THE Biomedical Engineering HANDBOOK

Editor-in-Chief
JOSEPH D. BRONZINO
Trinity College
Hartford, Connecticut

CRC PRESS
IEEE PRESS
A CRC Handbook Published in Cooperation with IEEE Press

Library of Congress Cataloging-in-Publication Data

The biomedical engineering handbook / editor-in-chief, Joseph D. Bronzino.
p. cm. — (The electrical engineering handbook series)
Includes bibliographical references and index.
ISBN 0-8493-8346-3
1. Biomedical engineering—Handbooks, manuals, etc. I. Bronzino, Joseph D., 1937– . II. Series.
[DNLM: 1. Biomedical Engineering—handbooks. QT 29 B615 1995]
R856.15.B86 1995
610′.28—dc20
DNLM/DLC
For Library of Congress 95-6294
CIP

No claim to original U.S. Government works
International Standard Book Number 0-8493-8346-3
Library of Congress Card Number 95-6294
Printed in the United States of America 1 2 3 4 5 6 7 8 9 0
Printed on acid-free paper

Introduction

In the 20th century, technological innovation has progressed at such an accelerated pace that it has permeated almost every facet of our lives. This is especially true in the field of medicine and the delivery of health care services. Today, in most developed countries, the modern hospital has emerged as the center of a technologically sophisticated health care system serviced by an equally technologically sophisticated staff.

With almost continual technological innovation driving medical care, engineering professionals have become intimately involved in many medical ventures. As a result, the discipline of biomedical engineering has emerged as an integrating medium for two dynamic professions, medicine and engineering. In the process, biomedical engineers have become actively involved in the design, development, and utilization of materials, devices (such as ultrasonic lithotripsy, pacemakers, etc.), and techniques (such as signal and image processing, artificial intelligence, etc.) for clinical research, as well as the diagnosis and treatment of patients. Thus many biomedical engineers now serve as members of health care delivery teams seeking new solutions for the difficult health care problems confronting our society. The purpose of this handbook is to *provide a central core of knowledge from those fields encompassed by the discipline of biomedical engineering.* Before presenting this detailed information, it is important to provide a sense of the evolution of the modern health care system and identify the diverse activities biomedical engineers perform to assist in the diagnosis and treatment of patients.

Evolution of the Modern Health Care System

Before 1900, medicine had little to offer the average citizen, since its resources consisted mainly of the physician, his education, and his "little black bag." In general, physicians seemed to be in short supply, but the shortage had rather different causes than the current crisis in the availability of health care professionals. Although the costs of obtaining medical training were relatively low, the demand for doctors' services also was very small, since many of the services provided by the physician also could be obtained from experienced amateurs in the community. The home was typically the site for treatment and recuperation, and relatives and neighbors constituted an able and willing nursing staff. Babies were delivered by midwives, and those illnesses not cured by home remedies were left to run their natural, albeit frequently fatal, course. The contrast with contemporary health care practices, in which specialized physicians and nurses located within the hospital provide critical diagnostic and treatment services, is dramatic.

The changes that have occurred within medical science originated in the rapid developments that took place in the applied sciences (chemistry, physics, engineering, microbiology, physiology, pharmacology, etc.) at the turn of the century. This process of development was characterized by intense

interdisciplinary cross-fertilization, which provided an environment in which medical research was able to take giant strides in developing techniques for the diagnosis and treatment of disease. For example, in 1903, Willem Einthoven, the Dutch physiologist, devised the first electrocardiograph to measure the electrical activity of the heart. In applying discoveries in the physical sciences to the analysis of a biologic process, he initiated a new age in both cardiovascular medicine and electrical measurement techniques.

New discoveries in medical sciences followed one another like intermediates in a chain reaction. However, the most significant innovation for clinical medicine was the development of x-rays. These "new kinds of rays," as their discoverer W. K. Roentgen described them in 1895, opened the "inner man" to medical inspection. Initially, x-rays were used to diagnose bone fractures and dislocations, and in the process, x-ray machines became commonplace in most urban hospitals. Separate departments of radiology were established, and their influence spread to other departments throughout the hospital. By the 1930s, x-ray visualization of practically all organ systems of the body had been made possible through the use of barium salts and a wide variety of radiopaque materials.

X-ray technology gave physicians a powerful tool that, for the first time, permitted accurate diagnosis of a wide variety of diseases and injuries. Moreover, since x-ray machines were too cumbersome and expensive for local doctors and clinics, they had to be placed in health care centers or hospitals. Once there, x-ray technology essentially triggered the transformation of the hospital from a passive receptacle for the sick to an active curative institution for all members of society.

For economic reasons, the centralization of health care services became essential because of many other important technological innovations appearing on the medical scene. However, hospitals remained institutions to dread, and it was not until the introduction of sulfanilamide in the mid-1930s and penicillin in the early 1940s that the main danger of hospitalization, i.e., cross-infection among patients, was significantly reduced. With these new drugs in their arsenals, surgeons were permitted to perform their operations without prohibitive morbidity and mortality due to infection. Furthermore, even though the different blood groups and their incompatibility were discovered in 1900 and sodium citrate was used in 1913 to prevent clotting, full development of blood banks was not practical until the 1930s, when technology provided adequate refrigeration. Until that time, "fresh" donors were bled and the blood transfused while it was still warm.

Once these surgical suites were established, the employment of specifically designed pieces of medical technology assisted in further advancing the development of complex surgical procedures. For example, the Drinker respirator was introduced in 1927 and the first heart-lung bypass in 1939. By the 1940s, medical procedures heavily dependent on medical technology, such as cardiac catheterization and angiography (the use of a cannula threaded through an arm vein and into the heart with the injection of radiopaque dye for the x-ray visualization of lung and heart vessels and valves), were developed. As a result, accurate diagnosis of congenital and acquired heart disease (mainly valve disorders due to rheumatic fever) became possible, and a new era of cardiac and vascular surgery was established.

Following World War II, technological advances were spurred on by efforts to develop superior weapon systems and establish habitats in space and on the ocean floor. As a by-product of these efforts, the development of medical devices accelerated and the medical profession benefited greatly from this rapid surge of "technological finds." Consider the following examples:

1. Advances in solid-state electronics made it possible to map the subtle behavior of the fundamental unit of the central nervous system—the neuron—as well as to monitor various physiologic parameters, such as the electrocardiogram, of patients in intensive care units.
2. New prosthetic devices became a goal of engineers involved in providing the disabled with tools to improve their quality of life.
3. Nuclear medicine—an outgrowth of the atomic age—emerged as a powerful and effective approach in detecting and treating specific physiologic abnormalities.

4. Diagnostic ultrasound based on sonar technology became so widely accepted that ultrasonic studies are now part of the routine diagnostic workup in many medical specialties.
5. "Spare parts" surgery also became commonplace. Technologists were encouraged to provide cardiac assist devices, such as artificial heart valves and artificial blood vessels, and the artificial heart program was launched to develop a replacement for a defective or diseased human heart.
6. Advances in materials have made the development of disposable medical devices, such as needles and thermometers, as well as implantable drug delivery systems, a reality.
7. Computers similar to those developed to control the flight plans of the *Apollo* capsule were used to store, process, and cross-check medical records, to monitor patient status in intensive care units, and to provide sophisticated statistical diagnoses of potential diseases correlated with specific sets of patient symptoms.
8. Development of the first computer-based medical instrument, the computerized axial tomography scanner, revolutionized clinical approaches to noninvasive diagnostic imaging procedures, which now include magnetic resonance imaging and positron emission tomography as well.

The impact of these discoveries and many others has been profound. The health care system consisting primarily of the "horse and buggy" physician is gone forever, replaced by a technologically sophisticated clinical staff operating primarily in "modern" hospitals designed to accommodate the new medical technology. This evolutionary process continues, with advances in biotechnology and tissue engineering altering the very nature of the health care delivery system itself.

The Field of Biomedical Engineering

Today, many of the problems confronting health professionals are of extreme interest to engineers because they involve the design and practical application of medical devices and systems—processes that are fundamental to engineering practice. These medically related design problems can range from very complex large-scale constructs, such as the design and implementation of automated clinical laboratories, multiphasic screening facilities (i.e., centers that permit many clinical tests to be conducted), and hospital information systems, to the creation of relatively small and "simple" devices, such as recording electrodes and biosensors, that may be used to monitor the activity of specific physiologic processes in either a research or clinical setting. They encompass the many complexities of remote monitoring and telemetry, including the requirements of emergency vehicles, operating rooms, and intensive care units. The American health care system, therefore, encompasses many problems that represent challenges to certain members of the engineering profession called *biomedical engineers*.

Biomedical Engineering: A Definition

Although what is included in the field of biomedical engineering is considered by many to be quite clear, there are some disagreements about its definition. For example, consider the terms *biomedical engineering, bioengineering*, and *clinical* (or *medical*) *engineering* which have been defined in Pacela's *Bioengineering Education Directory* [Quest Publishing Co., 1990]. While Pacela defines *bioengineering* as the broad umbrella term used to describe this entire field, *bioengineering* is usually defined as a basic research–oriented activity closely related to biotechnology and genetic engineering, i.e., the modification of animal or plant cells, or parts of cells, to improve plants or animals or to develop new microorganisms for beneficial ends. In the food industry, for example, this has meant the improvement of strains of yeast for fermentation. In agriculture, bioengineers may be concerned with the improvement of crop yields by treatment of plants with organisms to reduce frost damage. It is clear that bioengineers of the future will have a tremendous impact on the quality of human life, the potential of this specialty is difficult to imagine. Consider the following activities of bioengineers:

- Development of improved species of plants and animals for food production
- Invention of new medical diagnostic tests for diseases
- Production of synthetic vaccines from clone cells
- Bioenvironmental engineering to protect human, animal, and plant life from toxicants and pollutants
- Study of protein-surface interactions
- Modeling of the growth kinetics of yeast and hybridoma cells
- Research in immobilized enzyme technology
- Development of therapeutic proteins and monoclonal antibodies

In reviewing the above-mentioned terms, however, *biomedical engineering* appears to have the most comprehensive meaning. Biomedical engineers apply electrical, mechanical, chemical, optical, and other engineering principles to understand, modify, or control biologic (i.e., human and animal) systems, as well as design and manufacture products that can monitor physiologic functions and assist in the diagnosis and treatment of patients. When biomedical engineers work within a hospital or clinic, they are more properly called *clinical engineers.*

Activities of Biomedical Engineers

The breadth of activity of biomedical engineers is significant. The field has moved significantly from being concerned primarily with the development of medical devices in the 1950s and 1960s and to include a more wide-ranging set of activities. As illustrated below, the field of biomedical engineering now includes many new career areas, each of which is presented in this *Handbook.* These areas include

- Application of engineering system analysis (physiologic modeling, simulation, and control) to biologic problems
- Detection, measurement, and monitoring of physiologic signals (i.e., *biosensors* and *biomedical instrumentation*)
- Diagnostic interpretation via signal-processing techniques of bioelectric data
- Therapeutic and rehabilitation procedures and devices (rehabilitation engineering)
- Devices for replacement or augmentation of bodily functions (*artificial organs*)
- Computer analysis of patient-related data and clinical decision making (i.e., medical informatics and artificial intelligence)
- Medical imaging, i.e., the graphic display of anatomic detail or physiologic function
- The creation of new biologic products (i.e., *biotechnology* and *tissue engineering*)

Typical pursuits of biomedical engineers, therefore, include

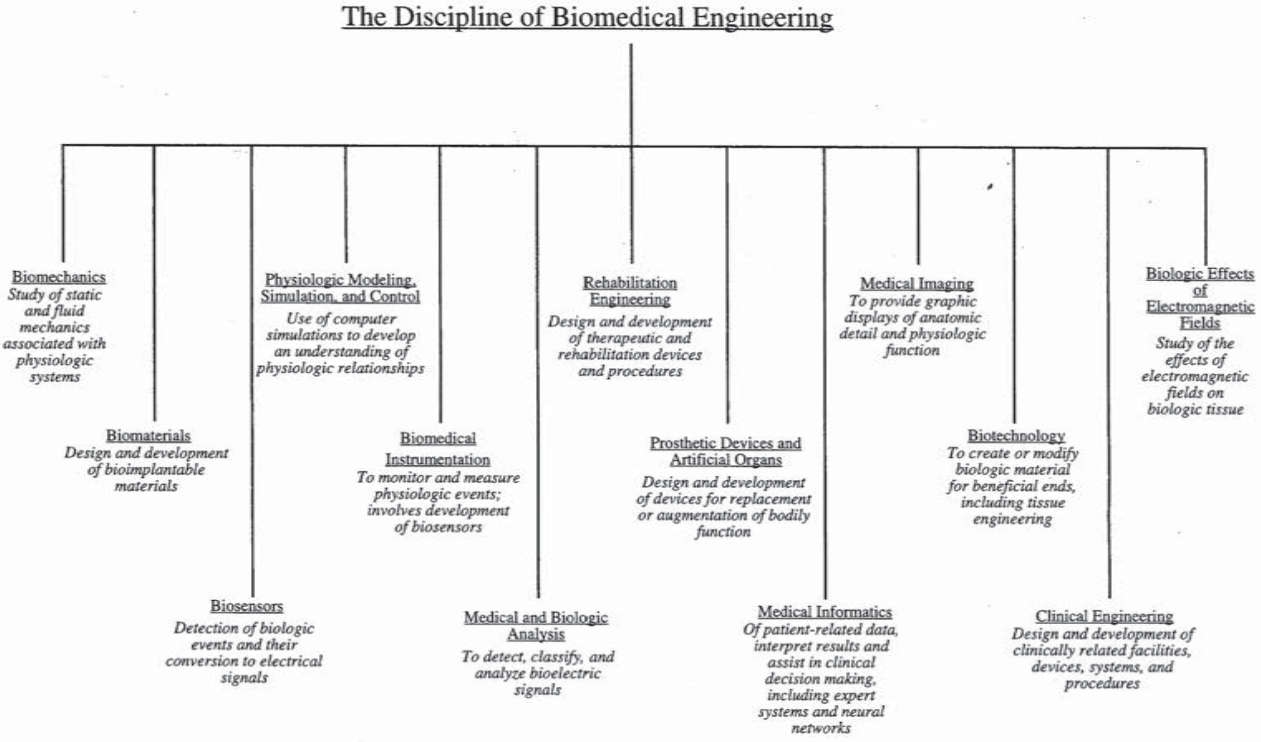
- Research in new materials for implanted artificial organs
- Development of new diagnostic instruments for blood analysis
- Computer modeling of the function of the human heart
- Writing software for analysis of medical research data
- Analysis of medical device hazards for safety and efficacy
- Development of new diagnostic imaging systems
- Design of telemetry systems for patient monitoring
- Design of biomedical sensors for measurement of human physiologic systems variables
- Development of expert systems for diagnosis of diseases
- Design of closed-loop control systems for drug administration

- Modeling of the physiologic systems of the human body
- Design of instrumentation for sports medicine
- Development of new dental materials
- Design of communication aids for the handicapped
- Study of pulmonary fluid dynamics
- Study of the biomechanics of the human body
- Development of material to be used as replacement for human skin

Biomedical engineering, then, is an interdisciplinary branch of engineering that ranges from theoretical, nonexperimental undertakings to state-of-the-art applications. It can encompass research, development, implementation, and operation. Accordingly, like medical practice itself, it is unlikely that any single person can acquire expertise that encompasses the entire field. Yet, because of the interdisciplinary nature of this activity, there is considerable interplay and overlapping of interest and effort between them. For example, biomedical engineers engaged in the development of biosensors may interact with those interested in prosthetic devices to develop a means to detect and use the same bioelectric signal to power a prosthetic device. Those engaged in automating the clinical chemistry laboratory may collaborate with those developing expert systems to assist clinicians in making decisions based on specific laboratory data. The possibilities are endless.

Perhaps a greater potential benefit occurring from the use of biomedical engineering is identification of the problems and needs of our present health care system that can be solved using existing engineering technology and systems methodology. Consequently, the field of biomedical engineering offers hope in the continuing battle to provide high-quality health care at a reasonable cost; and if properly directed toward solving problems related to preventive medical approaches, ambulatory care services, and the like, biomedical engineers can provide the tools and techniques to make our health care system more effective and efficient.

Joseph D. Bronzino
Editor-in-Chief



52
Optical Sensors

Yitzhak Mendelson
Worcester Polytechnic Institute

Optical methods are among the oldest and best-established techniques for sensing biochemical analytes. Instrumentation for optical measurements generally consists of a light source, a number of optical components to generate a light beam with specific characteristics and to direct this light to some modulating agent, and a photodetector for processing the optical signal. The central part of an optical sensor is the modulating component, and a major part of this chapter will focus on how to exploit the interaction of an analyte with optical radiation in order to obtain essential biochemical information.

The number of publications in the field of optical sensors for biomedical applications has grown significantly during the past two decades. Numerous scientific reviews and historical perspectives have been published, and the reader interested in this rapidly growing field is advised to consult these sources for additional details. This chapter will emphasize the basic concept of typical optical sensors intended for continuous in vivo monitoring of biochemical variables, concentrating on those sensors which have generally progressed beyond the initial feasibility stage and reached the promising stage of practical development or commercialization.

Optical sensors are usually based on optical fibers or on planar waveguides. Generally, there are three distinctive methods for quantitative optical sensing at surfaces:

1. The analyte directly affects the optical properties of a waveguide, such as evanescent waves (electromagnetic waves generated in the medium outside the optical waveguide when light is reflected from within) or surface plasmons (resonances induced by an evanescent wave in a thin film deposited on a waveguide surface).
2. An optical fiber is used as a plain transducer to guide light to a remote sample and return light from the sample to the detection system. Changes in the intrinsic optical properties of the medium itself are sensed by an external spectrophotometer.
3. An indicator or chemical reagent placed inside, or on, a polymeric support near the tip of the optical fiber is used as a mediator to produce an observable optical signal. Typically, conventional techniques, such as absorption spectroscopy and fluorimetry, are employed to measure changes in the optical signal.

0-8493-8346-3/95/$0.00+$.50

52.1 Instrumentation

The actual implementation of instrumentation designed to interface with optical sensors will vary greatly depending on the type of optical sensor used and its intended application. A block diagram of a generic instrument is illustrated in Fig. 52.1. The basic building blocks of such an instrument are the light source, various optical elements, and photodetectors.

Light Source

A wide selection of light sources are available for optical sensor applications. These include: highly coherent gas and semiconductor diode lasers, broad spectral band incandescent lamps, and narrow-band, solid-state, light-emitting diodes (LEDs). The important requirement of a light source is obviously good stability. In certain applications, for example in portable instrumentation, LEDs have significant advantages over other light sources because they are small and inexpensive, consume low power, produce selective wavelengths, and are easy to work with. In contrast, tungsten lamps produce a broader range of wavelengths, higher intensity, and better stability but require a sizable power supply and can cause heating problems inside the apparatus.

Optical Elements

Various optical elements are used routinely to manipulate light in optical instrumentation. These include lenses, mirrors, light choppers, beam splitters, and couplers for directing the light from the light source into the small aperture of a fiber optic sensor or a specific area on a waveguide surface and collecting the light from the sensor before it is processed by the photodetector. For wavelength selection, optical filters, prisms, and diffraction gratings are the most common components used to provide a narrow bandwidth of excitation when a broadwidth light source is utilized.

Photodetectors

In choosing photodetectors for optical sensors, a number of factors must be considered. These include sensitivity, detectivity, noise, spectral response, and response time. Photomultipliers and

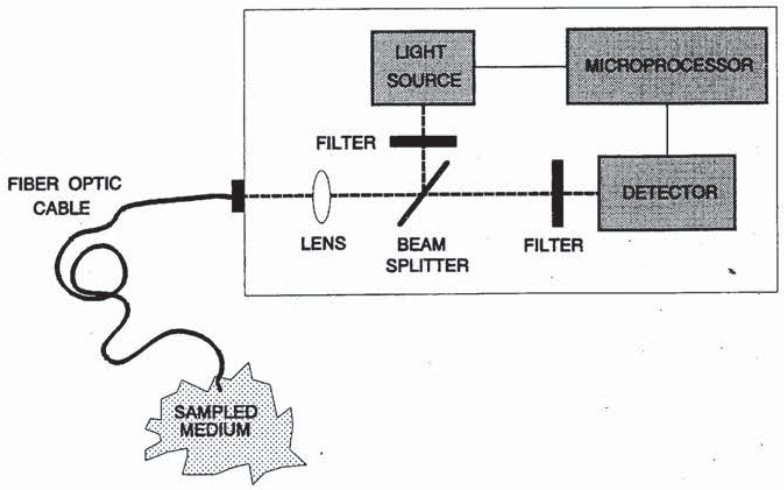


FIGURE 52.1 General diagram representing the basic building blocks of an optical instrument for optical sensor applications.

semiconductor quantum photodetectors, such as photoconductors and photodiodes, are both suitable. The choice, however, is somewhat dependent on the wavelength region of interest. Generally, both give adequate performance. Photodiodes are usually more attractive because of the compactness and simplicity of the circuitry involved.

Typically, two photodetectors are used in optical instrumentation because it is often necessary to include a separate reference detector to track fluctuations in source intensity and temperature. By taking a ratio between the two detector readings, whereby a part of the light that is not affected by the measurement variable is used for correcting any optical variations in the measurement system, a more accurate and stable measurement can be obtained.

Signal Processing

Typically, the signal obtained from a photodetector provides a voltage or a current proportional to the measured light intensity. Therefore, either simple analog computing circuitry (e.g., a current-to-voltage converter) or direct connection to a programmable gain voltage stage is appropriate. Usually, the output from a photodetector is connected directly to a preamplifier before it is applied to sampling and analog-to-digital conversion circuitry residing inside a computer.

Quite often two different wavelengths of light are utilized to perform a specific measurement. One wavelength is usually sensitive to changes in the species being measured, and the other wavelength is unaffected by changes in the analyte concentration. In this manner, the unaffected wavelength is used as a reference to compensate for fluctuations in instrumentation over time. In other applications, additional discriminations, such as pulse excitation or electronic background subtraction utilizing synchronized lock-in amplifier detection, are useful, allowing improved selectivity and enhanced signal-to-noise ratio.

52.2 Optical Fibers

Several types of biomedical measurements can be made by using either plain optical fibers as a remote device for detecting changes in the spectral properties of tissue and blood or optical fibers tightly coupled to various indicator-mediated transducers. The measurement relies either on direct illumination of a sample through the endface of the fiber or by excitation of a coating on the side wall surface through evanescent wave coupling. In both cases, sensing takes place in a region outside the optical fiber itself. Light emanating from the fiber end is scattered or fluoresced back into the fiber, allowing measurement of the returning light as an indication of the optical absorption or fluorescence of the sample at the fiber optic tip.

Optical fibers are based on the principle of total internal reflection. Incident light is transmitted through the fiber if it strikes the cladding at an angle greater than the so-called critical angle, so that it is totally internally reflected at the core/cladding interface. A typical instrument for performing fiber optic sensing consists of a light source, an optical coupling arrangement, the fiber optic light guide with or without the necessary sensing medium incorporated at the distal tip, and a light detector.

A variety of high-quality optical fibers are available commercially for biomedical sensor applications, depending on the analytic wavelength desired. These include plastic, glass, and quartz fibers which cover the optical spectrum from the UV through the visible to the near IR region. On one hand, plastic optical fibers have a larger aperture and are strong, inexpensive, flexible, and easy to work with but have poor UV transmission below 400 nm. On the other hand, glass and quartz fibers have low attenuation and better transmission in the UV but have small apertures, are fragile, and present a potential risk in in vivo applications.

Probe Configurations

There are many different ways to implement fiber optic sensors. Most fiber optic chemical sensors employ either a single-fiber configuration, where light travels to and from the sensing tip in one

fiber, or a double-fiber configuration, where separate optical fibers are used for illumination and detection. A single fiber optic configuration offers the most compact and potentially least expensive implementation. However, additional challenges in instrumentation are involved in separating the illuminating signal from the composite signal returning for processing.

The design of intravascular catheters require special considerations related to the sterility and biocompatibility of the sensor. For example, intravascular fiberoptic sensors must be sterilizable and their material nonthrombogenic and resistant to platelet and protein deposition. Therefore, these catheters are typically made of materials covalently bound with heparin or antiplatelet agents. The catheter is normally introduced into the jugular vein via a peripheral cut-down and a slow heparin flush is maintained until it is removed from the blood.

Optical Fiber Sensors

Advantages cited for fiber optic sensors include their small size and low cost. In contrast to electrical measurements, where the difference of two absolute potentials must be measured, fiber optics are self-contained and do not require an external reference signal. Because the signal is optical, there is no electrical risk to the patient, and there is no direct interference from surrounding electric or magnetic fields. Chemical analysis can be performed in real-time with almost an instantaneous response. Furthermore, versatile sensors can be developed that respond to multiple analytes by utilizing multiwavelength measurements.

Despite these advantages, optical fiber sensors exhibit several shortcomings. Sensors with immobilized dyes and other indicators have limited long-term stability, and their shelf life degrades over time. Moreover, ambient light can interfere with the optical measurement unless optical shielding or special time-synchronous gating is performed.

Indicator-Mediated Transducers

Only a limited number of biochemical analytes have an intrinsic optical absorption that can be measured with sufficient selectivity directly by spectroscopic methods. Other species, particularly hydrogen, oxygen, carbon dioxide, and glucose, which are of primary interest in diagnostic applications, are not susceptible to direct photometry. Therefore, indicator-mediated sensors have been developed using specific reagents that are properly immobilized on the surface of an optical sensor.

The most difficult aspect of developing an optical biosensor is the coupling of light to the specific recognition element so that the sensor can respond selectively and reversibly to a change in the concentration of a particular analyte. In fiber-optic-based sensors, light travels efficiently to the end of the fiber where it exits and interacts with a specific chemical or biologic recognition element that is immobilized at the tip of the fiber optic. These transducers may include indicators and ionophores (i.e., ion-binding compounds) as well as a wide variety of selective polymeric materials. After the light interacts with the sample, the light returns through the same or a different optical fiber to a detector which correlates the degree of change with the analyte concentration.

Typical indicator-mediated fiber-optic-sensor configurations are shown schematically in Fig. 52.2. In (*a*) the indicator is immobilized directly on a membrane positioned at the end of a fiber. An indicator in the form of a powder can be either glued directly onto a membrane, as shown in (*b*), or physically retained in position at the end of the fiber by a special permeable membrane (*c*), a tubular capillary/membrane (*d*), or a hollow capillary tube (*e*).

52.3 General Principles of Optical Sensing

Two major optical techniques are commonly available to sense optical changes at sensor interfaces. These are usually based on evanescent wave and surface plasmon resonance principles.

Evanescent Wave Spectroscopy

When light propagates along an optical fiber, it is not confined to the core region but penetrates to some extent into the surrounding cladding region. In this case, an electromagnetic component of the light penetrates a characteristic distance (on the order of one wavelength) beyond the reflecting surface into the less optically dense medium where it is attenuated exponentially according to Beer-Lambert's law (Fig. 52.3).

The evanescent wave depends on the angle of incidence and the incident wavelength. This phenomenon has been widely exploited to construct different types of optical sensors for biomedical applications. Because of the short penetration depth and the exponential decay of the intensity, the evanescent wave is absorbed mainly by absorbing compounds very close to the surface. In the case of particularly weak absorbing analytes, sensitivity can be enhanced by combining the evanescent wave principle with multiple internal reflections along the sides of an unclad portion of a fiber optic tip.

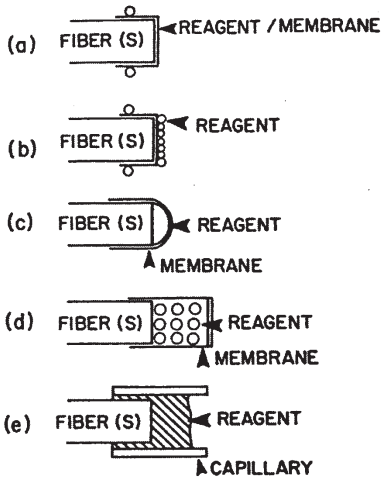


FIGURE 52.2 Typical configuration of different indicator-mediated fiber optic sensor tips (from Otto S. Wolfbeis, Fiber Optic Chemical Sensors and Biosensors, vol. 1, CRC Press, Boca Raton, 1990).

