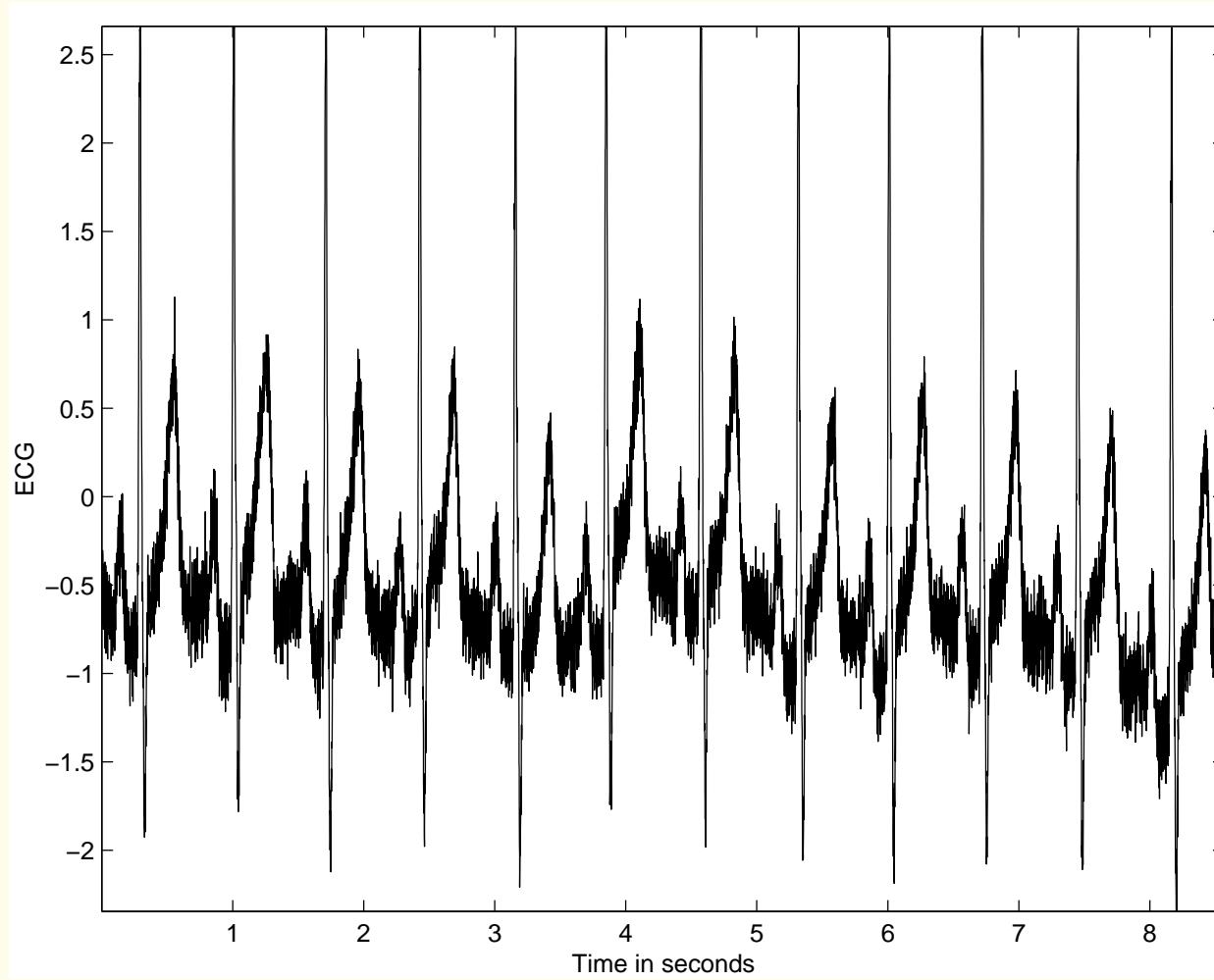


Figure 3.5: ECG signal with high-frequency noise.



3.2.3 *Motion artifact in the ECG*

Low-frequency artifacts and base-line drift

in chest-lead ECG due to coughing or

breathing with large movement of the chest,

or when an arm or leg is moved in the case of

limb-lead ECG.



EGG common source of artifact in chest-lead ECG.

Poor contact, polarization of electrodes:

low-frequency artifacts.



Base-line drift: variations in temperature

and bias in the instrumentation and amplifiers.

Figure 3.6: ECG with low-frequency artifact.

Base-line drift: affects analysis of the

isoelectric nature PQ and ST segments.

Large base-line drift: positive or negative peaks in the

ECG clipped by the amplifiers or the ADC.

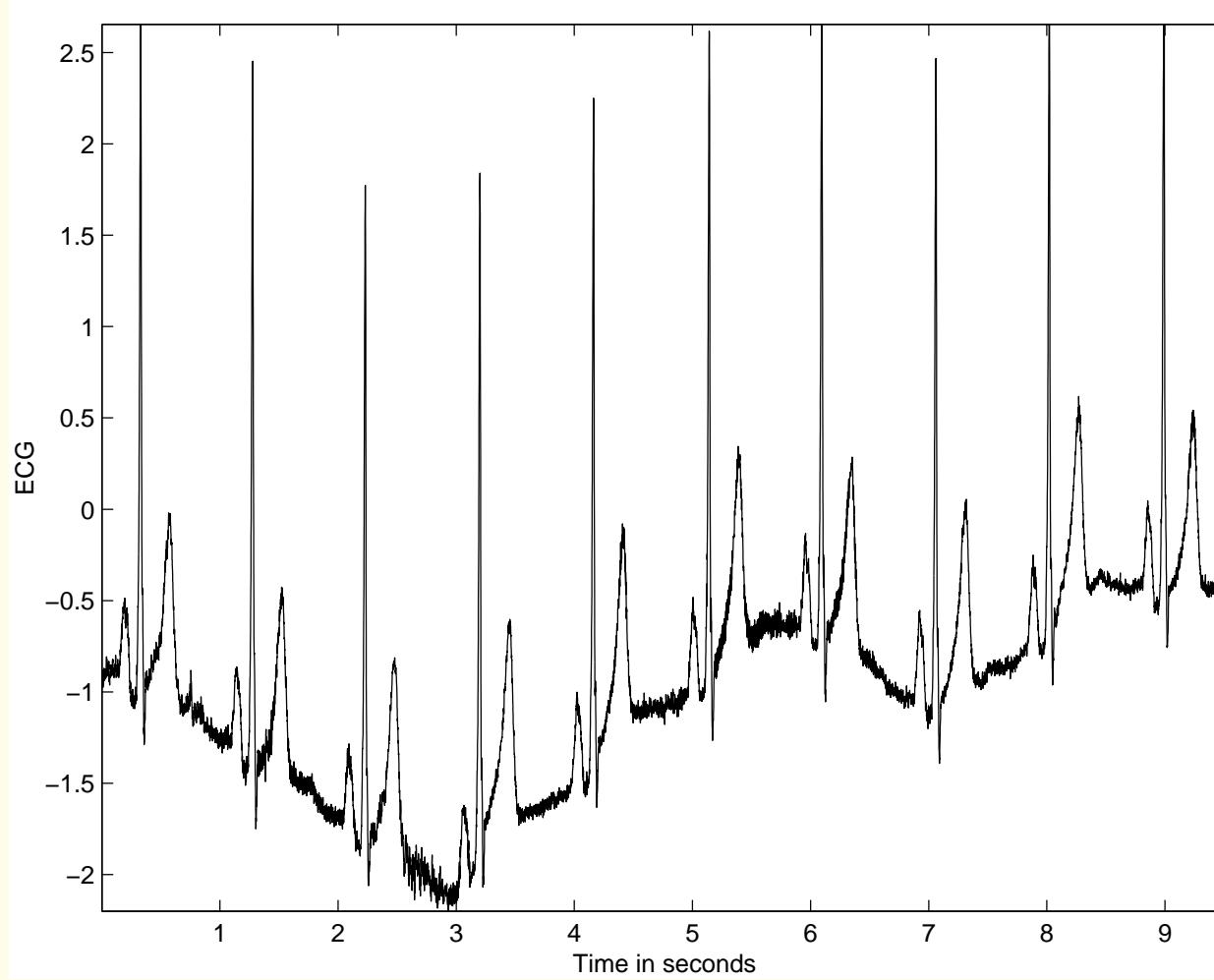


Figure 3.6: ECG signal with low-frequency artifact.



3.2.4 Power-line interference in ECG signals

Most common periodic artifact in biomedical signals:

power-line interference at 50 Hz or 60 Hz + harmonics.

Bandwidth of interest of ECG: $0.05 - 100\text{ Hz}$.

Lowpass filtering not appropriate for removal of

power-line interference:

will smooth and blur QRS, and affect PQ and ST segments.

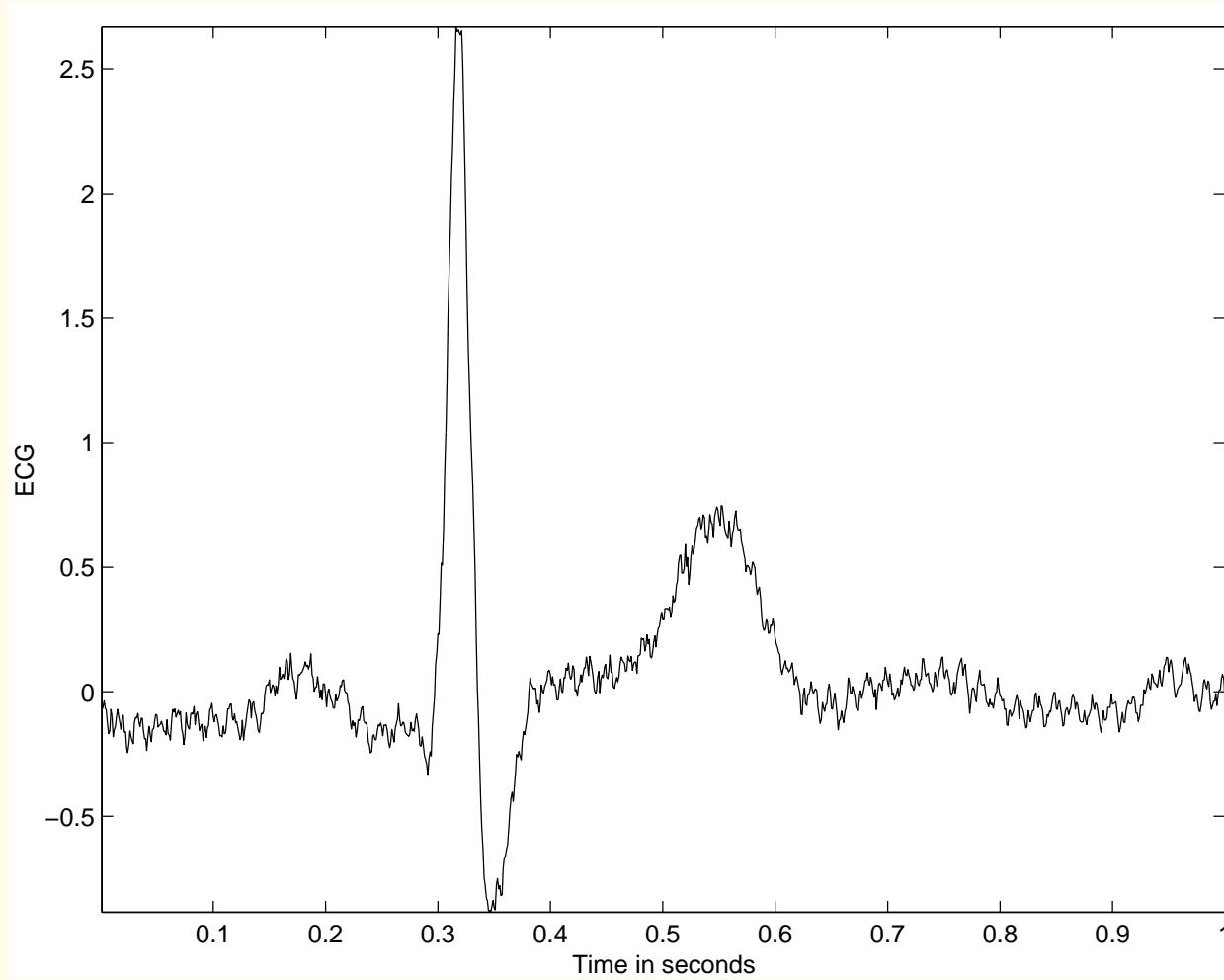


Figure 3.7: ECG signal with power-line (60 Hz) interference.

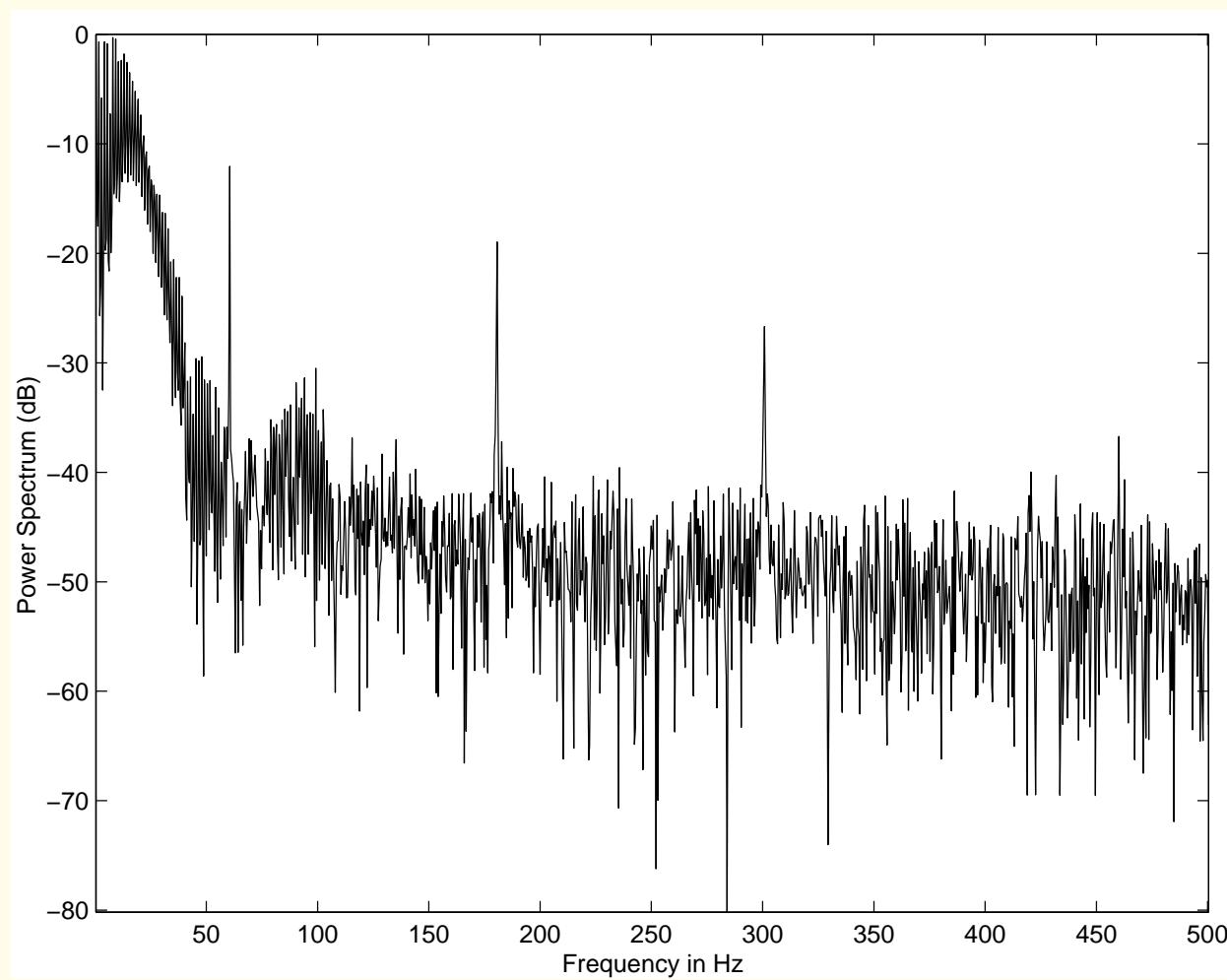


Figure 3.8: Power spectrum of the ECG signal in Figure 3.7 with power-line interference. The spectrum illustrates peaks at the fundamental frequency of 60 Hz as well as the third and fifth harmonics at 180 Hz and 300 Hz , respectively.



3.2.5 Maternal interference in fetal ECG

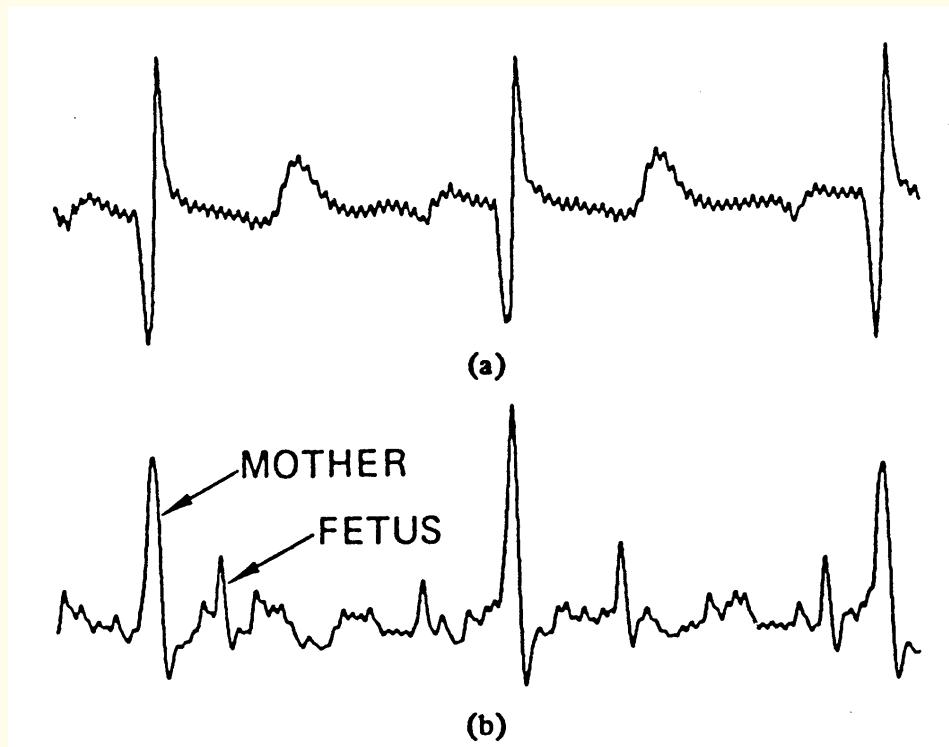


Figure 3.9: ECG signals of a pregnant woman from abdominal and chest leads: (a) chest-lead ECG, and (b) abdominal-lead ECG; the former presents the maternal ECG whereas the latter is a combination of the maternal and fetal ECG signals. (See also Figure 3.9.) Reproduced with permission from B. Widrow,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



3.2.6 Muscle-contraction interference in VAG signals

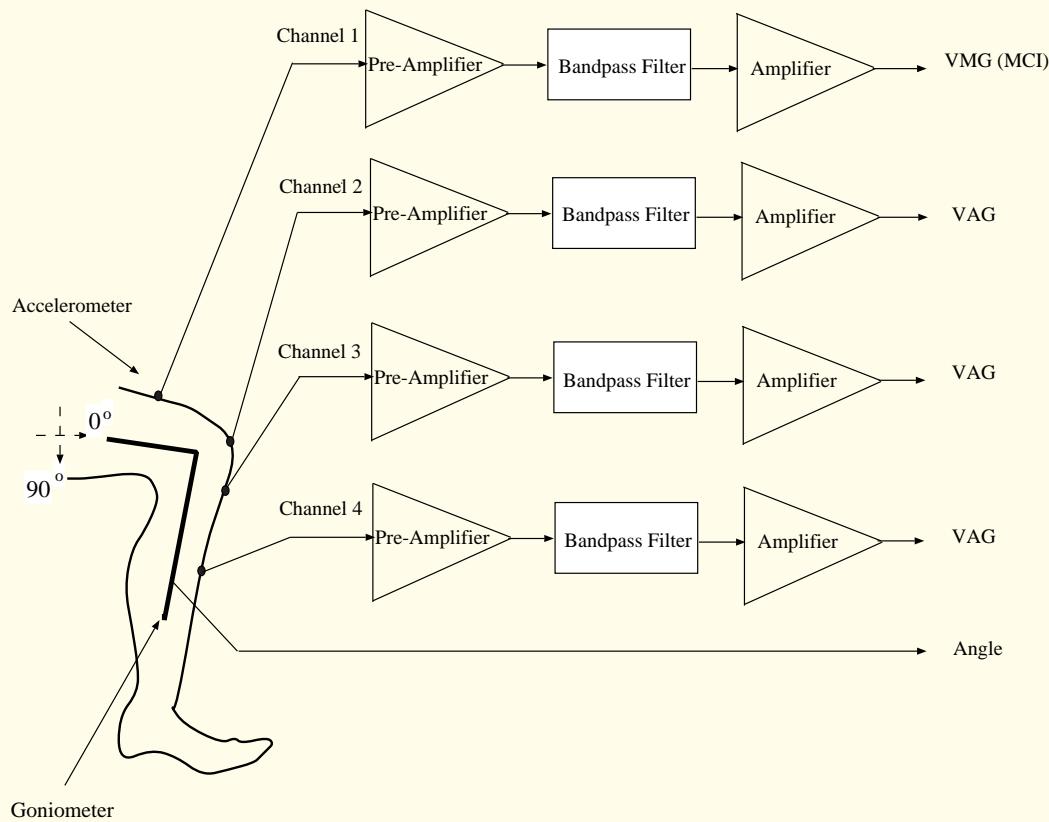
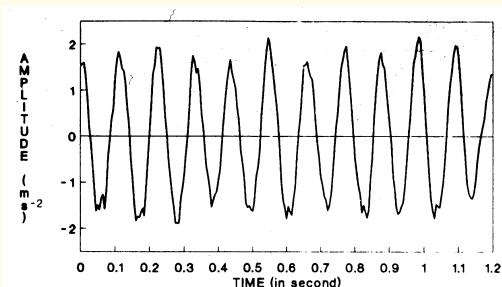
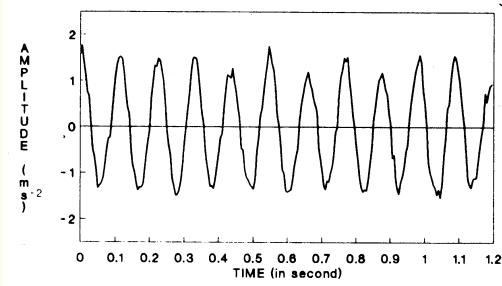


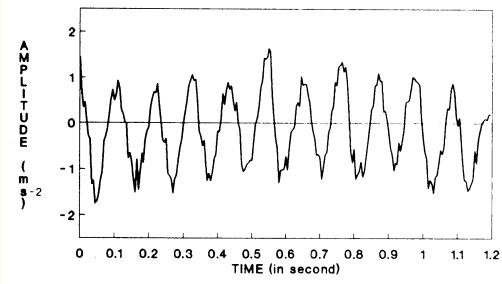
Figure 3.10: Experimental setup to measure VMG and VAG signals at different positions along the leg.



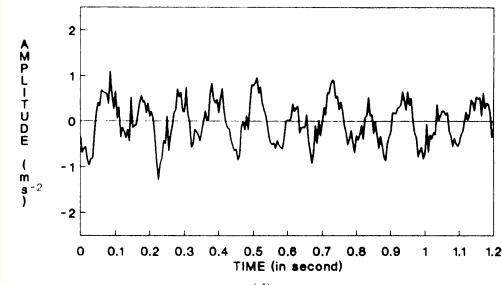
(a)



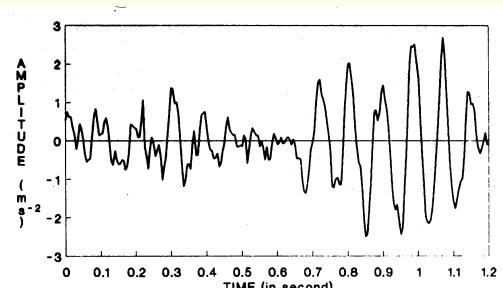
(b)



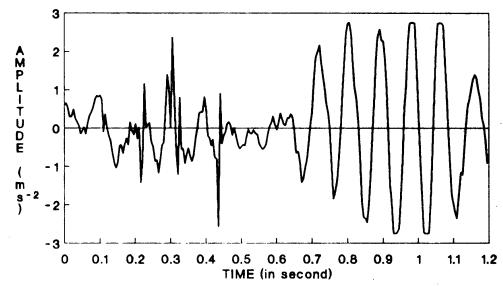
(c)



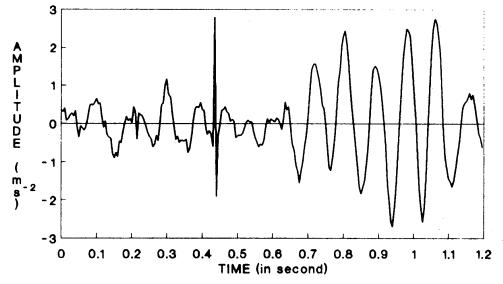
(d)



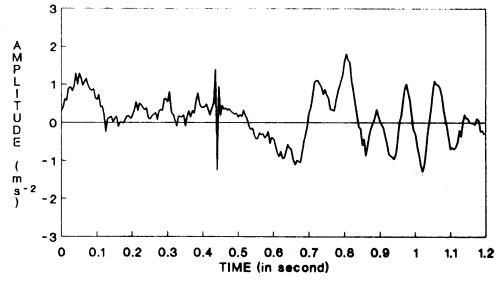
(a)



(b)



(c)



(d)



Figure 3.11: Left-hand column: VMG signals recorded simultaneously at (top-to-bottom) (a) the distal rectus femoris, (b) mid-patella, (c) tibial tuberosity, and (d) mid-tibial 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. 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.



3.2.7 Potential solutions to the problem

A practical problem encountered by an investigator

in the field may not precisely match

a specific textbook problem.

However, the knowledge of several techniques

and appreciation of the results of their application

could help in designing innovative and appropriate

solutions to new problems.



3.3 Time-domain Filters

Certain types of noise may be filtered directly

in the time domain using digital filters.

Advantage: spectral characterization of the signal and noise

not required — at least in a direct manner.

Time-domain processing may also be faster

than frequency-domain filtering.



3.3.1 *Synchronized averaging*

Linear filters fail when the signal and noise spectra overlap.

Synchronized signal averaging can separate a

repetitive signal from noise without distorting the signal.

ERP or SEP epochs may be obtained a number of times by

repeated application of the stimulus;

averaged using the stimulus as trigger to align the epochs.



If noise is random with zero mean, uncorrelated with signal, averaging will improve the SNR.

$y_k(n)$: one realization of a signal,

with $k = 1, 2, \dots, M$ representing the ensemble index,

$n = 1, 2, \dots, N$ representing the time-sample index.

M : number of copies, events, epochs, or realizations.

N : number of samples in each signal.



$$y_k(n) = x_k(n) + \eta_k(n), \quad (3.16)$$

$x_k(n)$: original uncorrupted signal,

$\eta_k(n)$: noise in k^{th} copy of signal.

For each instant of time n , add M copies of signal:

$$\sum_{k=1}^M y_k(n) = \sum_{k=1}^M x_k(n) + \sum_{k=1}^M \eta_k(n); \quad n = 1, 2, \dots, N. \quad (3.17)$$



If the repetitions of the signal are identical and aligned,

$$\sum_{k=1}^M x_k(n) = Mx(n).$$

If noise is random, has zero mean and variance σ_η^2 ,

$\sum_{k=1}^M \eta_k(n) \rightarrow 0$ as M increases,

with a variance of $M\sigma_\eta^2$.



RMS value of noise in the averaged signal = $\sqrt{M}\sigma_\eta$.

Thus SNR of signal increases by $\frac{M}{\sqrt{M}}$ or \sqrt{M} .

Larger the number of epochs or realizations averaged,

better the SNR of the result.

Synchronized averaging is a type of ensemble averaging.



