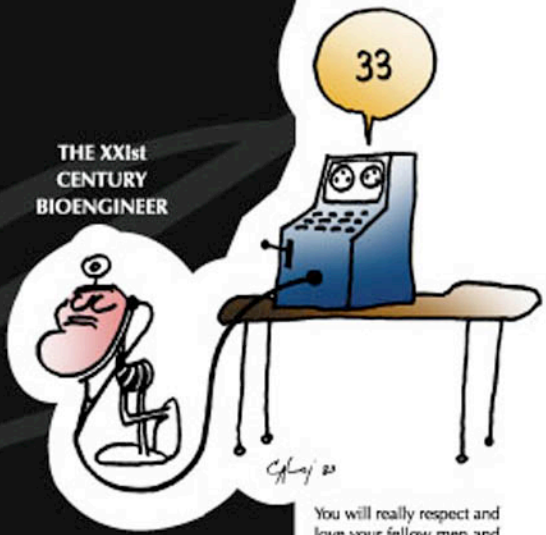
Series on Bioengineering & Biomedical Engineering – Vol. 4

UNDERSTANDING THE HUMAN MACHINE

A Primer for Bioengineering

Max E Valentinuzzi

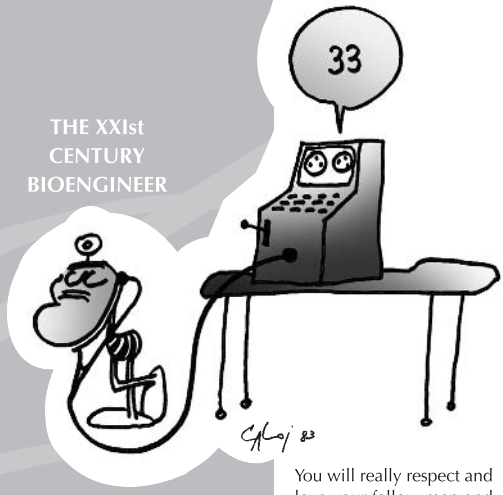


You will really respect and love your fellow men and women the very moment you learn how to laugh out of yourself

Series on Bioengineering & Biomedical Engineering – Vol. 4

UNDERSTANDING THE HUMAN MACHINE

A Primer for Bioengineering



You will really respect and love your fellow men and women the very moment you learn how to laugh out of yourself

Max E Valentinuzzi
Universidad Nacional de Tucumán, Argentina

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401–402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Cover design: Copyright Comic by Caloi, 1983.

Carlos Loiseau, Caloi, is a well-reputed and internationally-known graphic humorist (www.caloi.com.ar). In 1973, he created Clemente, a comic strip character with a keen human perception. Clemente met Carmelo J. Felice (contributor to this book), in Bariloche, Argentina, in 1983, and immediately became a bioengineer. Ever since, he has presided our Lab from a frame hanging on one wall of the author's office. Now, he is kind enough to joyfully lead you into this textbook, with his courtesy and wit, as the paradigmatic XXIst Century Bioengineer.

UNDERSTANDING THE HUMAN MACHINE
A Primer for Bioengineering
Series on Biomaterials and Bioengineering — Vol. 4

ISBN 981-238-930-X

Printed in Singapore.

Dedication

To my paternal Grandfather, Máximo I, carpenter,
for his everlasting kindness and generosity;

To my paternal Grandmother, Clementina, laundrywoman,
for the unrelenting concern aimed at her children's education;

To my maternal Grandmother, Concepción, seamstress, who could not read and write,
for her sweetness and constant love.

Their words, acts, faces, are still vivid in my memory.

My maternal Grandfather, Vicente, tailor,
died of lung cancer when I was only a few months old.