Instead of an absorbing species, a fluorophore can also be used. Light is absorbed by the fluorophore emitting detectable fluorescent light at a higher wavelength, thus providing improved sensitivity. Evanescent wave sensors have been applied successfully to measure the fluorescence of indicators in solution, for pH measurement, and in immunodiagnostics.

Surface Plasmon Resonance

Instead of the dielectric/dielectric interface used in evanescent wave sensors, it is possible to arrange a dielectric/metal/dielectric sandwich layer such that when monochromatic polarized light (e.g., from a laser source) impinges on a transparent medium having a metallized (e.g., Ag or Au) surface, light is absorbed within the plasma formed by the conduction electrons of the metal. This results in a phenomenon known as *surface plasmon resonance* (SPR). When SPR is induced, the effect is

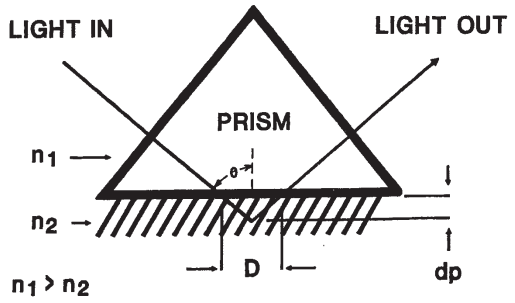


FIGURE 52.3 Schematic diagram of the path of a light ray at the interface of two different optical materials with index of refraction n_1 and n_2. The ray penetrates a fraction of a wavelength (*dp*) beyond the interface into the medium with the smaller refractive index.

observed as a minimum in the intensity of the light reflected off the metal surface.

As is the case with the evanescent wave, an SPR is exponentially decaying into solution with a penetration depth of about 20 nm. The resonance between the incident light and the plasma wave depends on the angle, wavelength, and polarization state of the incident light and the refractive indices of the metal film and the materials on either side of the metal film. A change in the dielectric constant or the refractive index at the surface causes the resonance angle to shift, thus providing a highly sensitive means of monitoring surface reactions.

The method of SPR is generally used for sensitive measurement of variations in the refractive index of the medium immediately surrounding the metal film. For example, if an antibody is bound to or absorbed into the metal surface, a noticeable change in the resonance angle can be readily observed because of the change of the refraction index at the surface, assuming all other parameters are kept constant (Fig. 52.4). The advantage of this concept is the improved ability to detect the direct interaction between antibody and antigen as an interfacial measurement.

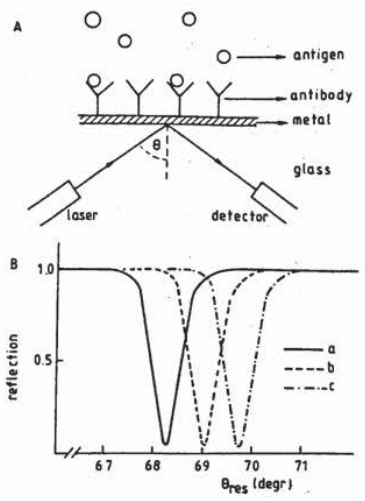


FIGURE 52.4 Surface plasmon resonance at the interface between a thin metallic surface and a liquid (A). A sharp decrease in the reflected light intensity can be observed in (B). The location of the resonance angle is dependent on the refractive index of the material present at the interface.

SPR has been used to analyze immunochemicals and to detect gases. The main limitations of SPR, however, is that the sensitivity depends on the optical thickness of the adsorbed layer, and, therefore, small molecules cannot be measured in very low concentrations.

52.4 Applications

Oximetry

Oximetry refers to the colorimetric measurement of the degree of oxygen saturation, that is, the relative amount of oxygen carried by the hemoglobin in the erythrocytes, by recording the variation in the color of deoxyhemoglobin (Hb) and oxyhemoglobin (HbO_2). A quantitative method for measuring blood oxygenation is of great importance in assessing the circulatory and respiratory status of a patient.

Various optical methods for measuring the oxygen saturation of arterial (SaO_2) and mixed venous (SvO_2) blood have been developed, all based on light transmission through, or reflecting from, tissue and blood. The measurement is performed at two specific wavelengths: λ_1, where there is a large difference in light absorbance between Hb and HbO_2 (e.g., 660 nm red light), and λ_2, which can be an isobestic wavelength (e.g., 805 nm infrared light), where the absorbance of light is independent of blood oxygenation, or a different wavelength in the infrared region $>$ 805 nm, where the absorbance of Hb is slightly smaller than that of HbO_2.

Assuming for simplicity that a hemolyzed blood sample consists of a two-component homogeneous mixture of Hb and HbO_2, and that light absorbance by the mixture of these two components is additive, a simple quantitative relationship can be derived for computing the oxygen saturation of blood:

$$\text{Oxygen saturation} = A - \text{B}\left[\frac{\text{OD}(\lambda_1)}{\text{OD}(\lambda_2)}\right]$$

where A and B are coefficients which are functions of the specific absorptivities of Hb and HbO_2, and OD is the corresponding absorbance (optical density) of the blood.

Since the original discovery of this phenomenon over 50 years ago, there has been progressive development in instrumentation to measure oxygen saturation along three different paths: bench-top oximeters for clinical laboratories, fiber optic catheters for invasive intravascular monitoring, and transcutaneous sensors, which are noninvasive devices placed against the skin.

Intravascular Fiber Optic SvO_2 Catheters

In vivo fiberoptic oximeters were first described in the early 1960s by Polanyi and Heir [1]. They demonstrated that in a highly scattering medium such as blood, where a very short path length is required for a transmittance measurement, a reflectance measurement was practical. Accordingly, they showed that a linear relationship exists between oxygen saturation and the ratio of the infrared-to-red (IR/R) light backscattered from the blood

$$\text{oxygen saturation} = \mathbf{a} - \mathbf{b}(\text{IR/R})$$

where **a** and **b** are catheter-specific calibration coefficients.

Fiber optic SvO_2 catheters consist of two separate optical fibers. One fiber is used for transmitting the light to the flowing blood, and a second fiber directs the backscattered light to a photodetector. In some commercial instruments (e.g., Oximetrix), automatic compensation for hematocrit is employed utilizing three, rather than two, infrared reference wavelengths. Bornzin and coworkers [2] and Mendelson and coworkers [3] described a 5-lumen, 7.5F thermodilution catheter that is comprised of three unequally spaced optical fibers, each fiber 250 μm in diameter, and provides continuous SvO_2 reading with automatic corrections for hematocrit variations (Fig. 52.5).

Intravenous fiberoptic catheters are utilized in monitoring SvO_2 in the pulmonary artery and can be used to indicate the effectiveness of the cardiopulmonary system during cardiac surgery and in the ICU. Several problems limit the wide clinical application of intravascular fiberoptic oximeters. These include the dependence of the individual red and infrared backscattered light intensities and their ratio on hematocrit (especially for SvO_2 below 80%), blood flow, motion artifacts due to catheter tip "whipping" against the blood vessel wall, blood temperature, and pH.

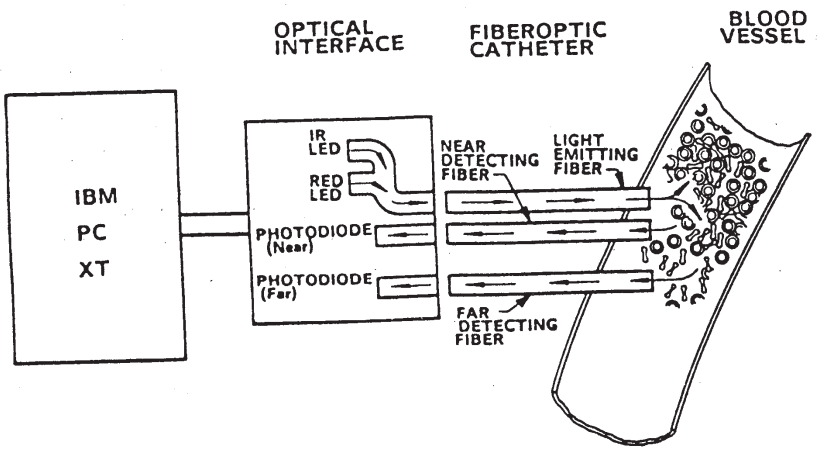


FIGURE 52.5 Principle of a three-fiber optical catheter for SvO_2/HCT measurement [2].

Noninvasive Pulse Oximetry

Noninvasive monitoring of SaO_2 by pulse oximetry is a rapidly growing practice in many fields of clinical medicine [4]. The most important advantage of this technique is the capability to provide continuous, safe, and effective monitoring of blood oxygenation at the patient's bedside without the need to calibrate the instrument before each use.

Pulse oximetry, which was first suggested by Aoyagi and colleagues [5] and Yoshiya and colleagues [6], relies on the detection of the time-variant photoplethysmographic signal, caused by changes in arterial blood volume associated with cardiac contraction. SaO_2 is derived by analyzing only the time-variant changes in absorbance caused by the pulsating arterial blood at the same red and infrared wavelengths used in conventional invasive type oximeters. A normalization process is commonly performed by which the pulsatile (*ac*) component at each wavelength, which results from the expansion and relaxation of the arterial bed, is divided by the corresponding nonpulsatile (*dc*) component of the photoplethysmogram, which is composed of the light absorbed by the blood-less tissue and the nonpulsatile portion of the blood compartment. This effective scaling process results in a normalized red/infrared ratio which is dependent on SaO_2 but is largely independent of the incident light intensity, skin pigmentation, skin thickness, and tissue vasculature.

Pulse oximeter sensors consist of a pair of small and inexpensive red and infrared LEDs and a single, highly sensitive, silicone photodetector. These components are mounted inside a reusable rigid spring-loaded clip, a flexible probe, or a disposable adhesive wrap (Fig. 52.6). The majority of the commercially available sensors are of the transmittance type in which the pulsatile arterial bed, e.g., ear lobe, fingertip, or toe, is positioned between the LEDs and the photodetector. Other probes are available for reflectance (backscatter) measurement where both the LEDs and photodetector are mounted side-by-side facing the skin [7, 8].

Noninvasive Cerebral Oximetry

Another substance whose optical absorption in the near infrared changes corresponding to its reduced and oxidized state is cytochrome aa3, the terminal member of the respiratory chain. Although the concentration of cytochrome aa3 is considerably lower than that of hemoglobin, advanced instrumentation including time-resolved spectroscopy and differential measurements is being used successfully to obtain noninvasive measurements of hemoglobin saturation and cytochrome aa3 by transilluminating areas of the neonatal brain [9–11].

FIGURE 52.6 Disposable finger probe of a noninvasive pulse oximeter.

Blood Gases

Frequent measurement of blood gases, i.e., oxygen partial pressure (pO_2), carbon dioxide partial pressure (pCO_2), and pH, is essential to clinical diagnosis and management of respiratory and metabolic problems in the operating room and the ICU. Considerable effort has been devoted over the last two decades to developing disposable extracorporeal and in particular intravascular fiber optic sensors which can be used to provide continuous information on the acid-base status of a patient.

In the early 1970s, Lubbers and Opitz [12] originated what they called *optodes* (from the Greek, *optical path*) for measurements of important physiologic gases in fluids and in gases. The principle upon which these sensors was designed was a closed cell containing a fluorescent indicator in solution, with a membrane permeable to the analyte of interest (either ions or gases) constituting one of the cell walls. The cell was coupled by optical fibers to a system that measured the fluorescence in the cell. The cell solution would equilibrate with the pO_2 or pCO_2 of the medium placed against it, and the fluorescence of an indicator reagent in the solution would correspond to the partial pressure of the measured gas.

Extracorporeal Measurement

Following the initial feasibility studies of Lubbers and Opitz, Cardiovascular Devices (CDI, USA) developed a GasStat extracorporeal system suitable for continuous online monitoring of blood gases ex vivo during cardiopulmonary bypass operations. The system consists of a disposable plastic sensor connected inline with a blood loop through a fiber optic cable. Permeable membranes separate the flowing blood from the sensor chemistry. The CO_2-sensitive indicator consists of a fine emulsion of a bicarbonate buffer in a two-component silicone. The pH-sensitive indicator is a cellulose material to which hydroxypyrene trisulfonate (HPTS) is bonded covalently. The O_2-sensitive chemistry is composed of a solution of oxygen-quenching decacyclene in a one-component silicone covered with a thin layer of black PTFE for optical isolation and to render the measurement insensitive to the halothane anesthetic.

The extracorporeal device has two channels, one for arterial blood and the other for venous blood, and is capable of recording the temperature of the blood for correcting the measurements to 37°C. Several studies have been conducted comparing the specifications of the GasStat with that of intermittent blood samples analyzed on bench-top blood gas analyzers [13–15].

Intravascular Catheters

During the past decade, numerous efforts have been made to develop integrated fiber optic sensors for intravascular monitoring of blood gases. A few commercial systems for monitoring blood gases and pH are currently undergoing extensive clinical testing. Recent literature reports of sensor performance show that considerable progress has been made mainly in improving the accuracy and reliability of these intravascular blood gas sensors [16–19].

Most fiber optic intravascular blood gas sensors employ either a single- or a double-fiber configuration. Typically, the matrix containing the indicator is attached to the end of the optical fiber as illustrated in Fig. 52.7. Since the solubility of O_2 and CO_2 gases, as well as the optical properties of the sensing chemistry itself, is affected by temperature variations, fiber optic intravascular sensors include a thermocouple or thermistor wire running alongside the fiber optic cable to monitor and correct for temperature fluctuations near the sensor tip. A nonlinear response is characteristic of most chemical indicator sensors, so they are designed to match the concentration region of the intended application. Also, the response time of the optode is somewhat slower compared to electrochemical sensors.

Intravascular fiber optic blood gas sensors are normally placed inside a standard 20-gauge catheter, which is sufficiently small to allow adequate spacing between the sensor and the catheter wall. The resulting lumen is large enough to permit the withdrawal of blood samples, introduction

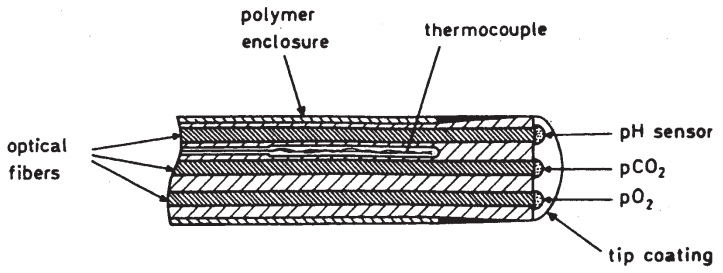


FIGURE 52.7 Principle diagram of an integrated fiber optic blood gas catheter (from Otto S. Wolfbeis, Fiber Optic Chemical Sensors and Biosensors, vol. 2, CRC Press, Boca Raton, 1990).

of a continuous heparin flush, and the recording of a blood pressure waveform. In addition, the optical fibers are encased in a protective tubing to contain any fiber fragments in case they break off.

pH Sensors

In 1976, Peterson and coworkers [20] originated the development of the first fiber optic chemical sensor for physiological pH measurement. The basic idea was to contain a reversible color-changing indicator at the end of a pair of optical fibers. The indicator, phenol red, was covalently bound to a hydrophilic polymer in the form of water-permeable microbeads. This technique stabilized the indicator concentration. The indicator beads were contained in a sealed hydrogen-ion-permeable envelope made out of a hollow cellulose tubing. In effect, this formed a miniature spectrophotometric cell at the end of the fibers and represented an early prototype of a fiber optic chemical sensor.

The phenol red dye indicator is a weak organic acid, and the acid form (un-ionized) and base form (ionized) are present in a concentration ratio determined by the ionization constant of the acid and the pH of the medium according to the familiar Henderson-Hasselbalch equation. The two forms of the dye have different optical absorption spectra, so the relative concentration of one of the forms, which varies as a function of pH, can be measured optically and related to variations in pH. In the pH sensor, green (560 nm) and red (longer than 600 nm) light emerging form the end of one fiber passes through the dye and is reflected back into the other fiber by light-scattering particles. The green light is absorbed by the base form of the indicator. The red light is not absorbed by the indicator and is used as an optical reference. The ratio of green to red light is measured and is related to pH by an *S*-shaped curve with an approximate high-sensitivity linear region where the equilibrium constant (pK) of the indicator matches the pH of the solution.

The same principle can also be used with a reversible fluorescent indicator, in which case the concentration of one of the indicator forms is measured by its fluorescence rather than absorbance intensity. Light in the blue or UV wavelength region excites the fluorescent dye to emit longer wavelength light, and the two forms of the dye may have different excitation or emission spectra to allow their distinction.

The original instrument design for a pH measurement was very simple and consisted of a tungsten lamp for fiber illumination, a rotating filter wheel to select the green and red light returning from the fiber optic sensor, and signal processing instrumentation to give a pH output based on the green-to-red ratio. This system was capable of measuring pH in the physiologic range between 7.0–7.4 with an accuracy and precision of 0.01 pH units. The sensor was susceptible to ionic strength variation in the order of 0.01 pH unit per 11% change in ionic strength.

Further development of the pH probe for practical use was continued by Markle and colleagues [21]. They designed the fiber optic probe in the form of a 25-gauge (0.5 mm o.d.) hypodermic needle, with an ion-permeable side window, using 75-mm-diameter plastic optical fibers. The sensor had a

90% response time of 30 s. With improved instrumentation and computerized signal processing, and with a three-point calibration, the range was extended to ±3 pH units, and a precision of 0.001 pH units was achieved.

Several reports have appeared suggesting other dye indicator systems that can be used for fiber optic pH sensing [22]. A classic problem with dye indicators is the sensitivity of their equilibrium constant to ionic strength. To circumvent this problem, Wolfbeis and Offenbacher [23] and Opitz and Lubbers [24] demonstrated a system in which a dual sensor arrangement can measure ionic strength and pH and simultaneously can correct the pH measurement for variations in ionic strength.

pCO_2 Sensors

The pCO_2 of a sample is typically determined by measuring changes in the pH of a bicarbonate solution which is isolated from the sample by a CO_2-permeable membrane but remains in equilibrium with the CO_2. The bicarbonate and CO_2, as carbonic acid, form a pH buffer system, and, by the Henderson-Hasselbalch equation, hydrogen ion concentration is proportional to the pCO_2 in the sample. This measurement is done with either a pH electrode or a dye indicator in solution.

Vurek [25] demonstrated that the same techniques can be used also with a fiber optic sensor. In his design, one plastic fiber carries light to the transducer, which is made of a silicone rubber tubing about 0.6 mm in diameter and 1.0 mm long, filled with a phenol red solution in a 35-mM bicarbonate. Ambient pCO_2 controls the pH of the solution which changes the optical absorption of the phenol red dye. The CO_2 permeates through the rubber to equilibrate with the indicator solution. A second optical fiber carries the transmitted signal to a photodetector for analysis. The design by Zhujun and Seitz [26] uses a pCO_2 sensor based on a pair of membranes separated from a bifurcated optical fiber by a cavity filled with bicarbonate buffer. The external membrane is made of silicone, and the internal membrane is HPTS immobilized on an ion-exchange membrane.

pO_2 Sensors

The development of an indicator system for fiber optic pO_2 sensing is challenging because there are very few known ways to measure pO_2 optically. Although a color-changing indicator would have been desirable, the development of a sufficiently stable indicator has been difficult. The only principle applicable to fiber optics appears to be the quenching effect of oxygen on fluorescence.

Fluorescence quenching is a general property of aromatic molecules, dyes containing them, and some other substances. In brief, when light is absorbed by a molecule, the absorbed energy is held as an excited electronic state of the molecule. It is then lost by coupling to the mechanical movement of the molecule (heat), reradiated from the molecule in a mean time of about 10 ns (fluorescence), or converted into another excited state with much longer mean lifetime and then reradiated (phosphorescence). Quenching reduces the intensity of fluorescence and is related to the concentration of the quenching molecules, such as O_2.

A fiber optic sensor for measuring pO_2 using the principle of fluorescence quenching was developed by Peterson and colleagues [27]. The dye is excited at around 470 nm (blue) and fluoresces at about 515 nm (green) with an intensity that depends on the pO_2. The optical information is derived from the ratio of green fluorescence to the blue excitation light, which serves as an internal reference signal. The system was chosen for visible light excitation, because plastic optical fibers block light transmission at wavelengths shorter than 450 nm, and glass fibers were not considered acceptable for biomedical use.

The sensor was similar in design to the pH probe continuing the basic idea of an indicator packing in a permeable container at the end of a pair of optical fibers. A dye perylene dibutyrate, absorbed on a macroreticular polystyrene adsorbent, is contained in a oxygen-permeable porous polystyrene envelope. The ratio of green to blue intensity was processed according to the Stren-Volmer equation

$$\frac{I_0}{I} = 1 + K\,pO_2$$

where I and I_0 are the fluorescence emission intensities in the presence and absence of quencher, respectively, and K is the Stern-Volmer quenching coefficient. This provides a nearly linear readout of pO_2 over the range of 0–150 mmHg (0–20 kPa), with a precision of 1 mm Hg (0.13 kPa). The original sensor was 0.5 mm in diameter, but it can be made much smaller. Although its response time in a gas mixture is a fraction of a second, it is slower in an aqueous system, about 1.5 min for 90% response.

Wolfbeis and coworkers [28] designed a system for measuring the widely used halothane anesthetic which interferes with the measurement of oxygen. This dual-sensor combination had two semipermeable membranes (one of which blocked halothane) so that the probe could measure both oxygen and halothane simultaneously. The response time of their sensor, 15–20 s for halothane and 10–15 s for oxygen, is considered short enough to allow gas analysis in the breathing circuit. Potential applications of this device include the continuous monitoring of halothane in breathing circuits and in the blood.

Glucose Sensors

Another important principle that can be used in fiber optic sensors for measurements of high sensitivity and specificity is the concept of competitive binding. This was first described by Schultz, Mansouri, and Goldstein [29] to construct a glucose sensor. In their unique sensor, the analyte (glucose) competes for binding sites on a substrate (the lectin concanavalin A) with a fluorescent indicator-tagged polymer [fluorescein isothiocyanate (FITC)-dextran]. The sensor, which is illustrated in Fig. 52.8, is arranged so that the substrate is fixed in a position out of the optical path of the fiber end. The substrate is bound to the inner wall of a glucose-permeable hollow fiber tubing (300 μ O.D.×200 μ I.D.) and fastened to the end of an optical fiber. The hollow fiber acts as the container and is impermeable to the large molecules of the fluorescent indicator. The light beam that extends from the fiber "sees" only the unbound indicator in solution inside the hollow fiber but not the indicator bound on the container wall. Excitation light passes through the fiber and into the solution, fluorescing the unbound indicator, and the fluorescent light passes back along the same fiber to a measuring system. The fluorescent indicator and the glucose are in competitive binding equilibrium with the substrate. The interior glucose concentration equilibrates with its concentration exterior to the probe. If the glucose concentration increases, the indicator is driven off the substrate to increase the concentration of the indicator. Thus, fluorescence intensity as seen by the optical fiber follows the glucose concentration.

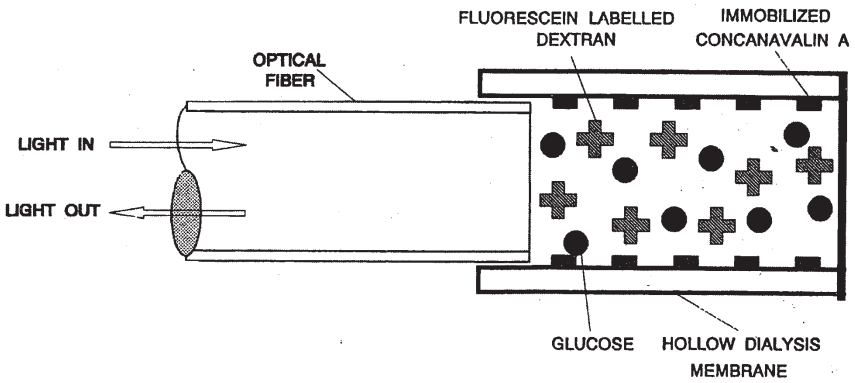


FIGURE 52.8 Schematic diagram of a competitive binding fluorescence affinity sensor for glucose measurement [29].

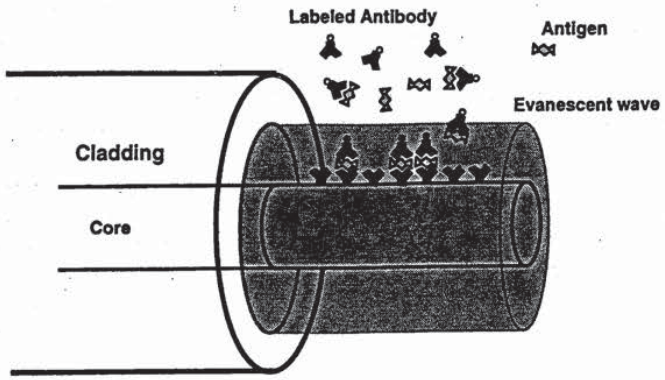


FIGURE 52.9 Basic principle of a fiber optic antigen-antibody sensor [33].

The response time of the sensor was found to be about 5 min. In vivo studies demonstrated fairly close correspondence between the sensor output and actual blood glucose levels. A time lag of about 5 min was found and is believed to be due to the diffusion of glucose across the hollow fiber membrane and the diffusion of FTIC-dextran within the tubing.

In principle, the concept of competitive binding can be applied to any analysis for which a specific reaction can be devised. However, long-term stability of these sensors remains the major limiting factor that needs to be solved.

Immunosensors

Immunologic techniques offer outstanding selectivity and sensitivity through the process of antibody-antigen interaction. This is the primary recognition mechanism by which the immune system detects and fights foreign matter and has therefore allowed the measurement of many important compounds at trace levels in complex biologic samples.

In principle, it is possible to design competitive binding optical sensors utilizing immobilized antibodies as selective reagents and detecting the displacement of a labeled antigen by the analyte. Therefore, antibody-based immunologic optical sensors have been the subject of considerable research in the past few years [30–34]. In practice, however, the strong binding of antigens to antibodies and vice versa causes difficulties in constructing reversible sensors with fast dynamic responses.

Several immunologic sensors based on fiber optic waveguides have been demonstrated for monitoring antibody-antigen reactions. Typically, several centimeters of cladding are removed along the fiber's distal end, and the recognition antibodies are immobilized on the exposed core surface. These antibodies bind fluorophore-antigen complexes within the evanescent wave as illustrated in Fig. 52.9. The fluorescent signal excited within the evanescent wave is then transmitted through the cladded fiber to a fluorimeter for processing.

Experimental studies have indicated that immunologic optical sensors can generally detect micromolar and even picomolar concentrations. However, the major obstacle that must be overcome to achieve high sensitivity in immunologic optical sensors is the nonspecific binding of immobilized antibodies.