Algorithmic description of synchronized averaging:

1. Obtain number of realizations of signal or event.
2. Determine reference point for each realization.
Trigger, stimulus, QRS in ECG, etc.
3. Extract parts of the signal corresponding to the events and add them to a buffer.
Various parts may have different durations.
Alignment of copies at trigger point is important;
the tail ends of all parts may not be aligned.
4. Divide the result by the number of events added.



Figure 3.12, upper two traces: two single-flash ERPs.

Averaging 10 and 20 flashes: third and fourth plots.

Visual ERPs: latencies of first major peak or positivity P120

(normal expected latency for adults is 120 *ms*),

trough or negativity before P120, labeled as N80;

trough following P120, labeled as N145.

N80, P120, and N145 latencies from averaged signal in

Trace 4 of Figure 3.12: 85.7, 100.7, and 117 *ms*.

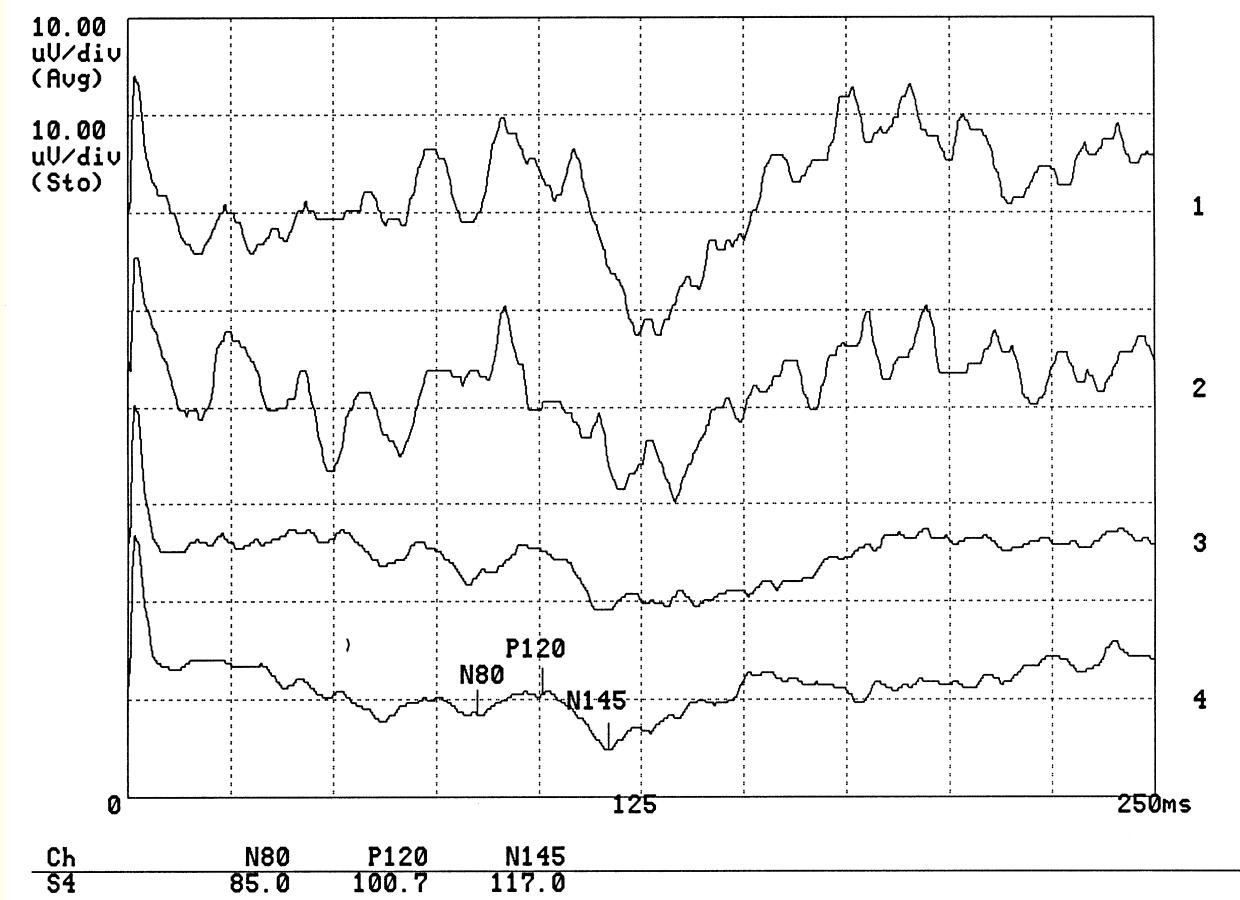


Figure 3.12: Traces 1 and 2: Two sample acquisitions of individual flash visual ERPs from the occipital midline (oz) position of a normal adult male. The ear lobes were used to form the reference lead (a1a2), and the left forehead was used as the reference (see Figure 1.20). Trace 3: Average of 10 ERPs. Trace 4: Average of 20 ERPs. The latencies of interest have been labeled on Trace 4 by an EEG technologist. See also Figure 3.2. Data courtesy of L. Alfaro and H. Darwish, Alberta Children's Hospital, Calgary.



Illustration of application:

Figure 3.13: noisy ECG signal over several beats.

To obtain trigger points:

sample QRS complex of *86 ms* duration

(86 samples at sampling rate $1,000 \text{ Hz}$)

extracted from first beat and used as a template.



Template matching using normalized correlation coefficient:

$$\gamma_{xy}(k) = \frac{\sum_{n=0}^{N-1} [x(n) - \bar{x}][y(k - N + 1 + n) - \bar{y}_k]}{\sqrt{\sum_{n=0}^{N-1} [x(n) - \bar{x}]^2 \sum_{n=0}^{N-1} [y(k - N + 1 + n) - \bar{y}_k]^2}}, \quad (3.18)$$

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



\bar{y}_k : average of part of y used in template matching.

Operation is causal, valid from $k = N$ to last sample of y .

x : template; \bar{x} : average value of x .

k : time index of y at which template is placed.

Averaging inherent in cross-correlation (over N samples)

reduces effect of noise on template matching.



Appropriate threshold: obtain trigger point

to extract QRS locations in the ECG.

Threshold of 0.9: all 12 beats detected.

Figure 3.14: two ECG cycles extracted

and result of averaging the first 11 cycles in the signal.

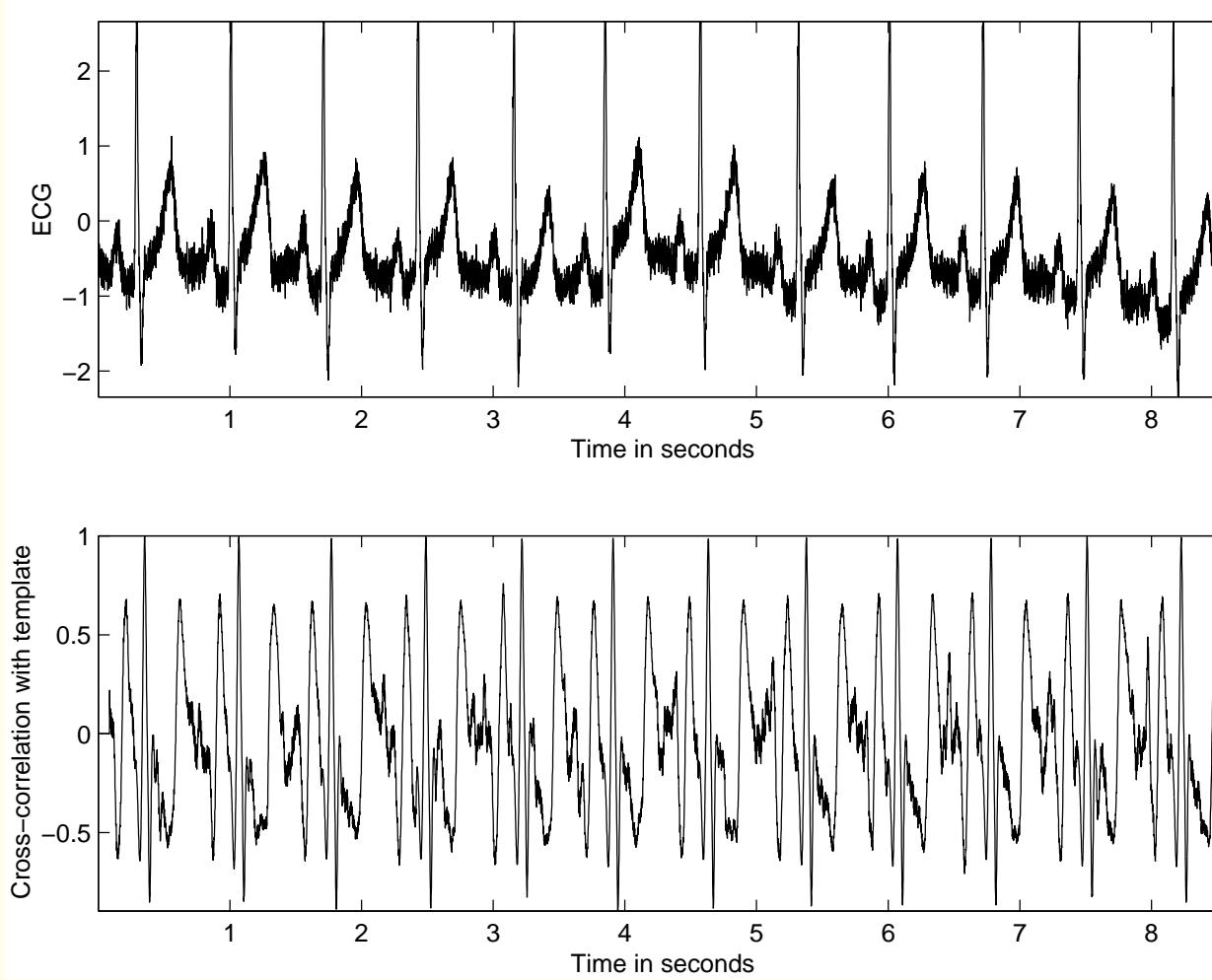


Figure 3.13: An ECG signal with noise (upper trace) and the result of cross-correlation (lower trace) with the QRS template selected from the first cycle. The cross-correlation coefficient is normalized to the range $(-1, 1)$.

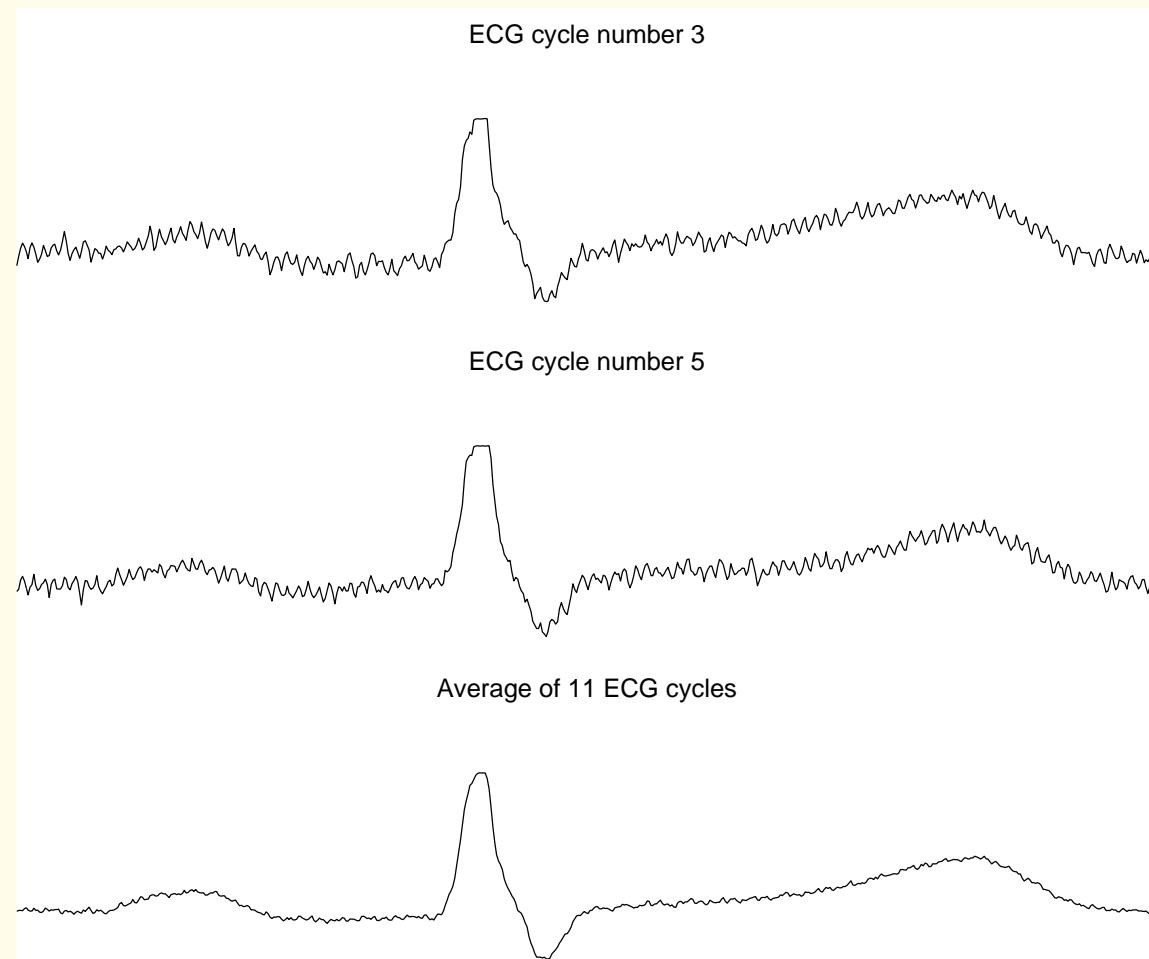


Figure 3.14: Upper two traces: two cycles of the ECG extracted from the signal in Figure 3.13. Bottom trace: the result of synchronized averaging of 11 cycles from the same ECG signal.



3.3.2 *Moving-average filters*

Problem:

Propose a time-domain technique to remove random noise

given only one realization of the signal or event of interest.

Solution:

Ensemble of several realizations not available —

synchronized averaging not possible.

Consider temporal averaging for noise removal.



Assumption: processes are ergodic.

Temporal statistics used instead of ensemble statistics.

Temporal window of samples moved to obtain

output at various points of time: moving-window averaging

or moving-average (MA) filter.

Average \sim weighted combination of samples.



General form of MA filter:

$$y(n) = \sum_{k=0}^N b_k x(n - k), \quad (3.19)$$

x and y : input and output of filter.

b_k : filter coefficients or tap weights.

N : order of filter.

Effect of division by the number of samples used ($N + 1$)

included in the values of the filter coefficients.

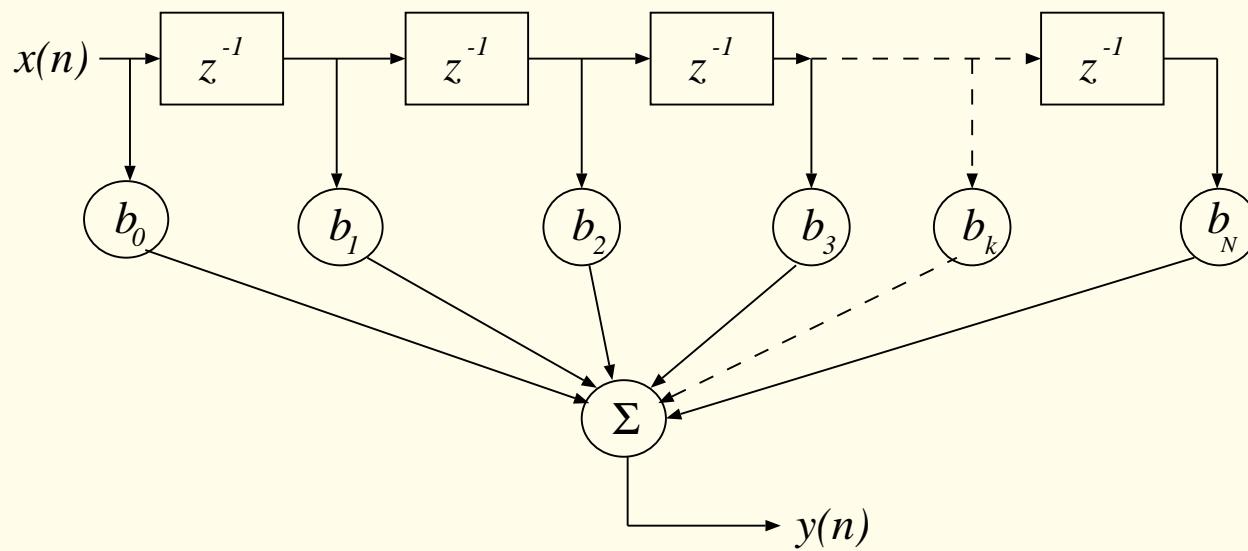


Figure 3.15: Signal-flow diagram of a moving-average filter of order N . Each block with the symbol z^{-1} represents a delay of one sample, and serves as a memory unit for the corresponding signal sample value.



Applying the z -transform, we get the transfer function:

$$H(z) = \frac{Y(z)}{X(z)} = \sum_{k=0}^N b_k z^{-k} = b_0 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_N z^{-N}, \quad (3.20)$$

$X(z)$ and $Y(z)$: z -transforms of $x(n)$ and $y(n)$.



MA filter to remove noise — von Hann or Hanning filter:

$$y(n) = \frac{1}{4}[x(n) + 2x(n - 1) + x(n - 2)]. \quad (3.21)$$

Impulse response: let $x(n) = \delta(n)$.

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



Transfer function of the Hanning filter:

$$H(z) = \frac{1}{4}[1 + 2z^{-1} + z^{-2}] = \frac{1}{4}[1 + z^{-1}]^2. \quad (3.22)$$

Double-zero at $z = -1$.

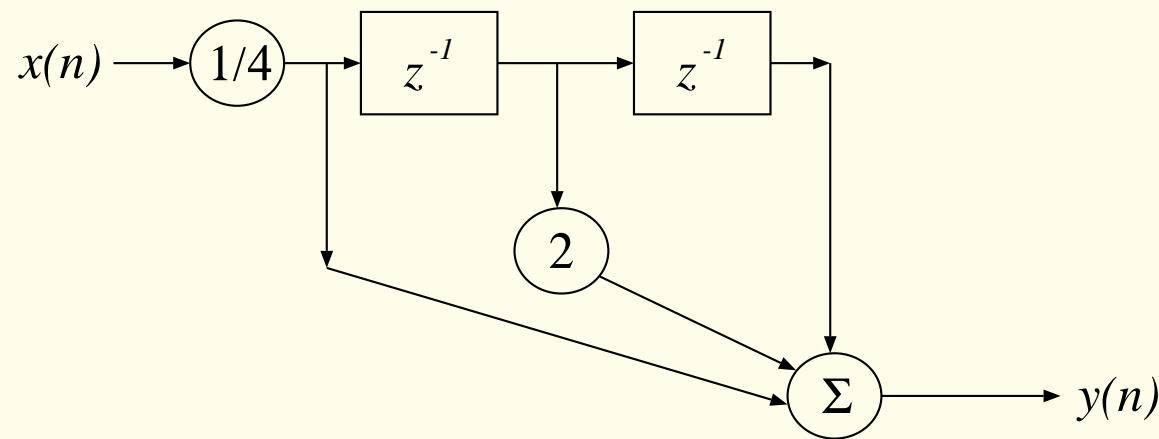


Figure 3.16: Signal-flow diagram of the Hanning filter.



An MA filter is a finite impulse response (FIR) filter:

- Impulse response $h(k)$ has a finite number of terms:
$$h(k) = b_k, \quad k = 0, 1, 2, \dots, N.$$
- An FIR filter may be realized non-recursively with no feedback.
- Output depends only on the present input sample and a few past input samples.



- Filter is a set of tap weights of the delay stages.
- Transfer function has no poles except at $z = 0$:
the filter is inherently stable.
- Filter has linear phase if the series of tap weights
is symmetric or antisymmetric.



Frequency response: substitute $z = e^{j\omega T}$ in $H(z)$,

T : sampling interval in seconds, $T = 1/f_s$,

f : frequency in Hz, f_s : sampling frequency,

ω : radian frequency, $\omega = 2\pi f$.



Set $T = 1$ and deal with normalized frequency in the range

$$0 \leq \omega \leq 2\pi \text{ or } 0 \leq f \leq 1;$$

then $f = 1$ or $\omega = 2\pi$ represents the sampling frequency,

lower frequency values represented as

normalized fraction of f_s .



Frequency response of the Hanning filter:

$$H(\omega) = \frac{1}{4}[1 + 2e^{-j\omega} + e^{-j2\omega}]. \quad (3.23)$$

Letting $e^{-j\omega} = \cos(\omega) - j \sin(\omega)$,

$$H(\omega) = \frac{1}{4}[\{2 + 2 \cos(\omega)\}e^{-j\omega}]. \quad (3.24)$$



Magnitude and phase responses:

$$|H(\omega)| = \left| \frac{1}{2} \{1 + \cos(\omega)\} \right| \quad (3.25)$$

and

$$\angle H(\omega) = -\omega. \quad (3.26)$$

Lowpass filter with linear phase.

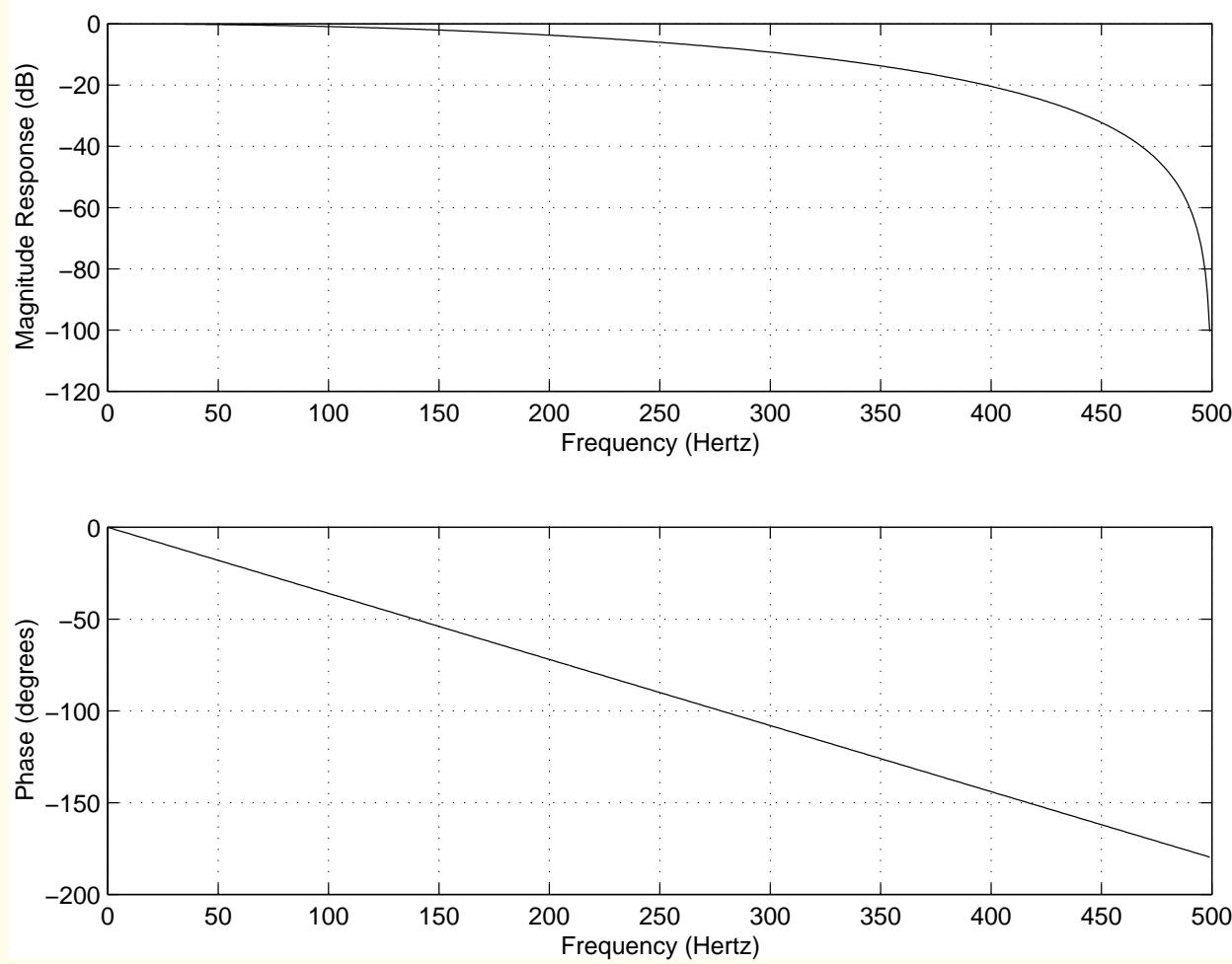


Figure 3.17: Magnitude and phase responses of the Hanning (smoothing) filter.



For more smoothing:

average signal samples over longer time windows;

expense of increased filter delay.

$$y(n) = \frac{1}{8} \sum_{k=0}^7 x(n-k). \quad (3.27)$$

Impulse response:

$$h(n) = \frac{1}{8} [\delta(n) + \delta(n-1) + \delta(n-2) + \delta(n-3)]$$

$$+ \delta(n-4) + \delta(n-5) + \delta(n-6) + \delta(n-7)].$$



Transfer function:

$$H(z) = \frac{1}{8} \sum_{k=0}^7 z^{-k}. \quad (3.28)$$

Frequency response:

$$\begin{aligned} H(\omega) &= \frac{1}{8} \sum_{k=0}^7 \exp(-j\omega k) \\ &= \frac{1}{8} [1 + \exp(-j4\omega) \\ &\quad \times \{1 + 2\cos(\omega) + 2\cos(2\omega) + 2\cos(3\omega)\}]. \end{aligned} \quad (3.29)$$



Zeros at $\frac{f_s}{8} = 125 \text{ Hz}$, $\frac{f_s}{4} = 250 \text{ Hz}$, $\frac{3f_s}{8} = 375 \text{ Hz}$,

and $\frac{f_s}{2} = 500 \text{ Hz}$.

More attenuation over $90 - 400 \text{ Hz}$ than the Hanning filter.

Attenuation after 100 Hz is nonuniform.

Phase response not linear, but piece-wise linear.

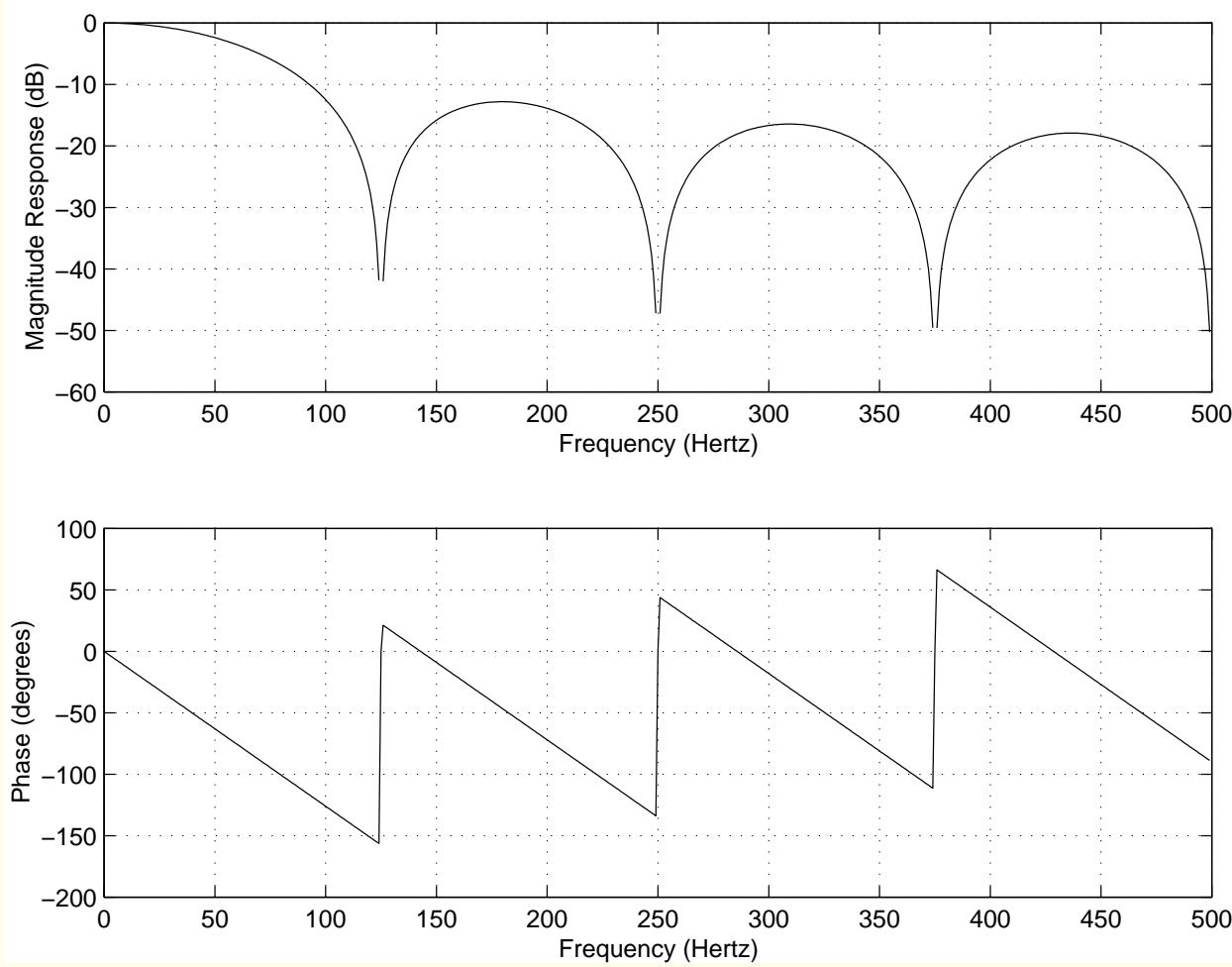


Figure 3.18: Magnitude and phase responses of the 8-point moving-average (smoothing) filter.