They gave me my parents, Emma and Máximo II;
my mother had only fourth grade, was raised in poverty and offered everything she could for the education of her children;

my father, who was also raised in poverty, and after a life spent with medicine and biophysics, as physician first and later on as teacher and researcher, said in his farewell words to his last students at the Chicago College of Osteopathic Medicine, in Chicago, on May 3, 1972, coincidentally and exactly 30 years ago:

"*I wish you luck*, *success and happiness. Study*, *work and enjoy life. Preserve ideals and dreams above everything. As members of a society*, *think and behave in terms of mankind. Strive for peace*, *toleration and mutual understanding. Your fellowmen possess as much truth as yourselves. Give your soul and a soul will be given to you. Someday*, *on the Earth*, *the Republic of Mankind will be established*".

Let me retransmit this message to you, young Men and Women of the XXIst Century.

Max E. Valentinuzzi

Tucumán, Argentina, May 3, 2002

Preface

This book is the result of over 35 years of teaching bioengineering and physiology at the undergraduate and graduate levels, in this country and abroad, in engineering and in medical schools, from the perspective of a simple teaching assistant under the supervision of great teachers like Hebbel E. Hoff and Leslie A. Geddes, at Baylor College of Medicine, in Houston, Texas, USA, to the level of full professorship and directorship of a graduate program, of collecting notes and material of different kinds, of actual research and technological development activities, of close interaction with students and collaborators alike, of joyful and sometimes painful work.

The basic idea for the eight proposed chapters is that of an **introduction intended for undergraduate students**; thus, they need an **overview**, **to grasp concepts**. The big question to answer is: **What is bioengineering all about**?

Sometimes, the subject will be treated in relative depth; sometimes the visit will be more superficial. I would like to stay away of encyclopedism, that is, trying to favor formation rather than mere information. Whenever possible, historical data would supply background material and spicy insights. Style should be light, sprinkled with a little humor. Smaller case will be used for the less important. Skip it, if you wish. However, **the core content will be preserved to avoid transforming a technical book into an easy-going novel**. There will be exercises to make the students think and search. He and she are supposed to work out the material, the suggested readings, the problems, they are supposed to search and dig. The student needs to learn how to learn by himself or herself, for in real life that is what counts. However, this sweating and hard process must always be an enjoyable one. If you cannot find the element of fun, then you are out of business. The book, if successful, should motivate even more the student in his/her endeavor to proceed further in the fascinating bioengineering/biomedical engineering enterprise. It should answer some questions and leave open many others; it should show new avenues to continue.

At the beginning of each chapter, there will be a short preface, some kind of an introduction that will help in its general overview. The bibliography, listed in alphabetical and chronological order at the end of the book, shall be selective and not excessive. The student gets overpowered when it is too much. Nontheless, do not fall into the mistake of thinking that the contents herein will suffice to satisfy all and every single doubt you may have. In that case, the student is invited to make his/her own searches in the WEB for any of the subjects presented in this text. Some specific addresses have been interspersed and checked in due time, however, they have a relative value because by the time you read them, they may be no longer available. Advances in science are happening extremely fast and are dramatic, and bioengineering is probably one of the leaders in this respect. Be prepared to the scenery and may this book serve you as a launching platform.

Any comment, observation, suggestion or criticism will be most welcome and can be transmitted to the author to the following e-mail addresses: maxvalentinuzzi@hotmail.com or maxvalentinuzzi@arnet.com.ar.

Max E. Valentinuzzi
Department of Bioengineering
Universidad Nacional de Tucumán
Cc 327, (4000)Tucumán, Argentina