References

1. Polanyi ML, Heir RM. 1962. In vivo oximeter with fast dynamic response. Rev Sci Instrum 33:1050.

2. Bornzin GA, Mendelson Y, Moran BL, et al. 1987. Measuring oxygen saturation and hematocrit using a fiberoptic catheter. Proc 9th Ann Conf Eng Med Bio Soc 807–809.
3. Mendelson Y, Galvin JJ, Wang Y. 1990. In vitro evaluation of a dual oxygen saturation/hematocrit intravascular fiberoptic catheter. Biomed Instrum Tech 24:199.
4. Mendelson Y. 1992. Pulse oximetry: Theory and applications for noninvasive monitoring. Clin Chem 28(9): 1601.
5. Aoyagi T, Kishi M, Yamaguchi K, et al. 1974. Improvement of the earpiece oximeter. Jpn Soc Med Electron Biomed Eng 90–91.
6. Yoshiya I, Shimada Y, Tanaka K. 1980. Spectrophotometric monitoring of arterial oxygen saturation in the fingertip. Med Biol Eng Comput 18:27.
7. Mendelson Y, Solomita MV. 1992. The feasibility of spectrophotometric measurements of arterial oxygen saturation from the scalp utilizing noninvasive skin reflectance pulse oximetry. Biomed Instrum Technol 26:215.
8. Mendelson Y, McGinn MJ. 1991. Skin reflectance pulse oximetry: In vivo measurements from the forearm and calf. J Clin Monit 7:7.
9. Chance B, Leigh H, Miyake H, et al. 1988. Comparison of time resolved and un-resolved measurements of deoxyhemoglobin in brain. Proc Nat Acad Sci 85:4971.
10. Jobsis FF, Keizer JH, LaManna JC, et al. 1977. Reflection spectrophotometry of cytochrome aa3 in vivo. Appl Physiol: Respirat Environ Excerc Physiol 43(5): 858.
11. Kurth CD, Steven IM, Benaron D, et al. 1993. Near-infrared monitoring of the cerebral circulation. J Clin Monit 9:163.
12. Lubbers DW, Opitz N. 1975. The pCO_2/pO_2-optode: A new probe for measurement of pCO_2 or pO_2 in fluids and gases. Z Naturforsch C: Biosci 30C:532.
13. Clark CL, O'Brien J, McCulloch J, et al. 1986. Early clinical experience with GasStat. J Extra Corporeal Technol 18:185.
14. Hill AG, Groom RC, Vinansky RP, et al. 1985. On-line or off-line blood gas analysis: Cost vs. time vs. accuracy. Proc Am Acad Cardiovasc Perfusion 6:148.
15. Siggaard-Andersen O, Gothgen IH, Wimberley, et al. 1988. Evaluation of the GasStat fluorescence sensors for continuous measurement of pH, pCO_2 and pO_2 during CPB and hypothermia. Scand J Clin Lab Invest 48 (Suppl. 189):77.
16. Zimmerman JL, Dellinger RP. 1993. Initial evaluation of a new intra-arterial blood gas system in humans. Crit Care Med 21(4):495.
17. Gottlieb A. 1992. The optical measurement of blood gases—approaches, problems and trends: Fiber optic medical and fluorescent sensors and applications. Proc SPIE 1648:4.
18. Barker SL, Hyatt J. 1991. Continuous measurement of intraarterial pHa, $PaCO_2$, and PaO_2 in the operation room. Anesth Analg 73:43.
19. Larson CP, Divers GA, Riccitelli SD. 1991. Continuous monitoring of PaO_2 and $PaCO_2$ in surgical patients. Abstr Crit Care Med 19:525.
20. Peterson JI, Goldstein SR, Fitzgerald RV. 1980. Fiber optic pH probe for physiological use. Anal Chem 52:864.
21. Markle DR, McGuire DA, Goldstein SR, et al. 1981. A pH measurement system for use in tissue and blood, employing miniature fiber optic probes, In DC Viano (ed), Advances in Bioengineering, p 123, New York, American Society of Mechanical Engineers.
22. Wolfbeis OS, Furlinger E, Kroneis H, et al. 1983. Fluorimetric analysis: 1. A study on fluorescent indicators for measuring near neutral (physiological) pH values. Fresenius' Z Anal Chem 314:119.
23. Wolfbeis OS, Offenbacher H. 1986. Fluorescence sensor for monitoring ionic strength and physiological pH values. Sens Actuators 9:85.
24. Opitz N, Lubbers DW. 1983. New fluorescence photomatrical techniques for simultaneous and continuous measurements of ionic strength and hydrogen ion activities. Sens Actuators 4:473.

25. Vurek GG, Feustel PJ, Severinghaus JW. 1983. A fiber optic pCO_2 sensor. Ann Biomed Eng 11:499.
26. Zhujun Z, Seitz WR. 1984. A carbon dioxide sensor based on fluorescence. Anal Chim Acta 160:305.
27. Peterson JI, Fitzgerald RV, Buckhold DK. 1984. Fiber-optic probe for in vivo measurement of oxygen partial pressure. Anal Chem 56:62.
28. Wolfbeis OS, Posch HE, Kroneis HW. 1985. Fiber optical fluorosensor for determination of halothane and/or oxygen. Anal Chem 57:2556.
29. Schultz JS, Mansouri S, Goldstein IJ. 1982. Affinity sensor: A new technique for developing implantable sensors for glucose and other metabolites. Diabetes Care 5:245.
30. Andrade JD, Vanwagenen RA, Gregonis DE, et al. 1985. Remote fiber optic biosensors based on evanescent-excited fluoro-immunoassay: Concept and progress. IEEE Trans Electron Devices ED-32: 1175.
31. Sutherland RM, Daehne C, Place JF, et al. 1984. Optical detection of antibody-antigen reactions at a glass-liquid interface. Clin Chem 30:1533.
32. Hirschfeld TE, Block MJ. 1984. Fluorescent immunoassay employing optical fiber in a capillary tube. US Patent No. 4,447,546.
33. Anderson GP, Golden JP, Ligler FS. 1994. An evanescent wave biosensor: Part I. Fluorescent signal acquisition from step-etched fiber optic probes. IEEE Trans Biomes Eng 41(6):578.
34. Golden JP, Anderson GP, Rabbany SY, et al. 1994. An evanescent wave biosensor: Part II. Fluorescent signal acquisition from tapered fiber optic probes. IEEE Trans Biomed Eng 41(6):585.

53

Bioanalytic Sensors

Richard P. Buck
University of North Carolina

53.1 Classification of Biochemical Reactions in the Context of Sensor Design and Development

Introduction and Definitions

Since sensors generate a measurable material property, they belong in some grouping of transducer devices. Sensors specifically contain a recognition process that is characteristic of a material sample at the molecular-chemical level, and a sensor incorporates a transduction process (step) to create a useful signal. Biomedical sensors include a whole range of devices that may be chemical sensors, physical sensors, or some kind of mixed sensor.

Chemical sensors use chemical processes in the recognition and transduction steps. Biosensors are also chemical sensors, but they use particular classes of biological recognition/transduction processes. A pure physical sensor generates and transduces a parameter that does not depend on the chemistry per se but is a result of the sensor responding as an aggregate of point masses or charges. All these when used in a biologic system (biomatrix) may be considered *bioanalytic* sensors without regard to the chemical, biochemical, or physical distinctions. They provide an "analytic signal of the biologic system" for some further use.

The chemical recognition process focuses on some molecular-level chemical entity, usually a kind of chemical structure. In classical analysis this structure may be a simple functional group: SiO^- in a glass electrode surface, a chromophore in an indicator dye, or a metallic surface structure, such as silver metal that recognizes Ag^+ in solution. In recent times, the biologic recognition processes have been better understood, and the general concept of recognition by *receptor* or *chemoreceptor* has come into fashion. Although these are often large molecules bound to cell membranes, they contain specific structures that permit a wide variety of different molecular recognition steps including recognition of large and small species and of charged and uncharged species. Thus, *chemoreceptor* appears in the sensor literature as a generic term for the principal entity doing the recognition. For a history and examples, see references [1–6].

0-8493-8346-3/95/$0.00+$.50

Biorecognition in biosensors has especially stressed "receptors" and their categories. Historically, application of receptors has not necessarily meant measurement directly of the receptor. Usually there are coupled chemical reactions, and the transduction has used measurement of the subsidiary products: change of pH, change of dissolved 0_2, generation of H_2O_2, changes of conductivity, changes of optical adsorption, and changes of temperature. Principal receptors are enzymes because of their extraordinary selectivity. Other receptors can be the more subtle species of biochemistry: antibodies, organelles, microbes, and tissue slices, not to mention the trace level "receptors" that guide ants such as pheromones and other unusual species. A sketch of a generic bioanalytic sensor is shown in Fig. 53.1.

Classification of Recognition Reactions and Receptor Processes

The concept of *recognition* in chemistry is universal. It almost goes without saying that all chemical reactions involved recognition and selection on the basis of size, shape, and charge. For the purpose of constructing sensors, general recognition based on these factors is not usually enough. Frequently in inorganic chemistry a given ion will react indiscriminantly with similar ions of the same size and charge. Changes in charge from unity to two, for example, do change the driving forces of some ionic reactions. By control of dielectric constant of phases, heterogeneous reactions can often be "tailored" to select divalent ions over monovalent ions and to select small versus large ions or vice versa.

Shape, however, has more special possibilities, and natural synthetic methods permit product control. Nature manages to use shape together with charge to build organic molecules, called enzymes, that have acquired remarkable selectivity. It is in the realm of biochemistry that these natural constructions are investigated and catalogued. Biochemistry books list large numbers of enzymes and other selective materials that direct chemical reactions. Many of these have been tried as the basis of selective sensors for bioanalytic and biomedical purposes. The list in Table 53.1 shows how some of these materials can be grouped into lists according to function and to analytic substrate, both organic and inorganic. The principles seem general, so there is no reason to discriminate against the inorganic substrates in favor or the organic substrates. All can be used in biomedical analysis.

53.2 Classification of Transduction Processes—Detection Methods

Some years ago, the engineering community addressed the topic of sensor classification—Richard M. White in IEEE Trans. Ultra., Ferro., Freq. Control (UFFC), UFFC-34 (1987) 124, and Wen E. Ko in IEEE/EMBS Symposium Abstract T.1.1 84CH2068-5 (1984). It is interesting because the physical

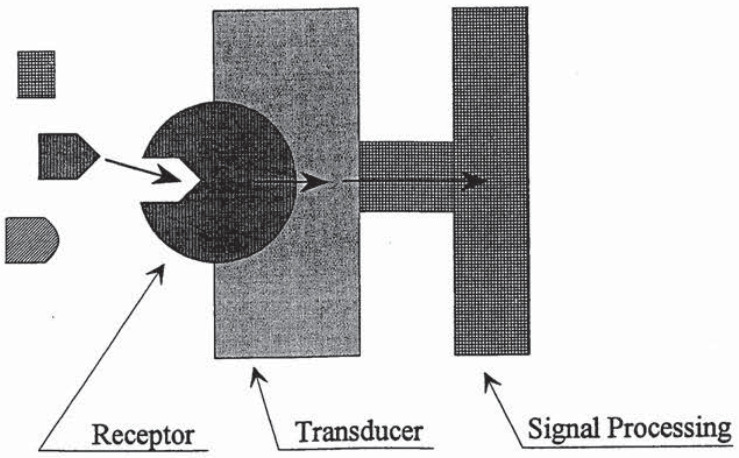


FIGURE 53.1 Generic bioanalytic sensor.

TABLE 53.1 Recognition Reactions and Receptor Processes

1. Insoluble salt-based sensors

a. $S^{+} + R^{-} \rightleftharpoons$ (insoluble salt)

Ion exchange with crystalline SR (homogeneous or heterogeneous crystals)

	chemical signal S^{+n}	receptor R^{-n}
	inorganic cations	inorganic anions
examples:	Ag^{+}, Hg_2^{2+}, Pb^{2+}, Cd^{2+}, Cu^{2+}	$S^{=}$, $Se^{2=}$, SCN^{-}, I^{-}, Br^{-}, Cl^{-}

b. $S^{-n} + R^{+n} \rightleftharpoons SR$ (insoluble salt)

Ion exchange with crystalline SR (homogeneous or heterogeneous crystals)

	chemical signal S^{-n}	receptor R^{+n}
	inorganic anions	inorganic cations
examples:	F^{-}, $S^{=}$, $Se^{2=}$, SCN^{-}, I^{-}, Br^{-}, Cl^{-}	LaF_2^{+}, Ag^{+}, Hg_2^{2+}, Pb^{2+}, Cd^{2+}, Cu^{2+}

2. Solid ion exchangers

a. $S^{+n} + R^{-n}$ (sites) $\rightleftharpoons S^{+n}R^{-n} = SR$ (in ion exchanger phase)

Ion exchange with synthetic ion exchangers containing negative fixed sites (homogeneous or heterogeneous, inorganic or organic materials)

	chemical signal S^{+n}	receptor R^{-n}
	inorganic and organic ions	inorganic and organic ion sites
examples:	H^{+}, Na^{+}, K^{+}	silicate glass Si-0^{-}
	H^{+}, Na^{+}, K^{+}, other M^{+n}	synthetic sulfonated, phosphorylated, EDTA-substituted polystyrenes

b. $S^{-n} + R^{+n}$(sites) $\rightleftharpoons S^{-n}R^{+n} = SR$ (in ion exchanger phase)

Ion exchange with synthetic ion exchangers containing positive fixed sites (homogeneous or heterogeneous, inorganic or organic materials)

	chemical signal S^{-n}	receptor R^{+n}
	organic and inorganic ions	organic and inorganic ion sites
examples:	hydrophobic anions	quaternized polystyrene

3. Liquid ion exchanger sensors with electrostatic selection

a. $S^{+n} + R^{-n}$(sites) $\rightleftharpoons S^{+n}R^{-n} = SR$ (in ion exchanger phase)

Plasticized, passive membranes containing mobile, trapped negative fixed sites (homogeneous or heterogeneous, inorganic or organic materials)

	chemical signal S^{+n}	receptor R^{-n}
	inorganic and organic ions	inorganic and organic ion sites
examples:	Ca^{2+}	diester of phosphoric acid or monoester of a phosphonic acid
	M^{+n}	dinonylnaphthalene sulfonate and other organic, hydrophobic anions
	$R_1R_2R_3R_4N^{+}$ and bis-Quaternary Cations cationic drugs tetrasubstituted arsonium^{+}	tetraphenylborate anion or substituted derivatives

b. $S^{-n} + R^{+n}$(sites) $\rightleftharpoons S^{-n}R^{+n} = SR$ (in ion exchanger phase)

Plasticized, passive membranes containing mobile, trapped negative fixed sites (homogeneous or heterogeneous, inorganic or organic materials)

	chemical signal S^{-n}	receptor R^{+n}
	inorganic and organic ions	inorganic and organic ion sites
examples:	anions, simple Cl^{-}, Br^{-}, ClO_4^{-}	quaternary ammonium cations: e.g. tridodecylmethyl-ammonium
	anions, complex, drugs	quaternary ammonium cations: e.g. tridodecylmethyl-ammonium

4. Liquid ion exchanger sensors with neutral (or charged) carrier selection

a. $S^{+n} + X$ and R^{-n}(sites) $\rightleftharpoons S^{+n}X\,R^{-n} = SXR$ (in ion exchanger phase)

Plasticized, passive membranes containing mobile, trapped negative fixed sites (homogeneous or heterogeneous, inorganic or organic materials)

	chemical signal S^{+n}	receptor R^{-n}
	inorganic and organic ions	inorganic and organic ion sites
examples:	Ca^{2+}	X = synthetic ionophore complexing agent selective to Ca^{2+}
		R^{-n} usually a substituted tetra phenylborate salt
	Na^{+}, K^{+}, H^{+}	X = selective ionophore complexing agent

(continued)

TABLE 53.1 Recognition Reactions and Receptor Processes *(continued)*

b. $S^{-n} + X$ and R^{+n}(sites) $\rightleftharpoons$ $S^{-n}X\,R^{+n} = SXR$ (in ion exchanger phase)

Plasticized, passive membranes containing mobile, trapped negative fixed sites (homogeneous or heterogeneous, inorganic or organic materials)

	chemical signal S^{-n}	receptor R^{+n}
	inorganic and organic ions	inorganic and organic ion sites
examples:	HPO_4^{2-}	R^{+n} = quaternary ammonium salt X = synthetic ionophore complexing agent: aryl organotin compound or suggested cyclic polyamido-polyamines
	HCO_3^-	X = synthetic ionophore: trifluoro acetophenone
	Cl^-	X = aliphatic organotin compound

5. Bioaffinity sensors based on change of local electron densities

$$S + R \rightleftharpoons SR$$

chemical signal S	receptor R
protein	dyes
saccharide	lectin
glycoprotein	
substrate	enzyme
inhibitor	
	Transferases Hydrolases (peptidases, esterases, etc.) Lyases Isomerases Ligases
prosthetic group	apoenzyme
antigen	antibody
hormone	"receptor"
sustrate analogue	transport system

6. Metabolism sensors based on substrate consumption and product formation

$$S + R \rightleftharpoons SR \rightarrow P + R$$

	chemical signal S	receptor R
	substrate	enzyme
examples:	lactate (SH_2)	hydrogenases catalyze hydrogen transfer from S to acceptor A (not molecular oxygen!) reversibly
	$SH_2 + A \rightleftharpoons S + AH_2$ lactate + NAD^+	pyruvate + NADH + H^+ using lactate dehydrogenase
	glucose (SH_2)	oxidases catalyze hydrogen transfer to molecular oxygen
	$SH_2 + \frac{1}{2}O_2 \rightleftharpoons S + H_2O$ or $SH_2 + O_2 \rightleftharpoons S + H_2O_2$ glucose + $O_2 \rightleftharpoons$ gluconolactone + H_2O_2	using glucose oxidase
	reducing agents (S) $2S + 2H^+ + H_2O_2 \rightleftharpoons 2S^+ + 2H_2O$ $Fe^{2+} + H_2O_2 + 2H^+ \rightleftharpoons Fe^{3+} + 2H_2O$	peroxidases catalyze oxidation of a substrate by H_2O_2 using horseradish peroxidase
	reducing agents L-lactate + $O_2 \rightleftharpoons$ acetate + $CO_2 + H_2O$	oxygenases catalyze substrate oxidations by molecular O_2
	cofactor	organelle
	inhibitor	microbe
	activator	tissue slice
	enzyme activity	

7. Coupled and hybrid systems using sequences, competition, anti-interference and amplification concepts and reactions
8. Biomimetic sensors

chemical signal S	receptor R
sound stress light	carrier-enzyme

Source: Adapted from [2, 6].

and chemical properties are given equal weight. There are many ideas given here that remain without embodiment. This list is reproduced as Table 53.2. Of particular interest in this section are "detection means used in sensors" and "sensor conversion phenomena." At present the principle transduction schemes use electrochemical, optical, and thermal detection effects and principles.

TABLE 53.2 Detection Means and Conversion Phenomena Used in Sensors

Detection means
Biologic
Chemical
Electric, magnetic, or electromagnetic wave
Heat, temperature
Mechanical displacement of wave
Radioactivity, radiation
Other
Conversion phenomena
Biologic
Biochemical transformation
Physical transformation
Effect on test organism
Spectroscopy
Other
Chemical
Chemical transformation
Physical transformation
Electrochemical process
Spectroscopy
Other
Physical
Thermoelectric
Photoelectric
Photomagnetic
Magnetoelectric
Elastomagnetic
Thermoelastic
Elastoelectric
Thermomagnetic
Thermooptic
Photoelastic
Other

Calorimetric, Thermometric, and Pyroelectric Transducers

Especially useful for enzymatic reactions, the generation of heat (enthalpy change) can be used easily and generally. The enzyme provides the selectivity and the reaction enthalpy cannot be confused with other reactions from species in a typical biologic mixture. The ideal aim is to measure total evolved heat, i.e., to perform a calorimetric measurement. In real systems there is always heat loss, i.e., heat is conducted away by the sample and sample container so that the process cannot be adiabatic as required for a total heat evolution measurement. As a result, temperature difference before and after evolution is measured most often. It has to be assumed that the heat capacity of the specimen and container is constant over the small temperature range usually measured.

The simplest transducer is a thermometer coated with the enzyme that permits the selected reaction to proceed. Thermistors are used rather than thermometers or thermocouples. The change of resistance of certain oxides is much greater than the change of length of a mercury column or the microvolt changes of thermocouple junctions.

Pyroelectric heat flow transducers are relatively new. Heat flows from a heated region to a lower temperature region, controlled to occur in one dimension. The lower temperature side can be coated with an enzyme. When the substrate is converted, the lower temperature side is warmed. The pyroelectric material is from a category of materials that develops a spontaneous voltage difference in a thermal gradient. If the gradient is disturbed by evolution or adsorption of heat, the voltage temporarily changes.

In biomedical sensing, some of the solid-state devices based on thermal sensing cannot be used effectively. The reason is that the sensor itself has to be heated or is heated quite hot by catalytic surface reactions. Thus pellistors (oxides with catalytic surfaces and embedded platinum wire thermometer), chemiresistors, and "Figaro" sensor "smoke" detectors have not found many biologic applications.

Optical, Optoelectronic Transducers

Most optical detection systems for sensors are small, i.e., they occupy a small region of space because the sample size and volume are themselves small. This means that common absorption spectrophotometers and photofluorometers are not used with their conventional sample-containing cells or with their conventional beam-handling systems. Instead light-conducting *optical fibers* are used to connect the sample with the more remote monochromator and optical readout system. The

techniques still remain absorption spectrophotometry, fluorimetry including fluorescence quenching, and reflectometry.

The most widely published optical sensors use a miniature reagent contained or immobilized at the tip of an optical fiber. In most systems a permselective membrane coating allows the detected species to penetrate the dye region. The corresponding absorption change, usually at a sensitive externally preset wavelength, is changed and correlated with the sample concentration. Similarly, fluorescence can be stimulated by the higher-frequency external light source and the lower-frequency emission detected. Some configurations are illustrated in references [1, 2]. Fluorimetric detection of coenzyme A, NAD^+/NADH, is involved in many so-called pyridine-linked enzyme systems. The fluorescence of NADH contained or immobilized can be a convenient way to follow these reactions. Optodes, miniature encapsulated dyes, can be placed in vivo. Their fluorescence can be enhanced or quenched and used to detect acidity, oxygen, and other species.

A subtle form of optical transduction uses the "peeled" optical fiber as a multiple reflectance cell. The normal fiber core glass has a refractive index greater than that of the exterior coating; there is a range of angles of entry to the fiber so that *all* the light beam remains inside the core. If the coating is removed and materials of lower index of refraction are coated on the exterior surface, there can be absorption by multiple reflections, since the evanescent wave can peretrate the coating. Chemical reagent can be added externally to create selective layers on the optical fiber.

Ellipsometry is a reflectance technique that depends on the optical constants and thickness of surface layer. For colorless layers, a polarized light beam will change its plane of polarization upon reflection by the surface film. The thickness can sometimes be determined when optical constants are known or approximated by constants of the bulk material. Antibody-antigen surface reaction can be detected this way.

Piezoelectric Transducers

Cut quartz crystals have characteristic modes of vibration that can be induced by painting electrodes on the opposite surfaces and applying a megaHertz ac voltage. The frequency is searched until the crystal goes into a resonance. The resonant frequency is very stable. It is a property of the material and maintains a value to a few parts per hundred million. When the surface is coated with a stiff mass, the frequency is altered. The shift in frequency is directly related to the surface mass for thin, stiff layers. The reaction of a substrate with this layer changes the constants of the film and further shifts the resonant frequency. These devices can be used in air, in vacuum, or in electrolyte solutions.

Electrochemical Transducers

Electrochemical transducers are commonly used in the sensor field. The main forms of electrochemistry used are potentiometry [zero-current cell voltage (potential difference measurements)], amperometry (current measurement at constant applied voltage at the working electrode), and ac conductivity of a cell.

Potentiometric Transduction

The classical generation of an activity-sensitive voltage is spontaneous in a solution containing both nonredox ions and redox ions. Classical electrodes of types 1, 2, and 3 respond by ion exchange directly or indirectly to ions of the same material as the electrode. Inert metal electrodes (sometimes called *type 0*)—Pt, Ir, Rh, and occasionally carbon C—respond by electron exchange from redox pairs in solution. Potential differences are interfacial and reflect ratios of activities of oxidized to reduced forms.

Amperometric Transduction

For dissolved species that can exchange electrons with an inert electrode, it is possible to force the transfer in one direction by applying a voltage very oxidizing (anodic) or reducing (cathodic). When the voltage is fixed, the species will be, by definition, out of equilibrium with the electrode at its

present applied voltage. Locally, the species (regardless of charge) will oxidize or reduce by moving from bulk solution to the electrode surface where they react. Ions do not move like electrons. Rather they diffuse from high to low concentration and do not usually move by drift or migration. The reason is that the electrolytes in solutions are at high concentrations, and the electric field is virtually eliminated from the bulk. The field drops through the first 1000 Angstroms at the electrode surface. The concentration of the moving species is from high concentration in bulk to zero at the electrode surface where it reacts. This process is called *concentration polarization.* The current flowing is limited by mass transport and so is proportional to the bulk concentration.

Conductometric Transducers

Ac conductivity (impedance) can be purely resistive when the frequency is picked to be about 1000 to 10,000 Hz. In this range the transport of ions is sufficiently slow that they never lose their uniform concentration. They simply quiver in space and carry current forward and backward each half cycle. In the lower and higher frequencies, the cell capacitance can become involved, but this effect is to be avoided.

53.3 Tables of Sensors from the Literature

The longest and most consistently complete references to the chemical sensor field is the review issue of *Analytical Chemistry Journal.* In the 1970s and 1980s these appeared in the April issue, but more recently they appear in the June issue. The editors are Jiri Janata and various colleagues [7–10]. Not all possible or imaginable sensors have been made according to the list in Table 53.2. A more realistic table can be constructed from the existing literature that describes actual devices. This list is Table 53.3. Book references are listed in Table 53.4 in reverse time order to about 1986. This list covers

TABLE 53.3 Chemical Sensors and Properties Documented in the Literature

I.	General topics including items II-V: selectivity, fabrication, data processing
II.	Thermal sensors
III.	Mass sensors
	Gas sensors
	Liquid sensors
IV.	Electrochemical sensors
	Potentiometric sensors
	References electrodes
	Biomedical electrodes
	Applications to cations, anions
	Coated wire/hybrids
	ISFETs and related
	Biosensors
	Gas sensors
	Amperometric sensors
	Modified electrodes
	Gas sensors
	Biosensors
	Direct electron transfer
	Mediated electron transfer
	Biomedical
	Conductimetric sensors
	Semiconducting oxide sensors
	Zinc oxide-based
	Chemiresistors
	Dielectrometers
V.	Optical sensors
	Liquid sensors
	Biosensors
	Gas sensors

most of the major source books and many of the symposium proceedings volumes. The reviews [7–10] are a principal source of references to the published research literature.

53.4 Applications of Microelectronics in Sensor Fabrication

The reviews of sensors since 1988 cover fabrication papers and microfabrication methods and examples [7–10]. A recent review by two of the few *chemical* sensor scientists (chemical engineers) who

TABLE 53.4 Books and Long Reviews Keyed to Items in Table 53.3
(Reviewed since 1988 in reverse time sequence)

I. Yamauchi S (ed). 1992. Chemical Sensor Technology, vol 4, Tokyo, Kodansha Ltd.
Flores JR, Lorenzo E. 1992. Amperometric Biosensors, In MR Smyth, JG Vos (eds), Comprehensive Analytical Chemistry Amsterdam, Elsevier.
Vaihinger S, Goepel W. 1991. Multicomponent analysis in chemical sensing. In W Goepel, J Hesse, J Zemel (eds), Sensors vol 2 Part 1, pp 191–237, Weinheim, Germany, VCH Publishers.
Wise DL (ed). 1991. Bioinstrumentation and Biosensors, New York, Marcel Dekker.
Scheller F, Schubert F. 1989. Biosensors, Basel, Switzerland, Birkhauser Verlag, see also [2].
Madou M, Morrison SR. 1989. Chemical Sensing with Solid State Devices, New York, Academic Press.
Janata J. 1989. Principles of Chemical Sensors, New York, Plenum Press.
Edmonds TE (ed). 1988. Chemical Sensors, Glasgow, Blackie.
Yoda K. 1988. Immobilized enzyme cells. Methods Enzymology, 137:61.
Turner APF, Karube I, Wilson GS eds. 1987. Biosensors: Fundamentals and Applications, Oxford, Oxford University Press.
Seiyama T (ed). 1986. Chemical Sensor Technology, Tokyo, Kodansha Ltd.

II. Thermal Sensor
There are extensive research and application papers and these are mentioned in books listed under I. However, the up-to-date lists of papers are given in references 7–10.

III. Mass Sensors
There are extensive research and application papers and these are mentioned in books listed under I. However, the up-to-date lists of papers are given in references 7–10. Fundamentals of this rapidly expanding field are recently reviewed:
Buttry DA, Ward MD. 1992. Measurement of Interfacial processes at Electrode Surfaces with the Electrochemical Quartz Crystal Microbalance, Chemical Reviews 92:1355.
Grate JW, Martin SJ, White RM. 1993. Acoustic Wave Microsensors, part 1, Analyt Chem 65:940A; part 2, Analyt Chem 65:987A.
Ricco AT. 1994. SAW Chemical Sensors, The Electrochemical Society Interface Winter: 38–44.