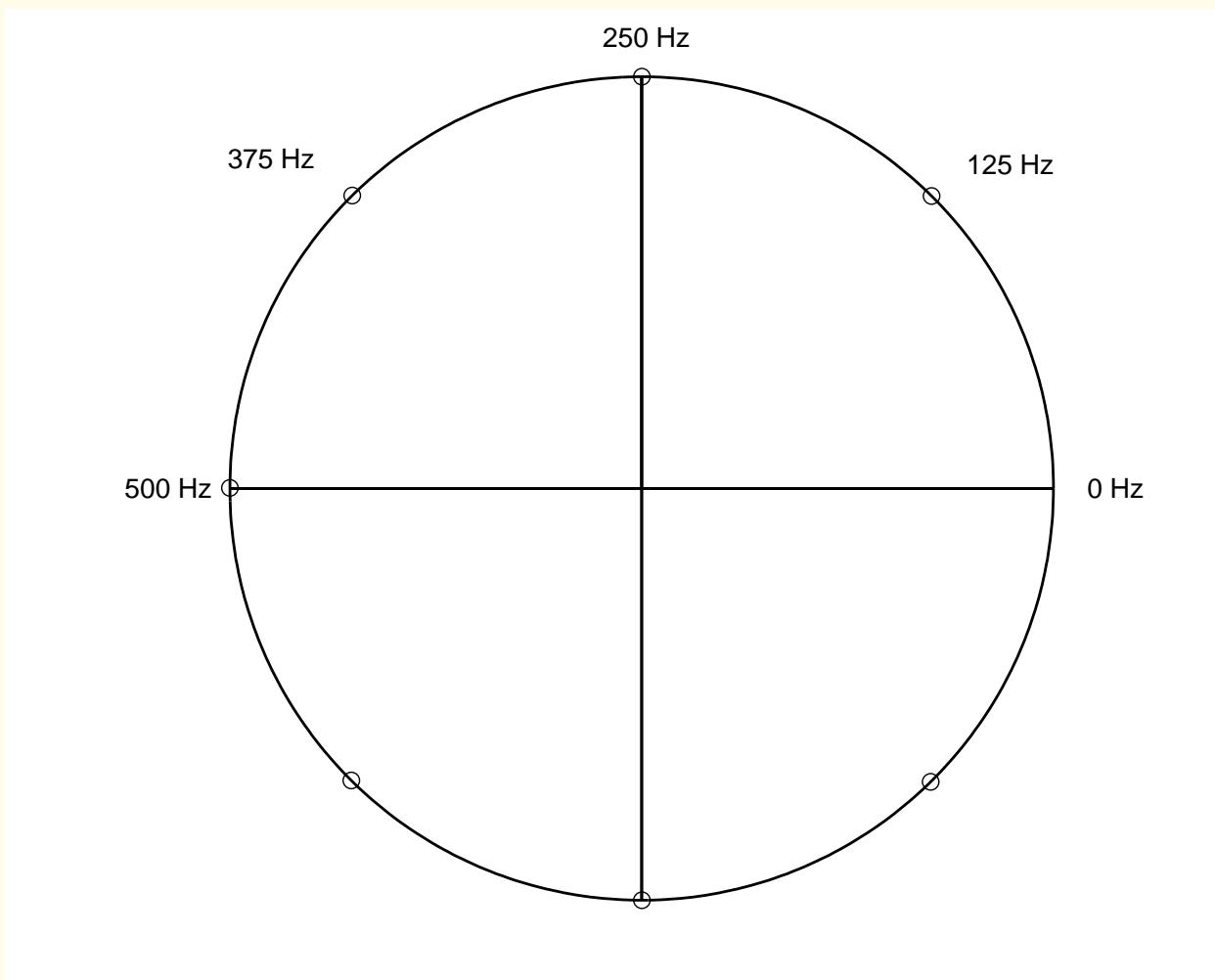


Figure 3.19: Pole-zero plot of the 8-point moving-average (smoothing) filter.



Relationship of moving-average filtering to integration:

Disregarding the $\frac{1}{8}$ scale factor,

the operation in Equation 3.27 may be interpreted as the

summation or integration of the signal from $n - 7$ to n .

Comparable integration of a continuous-time signal

$x(t)$ over the interval t_1 to t_2 :

$$y(t) = \int_{t_1}^{t_2} x(t) dt. \quad (3.30)$$



General definition of the integral of a signal:

$$y(t) = \int_{-\infty}^t x(t) dt, \quad (3.31)$$

or, if the signal is causal,

$$y(t) = \int_0^t x(t) dt. \quad (3.32)$$

Fourier transforms of the signals:

$$Y(\omega) = \frac{1}{j\omega} X(\omega) + \pi X(0)\delta(\omega). \quad (3.33)$$



Frequency response of the integration operator:

$$H(\omega) = \frac{1}{j\omega}. \quad (3.34)$$

Magnitude and phase responses:

$$|H(\omega)| = \left| \frac{1}{\omega} \right|, \quad (3.35)$$

$$\angle H(\omega) = -\frac{\pi}{2}. \quad (3.36)$$

Gain of the filter reduces nonlinearly as frequency increases: 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 $H(z) = \frac{1}{1-z^{-1}}$.

Such an operation is seldom used in practice.

Instead, a moving-window sum is computed as in

Equation 3.27.



The 8-point MA filter may be rewritten as

$$y(n) = y(n - 1) + \frac{1}{8}x(n) - \frac{1}{8}x(n - 8). \quad (3.37)$$

Recursive form clearly depicts integration.

Transfer function:

$$H(z) = \frac{1}{8} \left[\frac{1 - z^{-8}}{1 - z^{-1}} \right]. \quad (3.38)$$



Frequency response:

$$H(\omega) = \frac{1}{8} \left[\frac{1 - e^{-j8\omega}}{1 - e^{-j\omega}} \right] = \frac{1}{8} e^{-j\frac{7}{2}\omega} \left[\frac{\sin(4\omega)}{\sin(\frac{\omega}{2})} \right], \quad (3.39)$$

which is equivalent to that in Equation 3.29.

Summation over a limited discrete-time window results in a frequency response having sinc-type characteristics;

Figure 3.18.



Illustration of application:

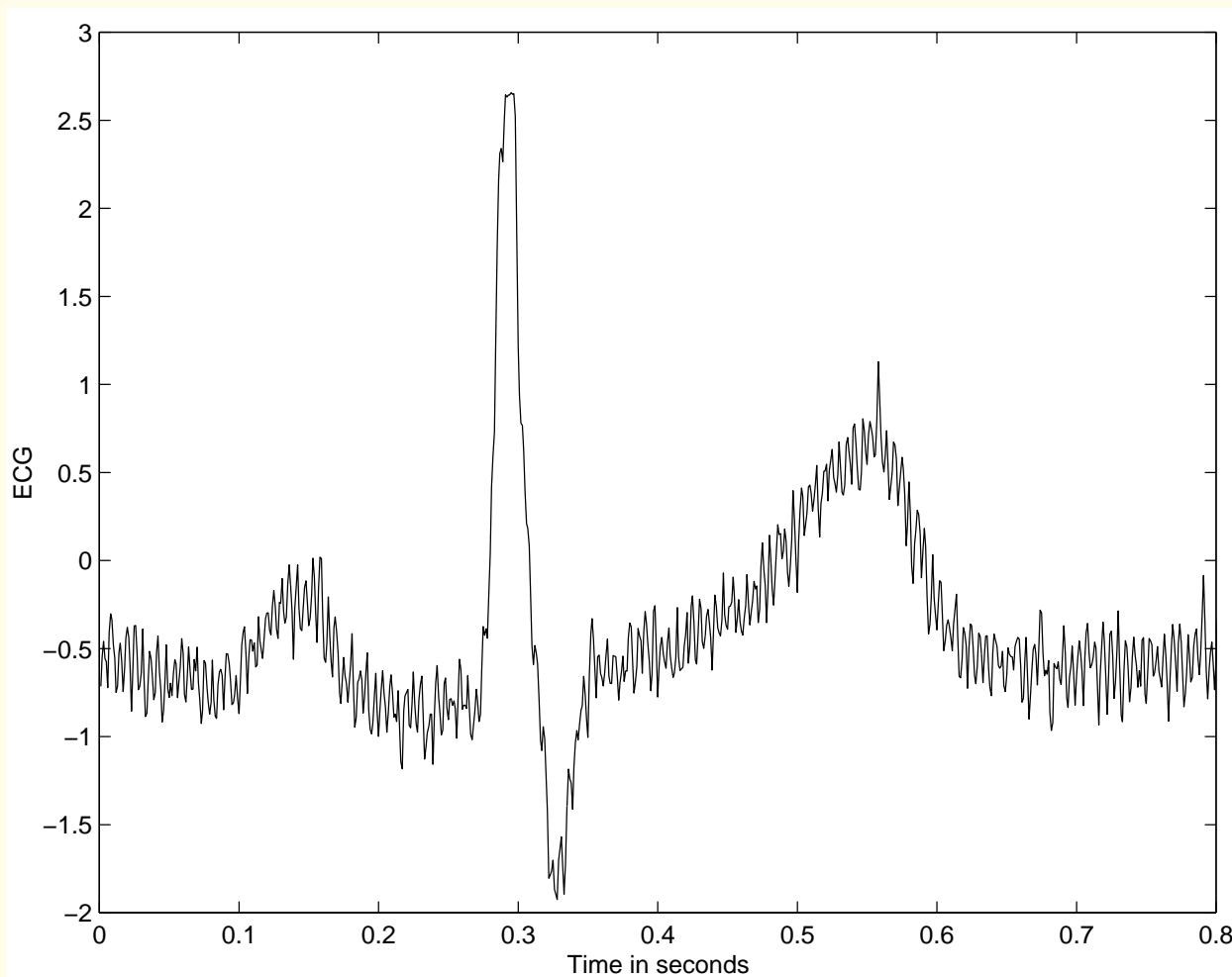


Figure 3.20: ECG signal with high-frequency noise; $f_s = 1,000 \text{ Hz}$.

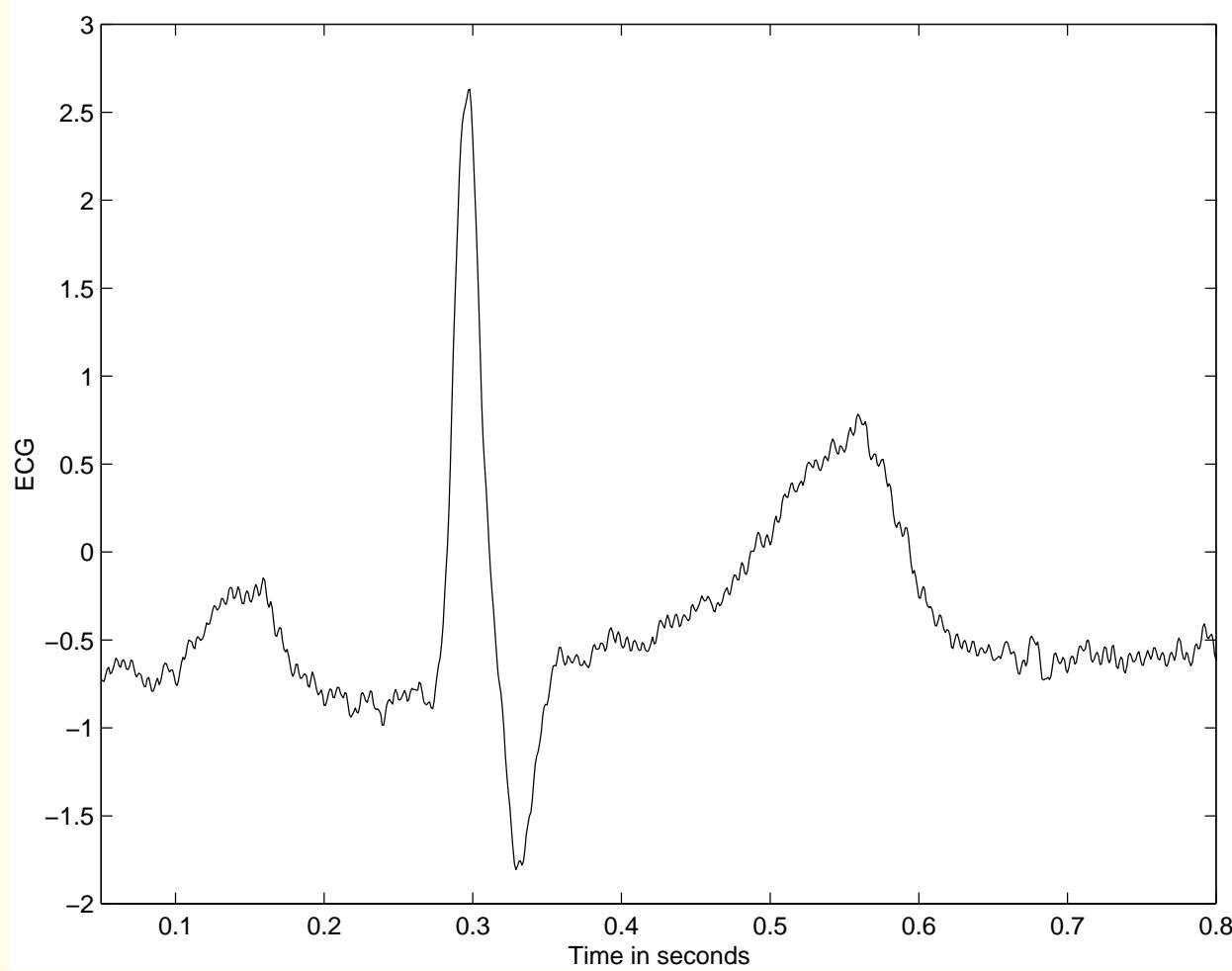


Figure 3.21: The ECG signal with high-frequency noise in Figure 3.20 after filtering by the 8-point MA filter shown in Figure 3.18.



3.3.3 *Derivative-based operators to remove low-frequency artifacts*

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

Solution: The derivative operator in the time domain

removes parts of the input that are constant.

Large changes in input → high values in output.



Ideal $\frac{d}{dt}$ operator in the time domain

results in multiplication of FT of original signal by

$j \omega = j 2\pi f$ in the frequency domain.

If $X(f) = \text{FT}[x(t)]$,

$$\text{FT}\left[\frac{dx}{dt}\right] = j 2\pi f X(f) = j \omega X(\omega).$$



Frequency response of the operation: $H(\omega) = j\omega$.

Gain increases linearly with frequency.

$$H(0) = 0:$$

DC component removed by the derivative operator.

Higher frequencies receive linearly increasing gain:

highpass filter.



The derivative operator may be used to remove DC,
suppress low-frequency components,
and boost high-frequency components.



Second-order derivative operator $\frac{d^2}{dt^2}$:

frequency response $H(\omega) = -\omega^2$,

quadratic increase in gain for higher frequencies.



In DSP, basic derivative = first-order difference:

$$y(n) = \frac{1}{T} [x(n) - x(n - 1)]. \quad (3.40)$$

Scale factor including the sampling interval T required to

obtain rate of change of signal with respect to true time.

Transfer function:

$$H(z) = \frac{1}{T} (1 - z^{-1}). \quad (3.41)$$

Zero at $z = 1$, the DC point.



Frequency response:

$$H(\omega) = \frac{1}{T} [1 - \exp(-j\omega)] = \frac{1}{T} \exp\left(-j\frac{\omega}{2}\right) \left[2j \sin\left(\frac{\omega}{2}\right)\right], \quad (3.42)$$

$$|H(\omega)| = \frac{2}{T} \left| \sin\left(\frac{\omega}{2}\right) \right|, \quad (3.43)$$

$$\angle H(\omega) = \frac{\pi}{2} - \frac{\omega}{2}. \quad (3.44)$$

High-frequency noise amplified significantly: noisy result.

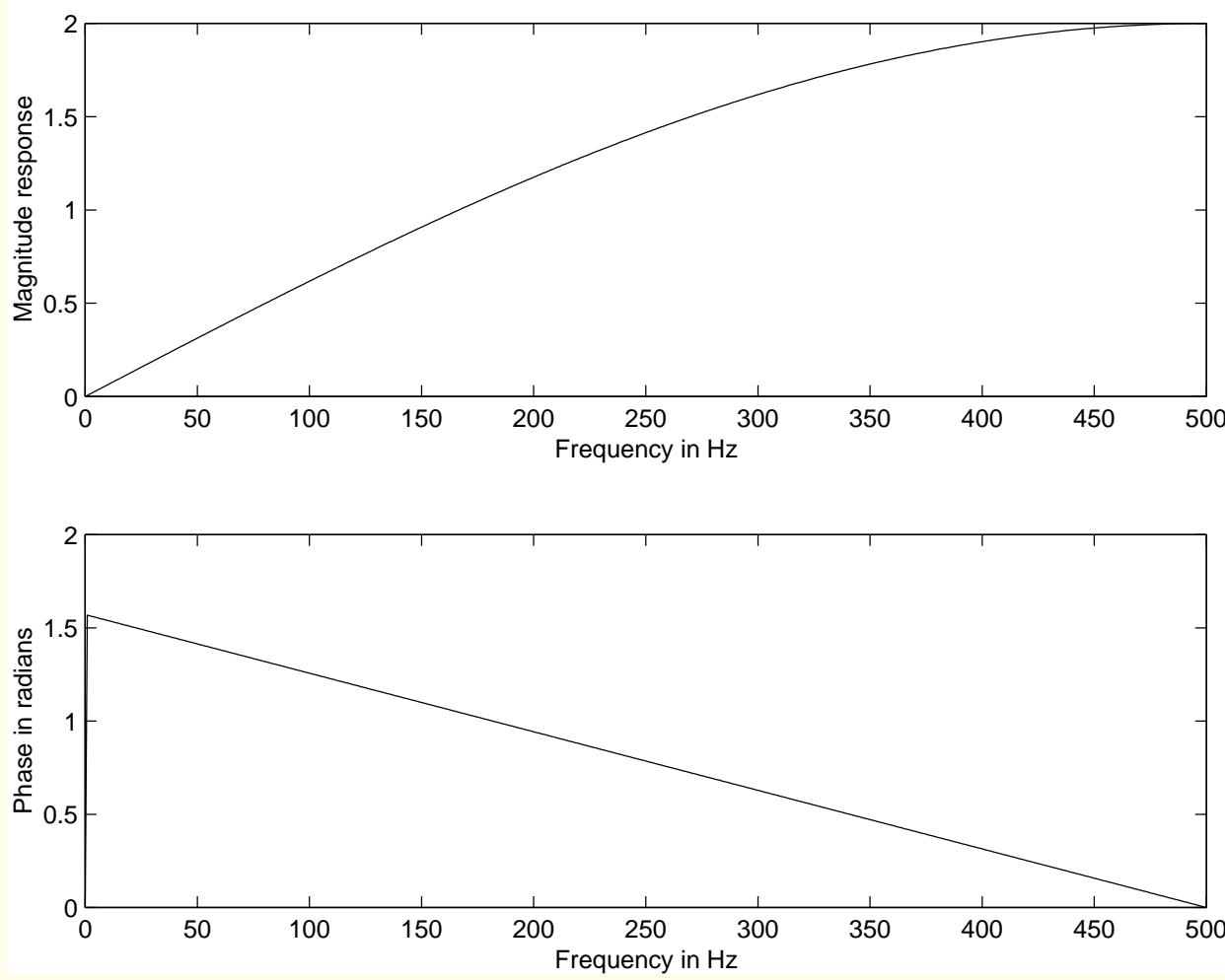


Figure 3.22: Magnitude and phase responses of the first-order difference operator. The magnitude response is shown on a linear scale in order to illustrate better its proportionality to frequency.



Three-point central difference:

Noise-amplification problem with the first-order difference

operator in Equation 3.40 may be controlled

by taking the average of two successive output values:

$$\begin{aligned}y_3(n) &= \frac{1}{2} [y(n) + y(n - 1)] \\&= \frac{1}{2T} [\{x(n) - x(n - 1)\} + \{x(n - 1) - x(n - 2)\}] \\&= \frac{1}{2T} [x(n) - x(n - 2)].\end{aligned}\tag{3.45}$$



$$H(z) = \frac{1}{2T} (1 - z^{-2}) = \left[\frac{1}{T} (1 - z^{-1}) \right] \left[\frac{1}{2} (1 + z^{-1}) \right]. \quad (3.46)$$

Transfer function of the three-point central-difference

operator = two filters in series (cascade)

= product of the transfer functions of the

first-order difference operator and a two-point MA filter.

Zeros at $z = 1$ and $z = -1$: bandpass filter.

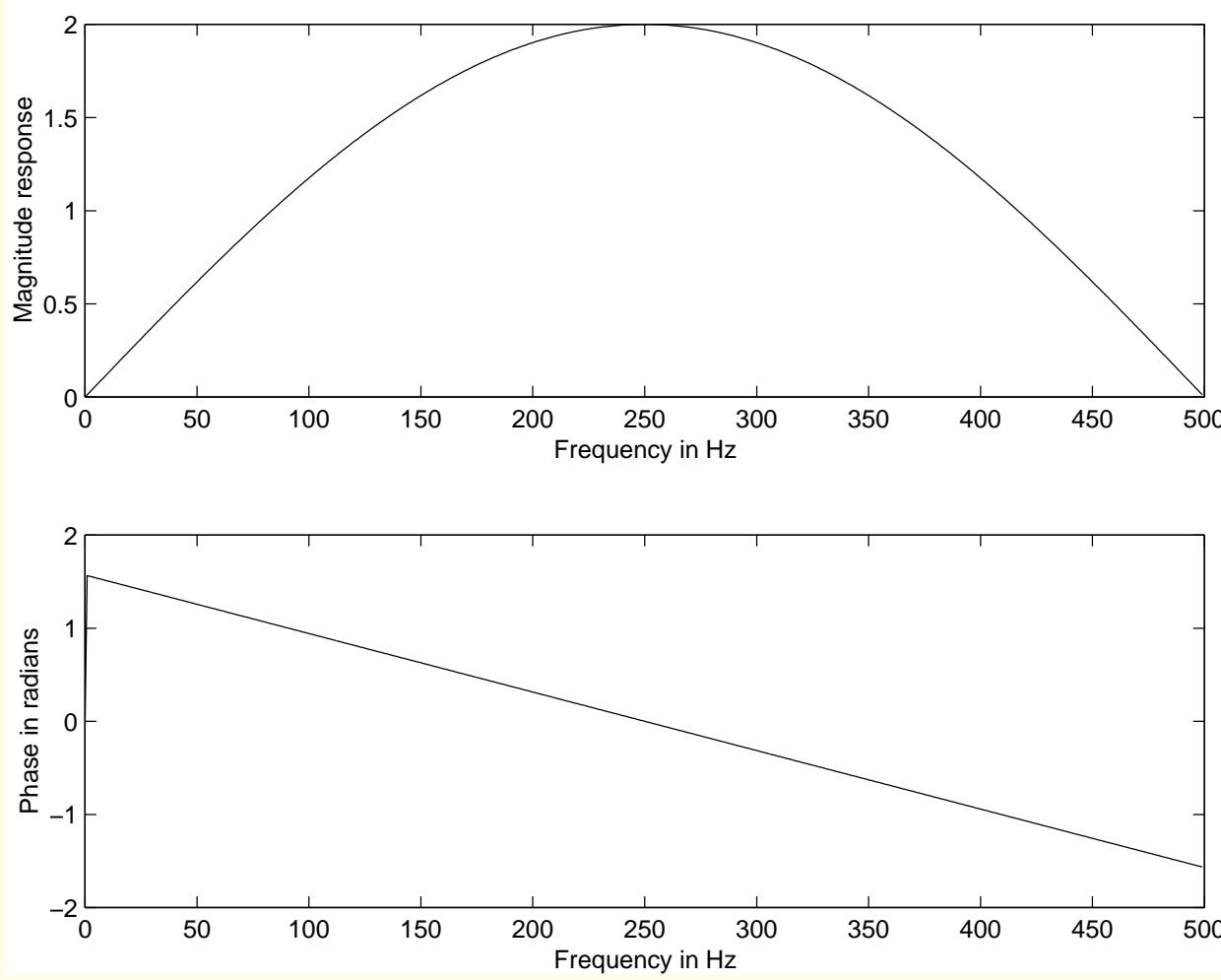


Figure 3.23: Magnitude and phase responses of the three-point central-difference operator. The magnitude response is shown on a linear scale.



Illustration of application: Figures 3.24 and 3.25.

Filtering the ECG signal with low-frequency noise in

Figure 3.6, using first-order difference and

three-point central difference.

Base-line drift has been removed,

but filters have also removed the slow P and T waves, and

altered the QRS complexes: results unlike ECG signals.