Acknowledgments

Many inputs can be directly traced back to former and present collaborators; some were students for relatively short periods while others stayed with me during many years. As in any family, most of them left searching for their own places, several enjoying successful professional and/or academic lives either in this country or abroad. Their names cannot be listed here but, if they ever run across this book, they will find themselves hidden in its lines. There are direct collaborators to this text: Veronica S. Valentinuzzi, my daughter, with the pineal section in Chapter 2, Section 6, Emilce Moler, with images in Chapter 3, Carmelo J. Felice and Rossana Madrid, with Chapter 4, and Myriam C. Herrera, with Chapter 5. Ron Leder, currently with the Universidad Autónoma Metropolitana Iztapalapa, generously and always full of good humor, passed me his inputs, too, especially for Chapter 6. They are specifically recognized in the proper places including also at the respective ends a short biographical note. Roger Guillemin, one of my dearest former professors at Baylor College, was kind enough to peruse the endocrinology section offering also his wise advice and comments. Frederik Nebeker, Senior Research Historian with the IEEE History Center, graciously gave me some linguistic counseling after reading samples of the book. **To all of them**, **my deep appreciation and recognition**. My gratefulness also to Dr. John K–J Li, editorial adviser from Rutgers University, to Yubing Zhai, Commissioning Editor at the New Jersey Office of World Scientific Publishing Company (WSPC), to the arbiters who reviewed my proposal and forwarded excellent comments and observations (hopefully well included in the book), to World Scientific Publishing Company which finally accepted the project. They all trusted in me and offered me the not-easy-to-get opportunity of addressing an audience so superb as the biomedical engineering studentship is. To Steven Patt, Desk Editor, and Yolande Koh, Production Manager, stationed at the Singapore Office of WSPC, who patiently answered very precisely my endless questions regarding formats, sizes, letter styles and so on, a special acknowledgment and a word of sincere friendly thanks. Finally, I must underline the

invaluable help of Ernesto Federico Treo, bioengineer from the Universidad Nacional de San Juan, Argentina, and our graduate student in the Department of Bioengineering of the Universidad Nacional de Tucumán. He actually put into practice what Yolande and Steve said I had to do, did read the whole text, checked numbers, figures, equations … many, many thanks, Federico, I just would not have been able to do it by myself.

Max E. Valentinuzzi
Department of Bioengineering
Universidad Nacional de Tucumán
Cc 327, (4000), Tucumán, Argentina

Contents

Chapter 1

Introduction

What is first goes first, what is last goes last, and often there is something in between.

1.1. Ancestor's Light: The Seven Lamps

A statement reaches its maximum clarity and beauty when it can be expressed in mathematical terms. Saint Mathesis, patron of mathematics, used to be surrounded by seven lamps: *lampas utilitatis*, *lampas imaginationis*, *lampas poesis*, *lampas infinitatis*, *lampas mysterii and lampas religionis*. Bioengineering, strongly supported by mathematics, searches the quantification of the biomedical sciences and thus, no doubt, these lamps also illuminate its course because this discipline is certainly useful (*utilitatis*) for it applies directly or indirectly to the human being, it requires tremendous imagination (*imaginationis*) and creativity (*poesis*), it extends unlimited in all directions (*infinitatis*) branching off here and there, while it shines with some degree of mystery (*mysterii*) when penetrating the darkness of the unknown, so spurring curiosity, and shows also even mystical and mythical edges when human, ethical and religious aspects are touched on (*religionis*). Thus, as an intellectual rainbow, the *Seven Lamps* spread their light over us. I hope and wish, you, **my young motivated student**, will find the proper and best path with it.

1.2. Definitions: Are They Really Necessary?

The métier of this book refers to the use and application of principles, laws, techniques and general knowledge, taken from the physical and engineering sciences, to the better understanding and solution of biological and medical problems at large. No one will argue this. No specific

name is given yet. We just realize how wide and ambitious the intent is. But we need names to identify our field and activity, and here is where discrepancies may begin.

Several terms dance around and seem to have relatively settled down in the last ten years or so, while other words have tried to find a place (such as bioelectronics or biocybernetics or electromedicine). However, a closer look immediately indicates that many of the latter really belong to the wider set given below:

Bioengineering, considered as the most general and scientific, refers to biology as a whole, tries to discover new phenomena in the biological processes and intends to clarify others already known. One difficulty: It is also associated with biotechnology, as for example production of combustible gas or hormones (say, insulin) from bacteria or with genetics, so broadening perhaps excessively the field. An alternative name sometimes used is **Biological Engineering**, however, an advantage of the former is to be brief and powerful.

Biomedical Engineering, contained in the previous one, more pragmatic, oriented toward human health, but still with a good dose of scientific curiosity where the basic biological or physiological process is present. **Medical Engineering** would walk a step away from biology dealing just with man.