IVA. Electrochemical Sensors—Liquid Samples
Scheller F, Schmid RD (eds). 1992. Biosensors: Fundamentals, Technologies and Applications, GBF Monograph Series, New York, VCH Publishers.
Erbach R, Vogel A, Hoffmann B. 1992. Ion-sensitive field-effect structures with Langmuir-Blodgett membranes, In F Scheller, RD Schmid (eds). Biosensors: Fundamentals, Technologies, and Applications, GBF Monograph 17, pp 353–357, New York, VCH Publishers.
Ho May YK, Rechnitz GA. 1992. An Introduction to Biosensors, In RM Nakamura, Y Kasahara, GA Rechnitz (eds), Immunochemical Assays and Biosensors Technology, pp 275–291, Washington DC, American Society Microbiology.
Mattiasson B, Haakanson H. Immunochemically-based assays for process control, 1992. Advances in Biochemical Engineering and Biotechnology 46:81.
Maas AH, Sprokholt R. 1990. Proposed IFCC Recommendations for electrolyte measurements with ISEs in clinical chemistry, In A Ivaska, A Lewenstam, R Sara (eds), Contemporary Electroanalytical Chemistry, Proceedings of the ElectroFinnAnalysis International Conference on Electroanalytical Chemistry, pp 311–315, New York, Plenum.
Vanrolleghem P, Dries D, Verstreate W. RODTOX: Biosensor for rapid determination of the biochemical oxygen demand, 1990. In C Christiansen, L Munck, J Villadsen, (eds), Proceedings of the 5th European Congress Biotechnology, vol 1, pp 161–164, Copenhagen, Denmark, Munksgaard.
Cronenberg C, Van den Heuvel H, Van den Hauw M, Van Groen B. Development of glucose microelectrodes for measurements in biofilms, 1990. In C Christiansen, L Munck, J Villadsen J, (eds), Proceedings of the 5th European Congress Biotechnology, vol 1, pp 548–551, Copenhagen, Denmark, Munksgaard.
Wise DL (ed). 1989. Bioinstrumentation Research, Development and Applications, Boston, MA, Butterworth-Heinemann.

Pungor E (ed). 1989. Ion-Selective Electrodes—Proceedings of the 5th Symposium (Matrafured, Hungary 1988), Oxford, Pergamon.
Wang J (ed). 1988. Electrochemical Techniques in Clinical Chemistry and Laboratory Medicine, New York, VCH Publishers.
Evans A. 1987. Potentiometry and Ion-selective Electrodes, New York, Wiley.
Ngo TT (ed) 1987. Electrochemical Sensors in Immunological Analysis, New York, Plenum.
IVB. Electrochemical Sensors—Gas Samples
Sberveglieri G (ed). 1992. Gas Sensors, Dordrecht , The Netherlands, Kluwer.
Moseley PT, Norris JOW, Williams DE. 1991. Technology and Mechanisms of Gas Sensors, Bristol, U.K., Hilger.
Moseley PT, Tofield BD (eds). 1989. Solid State Gas Sensors, Philadelphia, Taylor and Francis, Publishers.
V. Optical Sensors
Coulet PR, Blum LJ. Luminescence in Biosensor Design, 1991. In DL Wise, LB Wingard, Jr (eds). Biosensors with Fiberoptics, pp 293–324, Clifton, N.J., Humana.
Wolfbeis OS. 1991. Spectroscopic Techniques, In OS Wolfbeis (ed). Fiber Optic Chemical Sensors and Biosensors, vol 1, pp 25–60. Boca Raton, Fla, CRC Press.
Wolfbeis OS. 1987. Fibre-optic sensors for chemical parameters of interest in biotechnology, In RD Schmidt (ed). GBF (Gesellschaft fur Biotechnologische Forschung) Monogr. Series, vol 10, pp 197–206, New York, VCH Publishers.

also operate a microfabrication laboratory is C. C. Liu, Z.-R. Zhang. 1992. Research and development of chemical sensors using microfabrication techniques. Selective Electrode Rev 14:147.

References

1. Janata J. 1989. Principles of Chemical Sensors, New York, Plenum.
2. Scheller F, Schubert F. 1989. Biosensors, #18 in Advances in Research Technologies (Beitrage zur Forschungstechnologie), Berlin, Akademie-Verlag, Amsterdam, Elsevier (English translation).
3. Turner APF, Karube I, Wilson GS. 1987. Biosensors: Fundamentals and Applications, Oxford, Oxford University Press.
4. Hall EAH. 1990. Biosensors, Milton Keynes, England, Open University Press.
5. Eddoes MJ. 1990. Theoretical Methods for Analyzing Biosensor Performance. In AEG Cass (ed), Biosensor—A Practical Approach, Oxford, IRL Press at Oxford University Ch. 9 pp 211–262.
6. Cosofret VV, Buck RP. 1992. Pharmaceutical Applications of Membrane Sensors, Boca Raton, Fla, CRC Press.
7. Janata J, Bezegh A. 1988. Chemical sensors, Analyt Chem 60:62R.
8. Janata J. 1990. Chemical sensors, Analyt Chem 62:33R.
9. Janata J. 1992. Chemical sensors, Analyt Chem 66:196R.
10. Janata J, Josowicz M, DeVaney M. 1994. Chemical Sensors, Analyt Chem 66:207R.

55

Digital Biomedical Signal Acquisition and Processing

Luca T. Mainardi
Polytechnic University, Milan

Anna M. Bianchi
St. Raffaele Hospital, Milan

Sergio Cerutti
Polytechnic University, Milan

Biologic signals carry information that is useful for comprehension of the complex pathophysiologic mechanisms underlying the behavior of living systems. Nevertheless, such information cannot be available directly from the raw recorded signals; it can be masked by other biologic signals contemporaneously detected (endogenous effects) or buried in some additive noise (exogenous effects). For such reasons, some additional processing is usually required to enhance the relevant information and to extract from it parameters that quantify the behavior of the system under study, mainly for physiologic studies, or that define the degree of pathology for routine clinical procedures (diagnosis, therapy, or rehabilitation).

Several processing techniques can be used for such purposes (they are also called *preprocessing techniques*): time- or frequency-domain methods including filtering, averaging, spectral estimation, and others. Even if it is possible to deal with continuous time waveforms, it is usually convenient to convert them into a numerical form before processing. The recent progress of digital technology, in terms of both hardware and software, makes more efficient and flexible digital rather than analog processing. Digital techniques have several advantages: Their performance is generally powerful, being able to easily implement even complex algorithms, and accuracy depends only on the truncation and round-off errors, whose effects can be predicted and controlled by the designer and are largely unaffected by other unpredictable variables such as component aging and temperature, which can degrade the performances of analog devices. Moreover, design parameters can be more easily changed because they involve software rather than hardware modifications.

A few basic elements of signal acquisition and processing will be presented in the following; our aim is to stress mainly the aspects connected with acquisition and analysis of biologic signals, leaving to the cited literature a deeper insight into the various subjects for both the fundamentals of digital signal processing and the applications.

55.1 Acquisition

A schematic representation of a general acquisition system is shown in Fig. 55.1. Several physical magnitudes are usually measured from biologic systems. They include electromagnetic quantities

0-8493-8346-3/95/$0.00+$.50

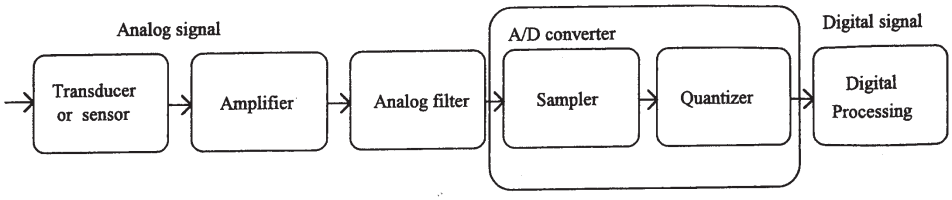


FIGURE 55.1 General block diagram of the acquisition procedure of a digital signal.

(currents, potential differences, field strengths, etc.), as well as mechanical, chemical, or generally nonelectrical variables (pressure, temperature, movements, etc.). Electric signals are detected by *sensors* (mainly electrodes), while nonelectric magnitudes are first converted by *transducers* into electric signals that can be easily treated, transmitted, and stored. Several books of biomedical instrumentation give detailed descriptions of the various transducers and the hardware requirements associated with the acquisition of the different biologic signals [Cobbold, 1988; Tompkins & Webster, 1981; Webster, 1992].

An analog preprocessing block is usually required to amplify and filter the signal (in order to make it satisfy the requirements of the following hardware such as the dynamic of the analog-to-digital converter), to compensate some unwanted sensor characteristics, or to reduce the portion of undesired noise. Moreover, the continuous-time signal should be bandlimited before analog-to-digital (A/D) conversion. Such an operation is needed to reduce the effect of aliasing induced by sampling, as will be described in the next section. Here it is important to remember that the acquisition procedure should preserve the information contained in the original signal waveform. This is a crucial point when recording biologic signals, whose characteristics often may be considered by physicians as indices of some underlying pathologies (i.e., the ST-segment displacement on an ECG signal can be considered a marker of ischemia, the peak-and-wave pattern on an EEG tracing can be a sign of epilepsy, and so on). Thus the acquisition system should not introduce any form of distortion that can be misleading or can destroy real pathologic alterations. For this reason, the analog prefiltering block should be designed with *constant modulus* and *linear phase* (or zero-phase) frequency response, at least in the passband, over the frequencies of interest. Such requirements make the signal arrive undistorted up to the A/D converter.

The analog waveform is then A/D converted into a digital signal; i.e., it is transformed into a series of numbers discretized both in time and amplitude that can be easily managed by digital processors. The A/D conversion ideally can be divided in two steps, as shown in Fig. 55.1: the *sampling* process, which converts the continuous signal in a discrete-time series and whose elements are named *samples,* and a *quantization* procedure, which assigns the amplitude value of each sample within a set of determined discrete values. Both processes modify the characteristics of the signal, and their effects will be discussed in the following sections.

The Sampling Theorem

The advantages of processing a digital series instead of an analog signal have been reported previously. Furthermore, the basic property when using a sampled series instead of its continuous waveform lies in the fact that the former, under certain hypotheses, is completely representative of the latter. When this happens, the continuous waveform can be perfectly reconstructed just from the series of sampled values. This is known as the *sampling theorem* (or *Shannon theorem*) [Shannon, 1949]. It states that a continuous-time signal can be completely recovered from its samples if, and only if, the sampling rate is greater than twice the signal bandwidth.

In order to understand the assumptions of the theorem, let us consider a continuous band-limited signal $x(t)$ (up to f_b) whose Fourier transform $X(f)$ is shown in Fig. 55.2*a* and suppose to uni-

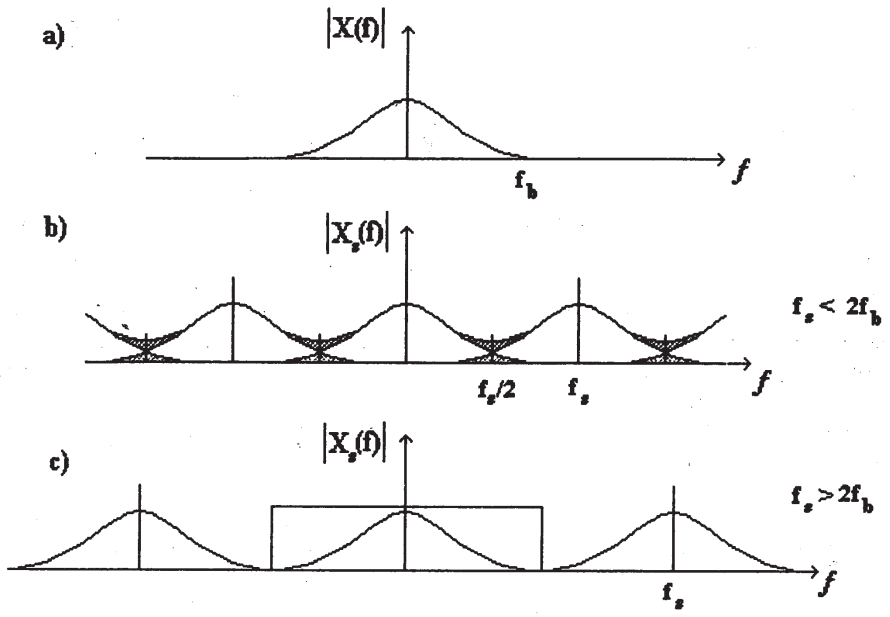


FIGURE 55.2 Effect of sampling frequency (f_s) on a band-limited signal (up to frequency f_b). Fourier transform of the original time signal (*a*), of the sampled signal when $f_s < 2f_b$ (*b*), and when $f_s > 2f_b$ (*c*). The dark areas in part *b* indicate the aliased frequencies.

formly sample it. The sampling procedure can be modeled by the multiplication of $x(t)$ with an impulse train

$$i(t) = \sum_{k=-\infty,\infty} \delta(t - kT_s) \tag{55.1}$$

where $\delta(t)$ is the delta (Dirac) function, k is an integer, and T_s is the sampling interval. The sampled signal becomes

$$x_s(t) = x(t) \cdot i(t) = \sum_{k=-\infty,\infty} x(t) \cdot \delta(t - kT_s) \tag{55.2}$$

Taking into account that multiplication in time domain implies convolution in frequency domain, we obtain

$$X_s(f) = X(f) * I(f) = X(f) * \frac{1}{T_s} \sum_{k=-\infty,\infty} \delta(f - kf_s) = \frac{1}{T_s} \sum_{k=-\infty,\infty} X(f - kf_s) \tag{55.3}$$

where $f_s = 1/T_s$ is the sampling frequency.

Thus $X_s(f)$, i.e., the Fourier transform of the sampled signal, is periodic and consists of a series of identical repeats of $X(f)$ centered around multiples of the sampling frequency, as depicted in Fig. 55.2*b,c*. It is worth noting in Fig. 55.2*b* that the frequency components of $X(f)$ placed above $f_s/2$ appears, when $f_s < 2f_b$, as folded back, summing up to the lower-frequency components. This phenomenon is known as *aliasing* (higher components look "alias" lower components). When aliasing occurs, the original information (Fig. 55.2*a*) cannot be recovered because the frequency components of the original signal are irreversibly corrupted by the overlaps of the shifted versions of $X(f)$.

A visual inspection of Fig. 55.2 allows one to observe that such frequency contamination can be avoided when the original signal is bandlimited [$X(f) = 0$ for $f > f_b$] and sampled at a frequency $f_s \geq 2f_b$. In this case, shown in Fig. 55.2*c*, no overlaps exist between adjacent replay of $X(f)$, and the original waveform can be retrieved by low-pass filtering the sampled signal [Oppenheim & Schafer, 1975]. Such observations are the basis of the sampling theorem previously reported.

The hypothesis of a bandlimited signal is hardly verified in practice, due to the signal characteristics or to the effect of superimposed wideband noise. It is worth noting that filtering before sampling is always needed even if we assume the incoming signal to be bandlimited. Let us consider the following example of an EEG signal whose frequency content of interest ranges between 0 and 40 Hz (the usual diagnostic bands are δ, 0 to 3.5 Hz; ϑ, 4 to 7 Hz; α, 8 to 13 Hz; β, 14 to 40 Hz). We may decide to sample it at 80 Hz, thus literarily respecting the Shannon theorem. If we do it without prefiltering, we could find some unpleasant results. Typically, the 50-Hz mains noise will replicate itself in the signal band (30 Hz, i.e., the β band), thus corrupting irreversibly the information, which is of great interest from a physiologic and clinical point of view. The effect is shown in Fig. 55.3*a* (before sampling) and Fig. 55.3*b* (after sampling). Generally, it is advisable to sample at a frequency greater than $2f_b$ [Gardenhire, 1964] in order to take into account the nonideal behaviour of the filter or the other preprocessing devices. Therefore, the prefiltering block of Fig. 55.1 is always required to bandlimit the signal before sampling and to avoid aliasing errors.

The Quantization Effects

The quantization produces a discrete signal, whose samples can assume only certain values according to the way they are coded. Typical step functions for a uniform quantizer are reported in Fig. 55.4*a,b*, where the quantization interval Δ between two quantization levels is evidenced in two cases: rounding and truncation, respectively.

Quantization is a heavily nonlinear procedure, but fortunately, its effects can be statistically modeled. Figure 55.4*c,d* shows it; the nonlinear quantization block is substituted by a statistical model in which the error induced by quantization is treated as an additive noise $e(n)$ (**quantization error**) to the signal $x(n)$. The following hypotheses are considered in order to deal with a simple mathematical problem:

1. $e(n)$ is supposed to be a white noise with uniform distribution.
2. $e(n)$ and $x(n)$ are uncorrelated.

First of all, it should be noted that the probability density of $e(n)$ changes according to the adopted coding procedure. If we decide to round the real sample to the nearest quantization level, we have

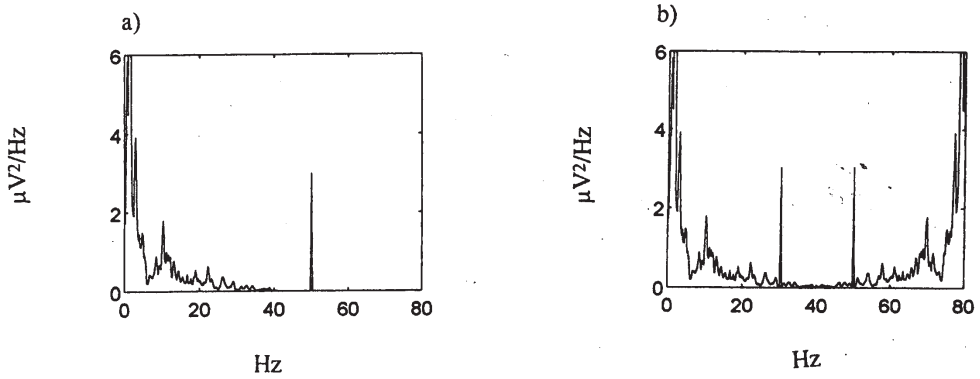


FIGURE 55.3 Power spectrum of an EEG signal (originally bandlimited up to 40 Hz). The presence of 50-Hz mains noise (*a*) causes aliasing error in the 30-Hz component (i.e., in the β diagnostic band) in the sampled signal (*b*) if $f_s = 80$ Hz.

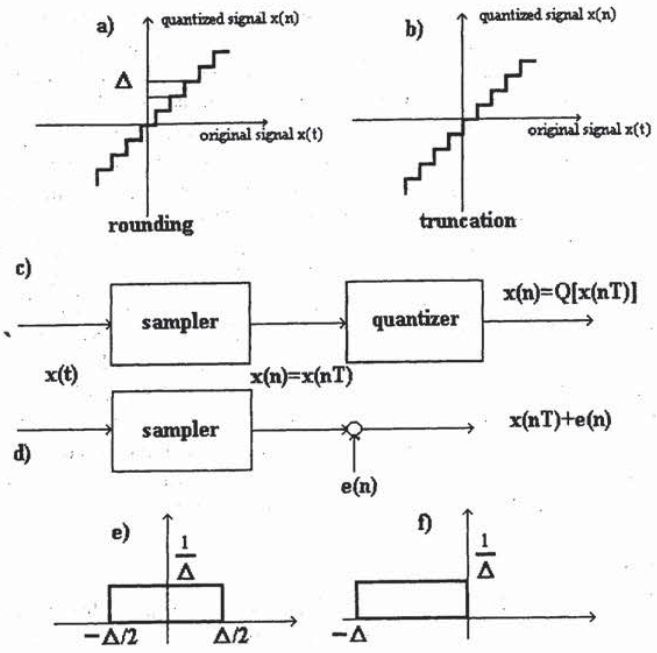


FIGURE 55.4 Nonlinear relationships for rounding (*a*) and truncation (*b*) quantization procedures. Description of quantization block (*c*) by a statistical model (*d*) and probability densities for the quantization noise $e(n)$ for rounding (*e*) and truncation (*f*). Δ is the quantization interval.

$-\Delta/2 \leq e(n) < \Delta/2$, while if we decide to truncate the sample amplitude, we have $-\Delta \leq e(n) < 0$. The two probability densities are plotted in Fig. 55.4*e,f*.

The two ways of coding yield processes with different statistical properties. In the first case the mean and variance value of $e(n)$ are

$$m_e = 0 \qquad \sigma_e^2 = \Delta^2/12$$

while in the second case $m_e = -\Delta/2$, and the variance is still the same. Variance reduces in the presence of a reduced quantization interval as expected.

Finally, it is possible to evaluate the signal-to-noise ratio (SNR) for the quantization process:

$$\text{SNR} = 10 \log_{10}\left(\frac{\sigma_x^2}{\sigma_e^2}\right) = 10 \log_{10}\left(\frac{\sigma_x^2}{2^{-2b}/12}\right) = 6.02b + 10.79 + 10 \log_{10}(\sigma_x^2) \quad (55.4)$$

having set $\Delta = 2^{-2b}$ and where σ_x^2 is the variance of the signal and b is the number of bits used for coding. It should be noted that the SNR increases by almost 6 dB for each added bit of coding. Several forms of quantization are usually employed: uniform, nonuniform (preceding the uniform sampler with a nonlinear block), or roughly (small number of quantization levels and high quantization step). Details can be found in Carassa [1983], Jaeger [1982], and Widrow [1956].

55.2 Signal Processing

A brief review of different signal-processing techniques will be given in this section. They include traditional filtering, averaging techniques, and spectral estimators.

Only the main concepts of analysis and design of digital filters are presented, and a few examples are illustrated in the processing of the ECG signal. Averaging techniques will then be described briefly and their usefulness evidenced when noise and signal have similar frequency contents but different statistical properties; an example for evoked potentials enhancement from EEG background noise is illustrated. Finally, different spectral estimators will be considered and some applications shown in the analysis of RR fluctuations [i.e., the heart rate variability (HRV) signal].

Digital Filters

A *digital filter* is a discrete-time system that operates some transformation on a digital input signal $x(n)$ generating an output sequence $y(n)$, as schematically shown by the block diagram in Fig. 55.5. The characteristics of transformation $T[\cdot]$ identify the filter. The filter will be *time-variant* if $T[\cdot]$ is a function of time or *time-invariant* otherwise, while is said to be *linear* if, and only if, having $x_1(n)$ and $x_2(n)$ as inputs producing $y_1(n)$ and $y_2(n)$, respectively, we have

$$T[ax_1 + bx_2] = aT[x_1] + bT[x_2] = ay_1 + by_2. \tag{55.5}$$

In the following, only linear, time-invariant filters will be considered, even if several interesting applications of nonlinear [Glaser & Ruchkin, 1976; Tompkins, 1993] or time-variant [Cohen, 1983; Huta & Webster, 1973; Thakor, 1987; Widrow et al., 1975] filters have been proposed in the literature for the analysis of biologic signals.

The behavior of a filter is usually described in terms of input-output relationships. They are usually assessed by exciting the filter with different inputs and evaluating which is the response (output) of the system. In particular, if the input is the impulse sequence $\delta(n)$, the resulting output, the impulse response, has a relevant role in describing the characteristic of the filter. Such a response can be used to determine the response to more complicated input sequences. In fact, let us consider a generic input sequence $x(n)$ as a sum of weighed and delayed impulses

$$x(n) = \sum_{k=-\infty,\infty} x(k) \cdot \delta(n-k) \tag{55.6}$$

and let us identify the response to $\delta(n-k)$ as $h(n-k)$. If the filter is time-invariant, each delayed impulse will produce the same response, but time-shifted; due to the linearity property, such responses will be summed at the output:

$$y(n) = \sum_{k=-\infty,\infty} x(k) \cdot h(n-k). \tag{55.7}$$

This convolution product links input and output and defines the property of the filter. Two of them should be recalled: *stability* and *causality*. The former ensures that bounded (finite) inputs will

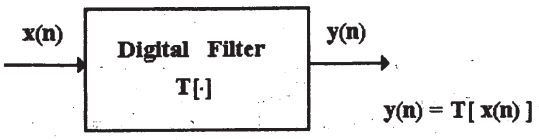


FIGURE 55.5 General block diagram of a digital filter. The output digital signal $y(n)$ is obtained from the input $x(n)$ by means of a transformation $T[\cdot]$ which identifies the filter.

produce bounded outputs. Such a property can be deduced by the impulse response; it can be proved that the filter is stable if and only if

$$\sum_{k=-\infty,\infty} |h(k)| < \infty \tag{55.8}$$

Causality means that the filter will not respond to an input before the input is applied. This is in agreement with our physical concept of a system, but it is not strictly required for a digital filter that can be implemented in a noncausal form. A filter is causal if and only if

$$h(k) = 0 \qquad \text{for } k < 0$$

Even if relation (55.7) completely describes the properties of the filter, most often it is necessary to express the input-output relationships of linear discrete-time systems under the form of the z-transform operator, which allows one to express relation (55.7) in a more useful, operative, and simpler form.

The z-Transform

The z-transform of a sequence $x(n)$ is defined by [Rainer et al., 1972]

$$X(z) = \sum_{k=-\infty,\infty} x(k) \cdot z^{-k} \tag{55.9}$$

where z is a complex variable. This series will converge or diverge for different z values. The set of z values which makes Eq. (55.9) converge is the region of convergence, and it depends on the series $x(n)$ considered.

Among the properties of the z-transform, we recall

- The delay (shift) property:
  If $w(n) = x(n - T)$ then $W(z) = X(z) \cdot z^{-T}$
- The product of convolution:
  If $w(n) = \sum_{k=-\infty,\infty} x(k) \cdot y(n - k)$ then $W(z) = X(z) \cdot Y(z)$

The Transfer Function in the z-Domain

Thanks to the previous property, we can express Eq. (55.7) in the z-domain as a simple multiplication:

$$Y(z) = H(z) \cdot X(z) \tag{55.10}$$

where $H(z)$, known as *transfer function* of the filter, is the z-transform of the impulse response. $H(z)$ plays a relevant role in the analysis and design of digital filters. The response to input sinusoids can be evaluated as follows: Assume a complex sinusoid $x(n) = e^{j\omega nT_s}$ as input, the correspondent filter output will be

$$y(n) = \sum_{k=0,\infty} h(k)e^{j\omega T_s(n-k)} = e^{j\omega nT_s} \sum_{k=0,\infty} h(k)e^{-j\omega kT_s} = x(n) \cdot H(z)|_{z=e^{j\omega T_s}} \tag{55.11}$$

Then a sinusoid in input is still the same sinusoid at the output, but multiplied by a complex quantity $H(\omega)$. Such complex function defines the response of the filter for each sinusoid of ω pulse in input, and it is known as the frequency response of the filter. It is evaluated in the complex z plane by computing $H(z)$ for $z = e^{j\omega T_s}$, namely, on the point locus that describes the unitary circle on the z plane ($|e^{j\omega T_s}| = 1$). As a complex function, $H(\omega)$ will be defined by its module $|H(\omega)|$ and by its

phase $\angle H(\omega)$ functions, as shown in Fig. 55.6 for a moving average filter of order 5. The figure indicates that the lower-frequency components will come through the filter almost unaffected, while the higher-frequency components will be drastically reduced. It is usual to express the horizontal axis of frequency response from 0 to π. This is obtained because only pulse frequencies up to $\omega_s/2$ are reconstructable (due to the Shannon theorem), and therefore, in the horizontal axis, the value of ωT_s is reported which goes from 0 to π. Furthermore, Fig. 55.6*b* demonstrates that the phase is piecewise linear, and in correspondence with the zeroes of $|H(\omega)|$, there is a change in phase of π value. According to their frequency response, the filters are usually classified as (1) low-pass, (2) high-pass, (3) bandpass, or (4) bandstop filters. Figure 55.7 shows the ideal frequency response for such filters with the proper low- and high-frequency cutoffs.

For a large class of linear, time-invariant systems, $H(z)$ can be expressed in the following general form:

$$H(z) = \frac{\sum_{m=0,M} b_m z^{-m}}{1 + \sum_{k=1,N} a_k z^{-k}} \qquad (55.12)$$

which describes in the z domain the following *difference equation* in the discrete time domain:

$$y(n) = -\sum_{k=1,N} a_k y(n-k) + \sum_{m=0,M} b_m x(n-m) \qquad (55.13)$$

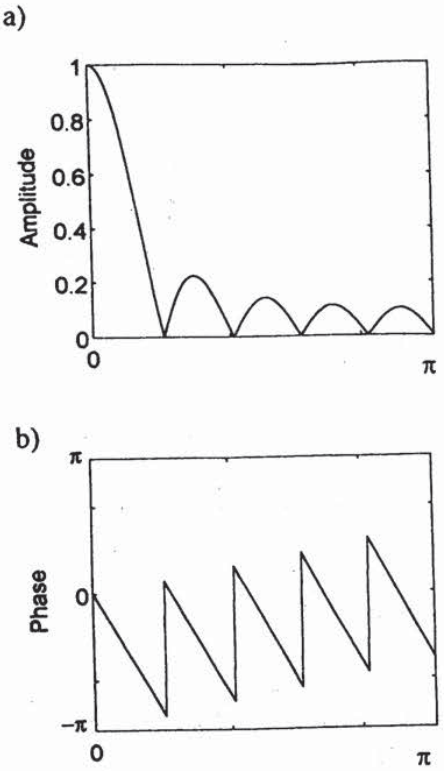


FIGURE 55.6 Modulus (*a*) and phase (*b*) diagrams of the frequency response of a moving average filter of order 5. Note that the frequency plots are depicted up to π. In fact, taking into account that we are dealing with a sampled signal whose frequency information is up to $f_s/2$, we have $\omega_{max} = 2\pi f_s/2 = \pi f_s$ or $\omega_{max} = \pi$ if normalized with respect to the sampling rate.