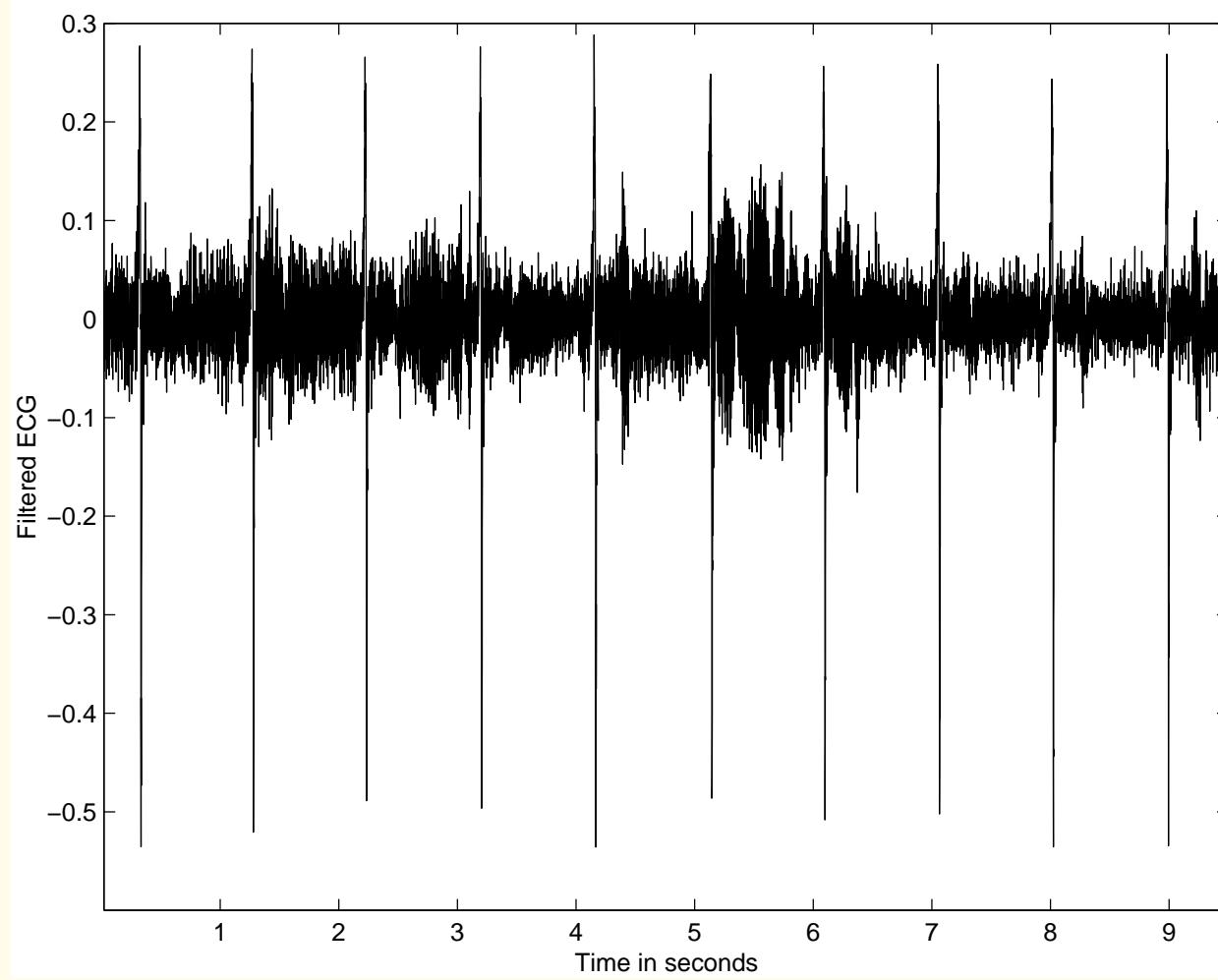


Figure 3.24: Result of filtering the ECG signal with low-frequency noise shown in Figure 3.6, using the first-order difference operator.

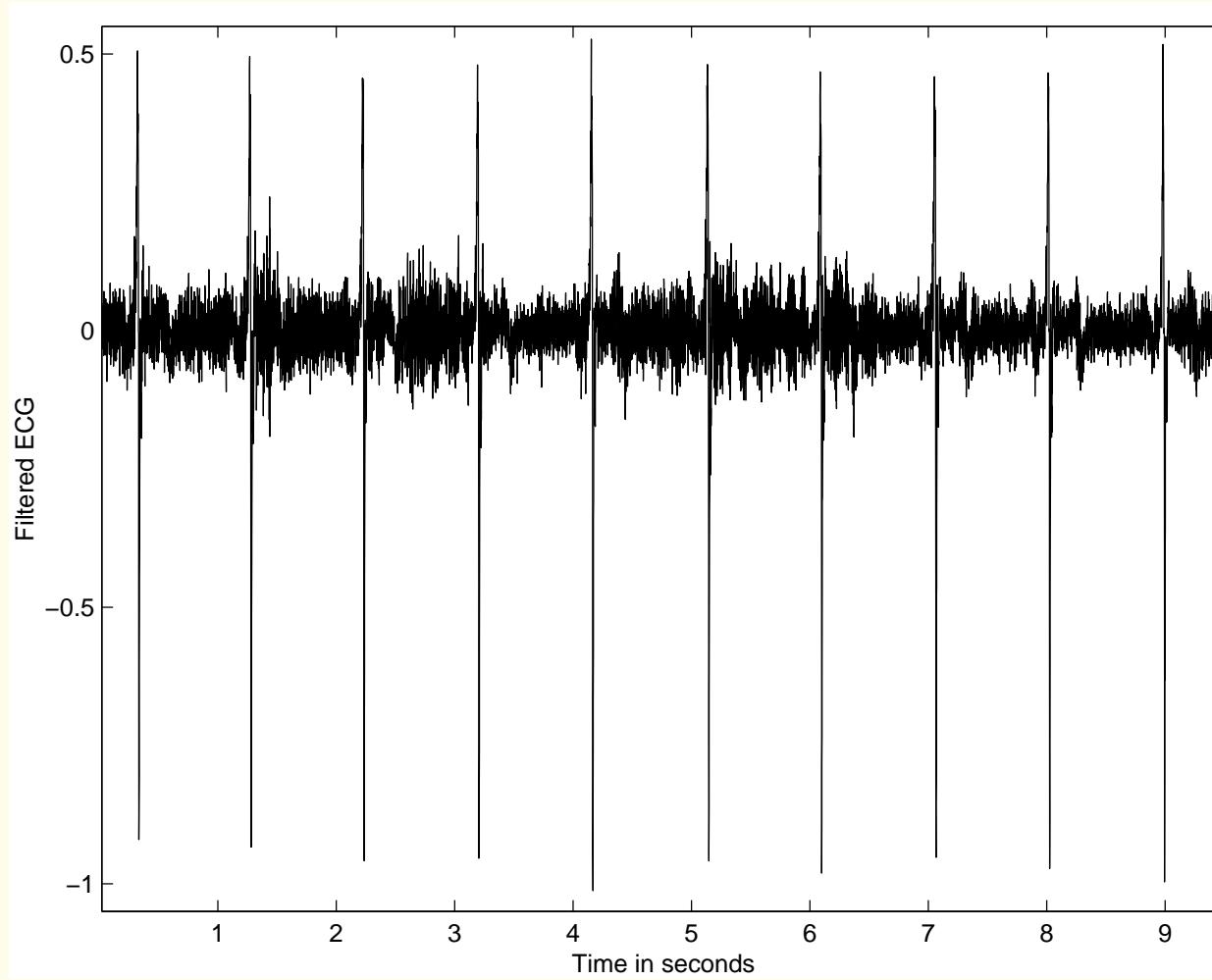


Figure 3.25: Result of filtering the ECG signal with low-frequency noise shown in Figure 3.6, using the three-point central-difference operator.



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

Solution: Drawback of the first-order difference and the three-point central-difference operators — magnitude response remains low for a wide range of frequencies.

Need gain of 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 related locations around the unit circle.

Stability: poles should be within the unit circle.

To maintain high gain at low frequencies,

place a pole on the real axis (zero frequency),

close to $z = 1$, e.g., $z = 0.995$:



$$H(z) = \frac{1}{T} \left[\frac{1 - z^{-1}}{1 - 0.995 z^{-1}} \right], \quad (3.47)$$

$$H(z) = \frac{1}{T} \left[\frac{z - 1}{z - 0.995} \right]. \quad (3.48)$$

$$y(n) = \frac{1}{T} [x(n) - x(n-1)] + 0.995 y(n-1). \quad (3.49)$$

Infinite impulse response filter!

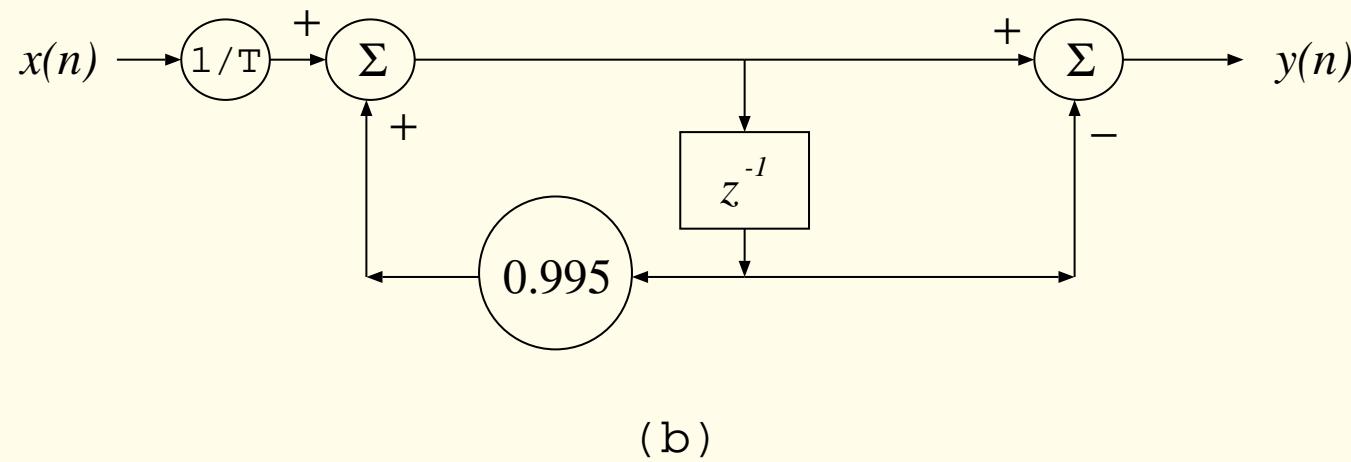
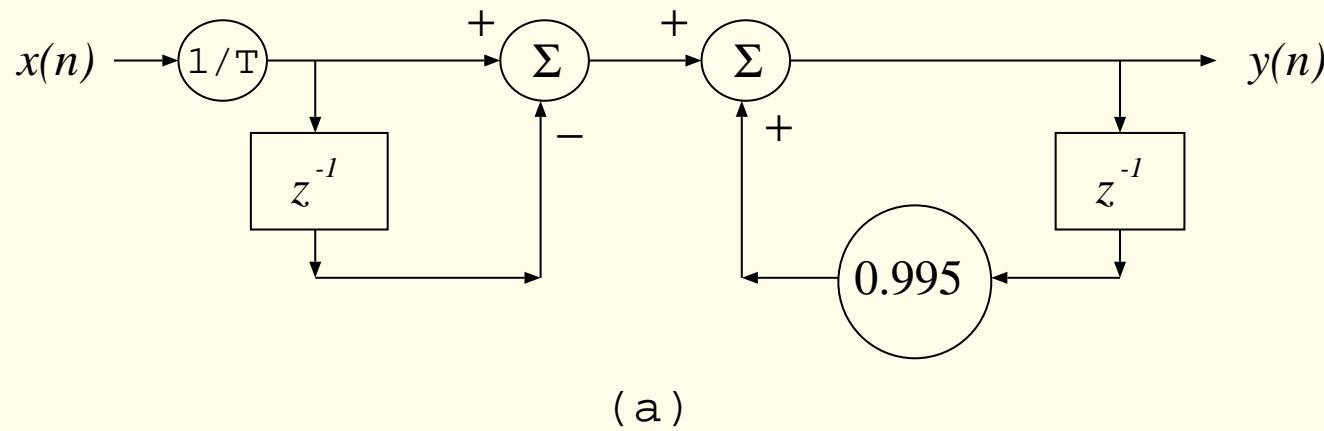


Figure 3.26: Two equivalent signal-flow diagrams of the filter to remove low-frequency noise or base-line wander. The form in (a) uses two delays, whereas that in (b) uses only one delay.



Graphical method for evaluation of response:

Evaluate transfer function at various points on the unit circle

in the z -plane, by letting $z = \exp(j\omega)$,

and evaluating $H(z)$ for various values of the frequency.

Magnitude transfer function $|H(z)| =$

product of the distances from z to all zeros of the system

divided by the product of the distances to the poles.



Phase response =

sum of the angles of the vectors from z to all the zeros

minus the sum of the angles to the poles.

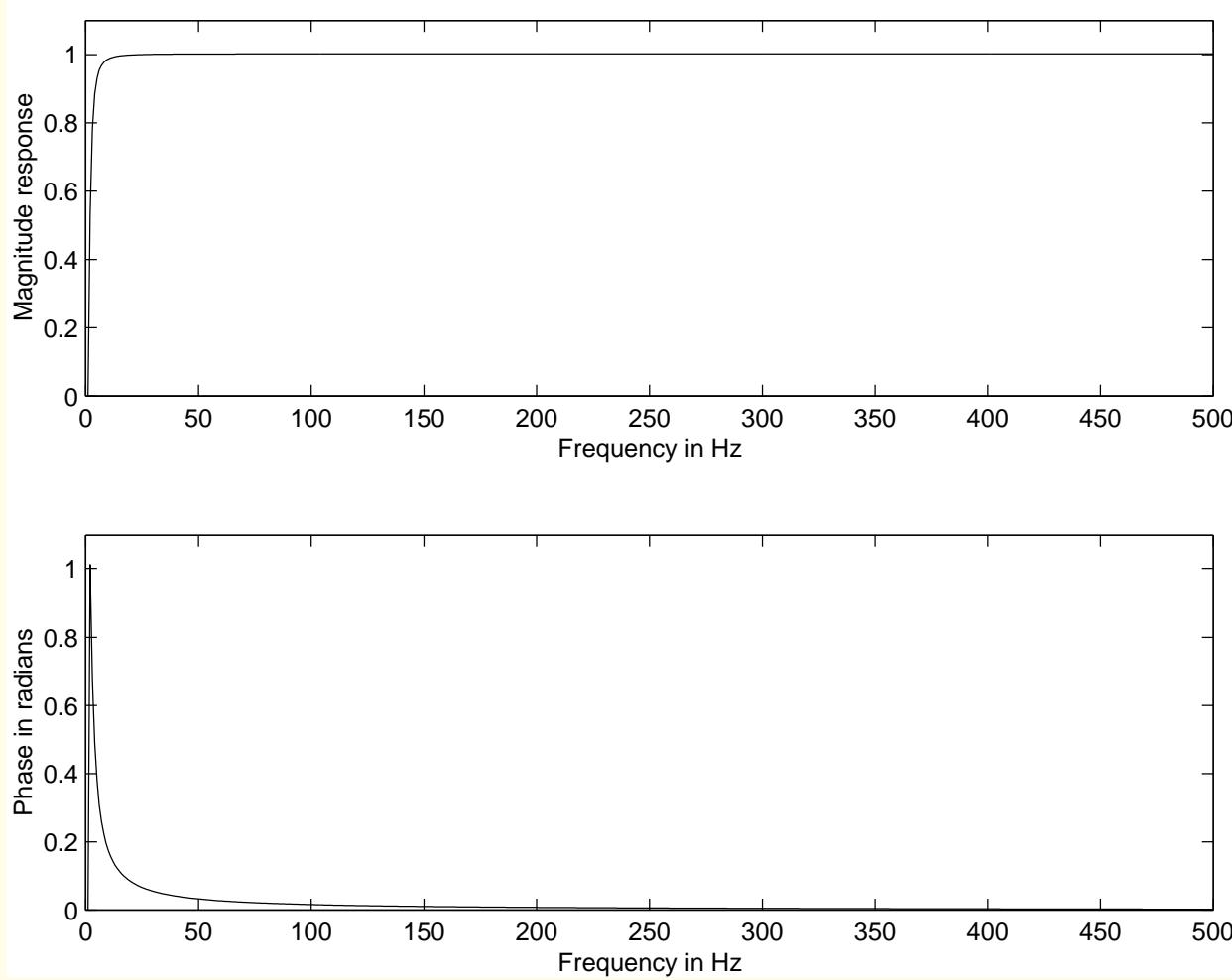


Figure 3.27: Normalized magnitude and phase responses of the filter to remove base-line wander as in Equation 3.47. The magnitude response is shown on a linear scale.

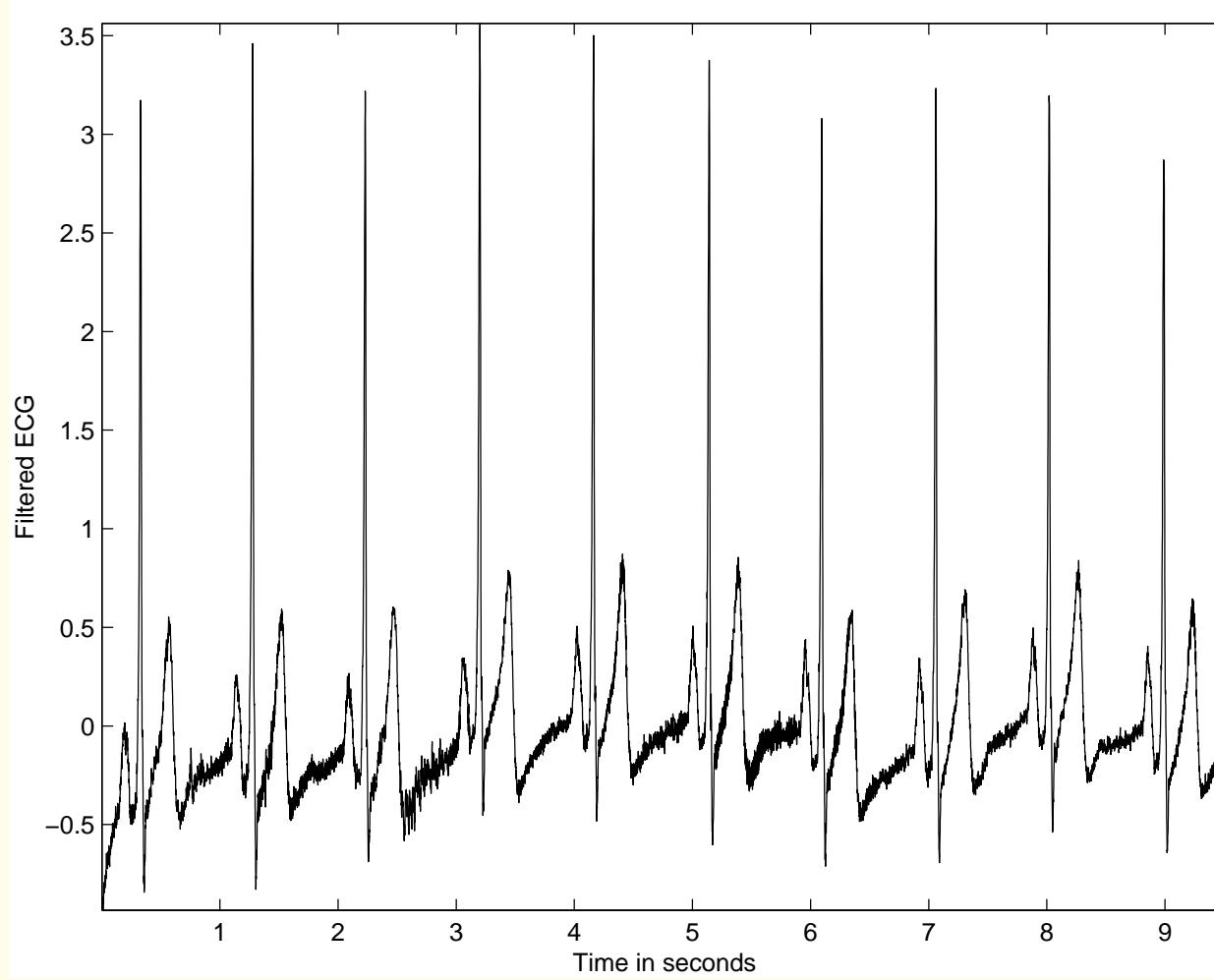


Figure 3.28: Result of processing the ECG signal with low-frequency noise shown in Figure 3.6, using the filter to remove base-line wander as in Equation 3.47. (Compare with the results in Figures 3.24 and 3.25.)



3.4 Frequency-domain Filters

Filters may be designed in the frequency domain to provide specific lowpass, highpass, bandpass, or band-reject (notch) characteristics.

Implemented in software after obtaining FT of input signal, or converted into equivalent time-domain filters.



Most commonly used designs:

Butterworth, Chebyshev, elliptic, and Bessel filters.

Well-established in the analog-filter domain:

commence with an analog design $H(s)$ and apply the

bilinear transformation to obtain a digital filter $H(z)$.



Design a lowpass filter with the desired pass-band, transition, and stop-band characteristics on a normalized-frequency axis, and then transform to the desired lowpass, highpass, bandpass, or band-reject filter.



Frequency-domain filters may also be specified directly

in terms of the values of the desired frequency response

at certain frequency samples only,

then transformed into the equivalent time-domain

filter coefficients via the inverse Fourier transform.



3.4.1 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 pass-band.*

Solution: Butterworth filter —

maximally flat magnitude response in the pass-band.



Butterworth lowpass filter of order N :

first $2N - 1$ derivatives of squared magnitude response

are zero at $\Omega = 0$,

where Ω = analog radian frequency.

Butterworth filter response is monotonic in the pass-band

as well as in the stop-band.



Basic Butterworth lowpass filter function:

$$|H_a(j\Omega)|^2 = \frac{1}{1 + \left(\frac{j\Omega}{j\Omega_c}\right)^{2N}}, \quad (3.50)$$

where H_a is the frequency response of the analog filter and

Ω_c is the cutoff frequency in *radians/s*.

Butterworth filter completely specified by

cutoff frequency Ω_c and order N .



As the order N increases, the filter response becomes

more flat in the pass-band, and

transition to the stop-band becomes faster or sharper.

$$|H_a(j\Omega_c)|^2 = \frac{1}{2} \text{ for all } N.$$



Changing to the Laplace variable s ,

$$H_a(s)H_a(-s) = \frac{1}{1 + \left(\frac{s}{j\Omega_c}\right)^{2N}}. \quad (3.51)$$

Poles of squared transfer function located with

equal spacing around a circle of radius Ω_c in the s -plane,

distributed symmetrically on either side of

imaginary axis $s = j\Omega$.



No pole on imaginary axis;

poles on real axis for odd N .

Angular spacing between 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.

For the filter coefficients to be real,

complex poles must appear in conjugate pairs.



To obtain a stable and causal filter, form $H_a(s)$ with

only the N poles on the left-hand side of the s -plane.

Pole positions in the s -plane given by

$$s_k = \Omega_c \exp \left[j\pi \left(\frac{1}{2} + \frac{(2k-1)}{2N} \right) \right], \quad (3.52)$$

$$k = 1, 2, \dots, 2N.$$



Once pole positions obtained in the s -plane,

derive the transfer function in the analog Laplace domain:

$$H_a(s) = \frac{G}{(s - p_1)(s - p_2)(s - p_3) \cdots (s - p_N)}, \quad (3.53)$$

where p_k , $k = 1, 2, \dots, N$, are the N poles

in the left-half of the s -plane,

and G is a gain factor specified as needed

or to normalize the gain at DC ($s = 0$) to be unity.



Bilinear transformation (BLT):

$$s = \frac{2}{T} \left[\frac{1 - z^{-1}}{1 + z^{-1}} \right]. \quad (3.54)$$

Butterworth circle in the s -plane maps to a

circle in the z -plane

with real-axis intercepts at $z = \frac{2-\Omega_c T}{2+\Omega_c T}$ and $z = \frac{2+\Omega_c T}{2-\Omega_c T}$.



Poles at $s = s_p$ and $s = -s_p$ in the s -plane map to

$z = z_p$ and $z = 1/z_p$.

Poles in the z -plane 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.



Unit circle in the z -plane: $z = e^{j\omega}$.

For points on the unit circle, we have

$$s = \sigma + j\Omega = \frac{2}{T} \left(\frac{1 - e^{-j\omega}}{1 + e^{-j\omega}} \right) = \frac{2j}{T} \tan \left(\frac{\omega}{2} \right). \quad (3.55)$$



For the unit circle, $\sigma = 0$; therefore,

continuous-time (s -domain) frequency variable Ω

related to discrete-time (z -domain) frequency variable ω as

$$\Omega = \frac{2}{T} \tan\left(\frac{\omega}{2}\right) \quad (3.56)$$

and

$$\omega = 2 \tan^{-1}\left(\frac{\Omega T}{2}\right). \quad (3.57)$$



Nonlinear relationship warps frequency values:

mapped from the imaginary (vertical) axis in the s -plane

to the unit circle in the z -plane —

should be taken into account in specifying

cutoff frequencies.



$H_a(s)$ mapped to the z -domain by the BLT: $s = \frac{2}{T} \frac{1-z^{-1}}{1+z^{-1}}$.

$H(z)$ simplified to

$$H(z) = \frac{G'}{\sum_{k=0}^N a_k z^{-k}} (1 + z^{-1})^N, \quad (3.58)$$

$a_k, k = 0, 1, 2, \dots, N$: filter coefficients or tap weights;

$a_0 = 1$;

G' : gain, normalized so that $|H(z)| = 1$ at DC ($z = 1$).

IIR filter; N zeros at $z = -1$ due to the BLT.

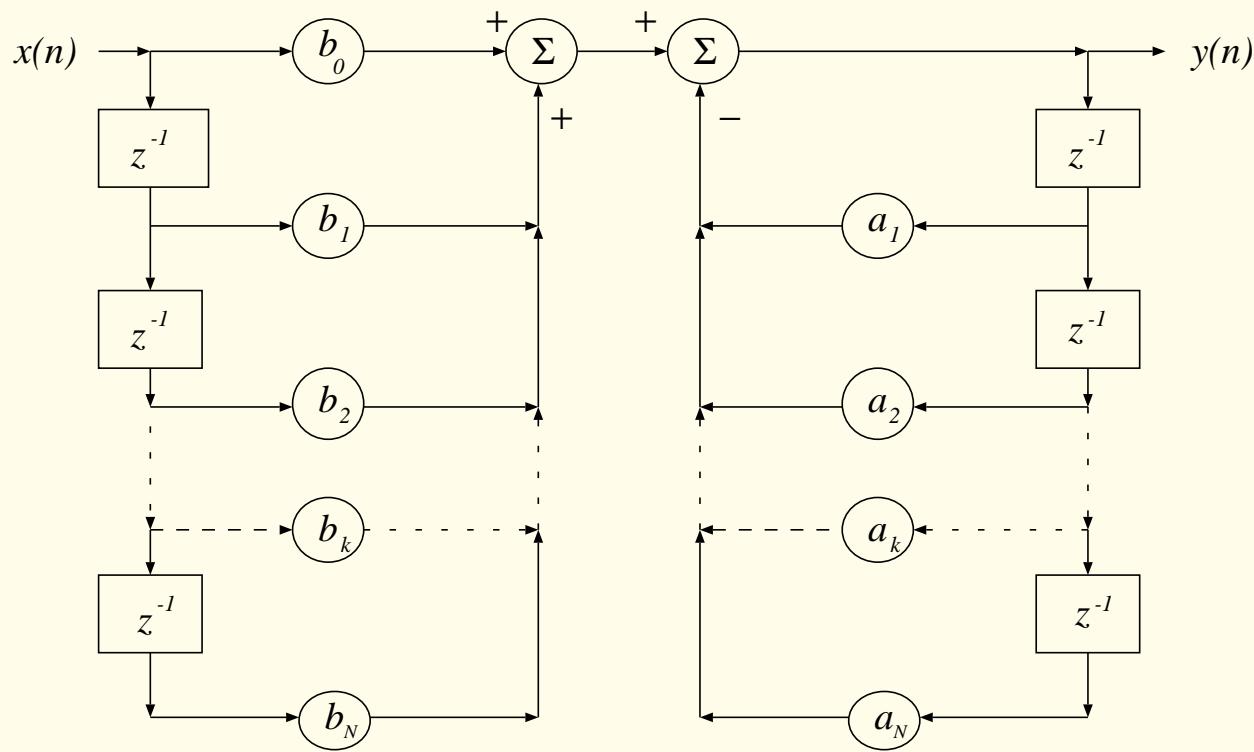


Figure 3.29: Signal-flow diagram of a direct realization of a generic infinite impulse response (IIR) filter. This form uses $2N$ delays and $2N + 1$ multipliers for a filter of order N .