Clinical Engineering, the newest of the three divisions, contained in **Medical Engineering**, directed to problems found in health care systems, hospitals and emergency services, and working side by side with medicine. It shows a well-defined personality with publications of its own.

Boundaries are not well delimited, nor they need be, and information and activities must flow freely from one division to another in a constant exchange if results are to be fruitful. In daily language, the names given above are frequently interchanged. Even the two biggest international organizations do not use them officially, preferring longer combinations of words: The *IEEE/Engineering in Medicine and Biology Society* and the *International Federation for Medical and Biological Engineering.*

Definitions are always controversial and, in this particular case, perhaps more for there is still a state of development and evolution of concepts,

making terms obsolete in a relatively short time. Recall that this discipline officially celebrated its first 50 years in 2002 (see *IEEE/EMB Magazine*, *May–June 2002*). Thus, **a flexible and open mind becomes mandatory** (quite an attractive characteristic of our discipline!). To answer the question by the subtitle: **Not quite**, **because definitions are not absolute**, **they are secondary**; **they can and should be changed as required**. **They are simply convenient**. Look with a serious frown at those people who stick stubbornly to definitions (and even worse, to standards!). Imagine what it would be like had we fastened ourselves to the definitions (and the concepts supporting them) of the Middle Age or, without going as far back, of just 100 years ago!

1.3. Quantification Process

> *I know about something when I can describe it with a number.*
> *Lord Kelvin*

Scientific fields and disciplines constantly evolve, usually starting at a *qualitative stage* (by being mostly *descriptive*, as the early anatomical or zoological knowledge), to enter later into more *quantitative stages* (like counting the number of lobes of an organ, weighing it and/or searching for a known geometrical shape to approximate it). It is obvious that some disciplines are more *quantifiable* and *quantified* than others. The cardiovascular and respiratory systems, for example, are easier in this respect because their variables have precise mathematical definitions. Psychophysiology, instead, does not have yet clear-cut variables to work with and, as a consequence, its quantification process is slower. Variables like anxiety, fear or anguish are rather elusive.

One of the objectives of Bioengineering is to bring quantification to the biological and biomedical sciences. A first step consists of breaking up a given system into several interconnected functional units, i.e., the *block diagram*, usually based on previous qualitative information. This is similar to studying, say, a super-heterodyne radio receptor (formed by well defined blocks: antenna, radio frequency unit, oscillator/mixer-converter, intermediate frequency unit, detector, audio stage, and loudspeaker). The second step is *the measurement of physiological events*,

which, in turn, requires an *exact*, *precise*, *mathematical definition of the variables* to be measured (as mentioned above) and, simultaneously, *the availability of an adequate technology* to develop the sensors or transducers to apply.

During the French Revolution, two astronomers, Pierre Méchain and Jean Baptiste Delambre, were commissioned by the National Assembly to determine a standard unit of length — **the meter** — *pour tous les hommes*, *pour tous les temps* (for all men, for all times), based on the measurement of the terrestrial meridian passing through the cities of Barcelona, Paris and Dunkerke [1 m = 10^{-7} × quarter of meridian = 25 × 10^{-9} × full 360° meridian]. It took them from June 1792 until June 1799, seven full years of exhausting fieldwork plus innumerable experiences and personal sacrifices, including a serious accident and even political and military threatening, to accomplish it, but in the end they supplied the basis of the present and universal Decimal System (Guedj, 1997). A transcendental highly significant step ahead had been taken in the quantification process of knowledge at large. In more recent times and under the light of new findings, that definition of meter was changed to 1/299,792,458 of the distance light travels in vacuum in 1 s.

Interpretation of the signals, by means of from simple visual inspection to sophisticated techniques, is essential for the understanding of normal phenomena and diagnosis of pathologies. In other words, signals must be perused. In the end, the development of mathematical models leads to *prediction*, which is the highest level of quantification. As such, astrophysics or atomic physics are well ahead of any other discipline. Recall, for example, how accurately the movements and positions of astronomical phenomena can be anticipated. And, after all, what is a physician trying to do when he/she examines a patient? He/she tries to determine what the disease is and, above all, to **predict its most probable course**. A veterinarian, an ecologist and other biologically related activities take a similar stand.