When at least one of the a_k coefficient is different from zero, some output values contribute to the current output. The filter contains some feedback, and it is said to be implemented in a *recursive* form. On the other hand, when the a_k values are all zero, the filter output is obtained only from the current or previous inputs, and the filter is said to be implemented in a *nonrecursive* form.

The transfer function can be expressed in a more useful form by finding the roots of both numerator and denominator:

$$H(z) = \frac{b_0 z^{N-M} \prod_{m=1,M} (z - z_m)}{\prod_{k=1,N} (z - p_k)} \qquad (55.14)$$

where z_m are the zeroes and p_k are the poles. It is worth nothing that $H(z)$ presents $N - M$ zeros in correspondence with the origin of the z plane and M zeroes elsewhere (N zeroes totally) and N poles. The pole-zero form of $H(z)$ is of great interest because several properties of the filter are immedi-

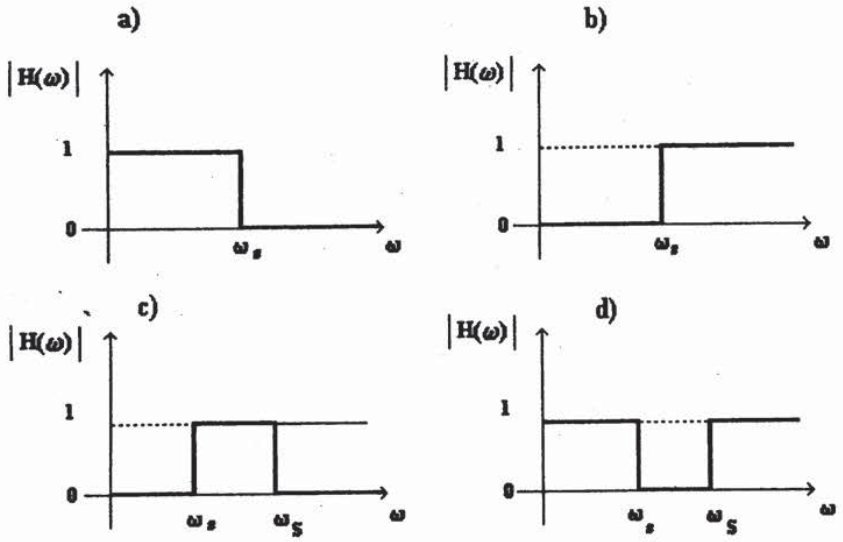


FIGURE 55.7 Ideal frequency-response moduli for low-pass (*a*), high-pass (*b*), bandpass (*c*), and bandstop filters (*d*).

ately available from the geometry of poles and zeroes in the complex z plane. In fact, it is possible to easily assess stability and by visual inspection to roughly estimate the frequency response without making any calculations.

Stability is verified when all poles lie inside the unitary circle, as can be proved by considering the relationships between the z-transform and the Laplace s-transform and by observing that the left side of the s plane is mapped inside the unitary circle [Jackson, 1986; Oppenheim & Schafer, 1975].

The frequency response can be estimated by noting that $(z - z_m)|_{z=e^{j\omega T_s}}$ is a vector joining the mth zero with the point on the unitary circle identified by the angle ωT_s. Defining

$$\begin{aligned} \vec{B}_m &= (z - z_m)|_{z=e^{j\omega T_s}} \\ \vec{A}_k &= (z - p_k)|_{z=e^{j\omega T_s}} \end{aligned} \tag{55.15}$$

we obtain

$$\begin{aligned} |H(\omega)| &= \frac{b_0 \Pi_{m=1,M} |\vec{B}_m|}{\Pi_{k=1,N} |\vec{A}_k|} \\ \angle H(\omega) &= \sum_{m=1,M} \angle \vec{B}_m - \sum_{k=1,N} \angle \vec{A}_k + (N - M)\omega T_s. \end{aligned} \tag{55.16}$$

Thus the modulus of $H(\omega)$ can be evaluated at any frequency ω° by computing the distances between poles and zeroes and the point on the unitary circle corresponding to $\omega = \omega^\circ$, as evidenced by Fig. 55.8, where a filter with two pairs of complex poles and three zeros is considered.

To obtain the estimate of $H(\omega)$, we move around the unitary circle and roughly evaluate the effect of poles and zeroes by keeping in mind a few rules [Challis & Kitney, 1982]: (1) when we are close to a zero, $|H(\omega)|$ will approach zero, and a positive phase shift will appear in $\angle H(\omega)$ as the vector from the zero reverses its angle; (2) when we are close to a pole, $|H(\omega)|$ will tend to peak, and a

negative phase change is found in $\angle H(\omega)$ (the closer the pole to unitary circle, the sharper is the peak until it reaches infinite and the filter becomes unstable); and (3) near a closer pole-zero pair, the response modulus will tends to zero or infinity if the zero or the pole is closer, while far from this pair, the modulus can be considered unitary. As an example, it is possible to compare the modulus and phase diagram of Fig. 55.8*b,c* with the relative geometry of poles and zeroes of Fig. 55.8*a*.

FIR and IIR Filters

A common way of classifying digital filters is based on the characteristics of their impulse response. For finite impulse response (FIR) filters, $h(n)$ is composed of a finite number of nonzero values, while for infinite impulse response (IIR) filters, $h(n)$ oscillates up to infinity with nonzero values. It is clearly evident that in order to obtain an infinite response to an impulse in input, the IIR filter must contain some feedback that sustains the output as the input vanishes. The presence of feedback paths requires to put particular attention to the filter stability.

Even if FIR filters are usually implemented in a nonrecursive form and IIR filters in a recursive form, the two ways of classification are not coincident. In fact, as shown by the following example, a FIR filter can be expressed in a recursive form

$$H(z) = \sum_{k=0,N-1} z^{-k} = \sum_{k=0,N-1} z^{-k} \frac{(1 - z^{-1})}{(1 - z^{-1})} = \frac{1 - z^{-N}}{1 - z^{-1}} \tag{55.17}$$

for a more convenient computational implementation.

As shown previously, two important requirements for filters are stability and linear phase response. FIR filters can be easily designed to fulfill such requirements; they are always stable (having no poles outside the origin), and the linear phase response is obtained by constraining the impulse response coefficients to have symmetry around their midpoint. Such constrain implies

$$b_m = \pm\, b^*_{M-m} \tag{55.18}$$

where the b_m are the M coefficients of an FIR filter. The sign + or − stays in accordance with the symmetry (even or odd) and M value (even or odd). This is a necessary and sufficient condition for FIR filters to have linear phase response. Two cases of impulse response that yield a linear phase filter are shown in Fig. 55.9.

It should be noted that condition (55.18) imposes geometric constrains to the zero locus of $H(z)$. Taking into account Eq. (55.12), we have

$$z^M H(z) = H\left(\frac{1}{z^*}\right) \tag{55.19}$$

Thus, both z_m and $1/z^*_m$ must be zeros of $H(z)$. Then the zeroes of linear phase FIR filters must lie on the unitary circle, or they must appear in pairs and with inverse moduli.

Design Criteria

In many cases, the filter is designed in order to satisfy some requirements, usually on the frequency response, which depend on the characteristic of the particular application the filter is intended for. It is known that ideal filters, like those reported in Fig. 55.7, are not physically realizable (they would require an infinite number of coefficients of impulse response); thus we can design FIR or IIR filters that can only mimic, with an acceptable error, the ideal response. Figure 55.10 shows a frequency response of a not ideal low-pass filter. Here, there are ripples in passband and in stopband, and there is a transition band from passband to stopband, defined by the interval $\omega_s - \omega_p$.

Several design techniques are available, and some of them require heavy computational tasks, which are capable of developing filters with defined specific requirements. They include window tech-

nique, frequency-sampling method, or equiripple design for FIR filters. Butterworth, Chebychev, elliptical design, and impulse-invariant or bilinear transformation are instead employed for IIR filters. For detailed analysis of digital filter techniques, see Antoniou [1979], Cerutti [1983], and Oppenheim and Schafer [1975].

Examples

A few examples of different kinds of filters will be presented in the following, showing some applications on ECG signal processing. It is known that the ECG contains relevant information over a wide range of frequencies; the lower-frequency contents should be preserved for correct measurement of the slow ST displacements, while higher-frequency contents are needed to correctly estimate amplitude and duration of the faster contributions, mainly at the level of the QRS complex. Unfortunately, several sources of noise are present in the same frequency band, such as, for example, higher-frequency noise due to muscle contraction (EMG noise), the lower-frequency noise due to motion artifacts (baseline wandering), the effect of respiration or the low-frequency noise in the skin-electrode interface, and others.

In the first example, the effect of two different low-pass filters will be considered. An ECG signal corrupted by an EMG noise (Fig. 55.11*a*) is low-pass filtered by two different low-pass filters whose frequency responses are shown in Fig. 55.11*b,c*. The two FIR filters have cutoff frequencies at 40 and 20 Hz, respectively, and were designed through window techniques (Weber-Cappellini window, filter length = 256 points) [Cappellini et al., 1978].

The output signals are shown in Fig. 55.11*d,e*. Filtering drastically reduces the superimposed noise but at the same time alters the original ECG waveform. In particular, the R wave amplitude is progressively reduced by decreasing the cutoff frequency, and the QRS width is progressively increased as well. On the other hand, P waves appears almost unaffected, having frequency components generally lower than 20 to 30 Hz. At this point, it is worth noting that an increase in QRS duration is generally associated with various pathologies, such as ventricular hypertrophy or bundle-branch block. It is therefore necessary to check that an excessive band limitation does not introduce a false-positive indication in the diagnosis of the ECG signal.

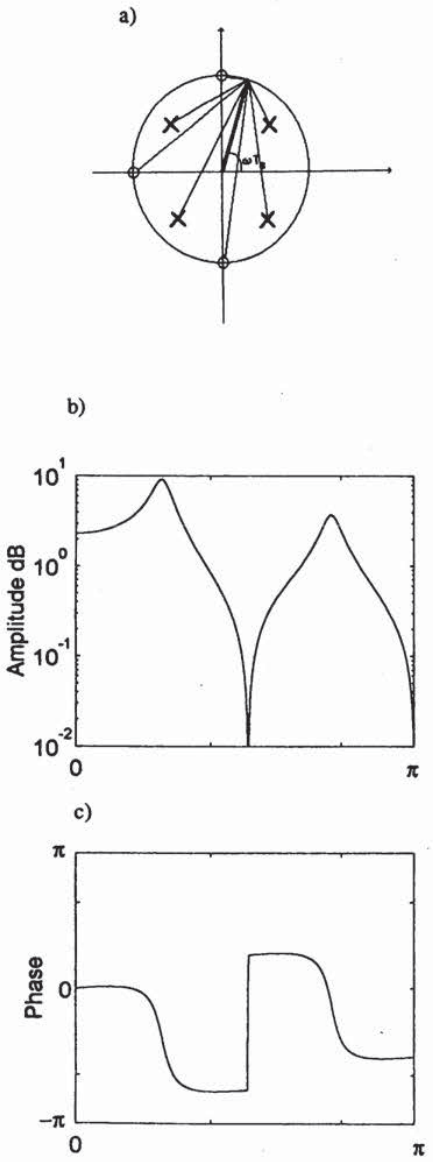


FIGURE 55.8 Poles and zeroes geometry (*a*) and relative frequency response modulus (*b*) and phase (*c*) characteristics. Moving around the unitary circle a rough estimation of $|H(\omega)|$ and $\angle H(\omega)$ can be obtained. Note the zeros' effects at π and $\pi/2$ and modulus rising in proximity of the poles. Phase shifts are cleary evident in part *c* closer to zeros and poles.

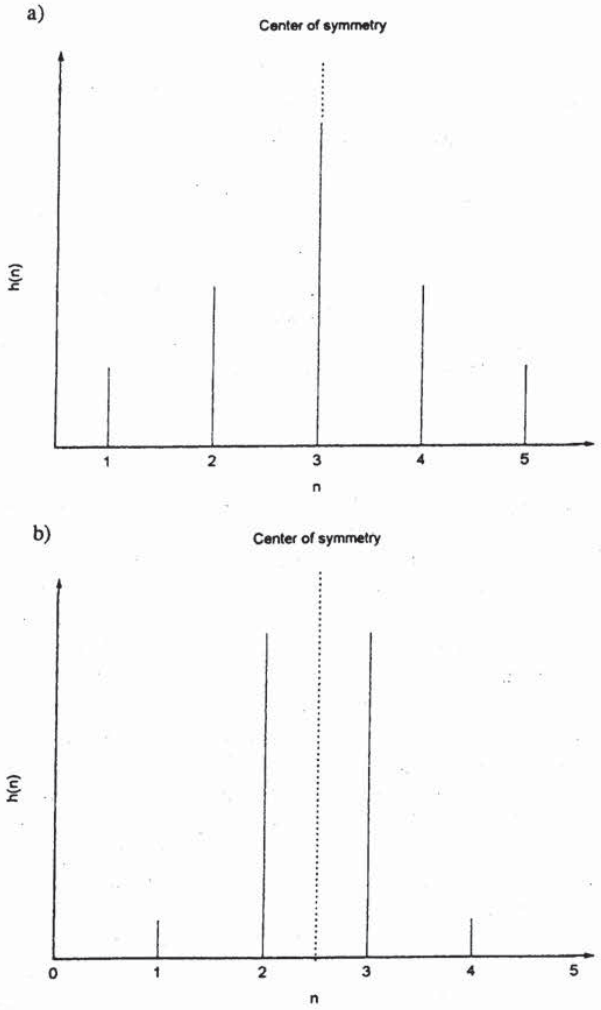


FIGURE 55.9 Examples of impulse response for linear phase FIR filters: odd (*a*) and even (*b*) number of coefficients.

An example of an application for stopband filters (notch filters) is presented in Fig. 55.12. It is used to reduce the 50-Hz mains noise on the ECG signal, and it was designed by placing a zero in correspondence of the frequency we want to suppress.

Finally, an example of a high-pass filter is shown for the detection of the QRS complex. Detecting the time occurrence of a fiducial point in the QRS complex is indeed the first task usually performed in ECG signal analysis. The QRS complex usually contains the higher-frequency components with respect to the other ECG waves, and thus such components will be enhanced by a high-pass filter. Figure 55.13 shows how QRS complexes (Fig. 55.13*a*) can be identified by a derivative high-pass filter with a cutoff frequency to decrease the effect of the noise contributions at high frequencies (Fig. 55.13*b*). The filtered signal (Fig. 55.13*c*) presents sharp and well-defined peaks that are easily recognized by a threshold value.

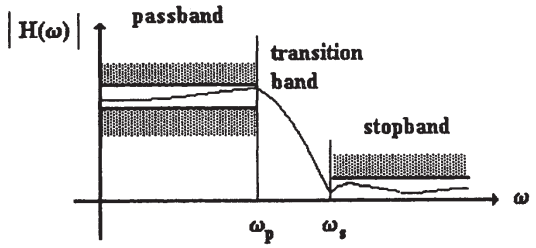


FIGURE 55.10 Amplitude response for a real low-pass filter. Ripples are admitted in both passband and stopband, but they are constrained into restricted areas. Limitations are also imposed to the width of the transition band.

Signal Averaging

Traditional filtering performs very well when the frequency content of signal and noise do not overlap. When the noise bandwidth is completely separated from the signal bandwidth, the noise can be decreased easily by means of a linear filter according to the procedures described earlier. On the other hand, when the signal and noise bandwidth overlap and the noise amplitude is enough to seriously corrupt the signal, a traditional filter, designed to cancel the noise, also will introduce signal cancellation or, at least, distortion. As an example, let us consider the brain potentials evoked by a sensory stimulation (visual, acoustic, or somatosensory) generally called *evoked potentials* (EP). Such a response is very difficult to determine because its amplitude is generally much lower than the background EEG activity. Both EP and EEG signals contain information in the same frequency range; thus the problem of separating the desired response cannot be approached via traditional digital filtering [Aunon et al., 1981]. Another typical example is in the detection of ventricular late potentials (VLP) in the ECG signal. These potentials are very small in amplitude and are comparable with the noise superimposed on the signal also for what concerns the frequency content [Simson, 1981]. In such cases, an increase in the SNR may be achieved on the basis of different statistical properties of signal and noise.

When the desired signal repeats identically at each iteration (i.e., the EP at each sensory stimulus, the VLP at each cardiac cycle), the averaging technique can satisfactorily solve the problem of separating signal from noise. This technique sums a set of temporal epochs of the signal together with the superimposed noise. If the time epochs are properly aligned, through efficient trigger-point recognition, the signal waveforms directly sum together. If the signal and the noise are characterized by the following statistical properties:

1. All the signal epochs contain a deterministic signal component $x(n)$ that does not vary for all the epochs.
2. The superimposed noise $w(n)$ is a broadband stationary process with zero mean and variance σ^2 so that

$$E[w(n)] = 0$$
$$E[w^2(n)] = \sigma^2 \tag{55.20}$$

3. Signal $x(n)$ and noise $w(n)$ are uncorrelated so that the recorded signal $y(n)$ at the ith iteration can be expressed as

$$y(n)_i = x(n) + w_i(n), \tag{55.21}$$

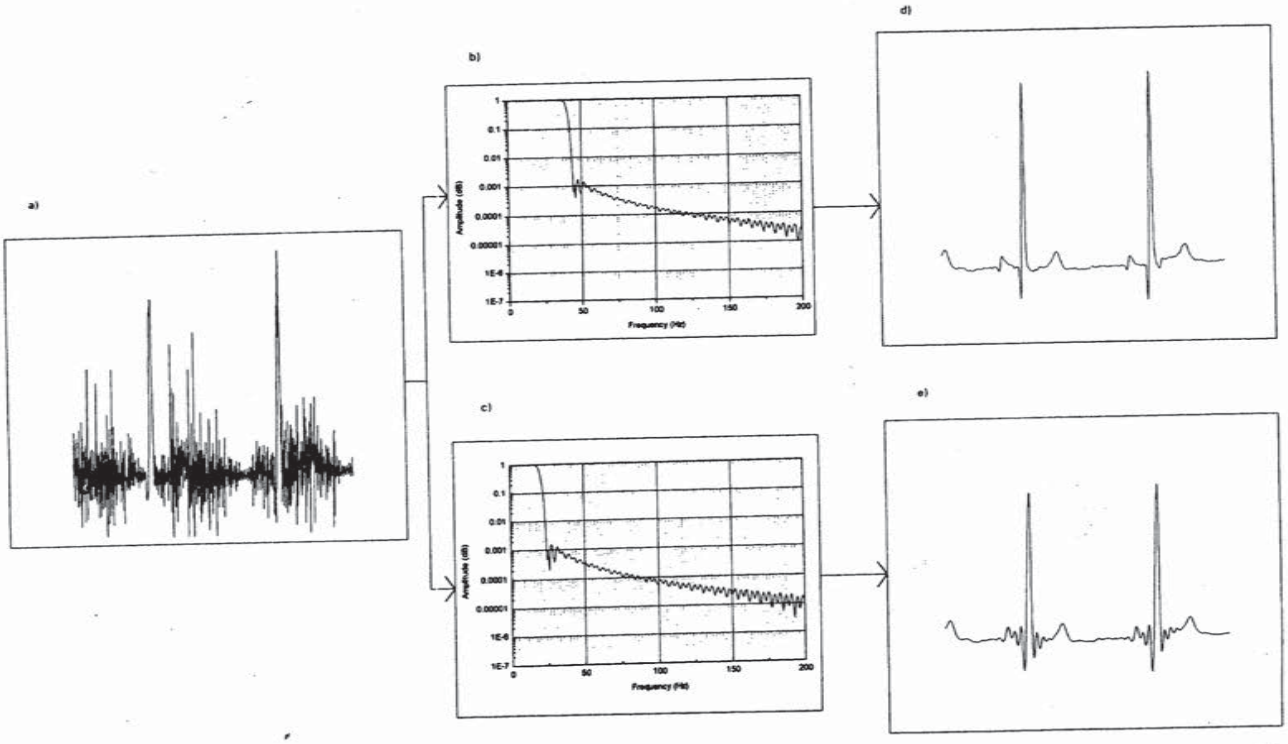


FIGURE 55.11 Effects of two different low-pass filters (*b*) and (*c*) on an ECG trace (*a*) corrupted by EMG noise. Both amplitude reduction and variation in the QRS width induced by too drastic lowpass filtering are evidenced.

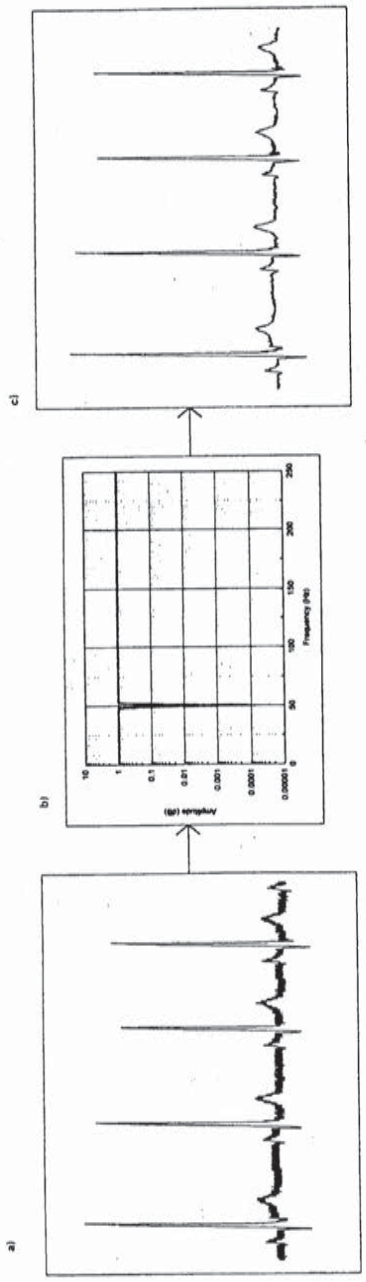


FIGURE 55.12 A 50-Hz noisy ECG signal (*a*); a 50-Hz rejection filter (*b*); a filtered signal (*c*).

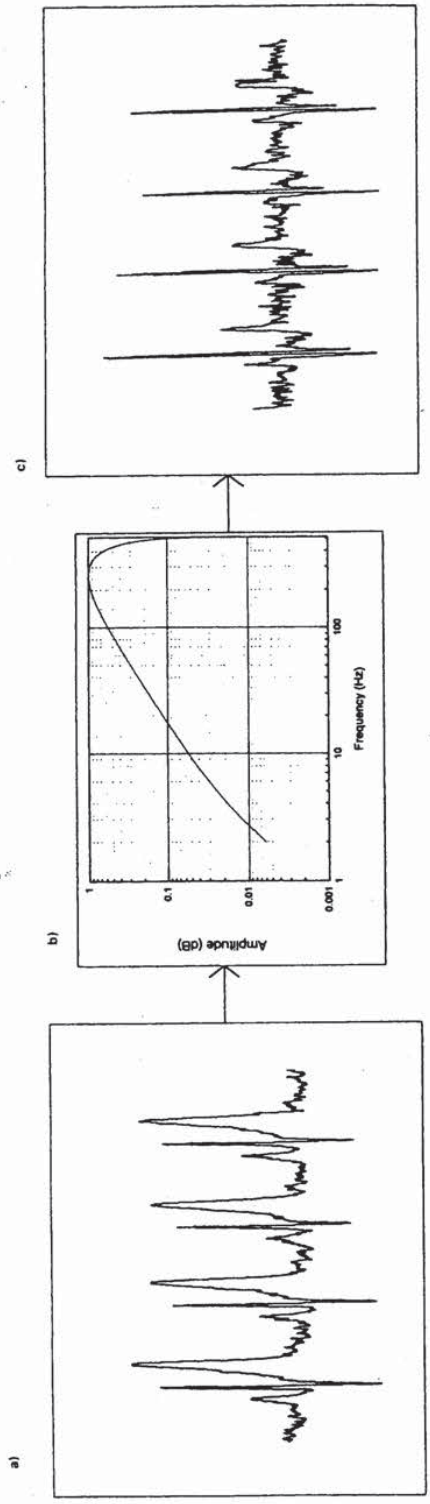


FIGURE 55.13 Effect of a derivative high-pass filter (*b*) on an ECG lead (*a*). (*c*) The output of the filter.

Then the averaging process yields y_t:

$$y_t(n) = \frac{1}{N}\sum_{i=1}^{N} y_i = x(n) + \sum_{i=1}^{N} w_i(n) \tag{55.22}$$

The noise term is an estimate of the mean by taking the average of N realizations. Such an average is a new random variable that has the same mean of the sum terms (zero in this case) and which has variance of σ^2/N. The effect of the coherent averaging procedure is then to maintain the amplitude of the signal and reduce the variance of the noise by a factor of N. In order to evaluate the improvement in the SNR (in rms values) in respect to the SNR; (at the generic ith sweep):

$$\text{SNR} = \text{SNR}_i \cdot \sqrt{N} \tag{55.23}$$

Thus signal averaging improves the SNR by a factor of $\sqrt{N}$ in rms value.

A coherent averaging procedure can be viewed as a digital filtering process, and its frequency characteristics can be investigated. From expression (55.17) through the z-transform, the transfer function of the filtering operation results in

$$H(z) = \frac{1 + z^{-h} + z^{-2h} + \cdots + z^{-(N-1)h}}{N} \tag{55.24}$$

where N is the number of elements in the average, and h is the number of samples in each response. An alternative expression for $H(z)$ is

$$H(z) = \frac{1}{N}\,\frac{1 - z^{-Nh}}{1 - z^{h}} \tag{55.25}$$

This is a moving average low-pass filter as discussed earlier, where the output is a function of the preceding value with a lag of h samples; in practice, the filter operates not on the time sequence but in the sweep sequence on corresponding samples.

The frequency response of the filter is shown in Fig. 55.14 for different values of the parameter N. In this case, the sampling frequency f_s is the repetition frequency of the sweeps, and we may assume it to be 1 without loss of generality. The frequency response is characterized by a main lobe with the first zero corresponding to $f = 1/N$ and by successive secondary lobes separated by zeroes at intervals $1/N$. The width of each tooth decreases as well as the amplitude of the secondary lobes when increasing the number N of sweeps.

The desired signal is sweep-invariant, and it will be unaffected by the filter, while the broadband noise will be decreased. Some leakage of noise energy takes place in the center of the sidelobes and, of course, at zero frequency. Under the hypothesis of zero mean noise, the dc component has no effect, and the diminishing sidelobe amplitude implies the leakage to be not relevant for high frequencies. It is important to recall that the average filtering is based on the hypothesis of broadband distribution of the noise and lack of correlation between signal and noise. Unfortunately, these assumptions are not always verified in biologic signals. For example, the assumption of independence of the background EEG and the evoked potential may be not completely realistic [Gevins & Remond, 1987]. In addition, much attention must be paid to the alignment of the sweeps; in fact, slight misalignments (fiducial point jitter) will lead to a low-pass filtering effect of the final result.