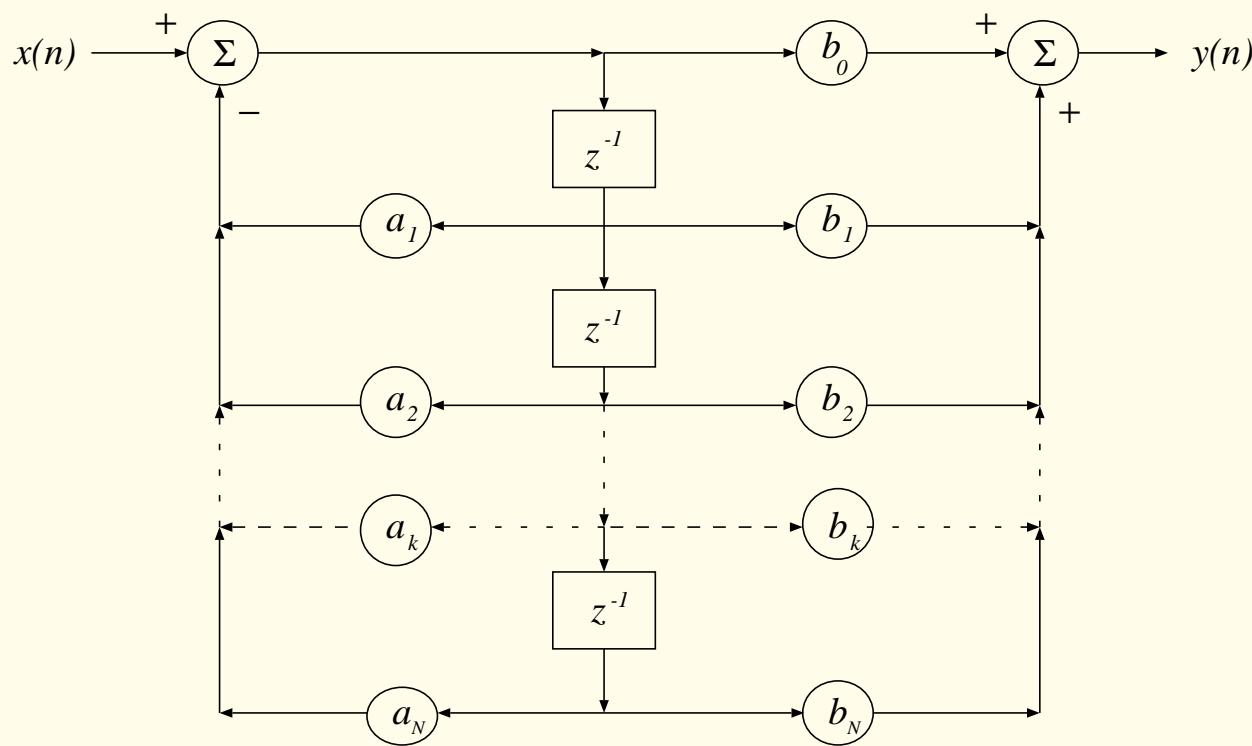


Figure 3.30: Signal-flow diagram of a realization of an IIR filter that uses only N delays and $(2N + 1)$ multipliers for a filter of order N .



Time-domain representation required if

filter applied to data samples directly in time domain.

From transfer function $H(z)$ in Equation 3.58:

$$y(n) = \sum_{k=0}^N b_k x(n-k) - \sum_{k=1}^N a_k y(n-k). \quad (3.59)$$

Coefficients b_k given by coefficients of expansion

of $G'(1 + z^{-1})^N$.

MATLAB command *butter*.



Also possible to specify the Butterworth filter directly as

$$|H(\omega)|^2 = \frac{1}{1 + \left(\frac{\omega}{\omega_c}\right)^{2N}}, \quad (3.60)$$

ω normalized to $(0, 2\pi)$ for sampled signals;

equation valid only for the range $(0, \pi)$;

function over $(\pi, 2\pi)$ = reflection of function over $(0, \pi)$.

Cutoff frequency ω_c should be specified in the range $(0, \pi)$.



If the discrete Fourier transform (DFT) is used:

$$|H(k)|^2 = \frac{1}{1 + \left(\frac{k}{k_c}\right)^{2N}}, \quad (3.61)$$

k : index of DFT array for discretized frequency.

K : number of points in DFT array,

k_c index corresponding to cutoff frequency ω_c , $k_c = K \frac{\omega_c}{\omega_s}$;

Equation valid for $k = 0, 1, 2, \dots, \frac{K}{2}$;

$H(k) = H(K - k)$, $k = \frac{K}{2} + 1, \dots, K - 1$.



Note: DFT includes two unique values — DC in $H(0)$ and folding-frequency component in $H(\frac{K}{2})$.

k : also used to represent normalized frequency in $(0, 1)$;

unity: sampling frequency;

0.5: maximum frequency in sampled signal

(folding frequency);

k_c specified in the range $(0, 0.5)$.



Note: MATLAB normalizes

half the sampling frequency to unity;

maximum normalized frequency in sampled signal = 1.

MATLAB and some other languages do not allow

array index = 0:

the indices mentioned above must be incremented by one.



Filtering in the frequency domain using the DFT:

Compute the DFT of the given signal,

multiply the result by $|H(k)|$, and

compute the inverse DFT to obtain the filtered signal.

Note: Pay attention to shifting or folding of spectrum.

Advantage: no phase change is involved.



Filter is a strictly magnitude-only transfer function.

Time-domain implementation required for

on-line real-time applications.



Butterworth lowpass filter design example:

Specify two parameters: ω_c and N ,

based on knowledge of the filter, signal, and noise.

Also possible to specify required minimum gain

at a certain frequency in the pass-band and

minimum attenuation at a frequency in the stop-band.

With Equation 3.50, obtain two equations;

solve for the filter parameters ω_c and N .



Given the 3 dB cutoff frequency f_c and order N :

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 3.56 and obtain Ω_c .
2. Derive the positions of the poles of the filter in the s -plane as given by Equation 3.52.
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 3.53.



4. Apply the bilinear transformation as per Equation 3.54 and obtain the transfer function $H(z)$ as in Equation 3.58.
5. Convert to the coefficients b_k and a_k as in Equation 3.59.



Design Butterworth lowpass filter with

$f_c = 40 \text{ Hz}$, $f_s = 200 \text{ Hz}$, and $N = 4$.

$$\omega_c = \frac{40}{200} 2\pi = 0.4\pi \text{ radians/s.}$$

Prewarped s -domain cutoff frequency is

$$\Omega_c = \frac{2}{T} \tan\left(\frac{\omega_c}{2}\right) = 1.453085 \text{ radians/s.}$$

Poles of $H_a(s)H_a(-s)$ placed around circle of

radius 1.453085

with angular separation of $\frac{\pi}{N} = \frac{\pi}{4} \text{ radians.}$



Poles of interest located at angles $\frac{5}{8}\pi$ and $\frac{7}{8}\pi$

and the corresponding conjugate positions.

Coordinates of the poles of interest:

$(-0.556072 \pm j 1.342475)$ and $(-1.342475 \pm j 0.556072)$.

$$H_a(s) = \quad (3.62)$$

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

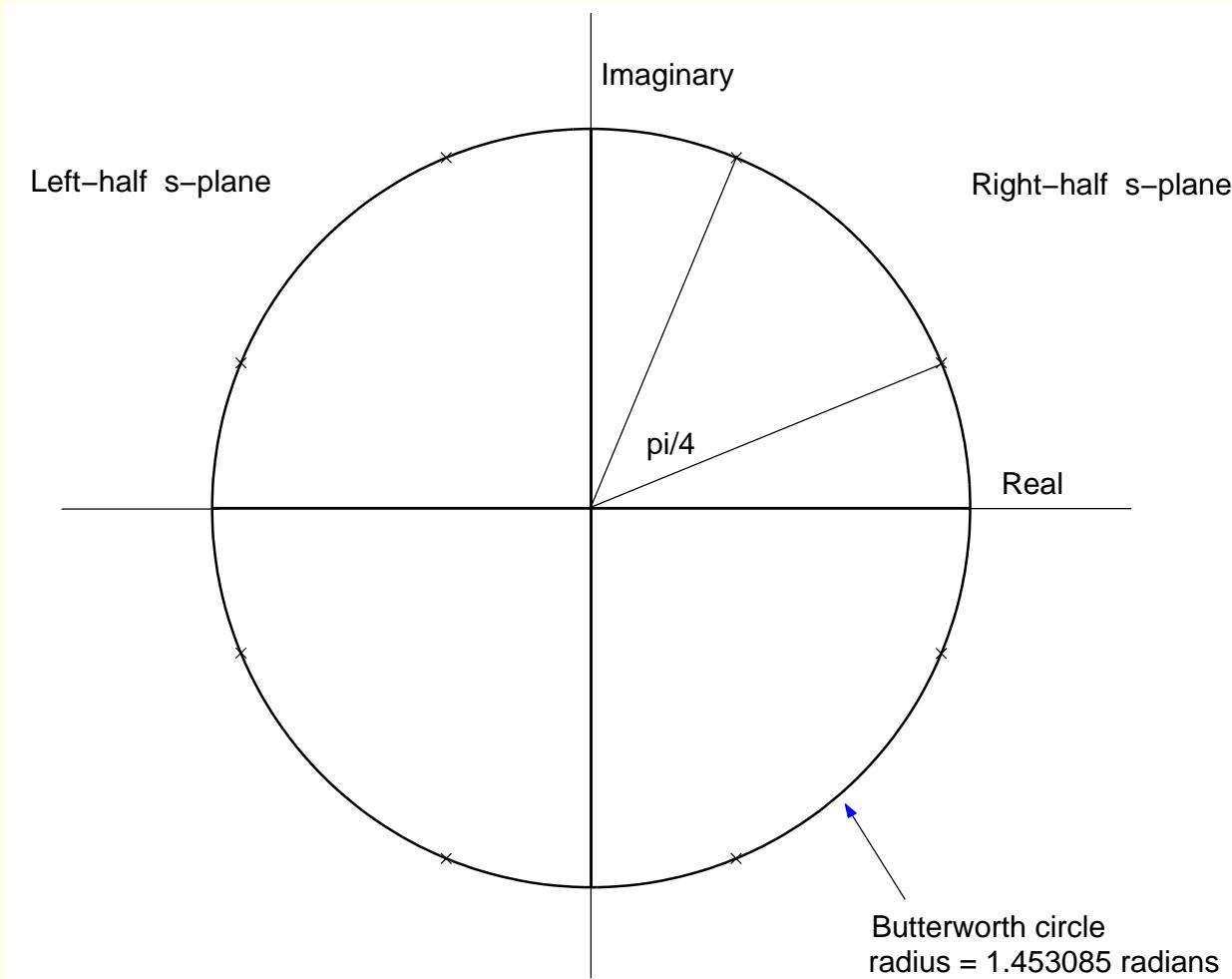


Figure 3.31: Pole positions in the s -plane of the squared magnitude response of the Butterworth lowpass filter with $f_c = 40 \text{ Hz}$, $f_s = 200 \text{ Hz}$, and $N = 4$.



Applying the bilinear transformation:

$$H(z) = \quad (3.63)$$

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

The filter has four poles at

$(0.223882 \pm j 0.640852)$ and $(0.164488 \pm j 0.193730)$,

and four zeros at $-1 + j 0$.



b_k coefficients as in Equation 3.59:

$$\{0.0465829, 0.186332, 0.279497, 0.186332, 0.046583\},$$

a_k coefficients:

$$\{1, -0.776740, 0.672706, -0.180517, 0.029763\}.$$

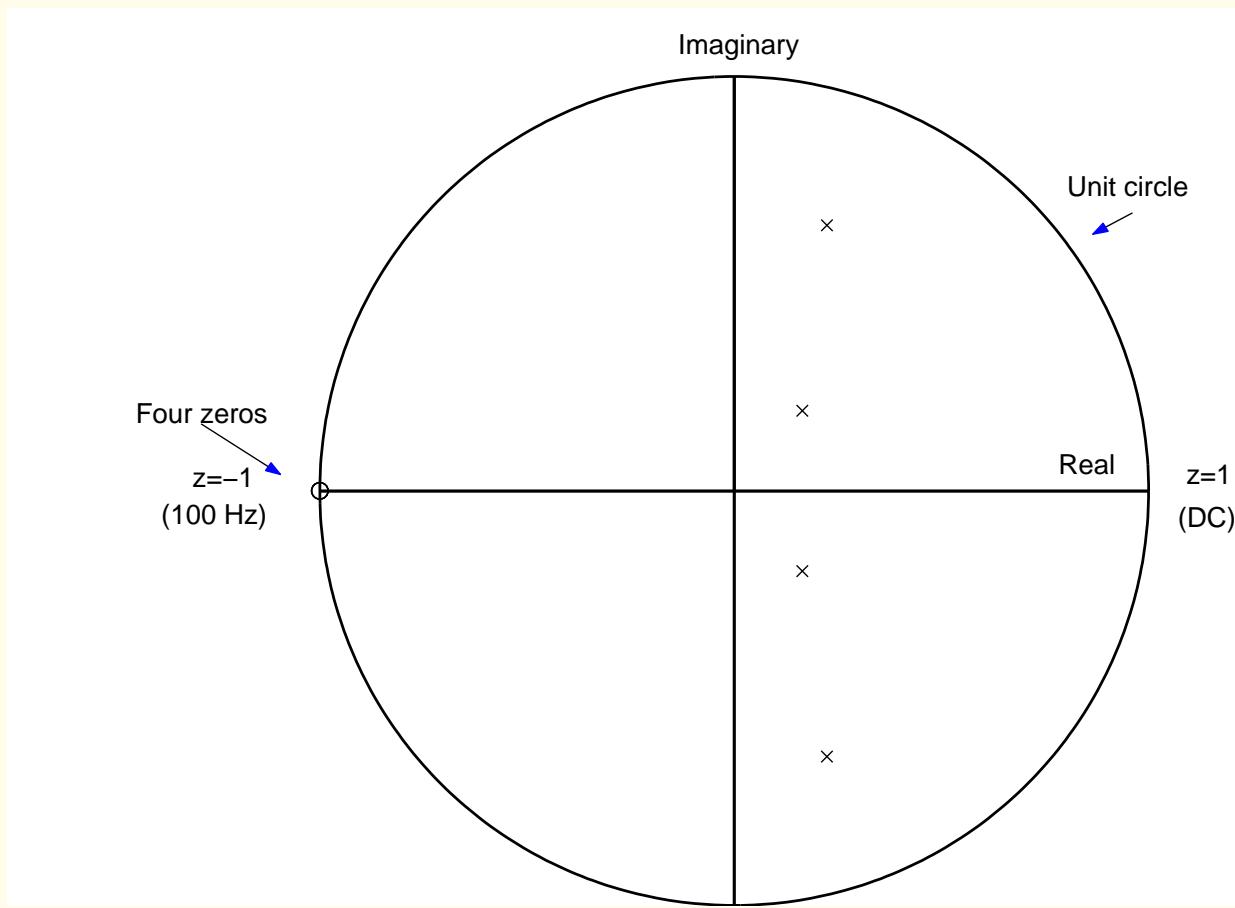


Figure 3.32: Positions of the poles and zeros in the z -plane of the Butterworth lowpass filter with $f_c = 40 \text{ Hz}$, $f_s = 200 \text{ Hz}$, and $N = 4$.

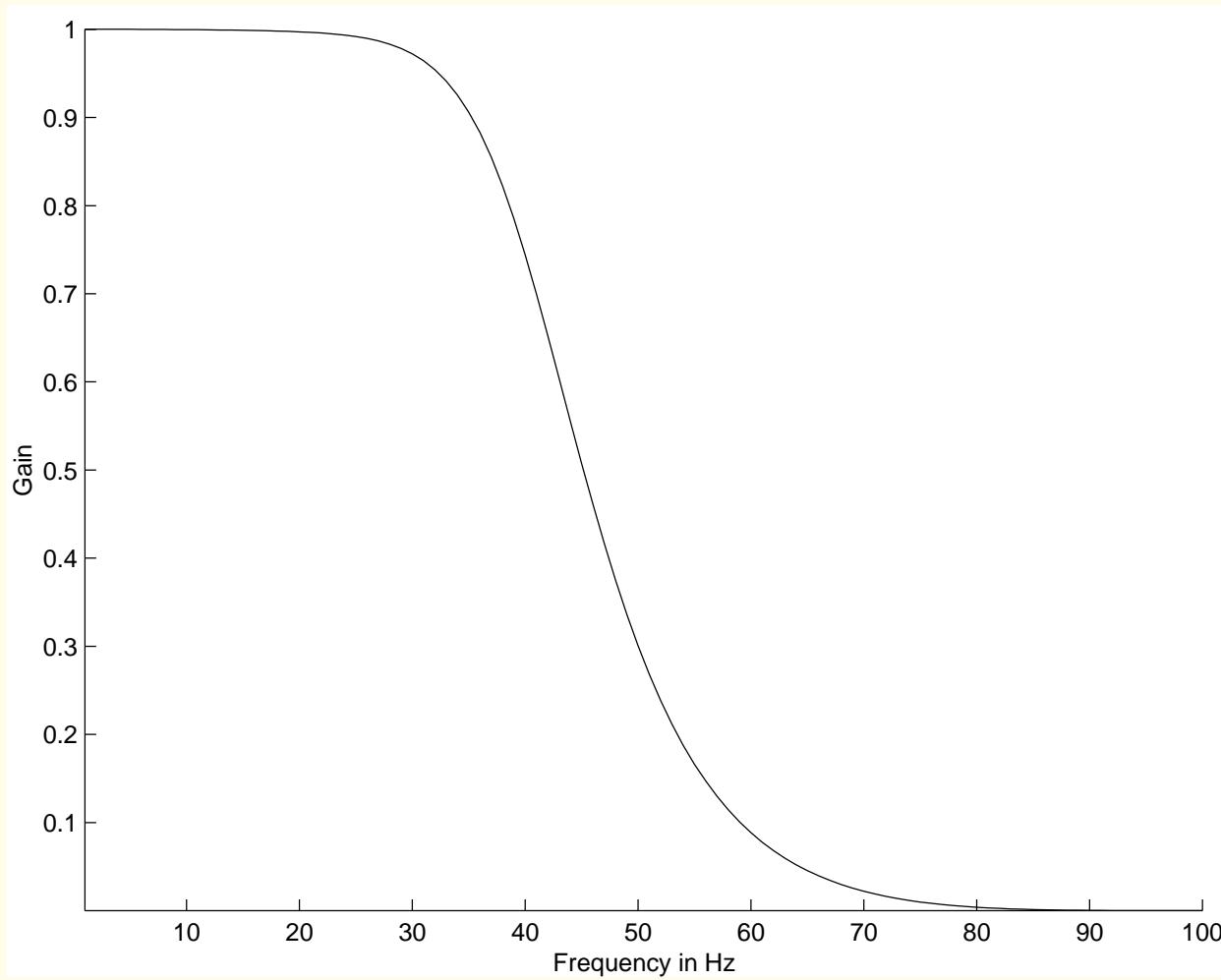


Figure 3.33: Magnitude response of the Butterworth lowpass filter with $f_c = 40 \text{ Hz}$, $f_s = 200 \text{ Hz}$, and $N = 4$.

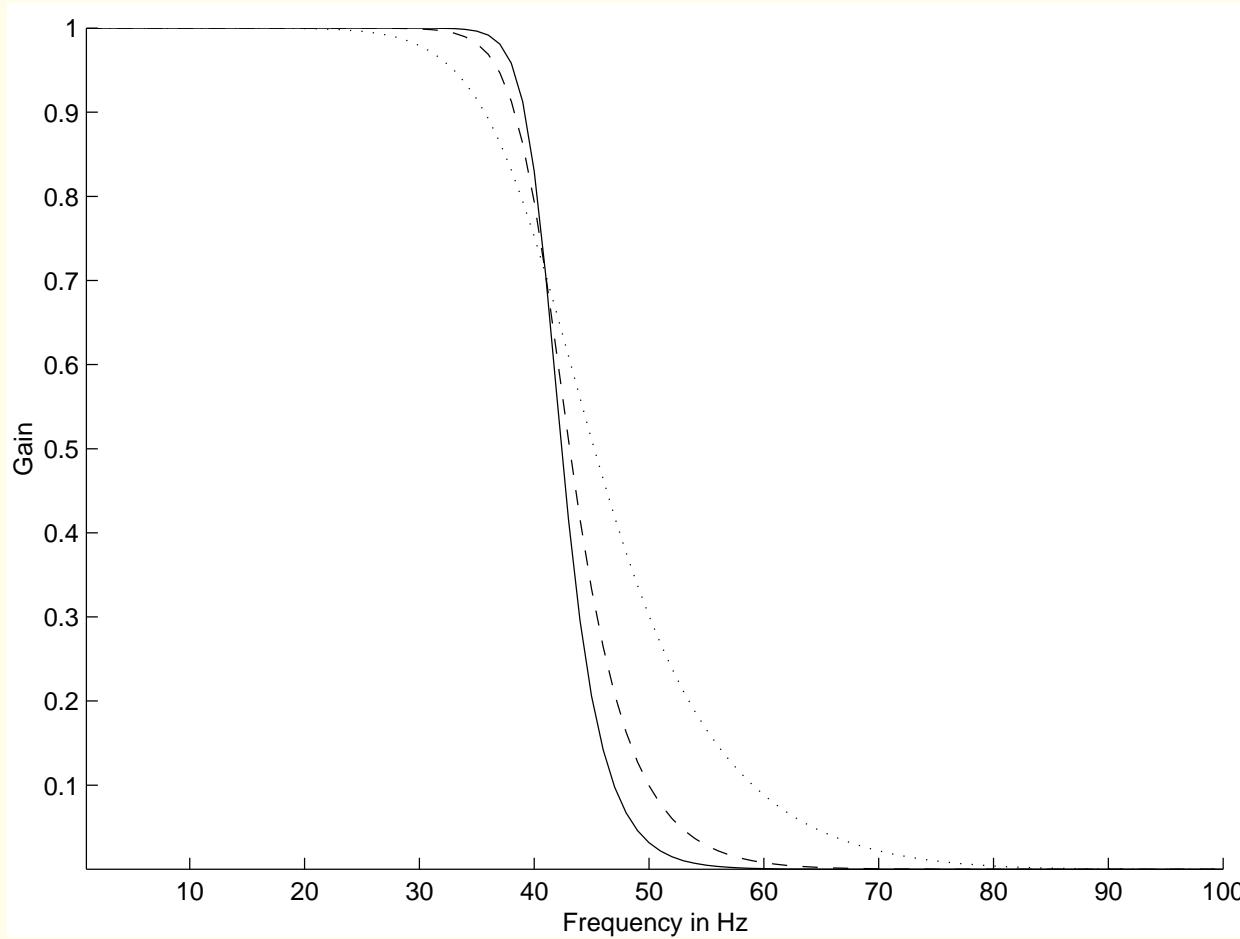


Figure 3.34: Magnitude responses of three Butterworth lowpass filters with $f_c = 40 \text{ Hz}$, $f_s = 200 \text{ Hz}$, and variable order: $N = 4$ (dotted), $N = 8$ (dashed), and $N = 12$ (solid).

All three filters: half power (gain = 0.707) at 40 Hz;

transition band becomes sharper as the order N is increased.

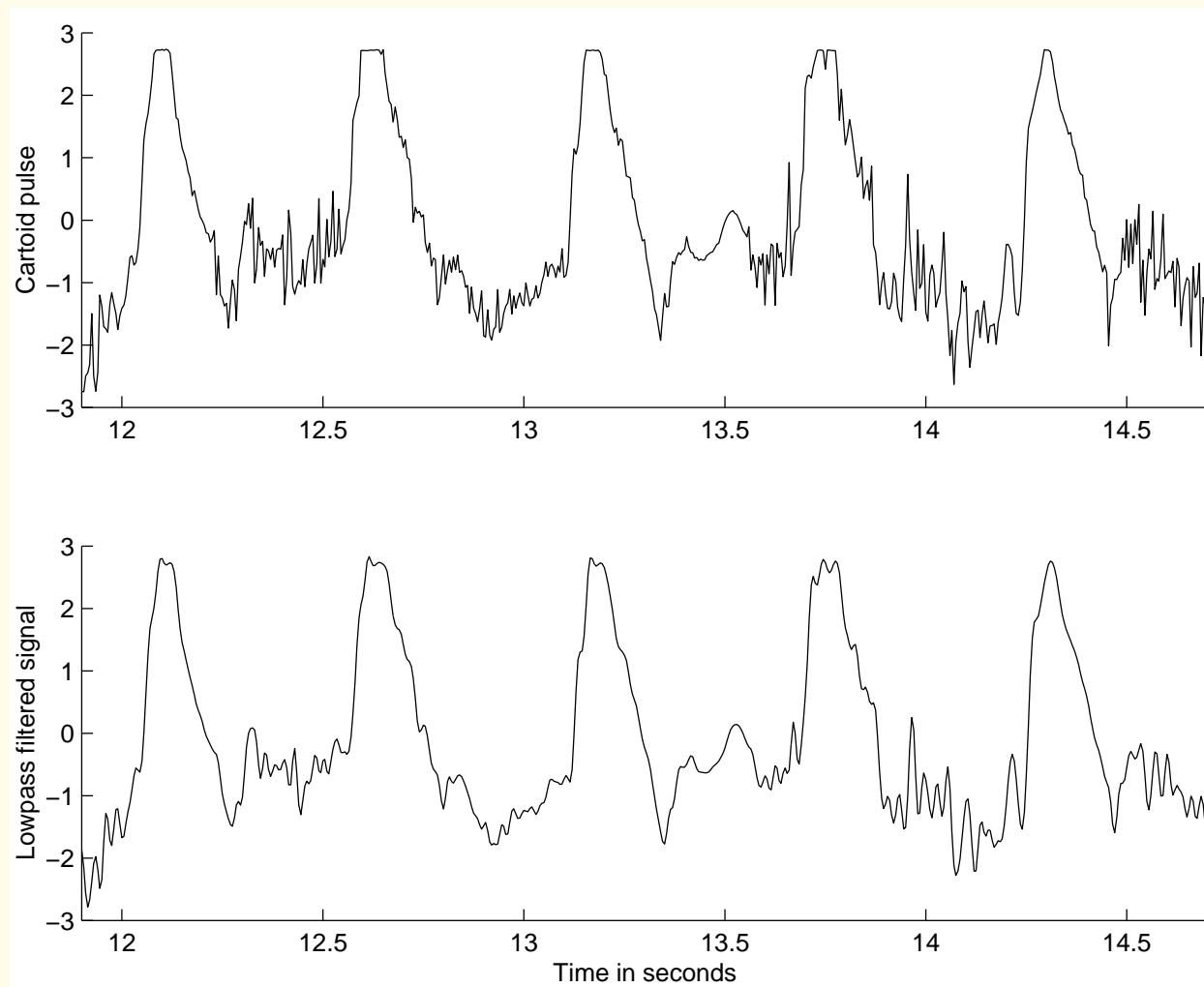


Figure 3.35: Upper trace: a carotid pulse signal with high-frequency noise and effects of clipping. Lower trace: result of filtering with a Butterworth lowpass filter with $f_c = 40 \text{ Hz}$, $f_s = 200 \text{ Hz}$, and $N = 4$. The filtering operation was performed in the time domain using the MATLAB `filter` command.

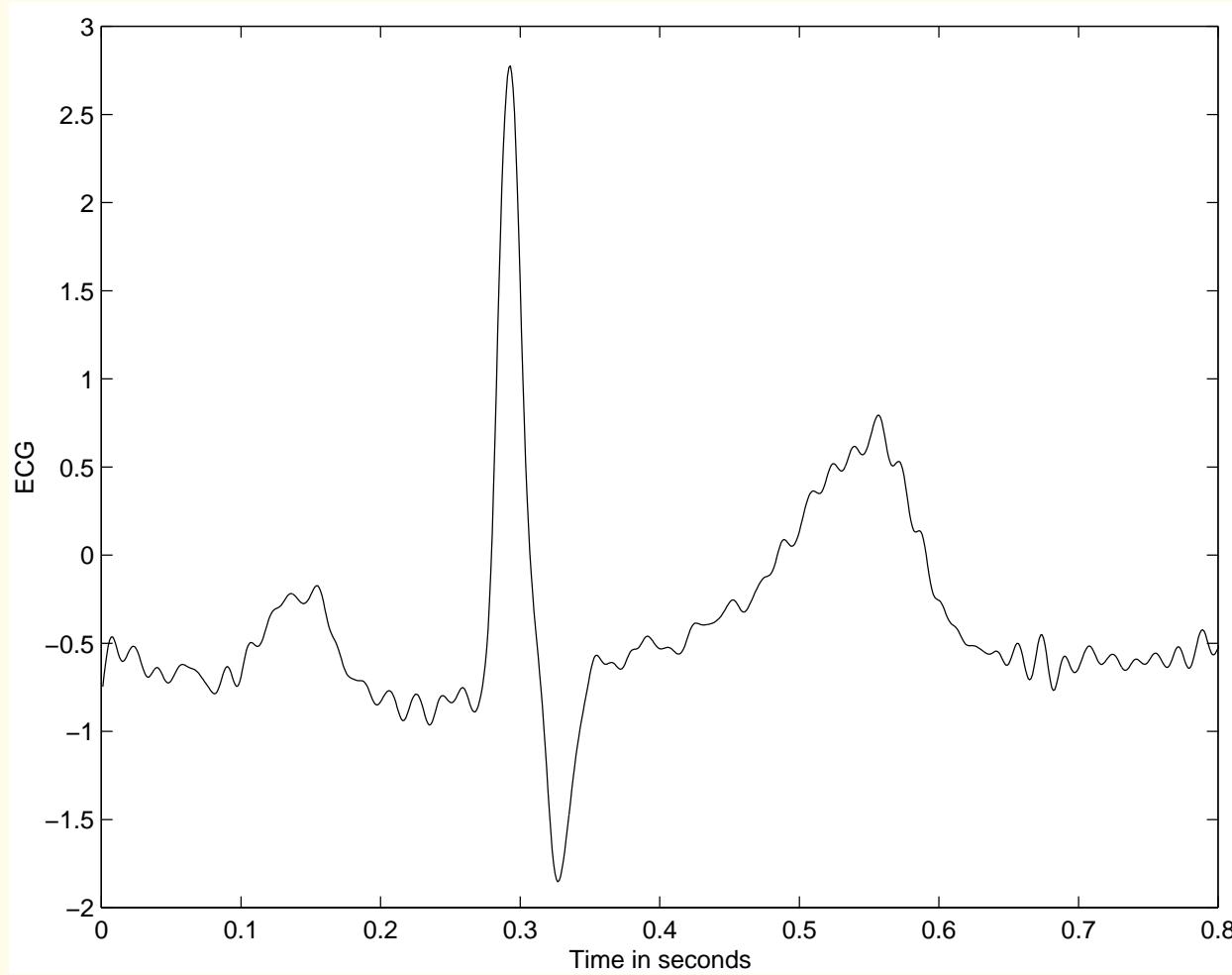


Figure 3.36: Result of frequency-domain filtering of the noisy ECG signal in Figure 3.20 with an eighth-order Butterworth lowpass filter with cutoff frequency = 70 Hz .