In short: Scientific disciplines show a slow, sometimes faster, but steady process of quantification. Biology, physiology and medicine are no exceptions. A still distant and well-yearned objective is to anticipate disease, as much in advance and as much quantifiably as possible, based on the current known condition of a given individual. For the time being,

even with the tools nowadays at hand of the medical profession, that prediction is still far from being exact and precise. Genetics and the sequencing of the Human Genome in 2000 certainly brought us closer to that possible prediction.

1.4. Witchery, Charlatanism, Frankenstein, Science Fiction

From the witch doctor to the biomedical engineer
When suffering hits, people seek relief wherever and from any hand:
From the real and the illusion, from the truth and from lies,
in spells and in magic, in praying and faith.
They're hopes and beliefs of so many
that few, just barely, are able to grasp.

Ever since the appearance of man on the surface of the Earth, diseases along with natural phenomena and catastrophes have interfered with his life. Sorcerers, witch doctors and their witchery were the only way to fight them off. As knowledge slowly and painfully advanced (discovering, for example, the effects of water, heat, cold and herbs or foods), the former practice freely interlaced with charlatanism or quackery (people who take actual scientific facts, perhaps distorting them, claiming to cure this or that illness). In between, we find the "old mothers", "barefoot doctors", medicine men and healers of all sorts — many times applying surprisingly good and efficient concoctions or procedures, so indicating keen observational abilities — and still acting all over the world, mainly in the country or in mountainous areas. Man is always attracted by the supernatural mysterious action.

Old Mother Hutton, in Shropshire, England, used to cure the dropsy, which highly qualified doctors did not know how to handle, with a decoction of herbs. William Withering (1741–1799) took the recipe from her, even when he did not explicitly recognize it, and experimented extensively with it in his medical practice publishing, later on in 1785, the famous book entitled *An Account of the Foxglove*. So was digitalis introduced (*Digitalis purpurea*, of Linnaeus, who taxonomically classified it), first cardiac drug already well beyond its bicentennial anniversary and still widely used. He never understood its action but he consistently

found a dramatic slowing of the heart rate clearly stating "*... it has a power over the motion of the heart*, *to a degree yet unobserved in any other medicine*" (Willius and Keys, 1941).

In China, the barefoot doctors are peasants who have gained some prestige and respect as healers among his village co-citizens. They are recommended by the communal brigades to spend a training period in the city, between three and six months, on elements of modern medicine **but without abandoning the traditional methods** (people would not accept it, either). They must return to their original places. China has in the order of more than one million barefoot doctors that significantly and efficiently reinforce the health care system. The use of herbs and the millenary acupuncture are nowadays academically recognized.

The advent of electricity (a good threatening unknown phenomenon!), with Luigi Galvani, Alessandro Volta, and their famous controversy in the middle, brought along the sweet story of Mary Godwin Wollaston (Shelley's very young and beautiful lover), Percy Shelley (the English poet), George Gordon (Lord Byron, a superb impetuous writer and unfillable drunkard) and their friends in one of those 1816 summer *soirees*, giving birth to the *Frankenstein* novel and, thus, inaugurating the now widely popular science fiction style. Apparently, that evening the young girl came up with the idea of a scientist and his monstrous creation, but the two literary geniuses turned it down for themselves and told her to go ahead and write it as her own ... what she did, and with extraordinary success, for the *Frankenstein* tale projected far ahead into the years with a thousand and one modifications and updatings! During that romantic period, Aldini (who was Galvani's nephew) proclaimed the use of electrical discharges to "resuscitate the dead or the quasi-dead", as some kind of quack anticipation of the present and daily defibrillation shocks in emergency rooms, operating theaters, industrial and street heart episodes.

Suggested study subject: What was the Galvani–Volta controversy? Search in the Web for The Bakken Library and Museum, in Minneapolis, MN, to find out more about the Frankenstein story and the medical uses of electricity during the XIXth Century.