Example

As mentioned previously, one of the fields in which signal-averaging technique is employed extensively is in the evaluation of cerebral evoked response after a sensory stimulation. Figure 55.15*a* shows

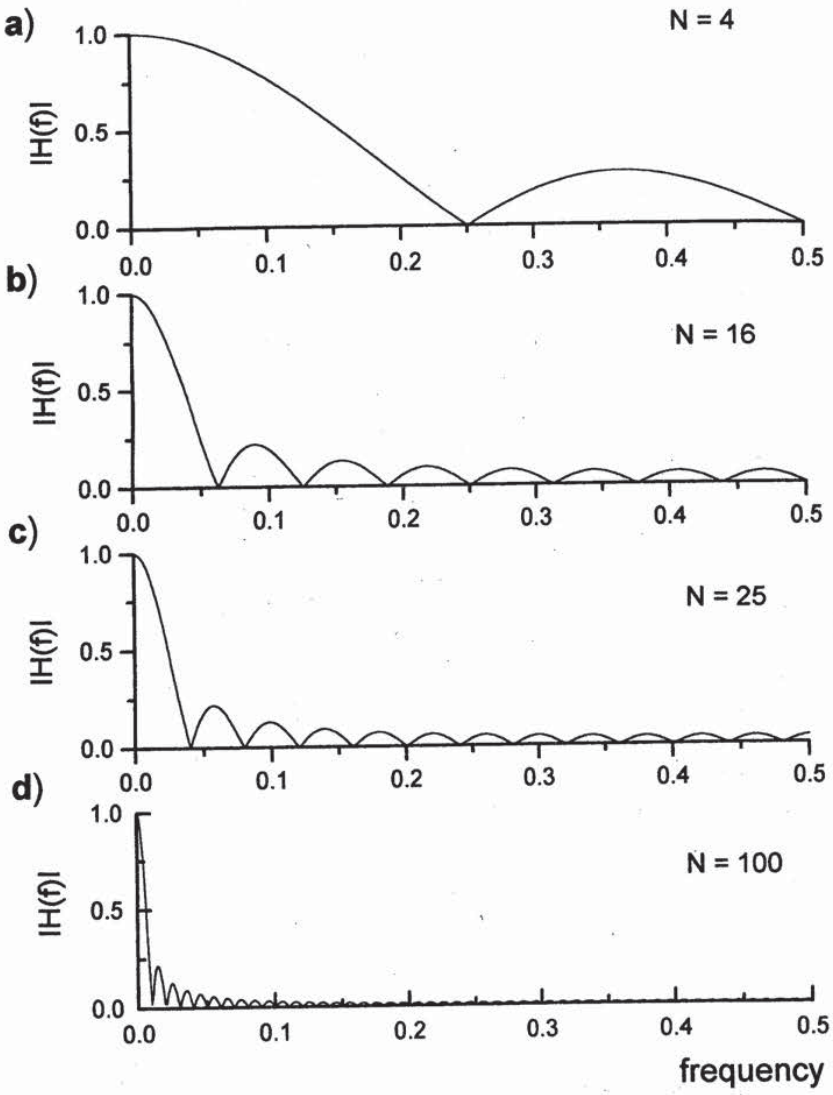


FIGURE 55.14 Equivalent frequency response for the signal-averaging procedure for different values of N (see text).

the EEG recorded from the scalp of a normal subject after a somatosensory stimulation released at time $t = 0$. The evoked potential (N = 1) is not visible because it is buried in the background EEG (upper panel). In the successive panels there is the same evoked potential after averaging different numbers of sweeps corresponding to the frequency responses shown in Fig. 55.14. As N increases, the SNR is improved by a factor $\sqrt{N}$ (in rms value), and the morphology of the evoked potential becomes more recognizable while the EEG contribution is markedly diminished. In this way it is easy to evaluate the quantitative indices of clinical interest, such as the amplitude and the latency of the relevant waves.

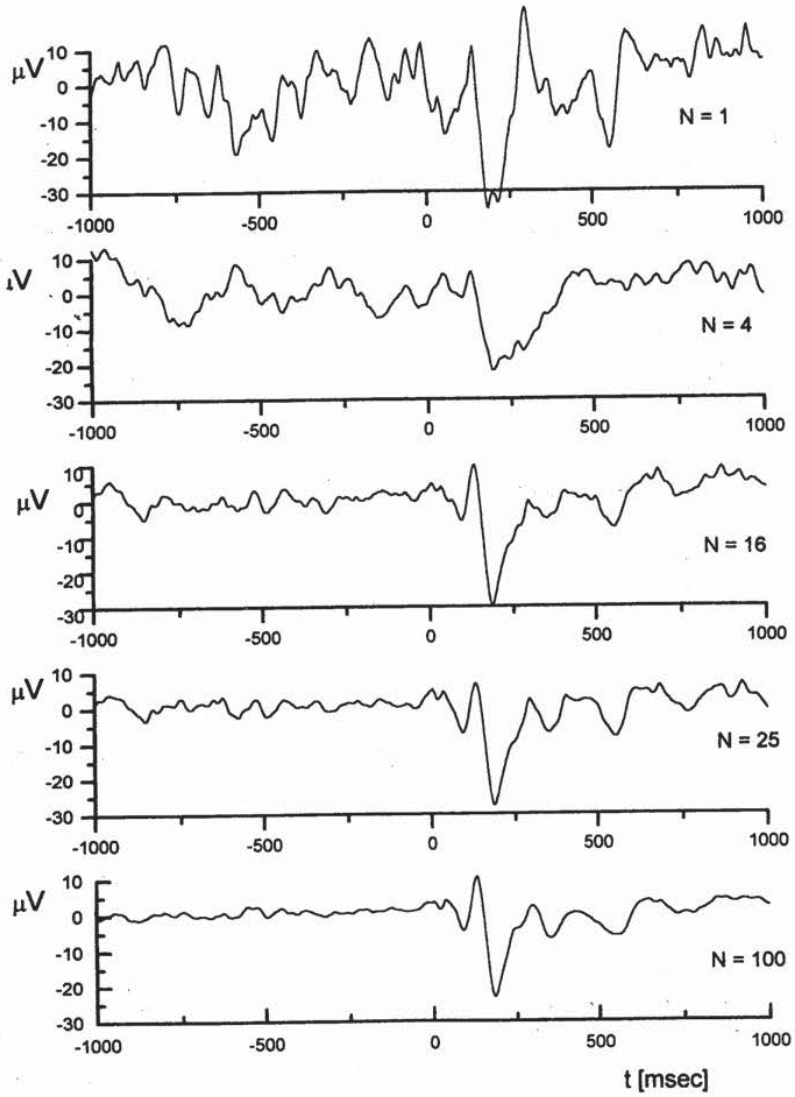


FIGURE 55.15 Enhancement of evoked potential (EP) by means of averaging technique. The EEG noise is progressively reduced, and the EP morphology becomes more recognizable as the number of averaged sweeps (N) is increased.

Spectral Analysis

The various methods to estimate the power spectrum density (PSD) of a signal may be classified as *nonparametric* and *parametric.*

Nonparametric Estimators of PSD

This is a traditional method of frequency analysis based on the Fourier transform that can be evaluated easily through the fast Fourier transform (FFT) algorithm [Marple, 1987]. The expression

of the PSD as a function of the frequency $P(f)$ can be obtained directly from the time series $y(n)$ by using the periodogram expression

$$P(f) = \frac{1}{T_s}\left|T_s \sum_{k=0}^{N-1} y(k)e^{-j2\pi fkT_s}\right|^2 = \frac{1}{NT_s}|Y(f)|^2 \tag{55.26}$$

where T_s is the sampling period, N is the number of samples, and $Y(f)$ is the discrete time Fourier transform of $y(n)$.

On the basis of the Wiener-Khintchin theorem, PSD is also obtainable in two steps from the FFT of the autocorrelation function $R_{yy}(k)$ of the signal, where $R_{yy}(k)$ is estimated by means of the following expression:

$$\hat{R}_{yy}(k) = \frac{1}{N} \sum_{i=0}^{N-k-1} y(i)y^*(i+k) \tag{55.27}$$

where * denotes the complex conjugate. Thus the PSD is expressed as

$$P(f) = T_s \cdot \sum_{k=-N}^{N} \hat{R}_{yy}(k)e^{-j2\pi fkT_s} \tag{55.28}$$

based on the available lag estimates $\hat{R}_{yy}(k)$, where $-(1/2T_s) \leq f \leq (1/2T_s)$

FFT-based methods are widely diffused, for their easy applicability, computational speed, and direct interpretation of the results. Quantitative parameters are obtained by evaluating the power contribution at different frequency bands. This is achieved by dividing the frequency axis in ranges of interest and by integrating the PSD on such intervals. The area under this portion of the spectrum is the fraction of the total signal variance due to the specific frequencies. However, autocorrelation function and Fourier transform are theoretically defined on infinite data sequences. Thus errors are introduced by the need to operate on finite data records in order to obtain estimators of the true functions. In addition, for the finite data set it is necessary to make assumptions, sometimes not realistic, about the data outside the recording window; commonly they are considered to be zero. This implicit rectangular windowing of the data results in a spectral leakage in the PSD. Different windows that smoothly connect the side samples to zero are most often used in order to solve this problem, even if they may introduce a reduction in the frequency resolution [Harris, 1978]. Furthermore, the estimators of the signal PSD are not statistically consistent, and various techniques are needed to improve their statistical performances. Various methods are mentioned in the literature; the methods of Dariell [1946], Bartlett [1948], and Welch [1970] are the most diffused ones. Of course, all these procedures cause a further reduction in frequency resolution.

Parametric Estimators

Parametric approaches assume the time series under analysis to be the output of a given mathematical model, and no drastic assumptions are made about the data outside the recording window. The PSD is calculated as a function of the model parameters according to appropriate expressions. A critical point in this approach is the choice of an adequate model to represent the data sequence. The model is completely independent of the physiologic, anatomic, and physical characteristics of the biologic system but provides simply the input-output relationships of the process in the so-called black-box approach.

Among the numerous possibilities of modeling, linear models, characterized by a rational transfer function, are able to describe a wide number of different processes. In the most general case, they are represented by the following linear equation that relates the input-driving signal $w(k)$ and the output of an autoregressive moving average (ARMA) process:

$$y(k) = -\sum_{i=1}^{p} a_i y(k-i) + \sum_{j=1}^{q} b_j w(k-j) + w(k) \tag{55.29}$$

where $w(k)$ is the input white noise with zero mean value and variance λ^2, p and q are the orders of AR and MA parts, respectively, and a_i and b_j are the proper coefficients.

The ARMA model may be reformulated as an AR or an MA if the coefficients b_j or a_i are, respectively, set to zero. Since the estimation of the AR parameters results in linear equations, AR models are usually employed in place of ARMA or MA models, also on the basis of the Wold decomposition theorem [Marple, 1987] that establishes that any stationary ARMA or MA process of finite variance can be represented as a unique AR model of appropriate order, even infinite; likewise, any ARMA or AR process can be represented by an MA model of sufficiently high order.

The AR PSD is then obtained from the following expression:

$$P(f) = \frac{\lambda^2 T_s}{\left|1 + \sum_{i=1}^{p} a_i z^{-i}\right|^2_{z=\exp(j2\pi f T_s)}} = \frac{\lambda^2 T_s}{\left|\prod_{l=1}^{p} (z - z_l)\right|^2_{z=\exp(j2\pi f T_s)}} \tag{55.30}$$

The right side of the relation puts into evidence the poles of the transfer function that can be plotted in the z-transform plane. Figure 55.16*b* shows the PSD function of the HRV signal depicted in Fig. 55.16*a*, while Fig. 55.16*c* displays the corresponding pole diagram obtained according to the procedure described in the preceding section.

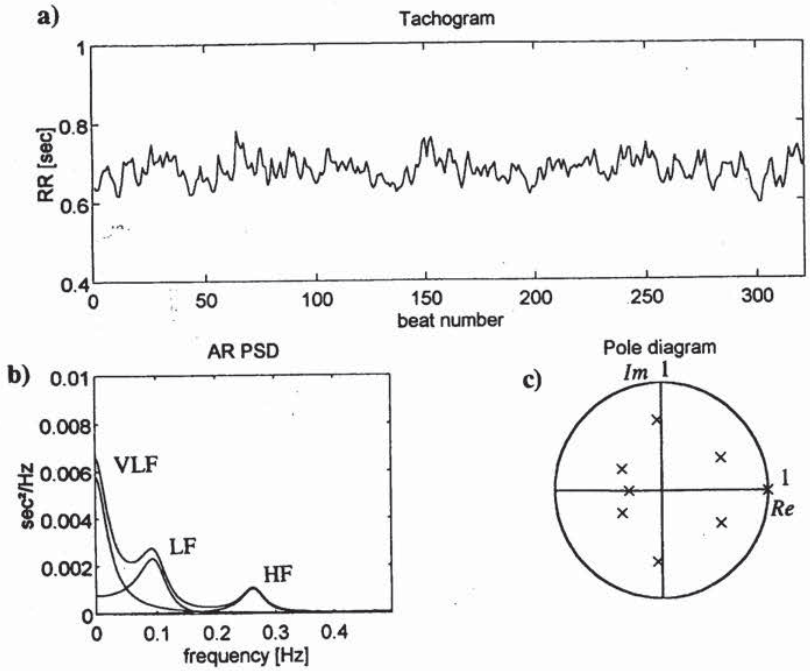


FIGURE 55.16 (*a*) Interval tachogram obtained from an ECG recording as the sequence of the RR time intervals expressed in seconds as a function of the beat number. (*b*) PSD of the signal (*a*) evaluated by means of an AR model (see text). (*c*) Pole diagram of the PSD shown in (*b*).

Parametric methods are methodologically and computationally more complex than the nonparametric ones, since they require an a priori choice of the structure and of the order of the model of the signal-generation mechanism. Some tests are required a posteriori to verify the whiteness of the prediction error, such as the Anderson test (autocorrelation test) [Box & Jenkins, 1976] in order to test the reliability of the estimation. Postprocessing of the spectra can be performed as well as for nonparametric approaches by integrating the $P(f)$ function in predefined frequency ranges; however, the AR modeling has the advantage of allowing a spectral decomposition for a direct and automatic calculation of the power and frequency of each spectral component. In the z-transform domain, the autocorrelation function (ACF) $R(k)$ and the P(z) of the signal are related by the following expression:

$$R(k) = \frac{1}{2\pi j} \int_{|z|=1} P(z) z^{k-1} dz \tag{55.31}$$

If the integral is calculated by means of the residual method, the ACF is decomposed into a sum of dumped sinusoids, each one related to a pair of complex conjugate poles, and of dumped exponential functions, related to the real poles [Zetterberg, 1969]. The Fourier transform of each one of these terms gives the expression of each spectral component that fits the component related to the relevant pole or pole pair. The argument of the pole gives the central frequency of the component, while the ith spectral component power is the residual γ_i in case of real poles and $2Re(\gamma_i)$ in case of conjugate pole pairs. γ_i is computed from the following expression:

$$\gamma_i = z^{-1}(z - z_i)P(z)\big|_{z=z_i} \tag{55.32}$$

It is advisable to point out the basic characteristics of the two approaches that have been described above: the nonparametric and the parametric. The latter (parametric) has evident advantages with respect to the former, which can be summarized in the following:

- It has a more statistical consistency even on short segments of data; i.e., under certain assumptions, a spectrum estimated through autoregressive modeling is a maximum entropy spectrum (MES).
- The spectrum is more easily interpretable with an "implicit" filtering of what is considered random noise.
- An easy and more reliable calculation of the spectral parameters (postprocessing of the spectrum), through the spectral decomposition procedure, is possible. Such parameters are directly interpretable from a physiologic point of view.
- There is no need to window the data in order to decrease the spectral leakage.
- The frequency resolution does not depend on the number of data.

On the other hand, the parametric approach

- Is more complex from a methodologic and computational point of view.
- Requires an a priori definition of the kind of the model (AR, MA, ARMA, or other) to be fitted and mainly its complexity defined (i.e., the number of parameters).

Some figures of merit introduced in literature may be of help in determining their value [Akaike, 1974]. Still, this procedure may be difficult in some cases.

Example

As an example, let us consider the frequency analysis of the heart rate variability (HRV) signal. In Fig. 55.16*a*, the time sequence of the RR intervals obtained from an ECG recording is shown. The RR intervals are expressed in seconds as a function of the beat number in the so-called interval

tachogram. It is worth noting that the RR series is not constant but is characterized by oscillations of up to the 10% of its mean value. These oscillations are not casual but are the effect of the action of the autonomic nervous system in controlling heart rate. In particular, the frequency analysis of such a signal (Fig. 55.16*b* shows the PSD obtained by mean of an AR model) has evidenced three principal contributions in the overall variability of the HRV signal. A very low frequency (VLF) component is due to the long-term regulation mechanisms that cannot be resolved by analyzing a few minutes of signal (3 to 5 minutes are generally studied in the traditional spectral analysis of the HRV signal). Other techniques are needed for a complete understanding of such mechanisms. The low-frequency (LF) component is centered around 0.1 Hz, in a range between 0.03 and 0.15 Hz. An increase in its power has always been observed in relation to sympathetic activations. Finally, the high-frequency (HF) component, in synchrony with the respiration rate, is due to the respiration activity mediated by the vagus nerve; thus it can be a marker of vagal activity. In particular, LF and HF power, both in absolute and in normalized units (i.e., as percentage value on the total power without the VLF contribution), and their ratio LF/HF are quantitative indices widely employed for the quantification of the sympathovagal balance in controlling heart rate [Malliani et al., 1991].

55.3 Conclusion

The basic aspects of signal acquisition and processing have been illustrated, intended as fundamental tools for the treatment of biologic signals. A few examples also were reported relative to the ECG signal, as well as EEG signals and EPs. Particular processing algorithms have been described that use digital filtering techniques, coherent averaging, and power spectrum analysis as reference examples on how traditional or innovative techniques of digital signal processing may impact the phase of informative parameter extraction from biologic signals. They may improve the knowledge of many physiologic systems as well as help clinicians in dealing with new quantitative parameters that could better discriminate between normal and pathologic cases.

Defining Terms

Aliasing: Phenomenon that takes place when, in A/D conversion, the sampling frequency f_s is lower than twice the frequency content f_b of the signal; frequency components above $f_s/2$ are folded back and are summed to the lower-frequency components, distorting the signal.

Averaging: Filtering technique based on the summation of N stationary waveforms buried in casual broadband noise. The SNR is improved by a factor of $\sqrt{N}$.

Frequency response: A complex quantity that, multiplied by a sinusoid input of a linear filter, gives the output sinusoid. It completely characterizes the filter and is the Fourier transform of the impulse response.

Impulse response: Output of a digital filter when the input is the impulse sequence $\delta(n)$. It completely characterizes linear filters and is used for evaluating the output corresponding to different kinds of inputs.

Notch filter: A stopband filter whose stopped band is very sharp and narrow.

Parametric methods: Spectral estimation methods based on the identification of a signal generating model. The power spectral density is a function of the model *parameters.*

Quantization error: Error added to the signal, during the A/D procedure, due to the fact that the analog signal is represented by a digital signal that can assume only a limited and predefined set of values.

Region of convergence: In the z-transform plane, the ensemble containing the z-complex points that makes a series converge to a finite value.

References

Akaike H. 1974. A new look at the statistical model identification. IEEE Trans Autom Contr (AC-19):716.

Antoniou A. 1979. Digital Filters: Analysis and Design. New York, McGraw-Hill.

Aunon JL, McGillim CD, Childers DG. 1981. Signal processing in evoked potential research: Averaging and modeling. CRC Crit Rev Bioing 5:323.

Bartlett MS. 1948. Smoothing priodograms from time series with continuous spectra. Nature 61:686.

Box GEP, Jenkins GM. 1976. Time Series Analysis: Forecasting and Control. San Francisco, Holden-Day.

Cappellini V, Constantinides AG, Emiliani P. 1978. Digital Filters and Their Applications. London, Academic Press.

Carassa F. 1983. Comunicazioni Elettriche. Torino, Boringhieri.

Cerutti S. 1983. Filtri numerici per l'eleborazione di segnali biologici. Milano, CLUP.

Challis RE, Kitney RI. 1982. The design of digital filters for biomedical signal processing: 1. Basic concepts. J Biomed Eng 5:267.

Cobbold RSC. 1988. Transducers for Biomedical Measurements. New York, Wiley.

Cohen A. 1983. Biomedical Signal Processing: Time and Frequency Domains Analysis. Boca Raton, Fla, CRC Press.

Dariell PJ. 1946. On the theoretical specification and sampling properties of autocorrelated time-series (discussion). JR Stat Soc 8:88.

Gardenhire LW. 1964. Selecting sample rate. ISA J 4:59.

Gevins AS, Remond A (eds). 1987. Handbook of Electrophysiology and Clinical Neurophysiology. Amsterdam, Elsevier.

Glaser EM, Ruchkin DS. 1976. Principles of Neurophysiological Signal Processing. New York, Academic Press.

Harris FJ, 1978. On the use of windows for harmonic analysis with the discrete Fourier transform. Proc IEEE 64(1):51.

Huta K, Webster JG. 1973. 60-Hz interference in electrocardiography. IEEE Trans Biomed Eng 20(2):91.

Jackson LB. 1986. Digital Signal Processing. Hingham, Mass, Kluer Academic.

Jaeger RC. 1982. Tutorial: Analog data acquisition technology: II. Analog to digital conversion. IEEE Micro 8:46.

Malliani A, Pagani M, Lombardi F, Cerutti S. 1991. Cardiovascular neural regulation explored in the frequency domain. Circulation 84:482.

Marple SL. 1987. Digital Spectral Analysis with Applications. Englewood Cliff, NJ, Prentice-Hall.

Oppenheim AV, Schafer RW. 1975. Digital Signal Processing. Englewood Cliffs, NJ, Prentice-Hall.

Rainer LR, Cooley JW, Helms HD, et al. 1972. Terminology in digital signal processing. IEEE Trans Audio Electroac AU-20:322.

Shannon CE. 1949. Communication in presence of noise. Proc IRE 37:10.

Simson MB. 1981. Use of signals in the terminal QRS complex to identify patients with ventricular tachycardia after myocardial infarction. Circulation 64:235.

Thakor NV. 1987. Adaptive filtering of evoked potential. IEEE Trans Biomed Eng 34:1706.

Tompkins WJ (ed). 1993. Biomedical Digital Signal Processing. Englewood Cliffs, NJ, Prentice-Hall.

Tompkins WJ, Webster JG (eds). 1981. Design of Microcomputer-Based Medical Instrumentation. Englewood Cliffs, NJ, Prentice-Hall.

Webster JG (ed). 1992. Medical Instrumentation, 2d ed. Boston, Houghton-Mufflin.

Welch DP. 1970. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short modified periodograms. IEEE Trans Acoust AU-15:70.

Widrow B. 1956. A study of rough amplitude quantization by means of Nyquist sampling theory. IRE Trans Cric Theory 3:266.

Widrow B, Glover JRJ, Kaunitz J, et al. 1975. Adaptive noise cancelling: Principles and applications. Proc IEEE 63(12):1692.

Zetterberg LH. 1969. Estimation of parameters for a linear difference equation with application to EEG analysis. Math Biosci 5:227.

Further Information

A book that provides a general overview of basic concepts in biomedical signal processing is *Digital Biosignal Processing,* by Rolf Weitkunat (ed) (Elsevier Science Publishers, Amsterdam, 1991). Contributions by different authors provide descriptions of several processing techniques and many applicative examples on biologic signal analysis. A deeper and more specific insight of actual knowledge and future perspectives on ECG analysis can be found in *Electrocardiography: Past and Future,* by Philippe Coumel and Oscar B. Garfein (eds) (Annals of the New York Academy Press, vol 601, 1990). Advances in signal processing are monthly published in the journal *IEEE Transactions on Signal Processing,* while the *IEEE Transaction on Biomedical Engineering* provides examples of applications in biomedical engineering fields.

88

Noninvasive Optical Monitoring

Ross Flewelling
Nellcor Incorporated

Optical measures of physiologic status are attractive because they can provide a simple, noninvasive, yet real-time assessment of medical condition. Noninvasive optical monitoring is here taken to mean the use of visible or near-infrared light to directly assess the internal physiologic status of a person without the need of extracting a blood or tissue sample or using a catheter. Liquid water strongly absorbs ultraviolet and infrared radiation, and thus these spectral regions are useful only for analyzing thin surface layers or respiratory gases, neither of which will be the subject of this review. Instead, it is the visible and near-infrared portions of the electromagnetic spectrum that provide a unique "optical window" into the human body, opening new vistas for noninvasive monitoring technologies.

Various molecules in the human body possess distinctive spectral absorption characteristics in the visible or near-infrared spectral regions and therefore make optical monitoring possible. The most strongly absorbing molecules at physiologic concentrations are the hemoglobins, myoglobins, cytochromes, melanins, carotenes, and bilirubin (see Fig. 88.1 for some examples). Perhaps less appreciated are the less distinctive and weakly absorbing yet ubiquitous materials possessing spectral characteristics in the near-infrared: water, fat, proteins, and sugars. Simple optical methods are now available to quantitatively and noninvasively measure some of these compounds directly in intact tissue. The most successful methods to date have used hemoglobins to assess the oxygen content of blood, cytochromes to assess the respiratory status of cells, and possibly near-infrared to assess endogenous concentrations of metabolytes, including glucose.

88.1 Oximetry and Pulse Oximetry

Failure to provide adequate oxygen to tissues—*hypoxia*—can in a matter of minutes result in reduced work capacity of muscles, depressed mental activity, and ultimately cell death. It is therefore of considerable interest to reliably and accurately determine the amount of oxygen in blood or tissues. *Oximetry* is the determination of the oxygen content of blood or tissues, normally by optical means. In the clinical laboratory the oxygen content of whole blood can be determined by a benchtop cooximeter or blood gas analyzer. But the need for timely clinical information and the desire to

0-8493-8346-3/95/$0.00+$.50

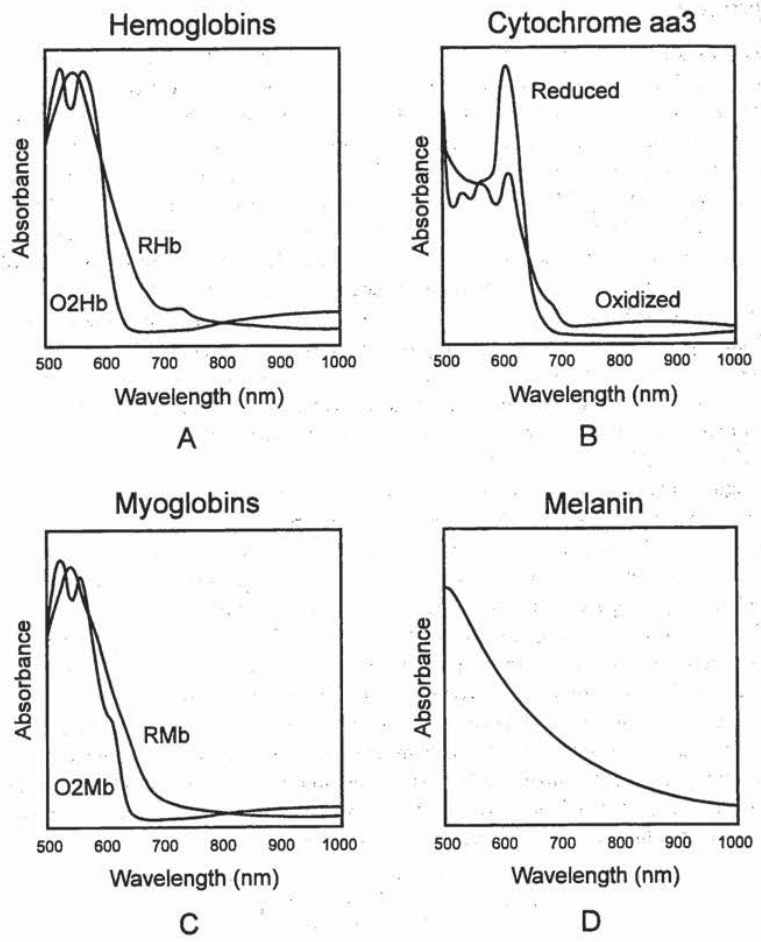


FIGURE 88.1 Absorption spectra of some endogenous biologic materials (*a*) hemoglobins, (*b*) cytochrome *aa3*, (*c*) myoglobins, and (*d*) melanin.

minimize the inconvenience and cost of extracting a blood sample and later analyze it in the lab has led to the search for alternative noninvasive optical methods. Since the 1930s, attempts have been made to use multiple wavelengths of light to arrive at a complete spectral characterization of a tissue. These approaches, although somewhat successful, have remained of limited utility, owing to the awkward instrumentation and unreliable results.

It was not until the invention of *pulse oximetry* in the 1970s and its commercial development and application in the 1980s that noninvasive oximetry became practical. Pulse oximetry is an extremely easy-to-use, noninvasive, and accurate measurement of real-time arterial oxygen saturation. Pulse oximetry is now used routinely in clinical practice, has become a standard of care in all U.S. operating rooms, and is increasingly used wherever critical patients are found. The explosive growth of this new technology and its considerable utility led John Severinghaus and Poul Astrup [1986] in an excellent historical review to conclude that pulse oximetry was "arguably the most significant technological advance ever made in monitoring the well-being and safety of patients during anesthesia, recovery and critical care."