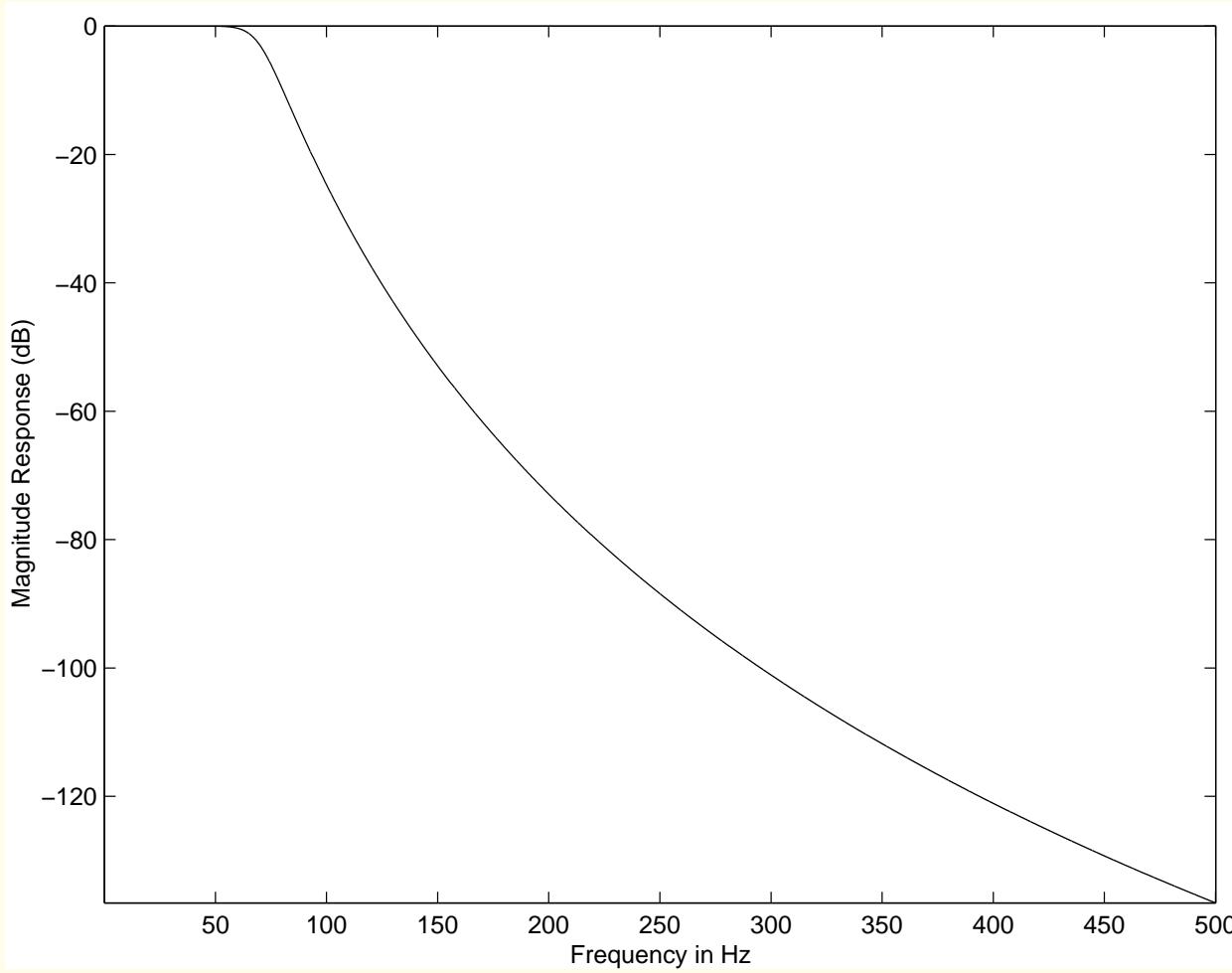


Figure 3.37: Frequency response of the eighth-order Butterworth lowpass filter with cutoff frequency = $f_c = 70 \text{ Hz}$ and $f_s = 1,000 \text{ Hz}$.



3.4.2 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 pass-band.



Solution:

Butterworth highpass filter specified directly

in the discrete-frequency domain as

$$|H(k)|^2 = \frac{1}{1 + \left(\frac{k_c}{k}\right)^{2N}}. \quad (3.64)$$

Illustration of application:

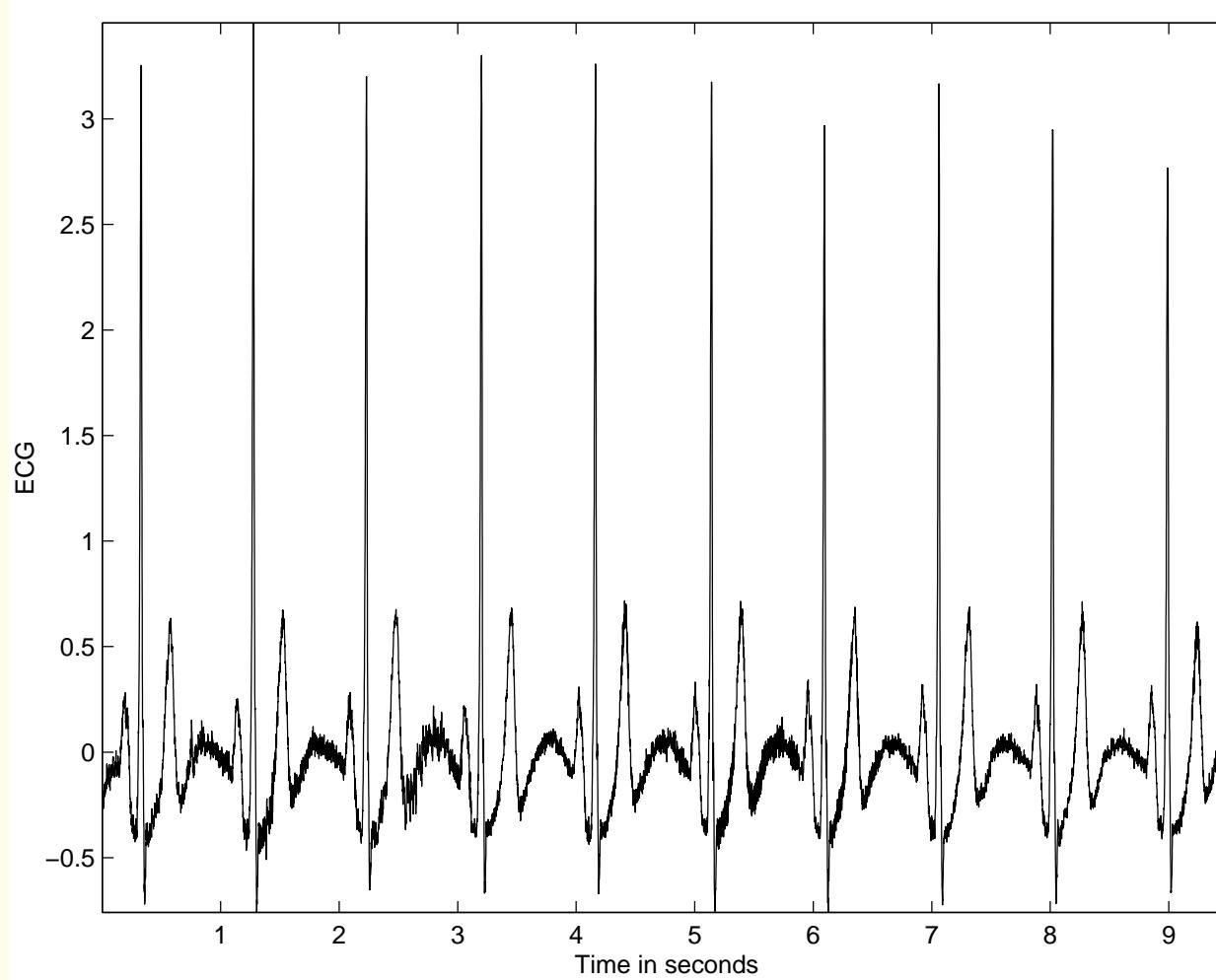


Figure 3.38: Result of frequency-domain filtering of the ECG signal with low-frequency noise in Figure 3.6 with an eighth-order Butterworth highpass filter with cutoff frequency = 2 Hz. (Compare with the results in Figures 3.24, 3.25, and 3.28.)

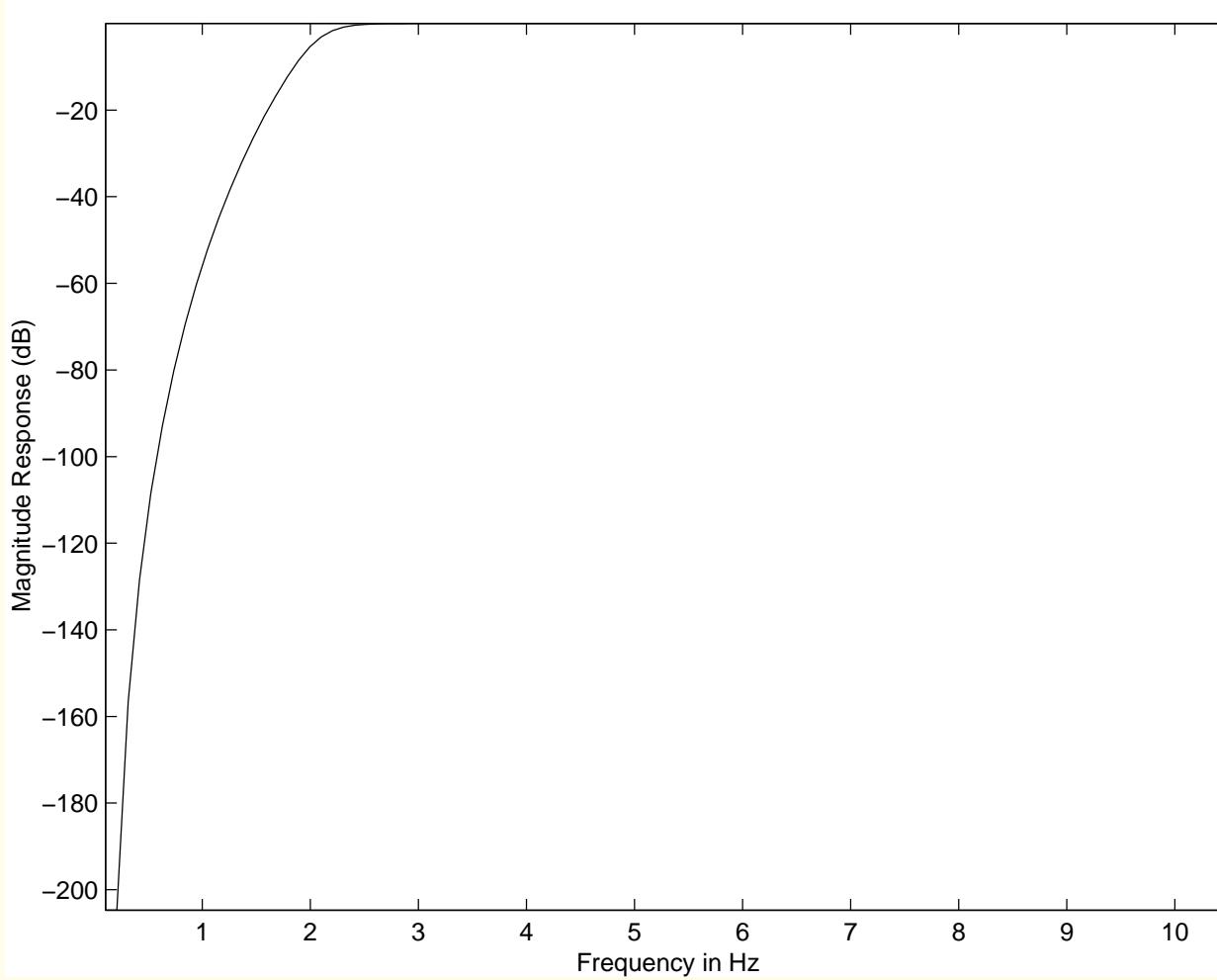


Figure 3.39: Frequency response of an eighth-order Butterworth highpass filter with cutoff frequency = 2 Hz. $f_s = 1,000$ Hz. The frequency response is shown on an expanded scale for the range 0 – 10 Hz only.



3.4.3 Removal of periodic artifacts: Notch and comb filters

Problem: *Design a frequency-domain filter to remove*

periodic artifacts such as power-line interference.

Solution: Periodic interference may be removed by

notch or comb filters with zeros on the unit circle

in the z -domain at the specific frequencies to be rejected.



f_o : interference frequency.

Angles of complex conjugate zeros required: $\pm \frac{f_o}{f_s}(2\pi)$.

Radius of the zeros: unity.

If harmonics present, multiple zeros required at $\pm \frac{n f_o}{f_s}(2\pi)$,

n : orders of all harmonics present.



Notch filter design example:

Power-line interference at $f_o = 60 \text{ Hz}$; $f_s = 1,000 \text{ Hz}$.

Notch filter zeros at

$$\omega_o = \pm \frac{f_o}{f_s} (2\pi) = \pm 0.377 \text{ radians} = \pm 21.6^\circ.$$

Zero locations: $\cos(\omega_o) \pm j \sin(\omega_o)$ or

$$z_1 = 0.92977 + j0.36812 \text{ and } z_2 = 0.92977 - j0.36812.$$



$$\begin{aligned} H(z) &= (1 - z^{-1} z_1)(1 - z^{-1} z_2) \\ &= 1 - 1.85955z^{-1} + z^{-2}. \end{aligned} \quad (3.65)$$

For $|H(1)| = 1$, divide $H(z)$ above by 0.14045.

Sharpness of notch may be improved by placing a few poles
symmetrically around the zeros and inside the unit circle.

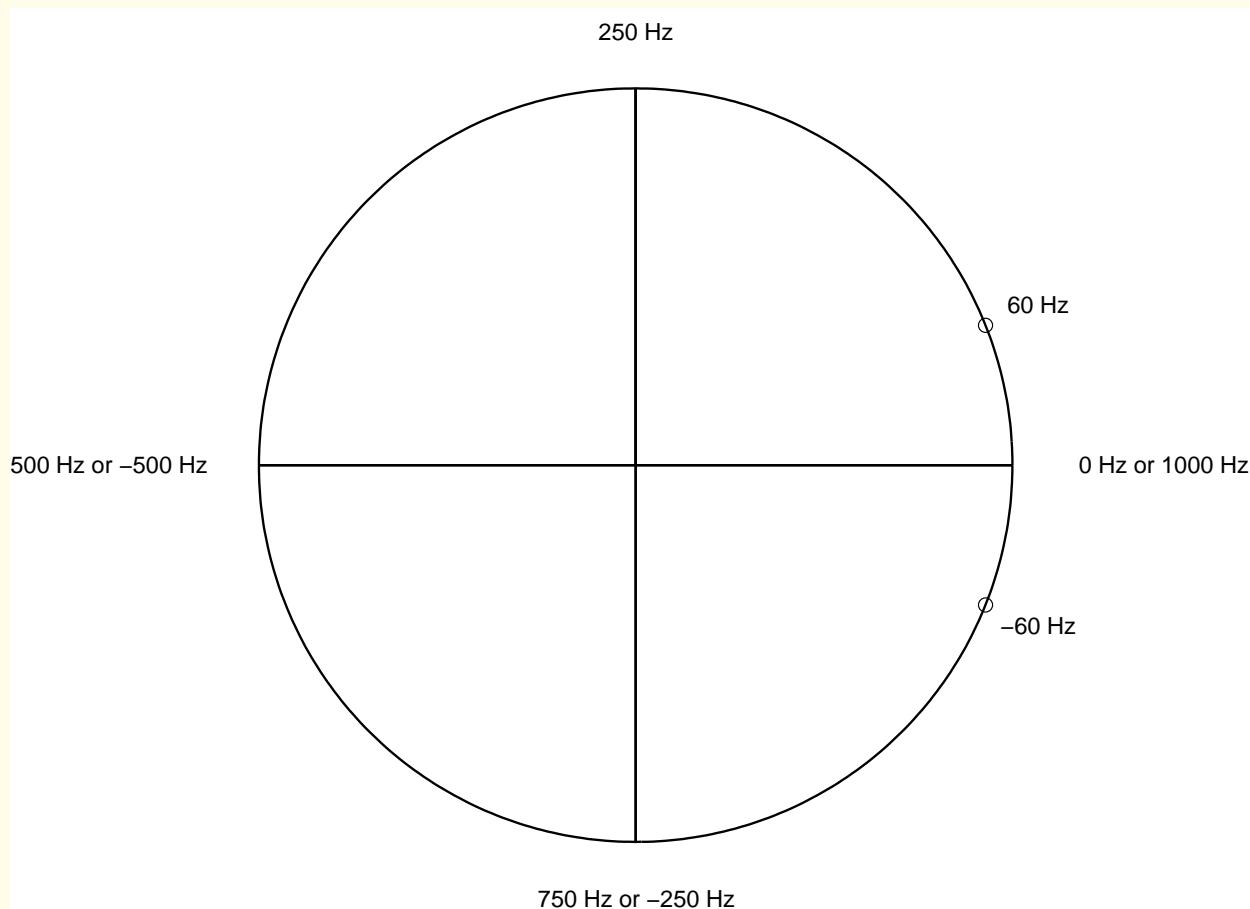


Figure 3.40: Zeros of the notch filter to remove 60 Hz interference, the sampling frequency being $1,000 \text{ Hz}$.

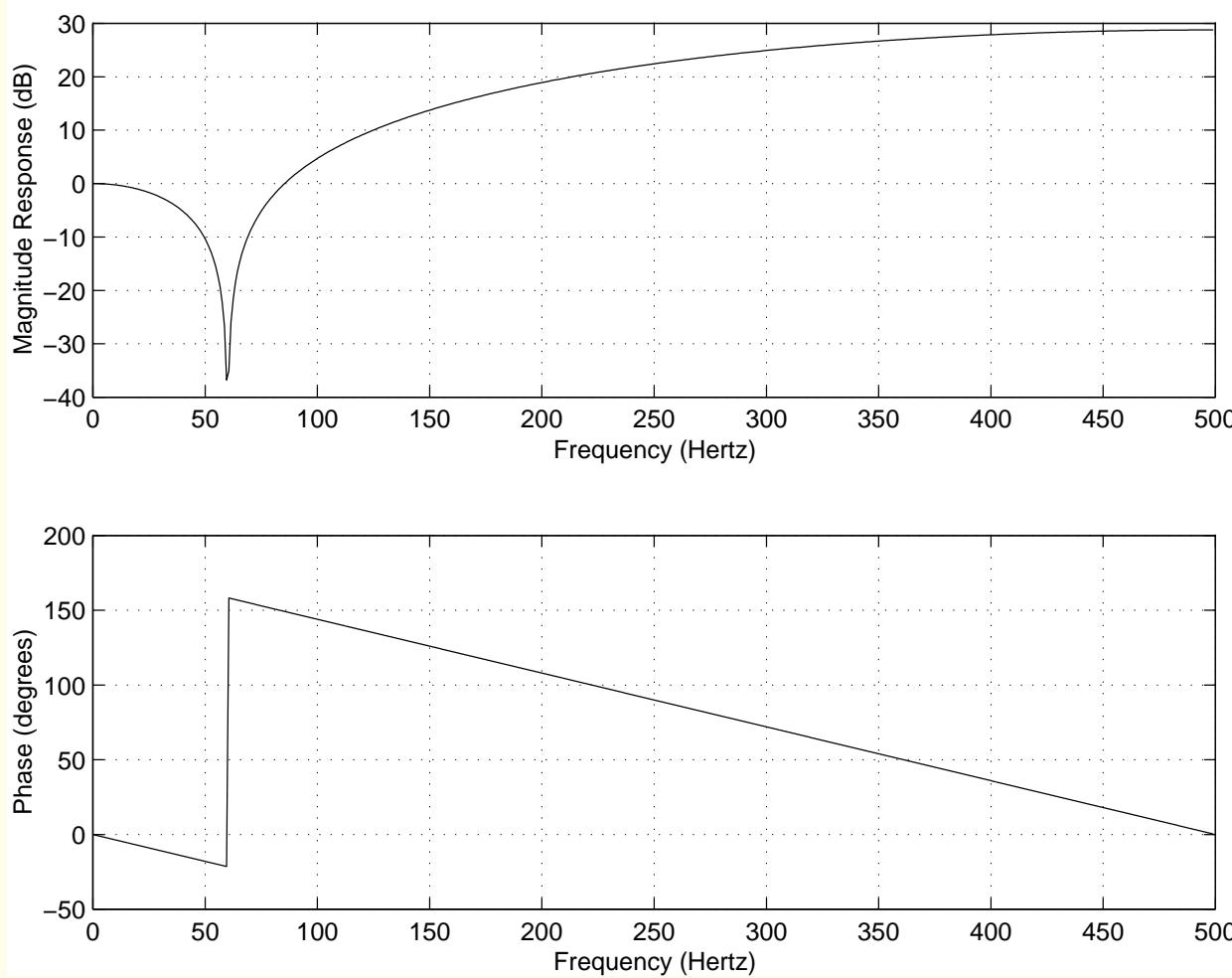


Figure 3.41: Magnitude and phase responses of the 60 Hz notch filter with zeros as shown in Figure 3.40.
 $f_s = 1,000 \text{ Hz}$.

Note: Large gain at $f_s/2$!



Comb filter design example:

Periodic artifact with fundamental frequency 60 Hz and

odd harmonics at 180 Hz , 300 Hz , and 420 Hz .

Zeros desired at 60 Hz , 180 Hz , 300 Hz , and 420 Hz ,

or $\pm 21.6^\circ$, $\pm 64.8^\circ$, $\pm 108^\circ$, and $\pm 151.2^\circ$,

with 360° corresponding to $f_s = 1,000\text{ Hz}$.



Coordinates of the zeros are $0.92977 \pm j0.36812$,

$0.42578 \pm j0.90483$, $-0.30902 \pm j0.95106$,

and $-0.87631 \pm j0.48175$.

$$\begin{aligned} H(z) = & G (1 - 1.85955z^{-1} + z^{-2})(1 - 0.85156z^{-1} + z^{-2}) \\ & \times (1 + 0.61803z^{-1} + z^{-2})(1 + 1.75261z^{-1} + z^{-2}) \end{aligned} \quad (3.66)$$

G : desired gain or scaling factor.



With G computed so that $|H(1)| = 1$:

$$\begin{aligned} H(z) = & 0.6310 - 0.2149z^{-1} + 0.1512z^{-2} \\ & - 0.1288z^{-3} + 0.1227z^{-4} - 0.1288z^{-5} \\ & + 0.1512z^{-6} - 0.2149z^{-7} + 0.6310z^{-8}. \quad (3.67) \end{aligned}$$

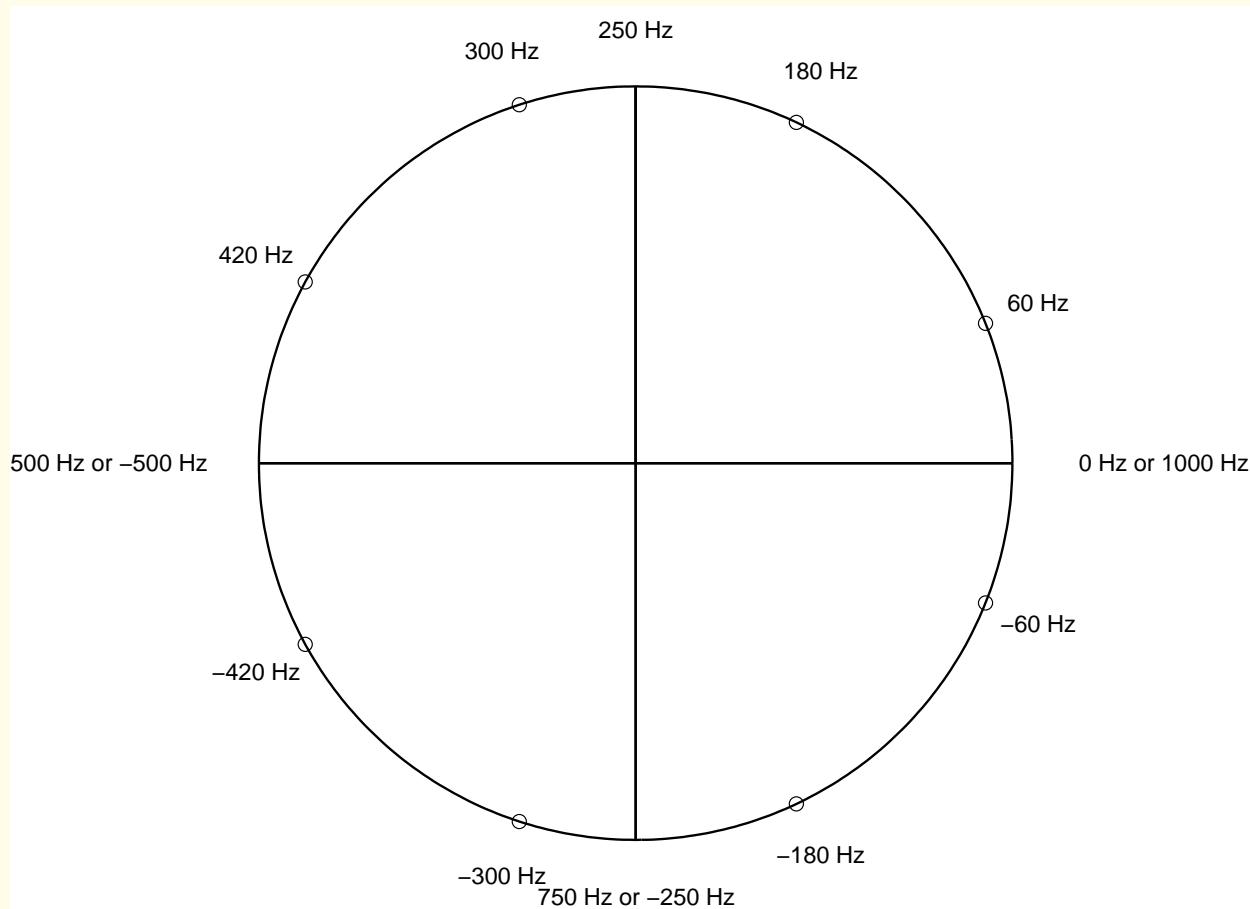


Figure 3.42: Zeros of the comb filter to remove 60 Hz interference with odd harmonics; the sampling frequency is $1,000 \text{ Hz}$.