S. Furman and Wilson Greatchbach, with the cardiac pacemaker (Furness, 1975) and William Kouwenhoven and Michael Mirowski, with the external and the implantable defibrillators (Tacker and Geddes, 1980), respectively, are quite recent names associated with the develop-

ment of these devices which have saved and prolonged the life of millions of patients all over the world. It even sounds as magic or supernatural or science-fictional: To order the heart when to contract! To revive it, when it stops! Does it not sound as *Frankensteinian?* After all, Frankenstein, a scientist, gave life to his monster by electric discharges!
All this is bioengineering ancestry: Galvani rocketing off electrophysiology with his concept of animal electricity, Volta starting up electrical engineering with his pile, Mary Wollaston launching a new literary style, hints of revolutionary medical electrotherapy by a charlatan … no ordinary light from the Seven Lamps, indeed, and just a mere 200 hundred years or less back!

1.5. Koestler's Creative Collisions

> *Two women meet at the supermarket. One looks cheerful, the other depressed. The following quick dialogue is started by the former:* "*What's eating you? Nothing's eating me. Death in the family? No, God forbid! Worried about money? No ... nothing like that. Trouble with the kids? Well, sort of, it's my little Jimmy. What's wrong with him? Nothing is wrong, just his teacher says he must see a psychiatrist. PAUSE. Well ... what's wrong with seeing a psychiatrist? Nothing ... just the psychiatrist says he's got an Oedipus complex. ANOTHER PAUSE. Well ... Oedipus, Shmoedipus, what the heck, I wouldn't worry so long as he's a good boy and loves his momma.*" *Koestler, A. The Logic of Laughter, in The Act of Creation, 1964.*

The creative act of the humorist consists in bringing about a momentary fusion between two habitually incompatible matrices. In the little story above, the cheerful woman's statement is ruled by the logic of common sense: if Jimmy is a good boy and loves his momma, there can't be much wrong. But in the context of Freudian psychiatry the relationship to the mother carries entirely different associations. Scientific discovery can be described, in very similar terms, as the permanent fusion of matrices of thought previously believed to be incompatible. A problem, when looked at from two or more points of view, may be understood and solved more easily. One reference frame may be, say, strictly biological, while another may be purely mathematical. They tend to enlighten each other so

favoring the intellectual act of creation. It is like taking photographs of a landscape from different positions. Some may show aspects that others may not. The creative act, by connecting previously unrelated dimensions of experience, enables the scientist to attain a higher level of mental evolution. It is an act of liberation, the defeat of the habit by originality. This is a theory put forward by Arthur Koestler —an Austrian journalist (who tragically died some years ago), also philosopher and thinker— in his book entitled *The Act of Creation* (1964). Koestler starts with a Theory of Humor: A joke would lead to laugh when two reference frames collide. An expert joke-teller should carefully prepare the first frame to suddenly exit into a totally unexpected situation, causing the collision and, thus, triggering laughter; in short, the stronger the collision, the higher the laugh intensity. Bioengineering, by its very definition and nature, favors all kinds and sorts of intellectual collisions. It brings together the improbable and the apparently unrelated, biology and mathematics, physiology and physics, medicine and engineering. Thus, through such cross-fertilization, it supplies a good culture broth or scenario for comprehension, discovery and invention. Is this not another strong *attractor* for the youngsters, enlightened by the *Lampas Poesis*?

Question for the curious inquisitive mind: What is an attractor? Where does this term come from? Does it have applications in biology?

1.6. Guiding Philosophy of this Book: The Recording Channel

> *'Verba volant*, *scripta manent'*, *which means words fly*, *scriptures remain.*

Johannes Gutenberg's printing machine (1398–1468) with movable types represented an essential step forward in the advancement of culture for it permitted permanent and easily obtained records of the human thought. By digging into historical accounts, Koestler found out that this invention sprang out as a bisociation or collision product of two unrelated practical concepts, namely, the wine-press (well-known in those days) and a coin or lead seal which, owing to pressure, would leave a trace on paper.