Background

The partial pressure of oxygen (pO_2) in tissues need only be about 3 mmHg to support basic metabolic demands. This tissue level, however, requires capillary pO_2 to be near 40 mmHg, with a corresponding arterial pO_2 of about 95 mmHg. Most of the oxygen carried by blood is stored in red blood cells reversibly bound to hemoglobin molecules. Oxygen saturation (SaO_2) is defined as the percentage of hemoglobin-bound oxygen compared to the total amount of hemoglobin available for reversible oxygen binding. The relationship between the oxygen partial pressure in blood and the oxygen saturation of blood is given by the hemoglobin oxygen dissociation curve as shown in Fig. 88.2. The higher the pO_2 in blood, the higher the SaO_2. But due to the highly cooperative binding of four oxygen molecules to each hemoglobin molecule, the oxygen binding curve is sigmoidal, and consequently the SaO_2 value is particularly sensitive to dangerously low pO_2 levels. With a normal arterial blood pO_2 above 90 mmHg, the oxygen saturation should be at least 95%, and a pulse oximeter can readily verify a safe oxygen level. If oxygen content falls, say to a pO_2 below 40 mmHg, metabolic needs may not be met, and the corresponding oxygen saturation will drop below 80%. Pulse oximetry therefore provides a direct measure of oxygen sufficiency and will alert the clinician to any danger of imminent hypoxia in a patient.

Although endogenous molecular oxygen is not optically observable, hemoglobin serves as an oxygen-sensitive "dye" such that when oxygen reversibly binds to the iron atom in the large heme prosthetic group, the electron distribution of the heme is shifted, producing a significant color change. The optical absorption of hemoglobin in its oxygenated and deoxygenated states is shown in Fig. 88.1. Fully oxygenated blood absorbs strongly in the blue and appears bright red; deoxygenated blood absorbs throughout the visible region and is very dark (appearing blue when observed through tissue due to light scattering effects). Thus the optical absorption spectra of oxyhemoglobin (O_2Hb) and "reduced" deoxyhemoglobin (RHb) differ substantially, and this difference provides the basis for spectroscopic determinations of the proportion of the two hemoglobin states. In addition to these two normal functional hemoglobins, there are also *dysfunctional hemoglobins*—carboxyhemoglobin, methemoglobin, and sulfhemoglobin—which are spectroscopically distinct but do not bind oxygen reversibly. Oxygen saturation is therefore defined in Eq. (88.1) only in terms of the *functional saturation* with respect to O_2Hb and RHb:

$$S_aO_2 = \frac{O_2Hb}{RHb + O_2Hb} \times 100\% \tag{88.1}$$

Cooximeters are bench-top analyzers that accept whole blood samples and utilize four or more wavelengths of monochromatic light, typically between 500 and 650 nm, to spectroscopically determine the various individual hemoglobins in the sample. If a blood sample can be provided, this spectroscopic method is accurate and reliable. Attempts to make an equivalent quantitative analysis noninvasively through intact tissue have been fraught with difficulty. The problem has been to contend with the wide variation in scattering and nonspecific absorption properties of very complex heterogeneous tissue. One of the more successful approaches, marketed by Hewlett-Packard, used eight optical wavelengths transmitted through the pinna of the ear. In this approach a "bloodless" measurement is first obtained by squeezing as much blood as possible from an area of tissue; the arterial blood is then allowed to flow back, and the oxygen saturation is determined by analyzing the change in the spectral absorbance characteristics of the tissue. While this method works fairly well, it is cumbersome, operator dependent, and does not always work well on poorly perfused or highly pigmented subjects.

In the early 1970s, Takuo Aoyagi recognized that most of the interfering nonspecific tissue effects could be eliminated by utilizing only the change in the signal during an arterial pulse. Although an early prototype was built in Japan, it was not until the refinements in implementation and application by Biox (now Ohmeda) and Nellcor Incorporated in the 1980s that the technology became widely adopted as a safety monitor for critical care use.

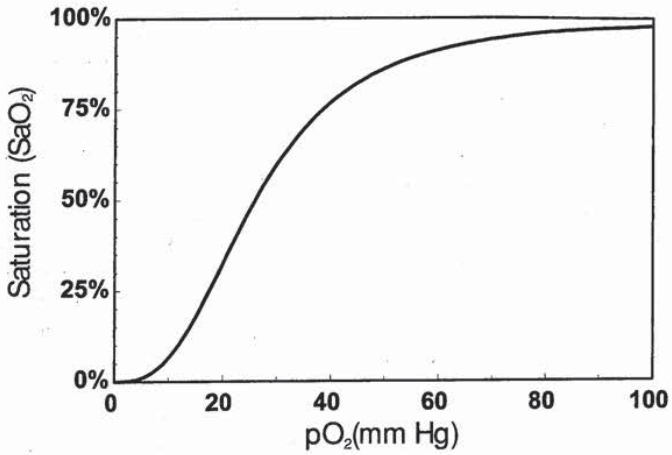


FIGURE 88.2 Hemoglobin oxygen dissociation curve showing the sigmoidal relationship between the partial pressure of oxygen and the oxygen saturation of blood. The curve is given approximately by $\%SaO_2 = 100\%/[1 + P_{50}/pO_2)^n]$, with $n = 2.8$ and $P_{50} = 26$ mm Hg.

Theory

Pulse oximetry is based on the fractional change in light transmission during an arterial pulse at two different wavelengths. In this method the fractional change in the signal is due only to the arterial blood itself, and therefore the complicated nonpulsatile and highly variable optical characteristics of tissue are eliminated. In a typical configuration, light at two different wavelengths illuminating one side of a finger will be detected on the other side, after having traversed the intervening vascular tissues (Fig. 88.3). The transmission of light at each wavelength is a function of the thickness, color, and structure of the skin, tissue, bone, blood, and other material through which the light passes. The

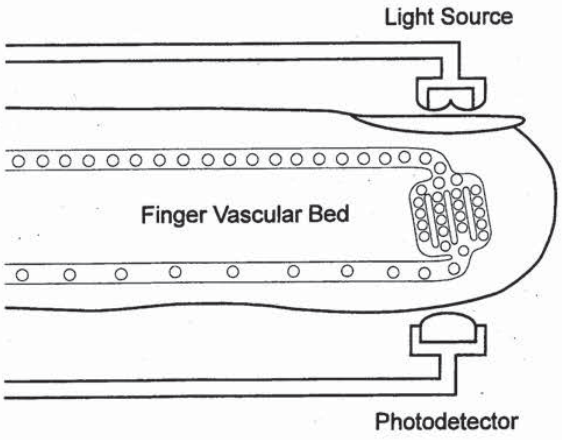


FIGURE 88.3 Typical pulse oximeter sensing configuration on a finger. Light at two different wavelengths is emitted by the source, diffusely scattered through the finger, and detected on the opposite side by a photodetector.

absorbance of light by a sample is defined as the negative logarithm of the ratio of the light intensity in the presence of the sample (I) to that without (I_o): $A = -\log(I/I_o)$. According to the *Beer-Lambert law*, the absorbance of a sample at a given wavelength with a molar absorptivity (ϵ) is directly proportional to both the concentration (c) and pathlength (l) of the absorbing material: $A = \epsilon cl$. (In actuality, biologic tissue is highly scattering, and the Beer-Lambert law is only approximately correct; see the references for further elaboration.) Visible or near-infrared light passing through about one centimeter of tissue (e.g., a finger) will be attenuated by about one or two orders of magnitude for a typical emitter-detector geometry, corresponding to an effective optical density (OD) of 1–2 OD (the detected light intensity is decreased by one order of magnitude for each OD unit). Although hemoglobin in the blood is the single strongest absorbing molecule, most of the total attenuation is due to the scattering of light away from the detector by the highly heterogeneous tissue. Since human tissue contains about 7% blood, and since blood contains typically about 14 g/dL hemoglobin, the effective hemoglobin concentration in tissue is about 1 g/dL (~150 uM). At the wavelengths used for pulse oximetry (650–950 nm), the oxy- and deoxyhemoglobin molar absorptivities fall in the range of 100–1000 $M^{-1}cm^{-1}$, and consequently hemoglobin accounts for less than 0.2 OD of the total observed optical density. Of this amount, perhaps only 10% is pulsatile, and consequently pulse signals of only about a few percent are ultimately measured, at times even one-tenth of this.

A mathematical model for pulse oximetry begins by considering light at two wavelengths, λ_1 and λ_2, passing through tissue and being detected at a distant location as in Fig. 88.3. At each wavelength the total light attenuation is described by four different component absorbances: oxyhemoglobin in the blood (concentration c_o, molar absorptivity ϵ_o, and effective pathlength l_o), "reduced" deoxyhemoglobin in the blood (concentration c_r, molar absorptivity ϵ_r, and effective pathlength l_r), specific variable absorbances that are not from the arterial blood (concentration c_x, molar absorptivity ϵ_x, and effective pathlength l_x), and all other non-specific sources of optical attenuation, combined as A_y, which can include light scattering, geometric factors, and characteristics of the emitter and detector elements. The total absorbance at the two wavelengths can then be written:

$$\begin{cases} A_{\lambda_1} = \epsilon_{o_1} c_o l_o + \epsilon_{r_1} c_r l_r + \epsilon_{x_1} c_x l_x + A_{y_1} \\ A_{\lambda_2} = \epsilon_{o_2} c_o l_o + \epsilon_{r_2} c_r l_r + \epsilon_{x_2} c_x l_x + A_{y_2} \end{cases} \tag{88.2}$$

The blood volume change due to the arterial pulse results in a modulation of the measured absorbances. By taking the time rate of change of the absorbances, the two last terms in each equation are effectively zero, since the concentration and effective pathlength of absorbing material outside the arterial blood do not change during a pulse [$d(c_x l_x)/dt = 0$], and all the nonspecific effects on light attenuation are also effectively invariant on the time scale of a cardiac cycle ($dA_y/dt = 0$). Since the extinction coefficients are constant, and the blood concentrations are constant on the time scale of a pulse, the time-dependent changes in the absorbances at the two wavelengths can be assigned entirely to the change in the blood pathlength (dl_o/dt and dl_r/dt). With the additional assumption that these two blood pathlength changes are equivalent (or more generally, their ratio is a constant), the ratio R of the time rate of change of the absorbance at wavelength *1* to that at wavelength *2* reduces to the following:

$$R = \frac{dA_{\lambda_1}/dt}{dA_{\lambda_2}/dt} = \frac{-d\log(I_1/I_o)/dt}{-d\log(I_2/I_o)/dt} = \frac{(\Delta I_1/I_1)}{(\Delta I_2/I_2)} = \frac{\epsilon_{o_1} c_o + \epsilon_{r_1} c_r}{\epsilon_{o_2} c_o + \epsilon_{r_2} c_r} \tag{88.3}$$

Observing that functional oxygen saturation is given by $S = c_o/(c_o + c_r)$, and that $(1\text{–}S) = c_r/(c_o + c_r)$, the oxygen saturation can then be written in terms of the ratio R as follows:

$$S = \frac{\epsilon_{r1} - \epsilon_{r2} R}{(\epsilon_{r1} - \epsilon_{o1}) - (\epsilon_{r2} - \epsilon_{o2}) R} \tag{88.4}$$

Equation (88.4) provides the desired relationship between the experimentally determined ratio R and the clinically desired oxygen saturation S. In actual use, commonly available LEDs are used as the light sources, typically a red LED near 660 nm and a near-infrared LED selected in the range 890–950 nm. Such LEDs are not monochromatic light sources, typically with bandwidths between 20 and 50 nm, and therefore standard molar absorptivities for hemoglobin cannot be used directly in Eq. (88.4). Further, the simple model presented above is only approximately true; for example, the two wavelengths do not necessarily have the exact same pathlength changes, and second-order scattering effects have been ignored. Consequently the relationship between S and R is instead determined empirically by fitting the clinical data to a generalized function of the form $S = (a - bR)/(c - dR)$. The final empirical calibration will ultimately depend on the details of an individual sensor design, but these variations can be determined for each sensor and included in unique calibration parameters. A typical empirical calibration for R versus S is shown in Fig. 88.4, together with the curve that standard molar absorptivities would predict.

In this way the measurement of the ratio of the fractional change in signal intensity of the two LEDs is used along with the empirically determined calibration equation to obtain a beat-by-beat measurement of the arterial oxygen saturation in a perfused tissue—continuously, noninvasively, and to an accuracy of a few percent.

Applications and Future Directions

Pulse oximetry is now routinely used in nearly all operating rooms and critical care areas in the United States and increasingly throughout the world. It has become so pervasive and useful that it is now being called the "fifth" vital sign (for an excellent review of practical aspects and clinical applications of the technology see Kelleher [1989]).

The principal advantages of pulse oximetry are that it provides continuous, accurate, and reliable monitoring of arterial oxygen saturation on nearly all patients, utilizing a variety of convenient sensors, reusable as well as disposable. Single-patient-use adhesive sensors can easily be applied to fingers for adults and children and to arms or legs for neonates. Surface reflectance sensors have also

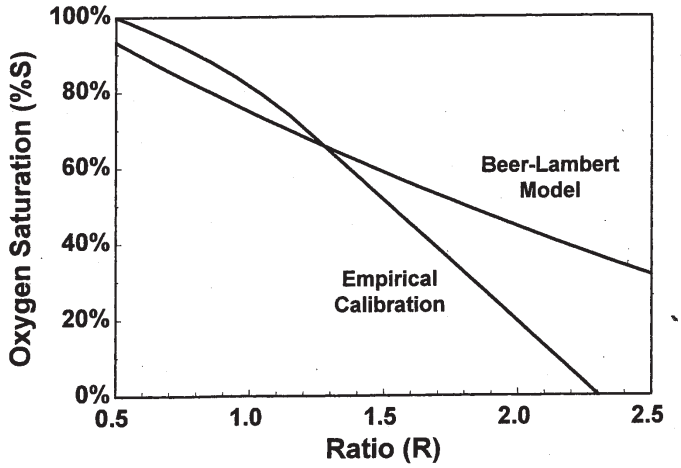


FIGURE 88.4 Relationship between the measured ratio of fractional changes in light intensity at two wavelengths, R, and the oxygen saturation S. Beer-Lambert model is from Eq. (88.4) with $\epsilon_{o_1} = 100$, $\epsilon_{o_2} = 300$, $\epsilon_{r_1} = 800$, and $\epsilon_{r_2} = 200$. Empirical calibration is based on $\%S = 100\% \times (a - bR)/(c - dR)$ with $a = 1000$, $b = 550$, $c = 900$, and $d = 350$, with a linear extrapolation below 70%.

been developed based on the same principles and offer a wider choice for sensor location, though they tend to be less accurate and prone to more types of interference.

Limitations of pulse oximetry include sensitivity to high levels of optical or electric interference, errors due to high concentrations of dysfunctional hemoglobins (methemoglobin or carboxyhemoglobin) or interference from physiologic dyes (such as methylene blue). Other important factors, such as total hemoglobin content, fetal hemoglobin, or sickle cell trait, have little or no effect on the measurement except under extreme conditions. Performance can also be compromised by poor signal quality, as may occur for poorly perfused tissues with weak pulse amplitudes or by motion artifact.

Hardware and software advances continue to provide more sensitive signal detection and filtering capabilities, allowing pulse oximeters to work better on more ambulatory patients. Already some pulse oximeters incorporate ECG synchronization for improved signal processing. A pulse oximeter for use in labor and delivery is currently under active development by several research groups and companies. A likely implementation may include use of a reflectance surface sensor for the fetal head to monitor the adequacy of fetal oxygenation. This application is still in active development, and clinical utility remains to be demonstrated.

88.2 Nonpulsatile Spectroscopy

Background

Nonpulsatile optical spectroscopy has been used for more than half a century for noninvasive medical assessment, such as in the use of multiwavelength tissue analysis for oximetry and skin reflectance measurement for bilirubin assessment in jaundiced neonates. These early applications have found some limited use, but with modest impact. Recent investigations into new nonpulsatile spectroscopy methods for assessment of deep-tissue oxygenation (e.g., cerebral oxygen monitoring), for evaluation of respiratory status at the cellular level, and for the detection of other critical analytes, such as glucose, may yet prove more fruitful. The former applications have led to spectroscopic studies of cytochromes in tissues, and the latter has led to considerable work into new approaches in near-infrared analysis of intact tissues.

Cytochrome Spectroscopy

Cytochromes are electron-transporting, heme-containing proteins found in the inner membranes of mitochondria and are required in the process of oxidative phosphorylation to convert metabolytes and oxygen into CO_2 and high-energy phosphates. In this metabolic process the cytochromes are reversibly oxidized and reduced, and consequently the oxidation-reduction states of cytochromes *c* and aa_3 in particular are direct measures of the respiratory condition of the cell. Changes in the absorption spectra of these molecules, particularly near 600 nm and 830 nm for cytochrome aa_3, accompany this shift. By monitoring these spectral changes, the cytochrome oxidation state in the tissues can be determined (see, for example, Jöbsis [1977] and Jöbsis et al. [1977]). As with all nonpulsatile approaches, the difficulty is to remove the dependence of the measurement on the various nonspecific absorbing materials and highly variable scattering effects of the tissue. To date, instruments designed to measure cytochrome spectral changes can successfully track relative changes in brain oxygenation, but absolute quantitation has not yet been demonstrated.

Near-Infrared Spectroscopy and Glucose Monitoring

Near-infrared (NIR), the spectral region between 780 nm and 3000 nm, is characterized by broad and overlapping spectral peaks produced by the overtones and combinations of infrared vibrational modes. Figure 88.5 shows typical NIR absorption spectra of fat, water, and starch. Exploitation of

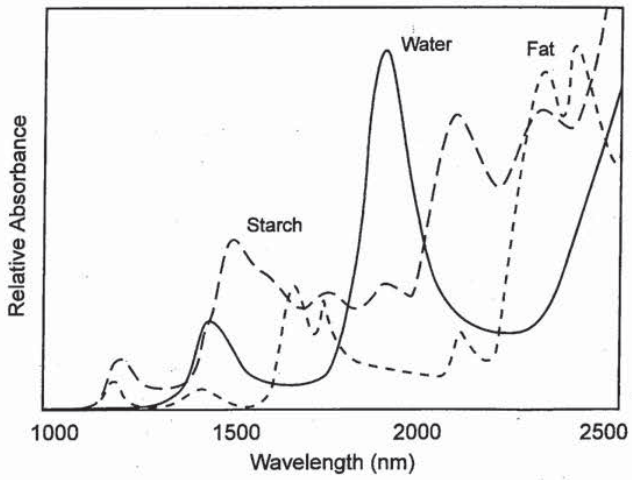


FIGURE 88.5 Typical near-infrared absorption spectra of several biologic materials.

this spectral region for *in vivo* analysis has been hindered by the same complexities of nonpulsatile tissue spectroscopy described above and is further confounded by the very broad and indistinct spectral features characteristic of the NIR. Despite these difficulties, NIR spectroscopy has garnered considerable attention, since it may enable the analysis of common analytes.

Karl Norris and coworkers pioneered the practical application of NIR spectroscopy, using it to evaluate water, fat, and sugar content of agricultural products (see Osborne et al. [1993] and Burns and Cuirczak [1992]). The further development of sophisticated *multivariate analysis* techniques, together with new scattering models (e.g., Kubelka-Munk theory) and high-performance instrumentation, further extended the application of NIR methods. Over the past decade, many research groups and companies have touted the use of NIR techniques for medical monitoring, such as for determining the relative fat, protein, and water content of tissue, and more recently for noninvasive glucose measurement. The body composition analyses are useful but crude and are mainly limited to applications in nutrition and sports medicine. Noninvasive glucose monitoring, however, is of considerable interest.

More than 2 million diabetics in the United States lance their fingers three to six times a day to obtain a drop of blood for chemical glucose determination. The ability of these individuals to control their glucose levels, and the quality of their life generally, would dramatically improve if a simple, noninvasive method for determining blood glucose levels could be developed. Among the noninvasive optical methods proposed for this purpose are optical rotation, NIR analysis, and raman spectroscopy. The first two have received the most attention. Optical rotation methods aim to exploit the small optical rotation of polarized light by glucose. To measure physiologic glucose levels in a 1-cm thick sample to an accuracy of 25 mg/dL would require instrumentation that can reliably detect an optical rotation of at least 1 millidegree. Finding an appropriate in vivo optical path for such measurements has proved most difficult, with most approaches looking to use either the aqueous humor or the anterior chamber of the eye [Coté et al., 1992; Rabinovitch et al., 1982]. Although several groups have developed laboratory analyzers that can measure such a small effect, so far in vivo measurement has not been demonstrated, due both to unwanted scattering and optical activity of biomaterials in the optical path and to the inherent difficulty in developing a practical instrument with the required sensitivity.

NIR methods for noninvasive glucose determination are particularly attractive, although the task is formidable. Glucose has spectral characteristics near 1500 nm and in the 2000–2500 nm band

where many other compounds also absorb, and the magnitude of the glucose absorbance in biologic samples is typically two orders of magnitude lower than those of water, fat, or protein. The normal detection limit for NIR spectroscopy is on the order of one part in 10^3, whereas a change of 25 mg/dL in glucose concentration corresponds to an absorbance change of 10^{-4} to 10^{-5}. In fact, the temperature dependence of the NIR absorption of water alone is at least an order of magnitude greater than the signal from glucose in solution. Indeed some have suggested that the apparent glucose signature in complex NIR spectra may actually be the secondary effect of glucose on the water.

Sophisticated chemometric (particularly, multivariate analysis) methods have been employed to try to extract the glucose signal out of the noise (for methods reviews see Martens and Næs [1989] and Haaland [1992]). Several groups have reported using multivariate techniques to quantitate glucose in whole blood samples, with encouraging results [Haaland et al., 1992]. And despite all theoretical disputations to the contrary, some groups claim the successful application of these multivariate analysis methods to noninvasive in vivo glucose determination in patients [Robinson et al., 1992]. Yet even with the many groups working in this area, much of the work remains unpublished, and few if any of the reports have been independently validated.

Time-Resolved Spectroscopy

The fundamental problem in making quantitative optical measurements through intact tissue is dealing with the complex scattering phenomena. This scattering makes it difficult to determine the effective pathlength for the light, and therefore attempts to use the Beer-Lambert law, or even to determine a consistent empirical calibration, continue to be thwarted. Application of new techniques in time-resolved spectroscopy may be able to tackle this problem. Thinking of light as a packet of photons, if a single packet from a light source is sent through tissue, then a distant receiver will detected a photon distribution over time—the photons least scattered arriving first and the photons most scattered arriving later. In principle, the first photons arriving at the detector passed directly through the tissue. For these first photons the distance between the emitter and the detector is fixed and known, and the Beer-Lambert law should apply, permitting determination of an *absolute* concentration for an absorbing component. The difficulty in this is, first, that the measurement time scale must be on the order of the photon transit time (subnanosecond), and second, that the number of photons getting through without scattering will be extremely small, and therefore the detector must be exquisitely sensitive. Although these considerable technical problems have been overcome in the laboratory, their implementation in a practical instrument applied to a real subject remains to be demonstrated. This same approach is also being investigated for noninvasive optical imaging, since the unscattered photons should produce sharp images (see Chance et al., [1988], Chance [1991], and Yoo and Alfano [1989]).

88.3 Conclusions

The remarkable success of pulse oximetry has established noninvasive optical monitoring of vital physiologic functions as a modality of considerable value. Hardware and algorithm advances in pulse oximetry are beginning to broaden its use outside the traditional operating room and critical care areas. Other promising applications of noninvasive optical monitoring are emerging, such as for measuring deep tissue oxygen levels, determining cellular metabolic status, or for quantitative determination of other important physiologic parameters such as blood glucose. Although these latter applications are not yet practical, they may ultimately impact noninvasive clinical monitoring just as dramatically as pulse oximetry.

Defining Terms

Beer-Lambert law: Principle stating that the optical absorbance of a substance is proportional to both the concentration of the substance and the pathlength of the sample.

Cytochromes: Heme-containing proteins found in the membranes of mitochondria and required for oxidative phosphorylation, with characteristic optical absorbance spectra.

Dysfunctional hemoglobins: Those hemoglobin species that cannot reversibly bind oxygen (carboxyhemoglobin, methemoglobin, and sulfhemoglobin).

Functional saturation: The ratio of oxygenated hemoglobin to total nondysfunctional hemoglobins (oxyhemoglobin plus deoxyhemoglobin).

Hypoxia: Inadequate oxygen supply to tissues necessary to maintain metabolic activity.

Multivariate analysis: Empirical models developed to relate multiple spectral intensities from many calibration samples to known analyte concentrations, resulting in an optimal set of calibration parameters.

Oximetry: The determination of blood or tissue oxygen content, generally by optical means.

Pulse oximetry: The determination of functional oxygen saturation of pulsatile arterial blood by ratiometric measurement of tissue optical absorbance changes.

References

Burns DA, Ciurczak EW (eds). 1992. Handbook of Near-Infrared Analysis. New York, Marcel Dekker.

Chance B. 1991. Optical method. Annu Rev Biophys Biophys Chem 20:1.

Chance B, Leigh JS, Miyake H, et al. 1988. Comparison of time-resolved and -unresolved measurements of deoxyhemoglobin in brain. Proc Natl Acad Sci USA 85(14):4971.

Coté GL, Fox MD, Northrop RB. 1992. Noninvasive optical polarimetric glucose sensing using a true phase measurement technique. IEEE Trans Biomed Eng 39(7):752.

Haaland DM. 1992. Multivariate calibration methods applied to the quantitative analysis of infrared spectra. In PC Jurs (ed), Computer-Enhanced Analytical Spectroscopy, vol 3, pp 1–30. New York Plenum.

Haaland DM, Robinson MR, Koepp GW, et al. 1992. Reagentless near-infrared determination of glucose in whole blood using multivariate calibration. Appl Spectros 46(10):1575.

Jöbsis FF. 1977. Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. Science 198(4323):1264.

Jöbsis FF, Keizer JH, LaManna JC, et al. 1977. Reflectance spectrophotometry of cytochrome aa_3 in vivo. J Appl Physiol 43(5):858.

Kelleher JF. 1989. Pulse oximetry. J Clin Monit 5(1):37.

Martens H, Næs T. 1989. Multivariate Calibration. New York, John Wiley.

Osborne BG, Fearn T, Hindle PH. 1993. Practical NIR Spectroscopy with Applications in Food and Beverage Analysis. Essex, England, Longman Scientific & Technical.

Payne JP, Severinghaus JW (eds). 1986. Pulse Oximetry. New York, Springer-Verlag.

Rabinovitch B, March WF, Adams RL. 1982. Noninvasive glucose monitoring of the aqueous humor of the eye: Part I. Measurement of very small optical rotations. Diabetes Care 5(3):254.

Robinson MR, Eaton RP, Haaland DM, et al. 1992. Noninvasive glucose monitoring in diabetic patients: a preliminary evaluation. Clin Chem 38(9):1618.

Severinghaus JW, Astrup PB. 1986. History of blood gas analysis. VI. Oximetry. J Clin Monit 2(4):270.

Severinghaus JW, Honda Y. 1987*a*. History of blood gas analysis. VII. Pulse oximetry. J Clin Monit 3(2):135.

Severinghaus JW, Honda Y. 1987*b*. Pulse oximetry. Int Anesthesiol Clin 25(4):205.

Severinghaus JW, Kelleher JF. 1992. Recent developments in pulse oximetry. Anesthesiology 76(6):1018.

Tremper KK, Barker SJ. 1989. Pulse oximetry. Anesthesiology 70(1):98.

Wukitsch MW, Petterson MT, Tobler DR, et al. 1988. Pulse oximetry: Analysis of theory, technology, and practice. J Clin Monit 4(4):290.

Yoo KM, Alfano RR. 1989. Photon localization in a disordered multilayered system. Phys Rev B 39(9):5806.