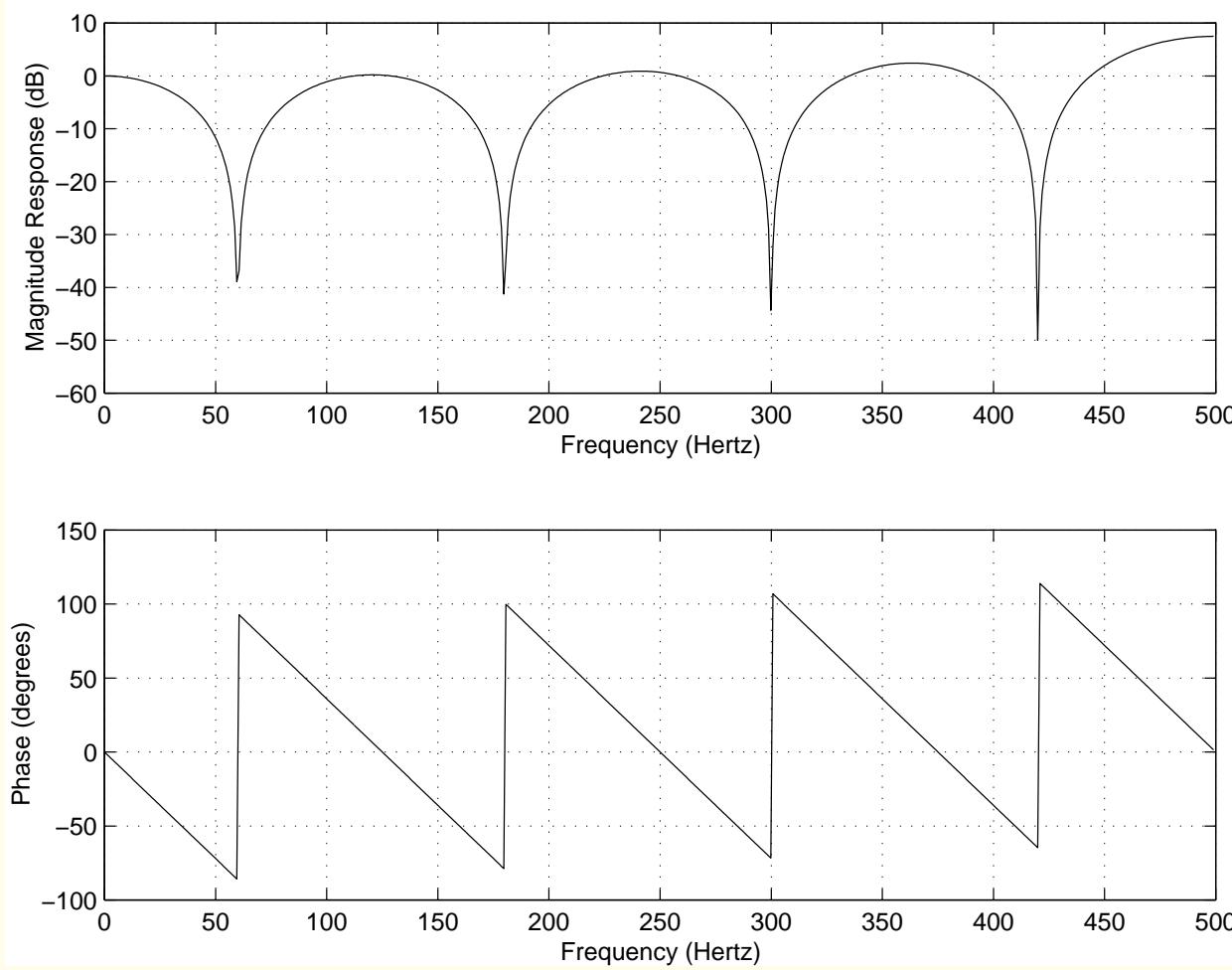


Figure 3.43: Magnitude and phase responses of the comb filter with zeros as shown in Figure 3.42.

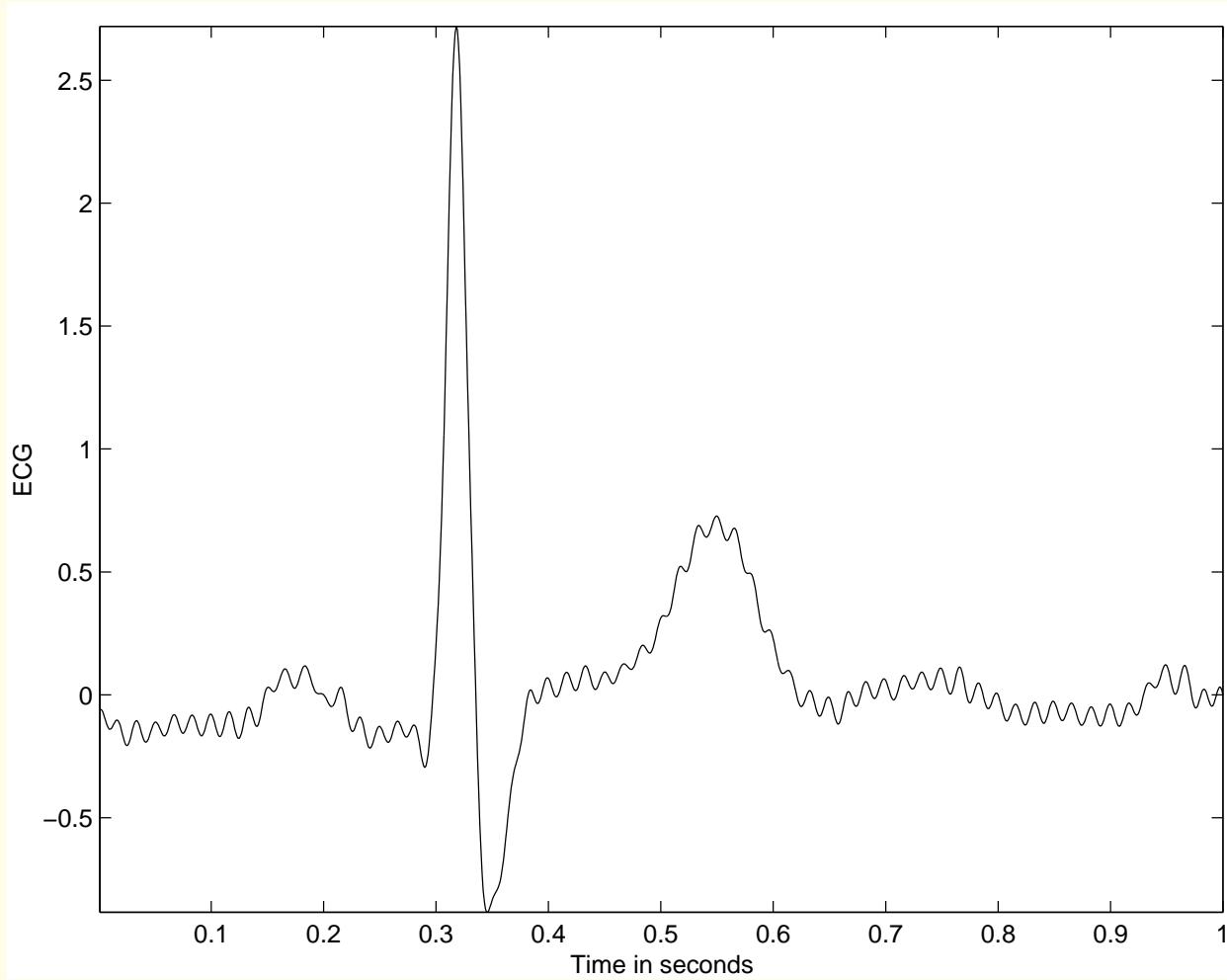


Figure 3.44: ECG signal with 60 Hz interference.

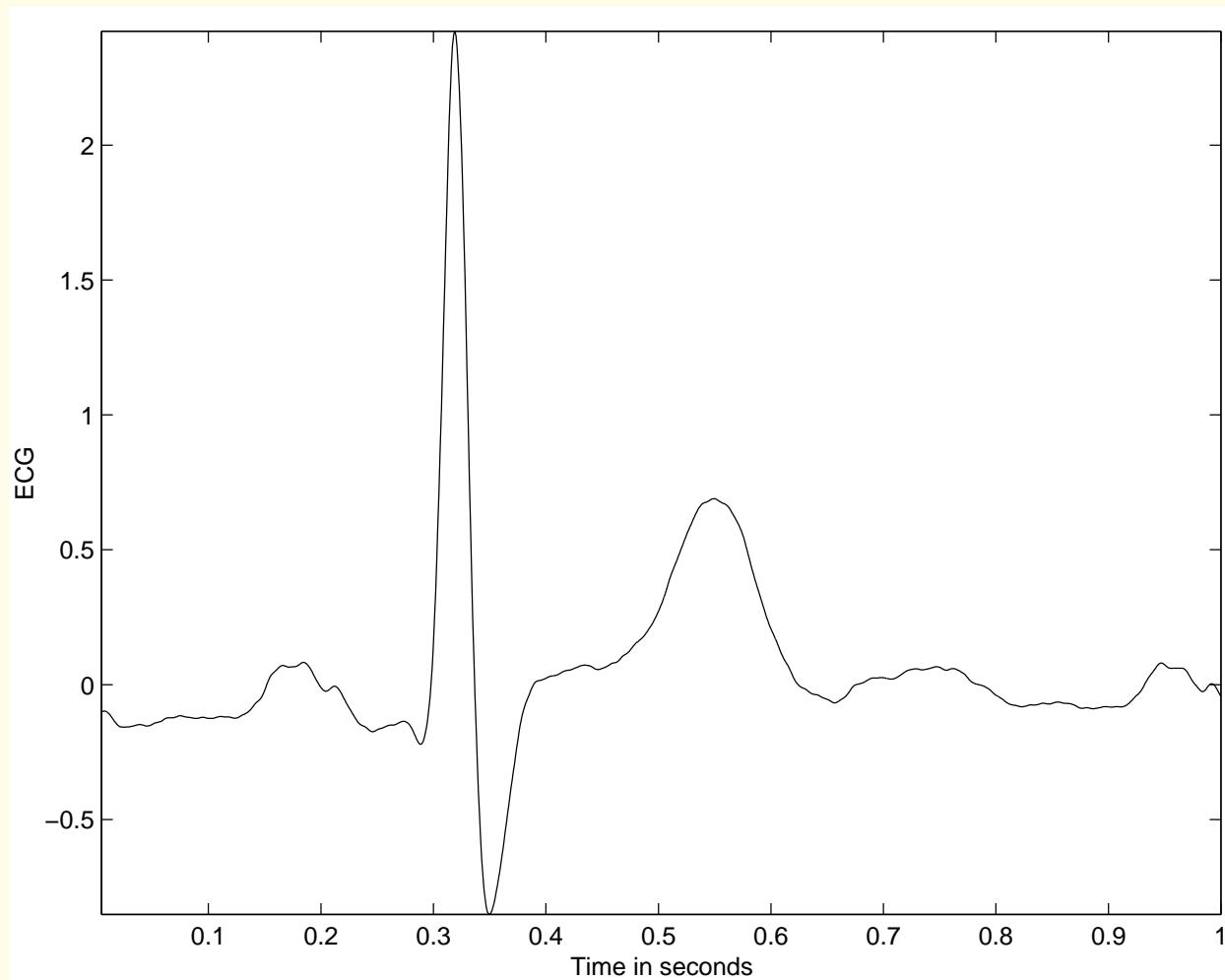


Figure 3.45: The ECG signal in Figure 3.44 after filtering with the 60 Hz notch filter shown in Figures 3.40 and 3.41.



3.5 Optimal Filtering: The Wiener Filter

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 the “desired” or ideal characteristics

of the uncorrupted signal to be 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.



Single-input, single-output, FIR filter with

real input signal values and real coefficients.

Figure 3.46: 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 $\tilde{d}(n)$.

Output $\tilde{d}(n)$ = an estimate of some “desired” signal

$d(n)$ that represents the ideal, uncorrupted signal.



If we assume that the desired signal is available,

estimation error between the output and the desired signal:

$$e(n) = d(n) - \tilde{d}(n). \quad (3.68)$$

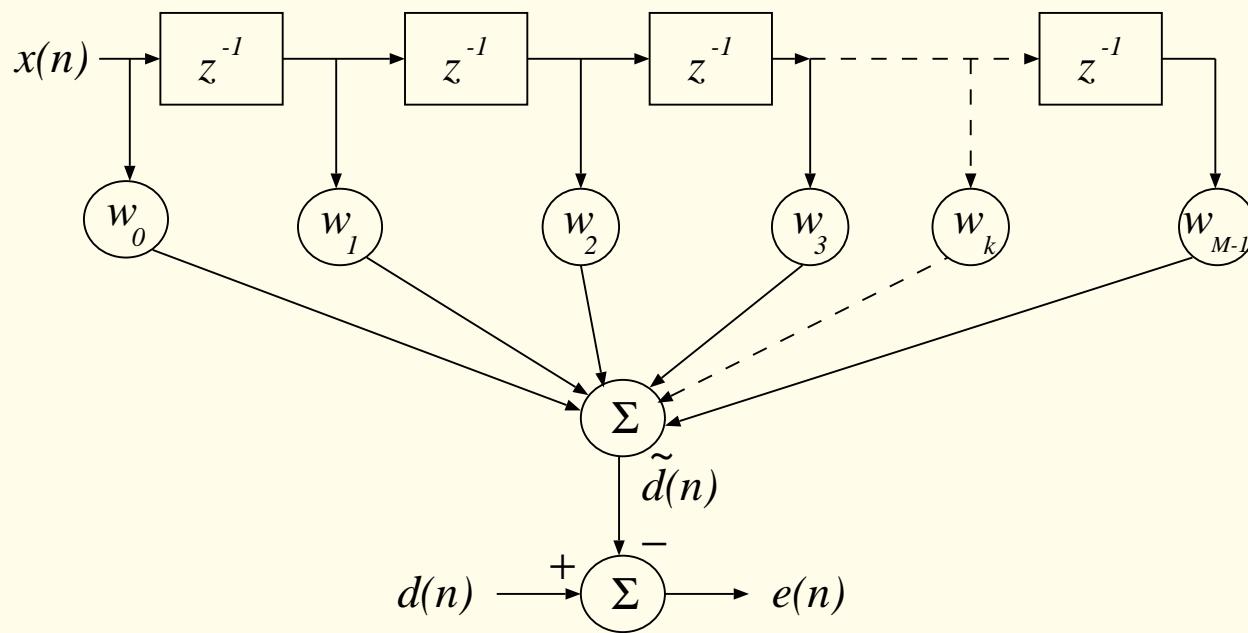


Figure 3.46: Block diagram of the Wiener filter.



$\tilde{d}(n)$ = output of a linear FIR filter

= convolution of the input $x(n)$

with the tap-weight sequence w_i :

$$\tilde{d}(n) = \sum_{k=0}^{M-1} w_k x(n - k). \quad (3.69)$$

w_i is also the impulse response of the filter.



For easier handling of the optimization procedures,

the tap-weight sequence may be written as an $M \times 1$

tap-weight vector:

$$\mathbf{w} = [w_0, w_1, w_2, \dots, w_{M-1}]^T, \quad (3.70)$$

where the bold-faced character \mathbf{w} represents a vector;

superscript T : vector transposition.



Tap weights convolved with M values of input.

Write the M input values as an $M \times 1$ vector:

$$\mathbf{x}(n) = [x(n), x(n - 1), \dots, x(n - M + 1)]^T. \quad (3.71)$$

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 $x(n - 1)$ to $x(n - M + 1)$.



Convolution in Equation 3.69 in a simpler form:

inner product or dot product of the vectors \mathbf{w} and $\mathbf{x}(n)$:

$$\tilde{d}(n) = \mathbf{w}^T \mathbf{x}(n) = \mathbf{x}^T(n) \mathbf{w} = \langle \mathbf{x}, \mathbf{w} \rangle. \quad (3.72)$$

Estimation error:

$$e(n) = d(n) - \mathbf{w}^T \mathbf{x}(n). \quad (3.73)$$



Wiener filter theory estimates the tap-weight sequence that minimizes the MS value of the estimation error; output = *minimum mean-squared error* (MMSE) estimate of the desired response: *optimal filter*.



Mean-squared error (MSE) defined as

$$\begin{aligned} J(\mathbf{w}) &= E[e^2(n)] \\ &= E[\{d(n) - \mathbf{w}^T \mathbf{x}(n)\}\{d(n) - \mathbf{x}^T(n)\mathbf{w}\}] \\ &= E[d^2(n)] - \mathbf{w}^T E[\mathbf{x}(n)d(n)] - E[d(n)\mathbf{x}^T(n)]\mathbf{w} \\ &\quad + \mathbf{w}^T E[\mathbf{x}(n)\mathbf{x}^T(n)]\mathbf{w}. \end{aligned} \tag{3.74}$$

Expectation operator not applicable to \mathbf{w} .



Assumption: input vector $\mathbf{x}(n)$ and desired response $d(n)$

are jointly stationary. Then:

- $E[d^2(n)] = \text{variance of } d(n) = \sigma_d^2$,
assuming that the mean of $d(n)$ is zero.



- $E[\mathbf{x}(n)d(n)] = M \times 1$ vector = cross-correlation between input vector $\mathbf{x}(n)$ and desired response $d(n)$:

$$\Theta = E[\mathbf{x}(n)d(n)]. \quad (3.75)$$

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

$$\theta(-k) = E[x(n - k)d(n)], \quad k = 0, 1, 2, \dots, M - 1. \quad (3.76)$$

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

$$\Theta^T = E[d(n)\mathbf{x}^T(n)]. \quad (3.77)$$



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

$$\Phi = E[\mathbf{x}(n)\mathbf{x}^T(n)] \quad (3.78)$$

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

$$\begin{bmatrix} \phi(0) & \phi(1) & \cdots & \phi(M-1) \\ \phi(-1) & \phi(0) & \cdots & \phi(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(-M+1) & \phi(-M+2) & \cdots & \phi(0) \end{bmatrix}. \quad (3.79)$$



The element in row k and column i of Φ is

$$\phi(i - k) = E[x(n - k)x(n - i)], \quad (3.80)$$

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 the M values of the autocorrelation $\phi(0), \phi(1), \dots, \phi(M - 1)$ for lags $0, 1, \dots, M - 1$.



MSE expression in Equation 3.74 is simplified to

$$J(\mathbf{w}) = \sigma_d^2 - \mathbf{w}^T \boldsymbol{\Theta} - \boldsymbol{\Theta}^T \mathbf{w} + \mathbf{w}^T \boldsymbol{\Phi} \mathbf{w}. \quad (3.81)$$

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 ,

differentiate $J(\mathbf{w})$ with respect to \mathbf{w} ,

set it to zero, and solve the resulting equation.



Note the following derivatives:

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

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

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



$$\frac{dJ(\mathbf{w})}{d\mathbf{w}} = -2\Theta + 2\Phi\mathbf{w} \rightarrow 0. \quad (3.82)$$

Condition for the optimal filter:

$$\Phi\mathbf{w}_o = \Theta. \quad (3.83)$$

Wiener-Hopf equation or the *normal equation*.

Optimal Wiener filter:

$$\mathbf{w}_o = \Phi^{-1} \Theta. \quad (3.84)$$



Wiener-Hopf equation in expanded form:

$$\begin{bmatrix} \phi(0) & \phi(1) & \cdots & \phi(M-1) \\ \phi(-1) & \phi(0) & \cdots & \phi(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(-M+1) & \phi(-M+2) & \cdots & \phi(0) \end{bmatrix} \begin{bmatrix} w_{o0} \\ w_{o1} \\ \vdots \\ w_{o(M-1)} \end{bmatrix} = \quad (3.85)$$

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



The matrix product is equivalent to the following:

$$\sum_{i=0}^{M-1} w_{oi} \phi(i-k) = \theta(-k), \quad k = 0, 1, 2, \dots, M-1. \quad (3.86)$$



Minimum MSE:

$$J_{\min} = \sigma_d^2 - \Theta^T \Phi^{-1} \Theta. \quad (3.87)$$

Given the condition that the signals are stationary, we have

$$\phi(i - k) = \phi(k - i) \text{ and } \theta(-k) = \theta(k).$$

Then, we may write Equation 3.86 as

$$\sum_{i=0}^{M-1} w_{oi} \phi(k - i) = \theta(k), \quad k = 0, 1, 2, \dots, M - 1. \quad (3.88)$$



This is equivalent to the convolution relationship

$$w_{ok} * \phi(k) = \theta(k). \quad (3.89)$$

Applying the Fourier transform:

$$W(\omega)S_{xx}(\omega) = S_{xd}(\omega). \quad (3.90)$$

Note: FT(ACF) = PSD and FT(CCF) = CSD.



Wiener filter frequency response:

$$W(\omega) = \frac{S_{xd}(\omega)}{S_{xx}(\omega)}. \quad (3.91)$$

$S_{xx}(\omega)$: PSD of the input signal and

$S_{xd}(\omega)$: cross-spectral density (CSD)

between the input signal and the desired signal.



Derivation of the optimal filter requires specific knowledge

about the input $x(n)$ and the desired response $d(n)$:

autocorrelation Φ of the input $x(n)$ and

cross-correlation Θ between $x(n)$ and $d(n)$.

In practice, desired response $d(n)$ not known,

but it is possible to obtain an estimate of its

temporal or spectral statistics \rightarrow estimate Θ .



Wiener filter to remove noise:

Input $x(n)$ = desired original signal $d(n)$ + noise $\eta(n)$:

$$x(n) = d(n) + \eta(n). \quad (3.92)$$

Vector notation:

$$\mathbf{x}(n) = \mathbf{d}(n) + \boldsymbol{\eta}(n). \quad (3.93)$$

$\boldsymbol{\eta}(n)$: vector representation of noise $\eta(n)$.



Autocorrelation matrix of input:

$$\Phi = E[\mathbf{x}(n)\mathbf{x}^T(n)] = E[\{\mathbf{d}(n) + \boldsymbol{\eta}(n)\}\{\mathbf{d}(n) + \boldsymbol{\eta}(n)\}^T]. \quad (3.94)$$

Assume that the noise process is statistically independent

of the signal process, and that the

mean of at least one of the processes is zero. Then,

$$E[\mathbf{d}(n)\boldsymbol{\eta}^T(n)] = E[\boldsymbol{\eta}^T(n)\mathbf{d}(n)] = E[\boldsymbol{\eta}(n)]E[\mathbf{d}(n)] = 0. \quad (3.95)$$



$$\Phi = E[\mathbf{d}(n)\mathbf{d}^T(n)] + E[\boldsymbol{\eta}(n)\boldsymbol{\eta}^T(n)] = \Phi_d + \Phi_\eta. \quad (3.96)$$

Φ_d, Φ_η : $M \times M$ autocorrelation matrices of signal and noise.



$$\begin{aligned}\Theta &= E[\mathbf{x}(n)d(n)] = E[\{\mathbf{d}(n) + \boldsymbol{\eta}(n)\}d(n)] \\ &= E[\mathbf{d}(n)d(n)] = \Phi_{1d}. \end{aligned} \tag{3.97}$$

Φ_{1d} : $M \times 1$ autocorrelation vector of the desired signal.



Optimal Wiener filter:

$$\mathbf{w}_o = (\Phi_d + \Phi_\eta)^{-1} \Phi_{1d}. \quad (3.98)$$

Frequency response of the Wiener filter:

$$S_{xx}(\omega) = S_d(\omega) + S_\eta(\omega), \quad (3.99)$$

$$S_{xd}(\omega) = S_d(\omega), \quad (3.100)$$

$$W(\omega) = \frac{S_d(\omega)}{S_d(\omega) + S_\eta(\omega)} = \frac{1}{1 + \frac{S_\eta(\omega)}{S_d(\omega)}}. \quad (3.101)$$



$S_d(\omega)$ and $S_\eta(\omega)$: PSDs of desired signal and noise.

Designing the optimal filter requires knowledge

of the PSDs of the desired signal and

the noise process (or models thereof).



Illustration of application:

Figure 3.47: ECG signal with noise.

Piece-wise linear model of the desired version of the signal

obtained by concatenating linear segments similar to

P, QRS, and T waves

in amplitude, duration, and interval in given noisy signal.

Base-line of the model signal set to zero.



Noise-free model signal: middle trace of Figure 3.47.

Log PSDs and Wiener filter frequency response:

Figure 3.48.

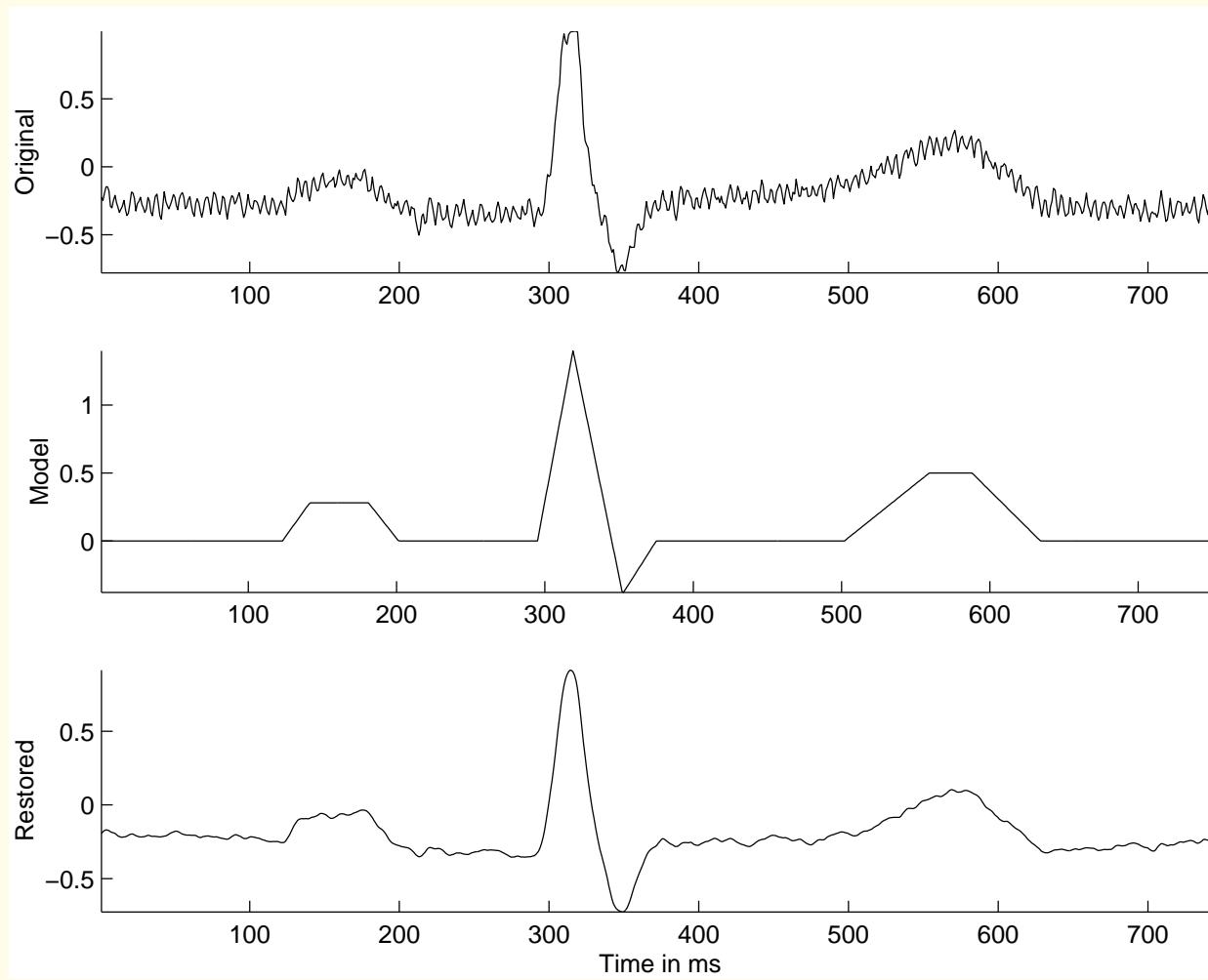


Figure 3.47: From top to bottom: one cycle of the noisy ECG signal in Figure 3.5 (labeled as Original); a piece-wise linear model of the desired noise-free signal (Model); and the output of the Wiener filter (Restored).