Before 1847, the physiological message given off by an animal or by man himself could not be graphically registered. Stephen Hales had to tell the ups and downs of the blood column (as it slowly and blandly coagulated), in a lengthy description, when he measured arterial blood pressure for the first time from a non-anesthetized mare, back in 1728. James Hope, in 1830, called for the presence of witnesses to certify (thus, "record") his description of the heart sounds, to prove their valvular nature, during experiments made in donkeys.

Suggested study subjects: Find out about Hales' and Hopes' doings. Questions: When did anesthesia show up? What was its origin? Any collision of frames? How important was it in the development of medicine? Do you see any possible engineering inputs to these subjects?

To pick up biological signals is essential because they carry, if not all, most of the information needed to understand the behavior of the system and, eventually, to take a decision (say, a therapy). Thus, the **recording channel** appears as a second system (a technological one) to be implemented, coupled to the biological system. **This book takes it as its blueprint**. Figure 1.1 explains its parts: The biological system, either as

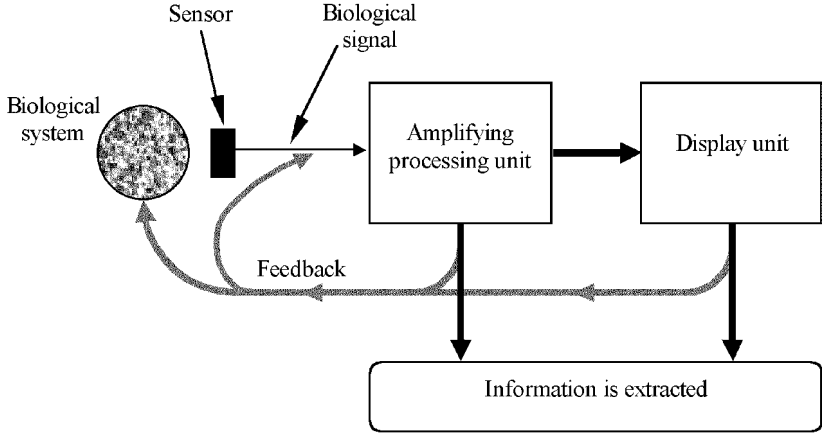


Figure 1.1. THE RECORDING CHANNEL. The biological system produces signals, which are picked up by specific sensors. They are amplified, processed and displayed. Eventually, feedback signals modify different input levels according to information obtained from the different outputs.

an individual, organ or cell, is the source producing signals (Chapter 2), **live signals**, which are picked up by specific sensors. There are very many biological signals (Chapter 3), however, their nature can always be summarized in a few kinds: electrical, mechanical, chemical and thermal. Accordingly, electrodes and transducers represent a key critical bottleneck (Chapter 4), permanently opened to research and new developments. Most of the time, if not always, the signal requires amplification and processing. This basic unit needs to fulfill unique characteristics not covered by the amplifier usually applied in other fields (Chapter 5). Once the signal is faithfully displayed (by and large, there is always some kind of visualization, on paper, oscilloscope screen or monitor), we have to read it to obtain a meaningful interpretation. This is, in simple words, all what signal analysis is about (Chapter 6). But do not get too excited about the content of the chapter, for we merely want to offer you a lead. You will need specific courses to go deeper. The information produced by the analysis of the biological signals, hopefully, will offer a variety of messages leading to one or more decisions: the signal pick up must be improved, signal processing has to be changed, stimulation of some kind is advisable to isolate a given response, a therapy is instituted, a relationship is uncovered or discovered, a mathematical model can be set, thus starting a fruitful feedback loop (Chapter 7). Thus, feedback is meant here in a very wide sense and context. The final chapter will try to put things together, if possible, to jet you off into higher levels with the help of your inspiration, sweat and, perhaps, some light from the *Seven Lamps.*

1.7. Objectives

> *Set always the goal. It will be easier to find the way. Be a Pathfinder!*

Leslie A. Geddes has been one of the outstanding contributors to Bioengineering. When I was a graduate student at Baylor College of Medicine, he told me: “Start the sentence with an infinitive. It means you are going to act. The sentence should be short and concise. If it is too long, either

you do not know well what you want to do or you have to split the objective in two or three." Let me follow the advice:

The general objective of this book is **to offer an overview of Bioengineering**. This is why it is simply an **introduction** at most, and more likely a **primer**, that is, a take off platform.