Further Information

Two collections of papers on pulse oximetry include a book edited by J. P. Payne and J. W. Severinghaus, *Pulse Oximetry* (New York, Springer-Verlag, 1986), and a journal collection—International Anesthesiology Clinics [25(4), 1987]. For technical reviews of pulse oximetry, see J. A. Pologe's 1987 "Pulse Oximetry" [Int Anesthesiol Clin 25(3):137], Kevin K. Tremper and Steven J. Barker's 1989 "Pulse Oximetry" [Anesthesiology 70(1):98], and Michael W. Wukitsch, Michael T. Patterson, David R. Tobler, and coworkers' 1988 "Pulse Oximetry: Analysis of Theory, Technology, and Practice" [J Clin Monit 4(4):290].

For a review of practical and clinical applications of pulse oximetry, see the excellent review by Joseph F. Kelleher [1989] and John Severinghaus and Joseph F. Kelleher [1992]. John Severinghaus and Yoshiyuki Honda have written several excellent histories of pulse oximetry [1987*a*, 1987*b*].

For an overview of applied near-infrared spectroscopy, see Donald A. Burns and Emil W. Ciurczak [1992] and B. G. Osborne, T. Fearn, and P. H. Hindle [1993]. For a good overview of multivariate methods, see Harald Martens and Tormod Næs [1989].

89

Medical Instruments and Devices Used in the Home

Bruce R. Bowman
EdenTec Corporation

Edward Schuck
EdenTec Corporation

89.1 Scope of the Market for Home Medical Devices

The market for medical devices used in the home and alternative sites has increased dramatically in the last 10 years and has reached an overall estimated size of more than $1.6 billion [FIND/SVP, 1992]. In the past, hospitals have been thought of as the only places to treat sick patients. But with the major emphasis on reducing healthcare costs, increasing numbers of sicker patients move from hospitals to their homes. Treating sicker patients outside the hospital places additional challenges on medical device design and patient use. Equipment designed for hospital use can usually rely on trained clinical personnel to support the devices. Outside the hospital, the patient and/or family members must be able to use the equipment, requiring these devices to have a different set of design and safety features. This chapter will identify some of the major market segments using medical devices in the home and discuss important design considerations associated with home use.

Table 89.1 outlines market segments where devices and products are used to treat patients outside the hospital [FIND/SVP, 1992]. The durable medical equipment market is the most established market providing aids for patients to improve access and mobility. These devices are usually not life supporting or sustaining, but in many cases they can make the difference in allowing a patient to be able to function outside a hospital or nursing or skilled facility. Other market segments listed employ generally more sophisticated solutions to clinical problems. These will be discussed by category of use.

The incontinence and ostomy area of products is one of the largest market segments and is growing in direct relationship to our aging society. Whereas sanitary pads and colostomy bags are not very "high-tech," well-designed aids can have a tremendous impact on the comfort and independence of these patients. Other solutions to incontinence are technically more sophisticated, such as use of electric stimulation of the sphincter muscles through an implanted device or a miniature stimulator inserted as an anal or vaginal plug to maintain continence [Wall et al., 1993].

Many forms of equipment are included in the Respiratory segment. These devices include those that maintain life support as well as those that monitor patients' respiratory function. These patients,

0-8493-8346-3/95/$0.00+$.50

TABLE 89.1 Major Market Segments Outside Hospitals

Market Segment	Estimated Equipment Size 1991	Device Examples
Durable medical equipment	$373 M*	Specialty beds, wheelchairs, toilet aids, ambulatory aids
Incontinence and ostomy products	$600 M*	Sanitary pads, electrical stimulators, colostomy bags
Respiratory equipment	$180 M*	Oxygen therapy, portable ventilators, nasal CPAP, monitors, apnea monitors
Drug infusion, drug measurement	$300 M	Infusion pumps, access ports, patient-controlled analgesia (PCA), glucose measurement, implantable pumps
Pain control and functional stimulation	$140 M	Transcutaneous electrical nerve stimulation (TENS), functional electrical nerve stimulation (FES)

**Source:* FIND/SVP [1992].

with proper medical support, can function outside the hospital at a significant reduction in cost and increased patient comfort [Pierson, 1994]. One area of this segment, infant apnea monitors, provides parents or caregivers the cardio/respiratory status of an at-risk infant so that intervention (CPR etc.) can be initiated if the baby has a life-threatening event. The infant monitor shown in Fig. 89.1 is an example of a patient monitor designed for home use and will be discussed in more detail later in this chapter. Pulse oximetry monitors are also going home with patients. They are used to measure noninvasively the oxygen level of patients receiving supplemental oxygen or ventilator-dependent patients to determine if they are being properly ventilated.

Portable infusion pumps are an integral part of providing antibiotics, pain management, chemotherapy, and parenteral and enteral nutrition. The pump shown in Fig. 89.2 is an example of technology that allows the patient to move about freely while receiving sometimes lengthy drug therapy. Implantable drug pumps are also available for special long-term therapy needs.

Pain control using electric stimulation in place of drug therapy continues to be an increasing market. The delivery of small electric impulses to block pain is continuing to gain medical accep-

FIGURE 89.1 Infant apnea monitor used in a typical home setting (photo courtesy of EdenTec Corporation).

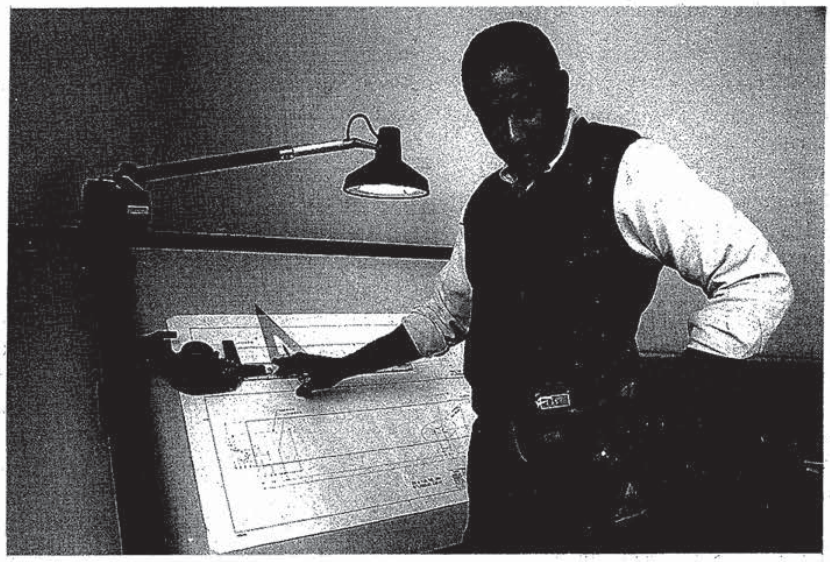


FIGURE 89.2 Portable drug pump used throughout the day (photo courtesy of Pharmacia Deltec Inc.).

tance for treatment outside the hospital setting. A different form of electric stimulation called functional electric stimulation (FES) applies short pulses of electric current to the nerves that control weak or paralyzed muscles. This topic is covered as a separate chapter in this book.

Growth of the homecare business has created problems in overall healthcare costs since a corresponding decrease in hospital utilization has not yet occurred. In the future, however, increased homecare will necessarily result in reassessment and downsizing in the corresponding hospital segment. There will be clear areas of growth and areas of consolidation in the new era of healthcare reform. It would appear, however, that homecare has a bright future of continued growth.

89.2 Unique Challenges to the Design and Implementation of High-Tech Homecare Devices

What are some of the unique requirements of devices that could allow more sophisticated equipment to go home with ordinary people of varied educational levels without compromising their care? Even though each type of clinical problem has different requirements for the equipment that must go home with the patient, certain common qualities must be inherent in most devices used in the home. Three areas to consider when equipment is used outside of the hospital are that the device (1) must provide a positive clinical outcome, (2) must be safe and easy to use, and (3) must be user-friendly enough so that it will be used.

The Device Must Provide a Positive Clinical Outcome

Devices cannot be developed any longer just because new technology becomes available. They must solve the problem for which they were intended and make a significant clinical difference in the outcome or management of the patient while saving money. These realities are being driven by those who reimburse for devices, as well as by the FDA as part of the submission for approval to market a new device.

The Device Must Be Safe to Use

Homecare devices may need to be even *more* reliable and even *safer* than hospital devices. We often think of hospitals as having the best quality and most expensive devices that money can buy. In addition to having the best equipment to monitor patients, hospitals have nurses and aids that keep an eye on patients so that equipment problems may be quickly discovered by the staff. A failure in the home may go unnoticed until it is too late. Thus systems for home use really need extra reliability with automatic backup systems and/or early warning signals.

Safety issues can take on a different significance depending on the intended use of the device. Certain safety issues are important regardless of whether the device is a critical device such as an implanted cardiac pacemaker or a noncritical device such as a bed-wetting alarm. No device should be able to cause harm to the patient regardless of how well or poorly it may be performing its intended clinical duties. Devices must be safe when exposed to all the typical environmental conditions to which the device could be exposed while being operated by the entire range of possible users of varied education and while exposed to siblings and other untrained friends or relatives. For instance, a bed-wetting alarm should not cause skin burns under the sensor if a glass of water spills on the control box. This type of safety issue must be addressed even when it significantly affects the final cost to the consumer.

Other safety issues are not obviously differentiated as to being actual safety issues or simply nuisances or inconveniences to the user. It is very important for the designer to properly define these issues; although some safety features can be included with little or no extra cost, other safety features may be very costly to implement. It may be a nuisance for the patient using a TENS pain control stimulator to have the device inadvertently turned off when its on/off switch is bumped while watching TV. In this case, the patient only experiences a momentary cessation of pain control until the unit is turned back on. But it could mean injuries or death to the same patient driving an automobile who becomes startled when his TENS unit inadvertently turns on and he causes an accident.

Reliability issues can also be mere inconveniences or major safety issues. Medical devices should be free of design and materials defects so that they can perform their intended functions reliably. Once again, reliability does not necessarily need to be expensive and often can be obtained with good design. Critical devices, i.e., devices that could cause death or serious injury if they stopped operating properly, may need to have redundant systems for backup, which likely will increase cost.

The Device Must Be Designed So That It *Will* Be Used

A great deal of money is being spent in healthcare on device for patients that end up not being used. There are numerous reasons for this happening including that the wrong device was prescribed for the patient's problem in the first place; the device works, but it has too many false alarms; the device often fails to operate properly; it is cumbersome to use or difficult to operate or too uncomfortable to wear.

Ease of Use

User-friendliness is one of the most important features in encouraging a device to be used. Technological sophistication may be just as necessary in areas that allow ease of use as in attaining accuracy and reliability in the device. The key is that the technologic sophistication be transparent to the user so that the device does not intimidate the user. Transparent features such as automatic calibration or automatic sensitivity adjustment may help allow successful use of a device that would otherwise be too complicated.

Notions of what makes a device easy to use, however, need to be thoroughly tested with the patient population intended for the device. Caution needs to be taken in defining what "simple" means to different people. A VCR may be simple to the designer because all features can be programmed with one button, but it may not be simple to users if they have to remember that it takes two long pushes and one short to get into the clock-setting program.

Convenience for the user is also extremely important in encouraging use of a device. Applications that require devices to be portable must certainly be light enough to be carried. Size is almost always important for anything that must fit within the average household. Either a device must be able to be left in place in the home or it must be easy to set up, clean, and put away. Equipment design can make the difference between the patient appropriately using the equipment or deciding that it is just too much hassle to bother.

Reliability

Users must also have confidence in the reliability of the device being used and must have confidence that if it is not working properly, the device will tell them that something is wrong. Frequent breakdowns or false alarms will result in frustration and ultimately in reduced compliance. Eventually patients will stop using the device altogether. Most often, reliability can be designed into a product with little or no extra cost in manufacturing, and everything that can be done at no cost to enhance reliability should be done. It is very important, however, to understand what level of additional reliability involving extra cost is necessary for product acceptance. Reliability can always be added by duplicated backup systems, but the market or application may not warrant such an approach. Critical devices which are implanted, such as cardiac pacemakers, have much greater reliability requirements, since they involve not only patient frustration but also safety.

Cost Reimbursement

Devices must be paid for before the patient can realize the opportunity to use new, effective equipment. Devices are usually paid for by one of two means. First, they are covered on an American Medical Association Current Procedural Terminology Code (CPT-code) which covers the medical, surgical, and diagnostic services provided by physicians. The CPT-codes are usually priced out by Medicare to establish a baseline reimbursement level. Private carriers usually establish a similar or different level of reimbursement based on regional or other considerations. Gaining new CPT-codes for new devices can take a great deal of time and effort. The second method is to cover the procedure and device under a capitated fee where the hospital is reimbursed a lump sum for a procedure including the device, hospital, homecare, and physician fees.

Every effort should be made to design devices to be low cost. Device cost is being scrutinized more and more by those who reimburse. It is easy to state, however, that a device needs to be inexpensive. Unfortunately the reality is that healthcare reforms and new regulations by FDA are making medical devices more costly to develop, to obtain regulatory approvals for [FDA, 1993], and to manufacture.

Professional Medical Service Support

The more technically sophisticated a device is, the more crucial that homecare support and education be a part of a program. In fact, in many cases, such support and education are as important as the device itself.

Medical service can be offered by numerous homecare service companies. Typically these companies purchase the equipment instead of the patient, and a monthly fee is charged for use of the equipment along with all the necessary service. The homecare company then must obtain reimbursement from third-party payers. Some of the services offered by the homecare company include training on how to use the equipment, CPR training, transporting the equipment to the home, servicing/repairing equipment, monthly visits, and providing on-call service 24 hours a day. The homecare provider must also be able to provide feedback to the treating physician on progress of the treatment. This feedback may include how well the equipment is working, the patient's medical status, and compliance of the patient.

89.3 Infant Monitor Example

Many infants are being monitored in the home using apnea monitors because they have been identified with breathing problems [Kelly, 1992]. Theses include newborn premature babies who have

apnea of prematurity [Henderson-Smart, 1992; NIH, 1987], siblings of babies who have died of sudden infant death syndrome (SIDS) [Hunt, 1992; NIH, 1987], or infants who have had an apparent life-threatening episode (ALTE) related to lack of adequate respiration [Kahn et al. 1992; NIH, 1987]. Rather than keeping infants in the hospital for a problem that they may soon outgrow (1–6 months), doctors often discharge them from the hospital with an infant apnea monitor that measures the duration of breathing pauses and heart rate and sounds an alarm if either parameter crosses limits prescribed by the doctor.

Infant apnea monitors are among the most sophisticated devices used routinely in the home. These devices utilize microprocessor control, sophisticated breath-detection and artifact rejection firmware algorithms, and internal memory that keeps track of use of the device as well as recording occurrence of events and the physiologic waveforms associated with the events. The memory contents can be downloaded directly to computer or sent via modem remotely where a complete 45-day report can be provided to the referring physician (see Fig. 89.3).

Most apnea monitors measure breathing effort through impedance pneumography. A small (100–200 uA) high-frequency (25–100 kHz) constant-current train of pulses is applied across the chest between a pair of electrodes. The voltage needed to drive the current is measured, and thereby the effective impedance between the electrodes can be calculated. Impedance across the chest increases as the chest expands and decreases as the chest contracts with each breath. The impedance change with each breath can be as low as 0.2 ohms on top of an electrode base impedance of 2000 ohms, creating some interesting signal-to-noise challenges. Furthermore, motion artifact and blood volume changes in the heart and chest can cause impedance changes of 0.6 ohms or more that can look just like breathing. Through the same pair of electrodes, heart rate is monitored by picking up the electrocardiogram (ECG) [AAMI, 1988].

Because the impedance technique basically measures motion of the chest, this technique can only be used to monitor central apnea or lack of breathing effort. Another less common apnea in infants called obstructive apnea results when an obstruction of the airway blocks air from flowing in spite of breathing effort. Obstructive apnea can not be monitored using impedance pneumography [Kelly, 1992].

There is a very broad socioeconomic and educational spectrum of parents or caregivers who may be monitoring their infants with an apnea monitor. This places an incredible challenge for the design of the device so that it is easy enough to be used by a variety of caregivers. It also puts special requirements on the homecare service company that must be able to respond to these patients within a matter of minutes, 24 hours a day.

The user-friendly monitor shown in Fig. 89.1 uses a two-button operation, the on/off switch, and a reset switch. The visual alarm indicators are invisible behind a back-lit panel except when an actual alarm occurs. A word describing the alarm then appears. By not showing all nine possible alarm conditions unless an alarm occurs, parent confusion and anxiety is minimized. Numerous safety features are built into the unit, some of which are noticeable but many of which are internal to the operation of the monitor. One useful safety feature is the self-check. When the device is turned on, each alarm LED lights in sequence, and the unit beeps once indicating that the self-check was completed successfully. This gives users the opportunity to confirm that all the alarm visual indicators and the audible indicator are working and provides added confidence for users leaving their baby on the monitor. A dual-level battery alarm gives an early warning that the battery will soon need charging. The weak battery alarm allows users to reset the monitor and continue monitoring their babies for several more hours before depleting the battery to the charge battery level where the monitor must be attached to the ac battery charger/adapter. This allows parents the freedom to leave their homes for a few hours knowing that their child can continue to be monitored.

A multistage alarm reduces the risk of parents sleeping through an alarm. Most parents are sleep-deprived with a new baby. Consequently, it can be easy for parents in a nearby room to sleep through a monitor alarm even when the monitor sounds at 85 dB. A three-stage alarm helps to reduce this risk. After 10 seconds of sounding at 1 beep per second, the alarm switches to 3 beeps per second for

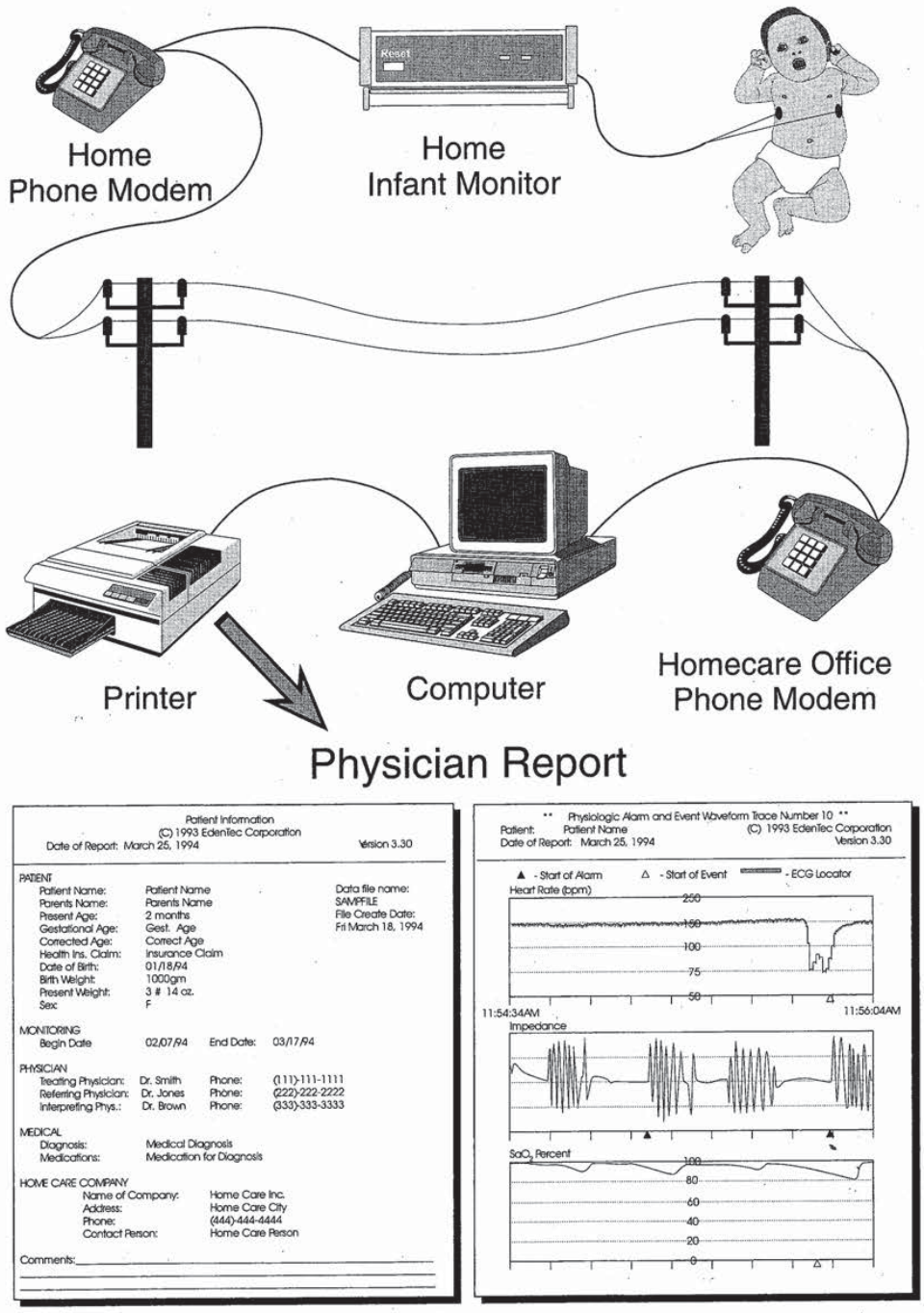


FIGURE 89.3 Infant apnea monitor with memory allows data to be sent by modem to generate physician report (drawing courtesy of EdenTec Corporation).

the next 10 seconds. Finally, if an alarm has not resolved itself after 20 seconds, the alarm switches to 6 beeps per second. Each stage of alarm sounds more intense than the previous one and offers the chance of jolting parents out of even the deepest sleep.

The physician always prescribes what alarm settings should be used by the homecare service company when setting up the monitor. As a newborn baby matures, these settings may need to be adjusted. Sometimes the parents can be relied upon for making these setting changes. To allow both accessibility to these switches as well as to keep them safe from unauthorized tampering from a helping brother or sister, a special tamper-resistant-adjustment procedure is utilized. Two simultaneous actions are required in order to adjust the alarm limit settings. The reset button must be continually pressed on the front of the unit while changing settings on the back of the unit. Heart rate levels are set in beats per minute, and apnea duration is set in single-second increments. Rather than using easy-to-set push-button switches, "pen-set" switches are used which require a pen or other sharp implement to make the change. If the proper switch adjustment procedure is not followed, the monitor alarms continuously and displays a switch alarm until the settings are returned to their original settings. A similar technique is used for turning the monitor Off. The reset button must first be pressed and then the on/off switch turned to the off position. Violation of this procedure will result in a switch alarm.

Other safety features are internal to the monitor and are transparent to the user. The monitor's alarm is designed to be normally on from the moment the device is turned on. Active circuitry controlled by the microprocessor turns the alarm off when there are no active alarm conditions. If anything hangs up the processor or if any of a number of components fail, the alarm will not turn off and will remain on in a fail-safe mode. This "alarm on unless turned off" technique is also used in a remote alarm unit for parents with their baby in a distant room. If a wire breakage occurs between the monitor and the remote alarm unit, or a connector pulls loose, or a component fails, the remote alarm no longer is turned off by the monitor and it alarms in a fail-safe condition.

Switches, connectors, and wires are prone to fail. One way to circumvent this potential safety issue is use of switches with a separate line for each possible setting. The monitor continuously polls every switch line of each switch element to check that "exactly" one switch position is making contact. This guards against misreading bad switch elements, a switch inadvertently being set between two positions, or a bad connector or cable. Violation of the exactly one contact condition results in a switch alarm.

It is difficult to manage an apnea monitoring program in rural areas where the monitoring family may be a hundred miles or more away from the homecare service company. There are numerous ways to become frustrated with the equipment and stop using the monitor. Therefore, simplicity of use and reliability are important. Storing occurrence of alarms and documenting compliance in internal memory in the monitor help the homecare service company and the remote family cope with the situation. The monitor shown in Fig. 89.1 stores in digital memory the time, date, and duration of (1) each use of the monitor; (2) occurrence of all equipment alarms; and (3) all physiologic alarms including respiratory waveforms, heart rate, and ECG for up to a 45-day period. These data in the form of a report (see Fig. 89.3) can be downloaded to a laptop PC or sent via modem to the homecare service company or directly to the physician.

89.4 Conclusions

Devices that can provide positive patient outcomes with reduced overall cost to the healthcare system while being safe, reliable, and user-friendly will succeed based on pending healthcare changes. Future technology in areas of sensors, communications, and memory capabilities should continue to increase the potential effectiveness of homecare management programs by using increasingly sophisticated devices. The challenge for the medical device designer is to provide cost-effective, reliable, and easy-to-use solutions that can be readily adopted by the multidisciplinary aspects of homecare medicine while meeting FDA requirements.

Defining Terms

Apnea: Cessation of breathing. Apnea can be classified as **central, obstructive,** or mixed, which is a combination.

Apnea of prematurity: Apnea in which the incidence and severity increases with decreasing gestational age attributable to immaturity of the respiratory control system. The incidence has increased due to improved survival rates for very-low-birth-weight premature infants.

Apparent life-threatening episode (ALTE): An episode characterized by a combination of apnea, color change, muscle tone change, choking, or gagging. To the observer it may appear the infant has died.

Capitated fee: A fixed payment for *total* program services versus the more traditional fee for service in which each individual service is charged.

Cardiac pacemaker: A device that electrically stimulates the heart at a certain rate used in absence of normal function of the heart's sino-atrial node.

Central apnea: Apnea secondary to lack of respiratory or diaphragmatic effort.

Chemotherapy: Treatment of disease by chemical agents. Term popularly used when fighting cancer chemically.

Colostomy: The creation of a surgical hole as an alternative opening of the colon.

CPR (cardiopulmonary resuscitation): Artificially replacing heart and respiration function through rhythmic pressure on the chest.

CPT-code (current procedural terminology code): A code used to describe specific procedures/tests developed by the AMA.

Electrocardiogram (ECG): The electric potential recorded across the chest due to depolarization of the heart muscle with each heartbeat.

Enteral nutrition: Chemical nutrition injected intestinally.

Food and Drug Administration (FDA): Federal agency that oversees and regulates foods, drugs, and medical devices.

Functional electrical stimulation (FES): Electric stimulation of peripheral nerves or muscles to gain functional, purposeful control over partially or fully paralyzed muscles.

Incontinence: Loss of voluntary control of the bowel or bladder.

Obstructive apnea: Apnea in which effort to breath continues but airflow ceases due to obstruction or collapse of the airway.

Ostomy: Surgical procedure that alters the bladder or bowel to eliminate through an artificial passage.

Parenteral nutrition: Chemical nutrition injected subcutaneously, intramuscularly, intrasternally, or intravenously.

Sphincter: A band of muscle fibers that constricts or closes an orifice.

Sudden infant death syndrome (SIDS): The sudden death of an infant which is unexplained by history or postmortem exam.

Transcutaneous electrical nerve stimulation (TENS): Electrical stimulation of sensory nerve fibers resulting in control of pain.

References

AAMI. 1988. Association for the Advancement of Medical Instrumentation Technical Information Report: Apnea Monitoring by Means of Thoracic Impedance Pneumography, Arlington, Virg.

FDA. November 1993. Reviewers Guidance for Premarket Notification Submissions (Draft), Anesthesiology and Respiratory Device Branch. Division of Cardiovascular, Respiratory, and Neurological Devices. Food and Drug Administration. Washington DC.

FIND/SVP. 1992. The Market for Home Care Products, a Market Intelligence Report. New York.

Henderson-Smart, DJ. 1992. Apnea of prematurity. In R Beckerman, R Brouillette, C Hunt (eds), Respiratory Control Disorders in Infants and Children, pp 161–177, Baltimore, Williams and Wilkins.

Hunt, CE. 1992. Sudden infant death syndrome. In R Beckerman, R Brouillette, C Hunt (eds), Respiratory Control Disorders in Infants and Children, pp 190–211, Baltimore, Williams and Wilkins.

Kahn A, Rebuffat E, Franco P, et al. 1992. Apparent life-threatening events and apnea of infancy. In R Beckerman, R Brouillette, C Hunt (eds), Respiratory Control Disorders in Infants and Children, pp 178–189, Baltimore, Williams and Wilkins.

Kelly DH. 1992. Home monitoring. In R Beckerman, R Brouillette, C Hunt (eds), Respiratory Control Disorders in Infants and Children, pp 400–412, Baltimore, Williams and Wilkins.

NIH. 1987. Infantile Apnea and Home Monitoring, Report of NIH Consensus Development Conference, US Department of Health and Human Services, NIH publication 87-2905.

Pierson DJ. 1994. Controversies in home respiratory care: Conference summary. Respir Care 39(4):294.

Wall LL, Norton PA, Dehancey JOL. 1993. Practical Urology, Baltimore, Williams and Wilkins.