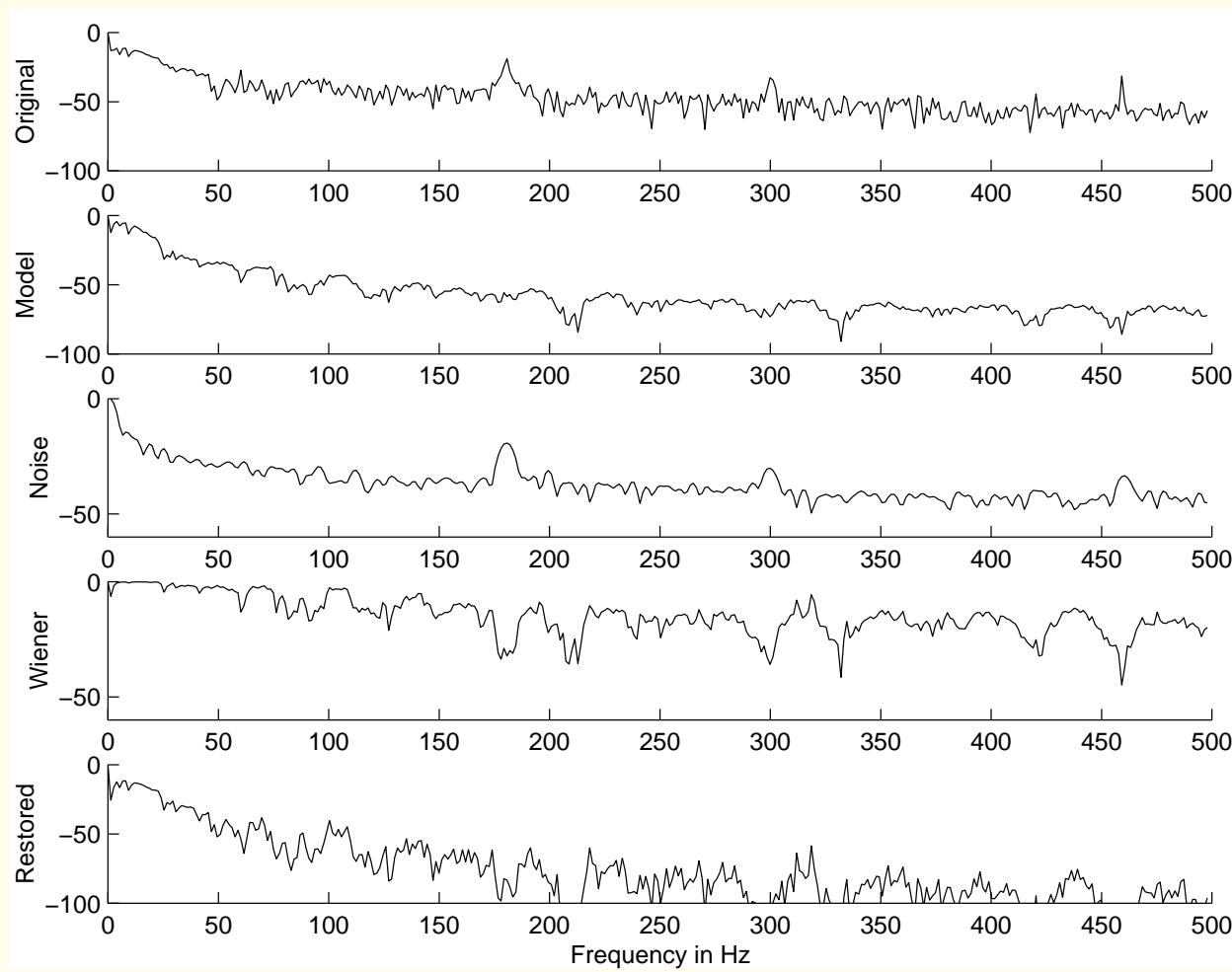


Figure 3.48: From top to bottom: log PSD (in dB) of the given noisy signal (labeled as Original); log PSD of the noise-free model (Model); estimated log PSD of the noise process (Noise); log frequency response of the Wiener filter (Wiener); and log PSD of the filter output (Restored).



T – P intervals between successive cardiac cycles:

inter-beat intervals typically iso-electric base-line.

Then, any activity 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 3.101).



Estimated log PSD of noise shown in Figure 3.48.

Wiener filter derived with models of noise and signal PSDs:

obtained from the given signal itself!

No cutoff frequency specified.



3.6 Adaptive Filters for Removal of Interference

Filters with fixed characteristics, tap weights, or coefficients

are suitable when the characteristics of the signal and noise

are stationary and known.



Such filters are not applicable when the characteristics of the

signal and/or noise vary with time:

when they are nonstationary.

They are also not suitable when the spectral contents of the

signal and the interference overlap significantly.



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:

Need to address two different concerns:

1. The filter should be *adaptive*;
the tap-weight vector of the filter varies with time:
adaptive filter or adaptive noise canceler (ANC).
2. The filter should be *optimal*: LMS and RLS algorithms.



3.6.1 The adaptive noise canceler

“Primary input” or observed signal $x(n)$ is a mixture of the signal of interest $v(n)$ and the “primary noise” $m(n)$:

$$x(n) = v(n) + m(n). \quad (3.102)$$

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 $v(n)$.

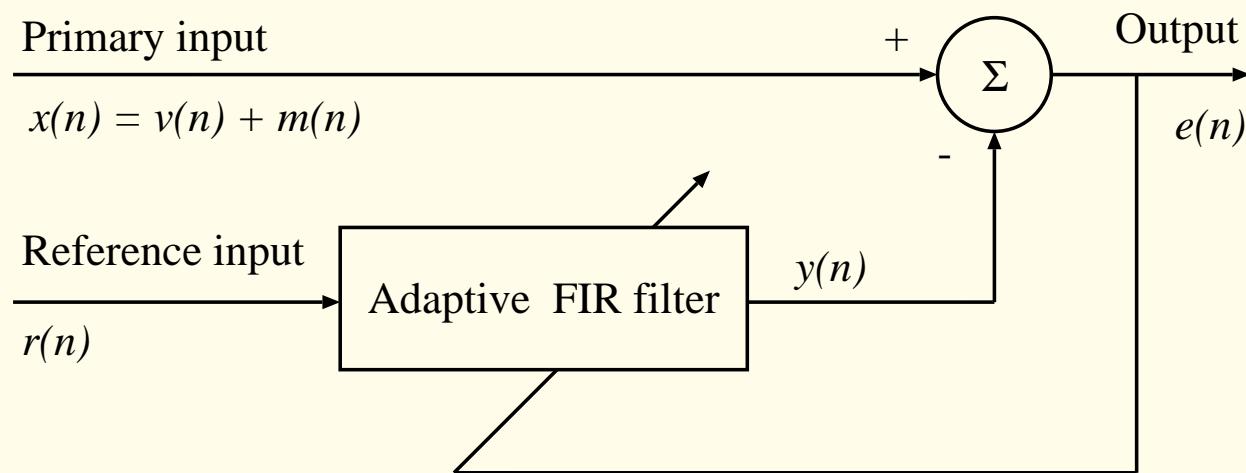


Figure 3.49: Block diagram of a generic adaptive noise canceler (ANC) or adaptive filter.



It is assumed that $v(n)$ and $m(n)$ are uncorrelated.

ANC requires a second input: “reference input” $r(n)$,

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 $y(n)$ as close to noise $m(n)$ as possible.



$y(n)$ is subtracted from primary input

to estimate desired signal:

$$\tilde{v}(n) = e(n) = x(n) - y(n). \quad (3.103)$$



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 will be removed later

when the expectations are computed in moving windows.



$v(n)$ is uncorrelated with $m(n)$ and $r(n)$, and

$r(n)$ is correlated with $m(n)$.

$$\begin{aligned} e(n) &= x(n) - y(n) \\ &= v(n) + m(n) - y(n). \end{aligned} \tag{3.104}$$

$y(n) = \tilde{m}(n)$: estimate of the primary noise

obtained at the output of the adaptive filter.



Taking the square and expectation of both sides

of Equation 3.104:

$$E[e^2(n)] = E[v^2(n)] + E[\{m(n) - y(n)\}^2] \quad (3.105)$$

$$+ 2E[v(n)\{m(n) - y(n)\}].$$



Because $v(n)$ is uncorrelated with $m(n)$ and $y(n)$ and

all of them have zero means, we have

$$\begin{aligned} E[v(n)\{m(n) - y(n)\}] &= E[v(n)] E[m(n) - y(n)] \\ &= 0. \end{aligned} \tag{3.106}$$

$$E[e^2(n)] = E[v^2(n)] + E[\{m(n) - y(n)\}^2]. \tag{3.107}$$



Output $e(n)$ used to control the adaptive filter.

In ANC, the objective is to obtain an output $e(n)$ that is

a least-squares fit to the desired signal $v(n)$.

Achieved by feeding the output back to the adaptive filter

and adjusting the filter to minimize the total output power.

System output: error signal for the adaptive process.



Signal power $E[v^2(n)]$ unaffected as the filter is adjusted

to minimize $E[e^2(n)]$; the minimum output power is

$$\min E[e^2(n)] = E[v^2(n)]$$

$$+ \min E[\{m(n) - y(n)\}^2]. \quad (3.108)$$

As the filter is adjusted so that $E[e^2(n)]$ is minimized,

$E[\{m(n) - y(n)\}^2]$ is 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 minimized, because

(from Equation 3.104)

$$e(n) - v(n) = m(n) - y(n). \quad (3.109)$$



Adapting the filter to minimize the total output power is

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 filter

and for the given reference input.



Output $e(n)$ contains signal of interest $v(n)$ and noise.

From Equation 3.109, the output noise is given by

$$e(n) - v(n) = \tilde{v}(n) - v(n) = m(n) - y(n).$$

Minimizing $E[e^2(n)]$ minimizes $E[\{m(n) - y(n)\}^2]$;

therefore

minimizing the total output power

minimizes the output noise power.



Because the signal component $v(n)$

in the output remains unaffected,

minimizing the total output power

maximizes the output SNR.



Note from Equation 3.107 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 $e(n) = v(n)$:

then, 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.

$$y(n) = \sum_{k=0}^{M-1} w_k r(n - k). \quad (3.110)$$

w_k , $k = 0, 1, 2, \dots, M - 1$, are the tap weights,

and M is the order of the filter.



Estimation error $e(n)$ or output of ANC:

$$e(n) = x(n) - y(n). \quad (3.111)$$

Define the tap-weight vector at time n as

$$\mathbf{w}(n) = [w_0(n), w_1(n), \dots, w_{M-1}(n)]^T. \quad (3.112)$$



Tap-input vector at each time instant n :

$$\mathbf{r}(n) = [r(n), r(n-1), \dots, r(n-M+1)]^T. \quad (3.113)$$

Then, estimation error $e(n)$:

$$e(n) = x(n) - \mathbf{w}^T(n) \mathbf{r}(n). \quad (3.114)$$



3.6.2 *The least-mean-squares adaptive filter*

Adjust the tap-weight vector to minimize the MSE.

Squaring the estimation error $e(n)$ in Equation 3.114:

$$\begin{aligned} e^2(n) &= x^2(n) - 2 \ x(n) \ \mathbf{r}^T(n) \ \mathbf{w}(n) \\ &\quad + \ \mathbf{w}^T(n) \ \mathbf{r}(n) \ \mathbf{r}^T(n) \ \mathbf{w}(n). \end{aligned} \tag{3.115}$$



Squared error is a second-order or quadratic function

of the tap-weight vector and the inputs, and may be

depicted as a concave hyper-paraboloidal, bowl-like surface.

Aim of optimization:

reach the bottom of the bowl-like function.

Gradient-based methods may be used for this purpose.



Taking the expected values of the entities in Equation 3.115

and the derivative with respect to the tap-weight vector,

we may derive the Wiener-Hopf equation for the ANC.

The LMS algorithm takes a simpler approach:

assume the square of the instantaneous error

in Equation 3.115 to stand for an estimate of the MSE.



LMS algorithm based on the method of steepest descent:

new tap-weight vector $\mathbf{w}(n + 1)$ given by the

present tap-weight vector $\mathbf{w}(n)$ plus a

correction proportional to the negative of the

gradient $\nabla(n)$ of the squared error:

$$\mathbf{w}(n + 1) = \mathbf{w}(n) - \mu \nabla(n). \quad (3.116)$$



Parameter μ controls stability and rate of convergence:

larger the value of μ , larger is the gradient of the error

that is introduced, and the faster is the convergence.

LMS algorithm approximates $\nabla(n)$ by the

derivative of the squared error in Equation 3.115

with respect to the tap-weight vector as

$$\begin{aligned}\tilde{\nabla}(n) &= -2 x(n) \mathbf{r}(n) + 2 \{ \mathbf{w}^T(n) \mathbf{r}(n) \} \mathbf{r}(n) \\ &= -2 e(n) \mathbf{r}(n).\end{aligned}\tag{3.117}$$



Using this estimate of the gradient in Equation 3.116, we get

$$\mathbf{w}(n+1) = \mathbf{w}(n) + 2 \mu e(n) \mathbf{r}(n). \quad (3.118)$$

This is known as the Widrow-Hoff LMS algorithm.



Advantages of LMS algorithm:

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.



Illustration of application:

Zhang et al. used a two-stage adaptive LMS filter to

cancel muscle-contraction interference from VAG signals.

First stage: to remove measurement noise;

second stage: to cancel the muscle signal.

Optimization of step size μ by using an

RMS-error-based misadjustment factor and a

time-varying estimate of input signal power.



$$\mathbf{w}(n+1) = \mathbf{w}(n) + 2 \mu(n) e(n) \mathbf{r}(n). \quad (3.119)$$

$$\mu(n) = \frac{\mu}{(M+1) \bar{x}^2(n) (\alpha, r(n), \bar{x}^2(n-1))}, \quad (3.120)$$

$$0 < \mu < 1.$$

$\bar{x}^2(n)$ time-varying function of $\alpha, r(n)$, and $\bar{x}^2(n-1)$:

$$\bar{x}^2(n) = \alpha r^2(n) + (1 - \alpha) \bar{x}^2(n-1).$$

Forgetting factor α in the adaptation process, $0 \leq \alpha \ll 1$.

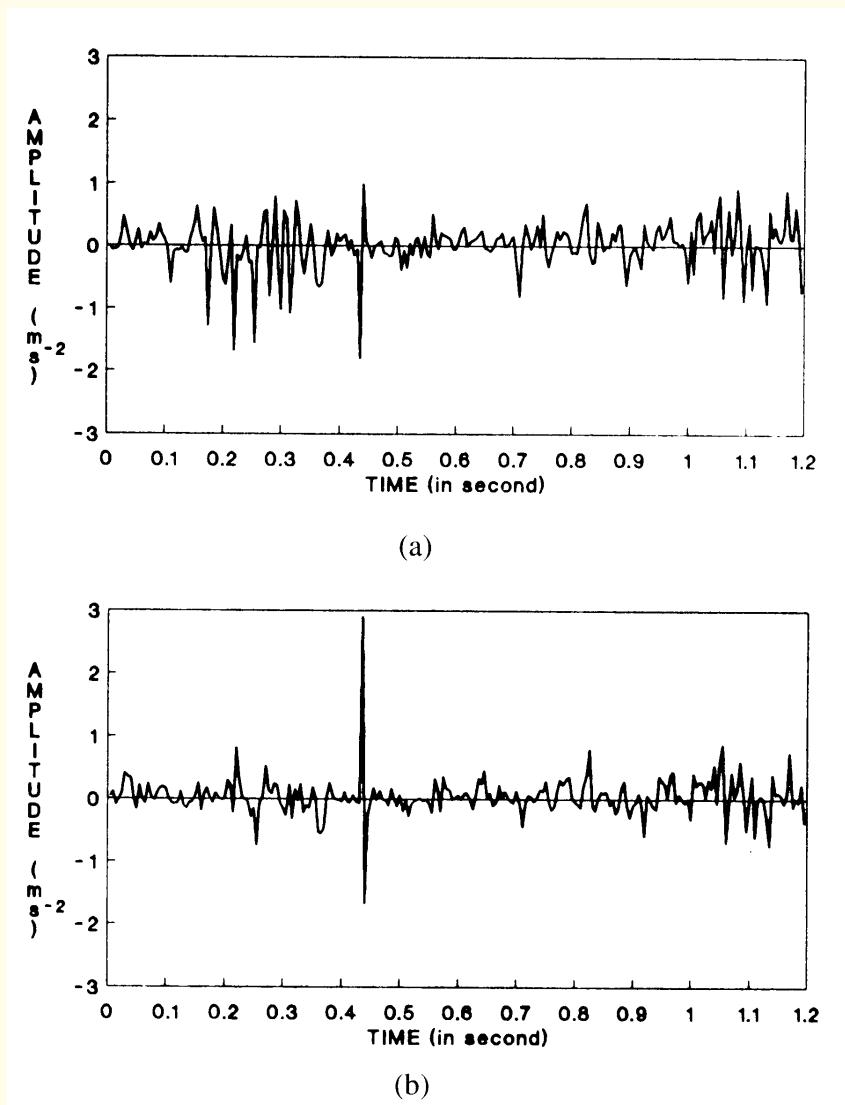


Figure 3.50: LMS-filtered versions of the VAG signals recorded from the mid-patella and the tibial tuberosity positions, as shown in Figure 3.11 (traces (b) and (c), right-hand column). The muscle-contraction signal recorded at the distal rectus femoris position was used as the reference input (Figure 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.



3.6.3 *The recursive least-squares adaptive filter*

Skip.



3.7 Selecting an Appropriate Filter

Synchronized or ensemble averaging is possible when:

- The signal is statistically stationary, (quasi-)periodic, or cyclo-stationary.
- Multiple realizations of signal of interest available.
- A trigger point or time marker is available, or can be derived to extract and align the copies of the signal.
- 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 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.
- The signal spectrum is limited in bandwidth compared to that of the noise (or vice-versa).
- 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.
- 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 strongly correlated with the noise but uncorrelated with the signal.



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.

All of the filters mentioned above are applicable

only when the noise is additive.

Homomorphic filtering may be used as a preprocessing step

if signals are combined with operations other than addition.



3.8 Application: Removal of Artifacts in the ECG

Problem: *Figure 3.8 (top trace) shows an ECG signal with a combination of base-line drift, high-frequency noise, and power-line interference.*

Design filters to remove the artifacts.

Solution: HPF, LPF, and comb filter in series (cascade).

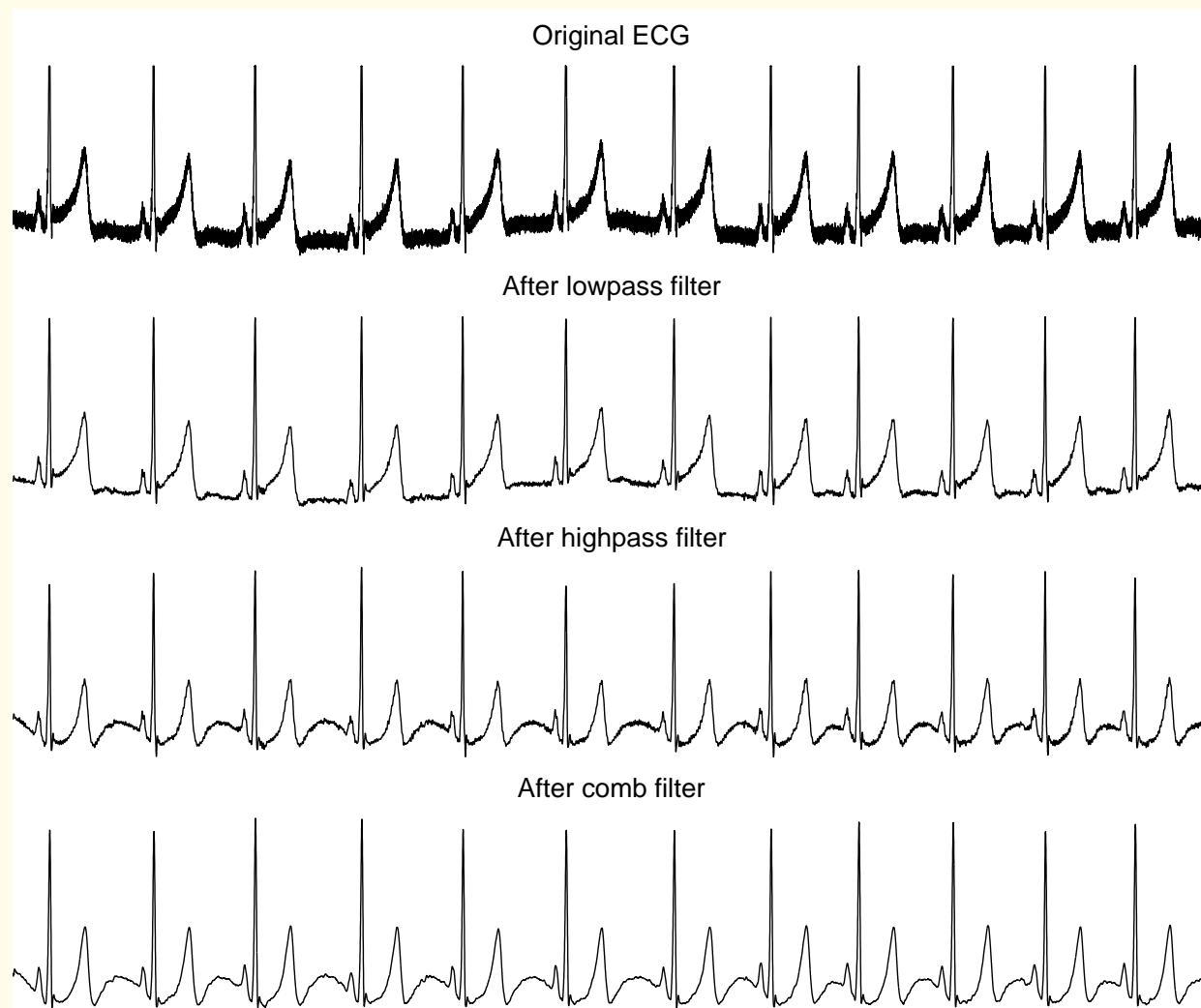


Figure 3.56 ECG signal with a combination of artifacts and its filtered versions. The duration of the signal is 10.7 s.

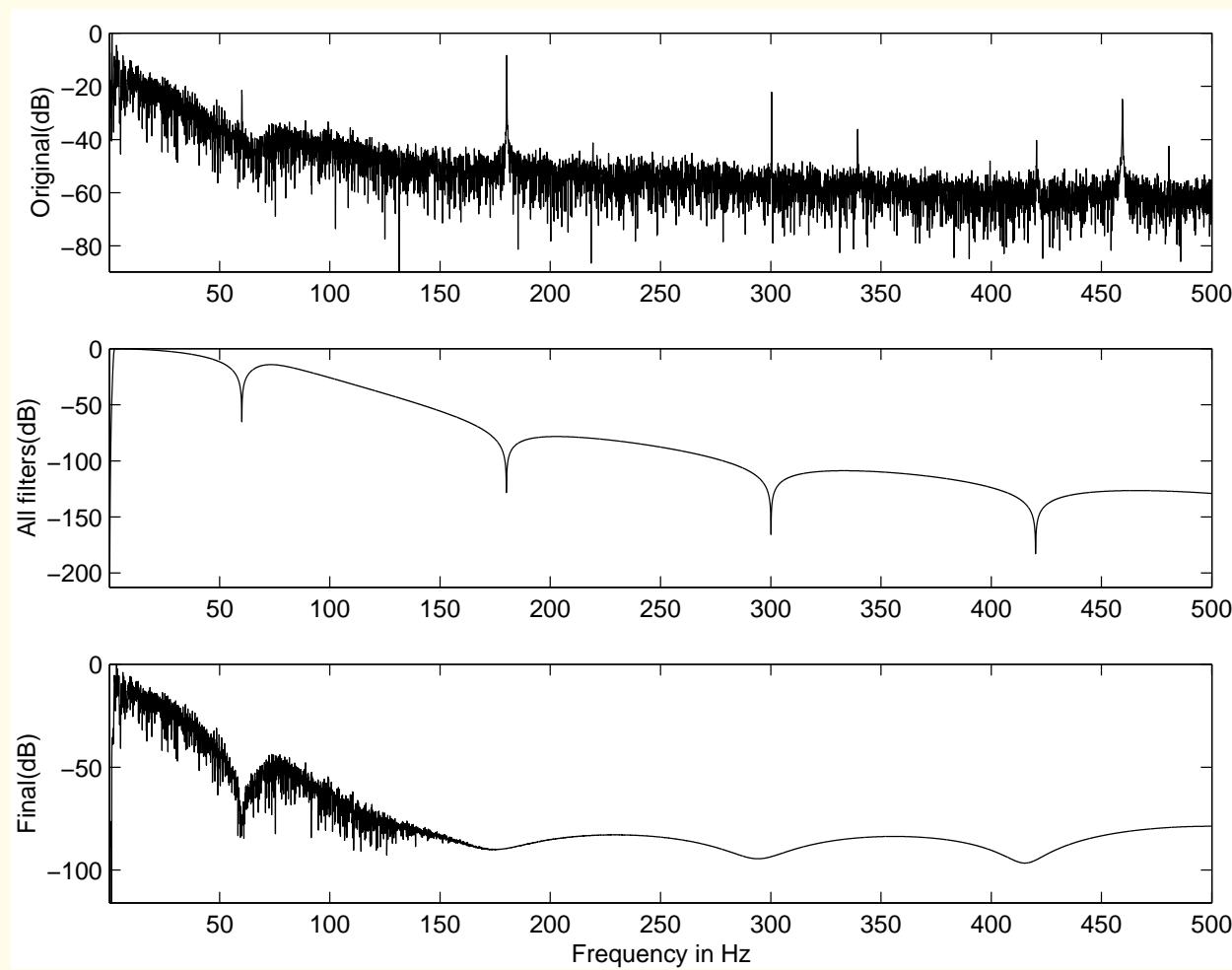


Figure 3.57 Top and bottom plots: Power spectra of the ECG signals in the top and bottom traces of Figure 3.8. Middle plot: Frequency response of the combination of lowpass, highpass, and comb filters. The cutoff frequency of the highpass filter is **2 Hz**; the highpass portion of the frequency response is not clearly seen in the plot.



3.9 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 mother

may be used for reference.



Solution: Widrow et al. proposed a multiple-reference

ANC for removal of the maternal ECG in order

to obtain the fetal ECG.

Combined ECG obtained from a single abdominal lead.

Maternal ECG was obtained via four chest leads.

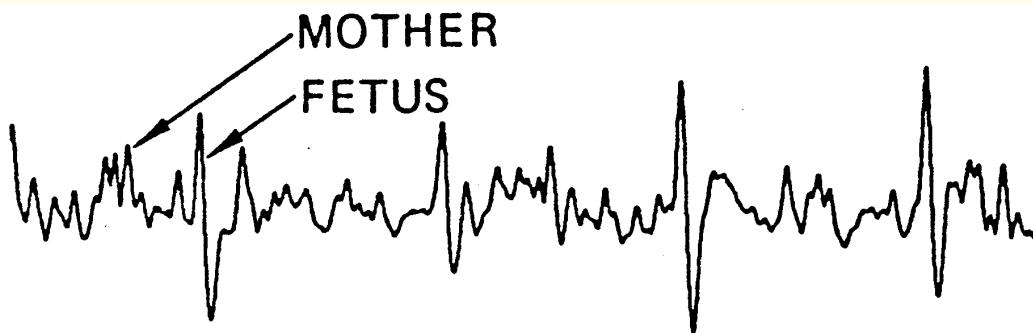


Figure 3.58 Result of adaptive cancellation of the maternal chest ECG from the abdominal ECG in Figure 3.9. The QRS complexes extracted correspond to the fetal ECG. Reproduced with permission from B. Widrow, J.R. Glover, Jr., J.M. McCool, J. Kaunitz, C.S. Williams, R.H. Hearn, J.R. Zeidler, E. Dong, Jr., R.C. Goodlin, Adaptive noise cancelling: Principles and applications, Proceedings of the IEEE, 63(12):1692–1716, 1975. ©IEEE.



3.10 Application: Adaptive Cancellation of Muscle-contraction Interference in Knee-joint Vibration Signals

Skip.