The **specific objectives** are:

1. **To describe somewhat critically the basic sections of Bioengineering**. **The rest springs up from them**.
2. **To give a historical feeling**. **It teaches humility**, **it may show the way and avoids repetitions**.
3. **To show you how much you can learn by yourself**. **It is extremely difficult**, **but let us try**. **Your personal effort is indispensable**.

In other words, after finishing it (perhaps a full semester work course), you should have the feeling of having caught almost everything and yet realizing that the true meat and potatoes and full taste of it is still to come, for the book is aimed at the **undergraduate bioengineering student**. Some mathematics is used, say, at the level of calculus, differential equations, and the concept of transform.
In the meantime, learn and put into practice as soon as possible the ABC of science (Bishop, 1997). The authors, believe me, are also trying hard in this respect. Unfortunately, the human being (scientists and engineers are also human beings) is not always too willing to comply with these nice suggestions:

Accept criticism graciously
Become focused and organized
Cope with the burgeoning scientific literature
Develop critical thinking and logical sequence of ideas
Experience satisfaction and joy from your endeavors
Finish tasks undertaken
Give the best at all times
Hone communication skills in writing and speaking.

Even though the word BIOENGINEERING may have conflicting meanings when its detailed coverage is required, as briefly referred to above, it is widely and loosely used in the daily language, probably because it is

shorter and powerfully says what the métier is. Therefore, we will stick to it in the text. From here on, it will not be capitalized, as the names of other disciplines will not either.

Chapter 2

Source: Physiological Systems and Levels

Know the ground thou set thy foot

Physiology, normal and pathological, is seminal to bioengineering. It is the ground we move about, source of our knowledge and recipient of our efforts. True, a bioengineer is first an engineer, but he/she must also be a physiologist and has to be fluent with its terminology. Otherwise, the bioengineer will not be able to communicate with his fellow physiologists, physicians, veterinarians and biologists. Thus, **the objective of this chapter is to teach SOME physiology**. It is assumed you know a little already and, more important, that you will study quite a bit in the near future. A few systems were left out (like the reproductive and bone systems), without meaning they are not important. At this level, it will not hurt. Those included are summarily and incompletely treated and probably will be criticized by the specialist or the prickly mind. Remember, **it is a bird's eye view** (bioengineers like to fly). The approach, however, intends to be from the eyes of an engineer also, thus stressing the block diagram and even introducing some simple mathematical models. Descriptions are minimal. If necessary, the student will have to complement and supplement with any of the many available good physiology textbooks.

2.1. Organism in a Block Diagram

With the Great Engineer's Forgiveness

To reduce a higher animal, or man, to a mere collection of interconnected functional blocks is a daring and even disrespectful task! However, and

very humbly, we just want to see what the main systems are, how they relate, and what perspective we may obtain from it. This is exactly what an engineer does when approaching an unknown equipment. Figure 2.1 underlines the two biological purposes: (1) **maintenance or survival of the individual** through maintenance of the tissues, and (2) **maintenance or survival of the species** by means of the reproductive system.

For the first essential task, the tissues take up oxygen and nutrients while releasing carbon dioxide and metabolites. These are passive transports occurring from and to the circulatory stream, via an exchanger, E1, represented by the peripheral capillary network closely associated to the tissue cells which, in the end, are the final users or customers. A second exchanger, E2, relating the pulmonary capillary bed to alveolar walls (in turn, connected to the external terrestrial atmosphere), allows the unloading of excess CO_2 and the re-oxygenation of blood, also passive processes. Passive means that substances move along concentration gradients without utilizing any energy. Blood replenishes its nutrient load via a third exchanger, E3, divided in two sections (hepatic and intestinal) and

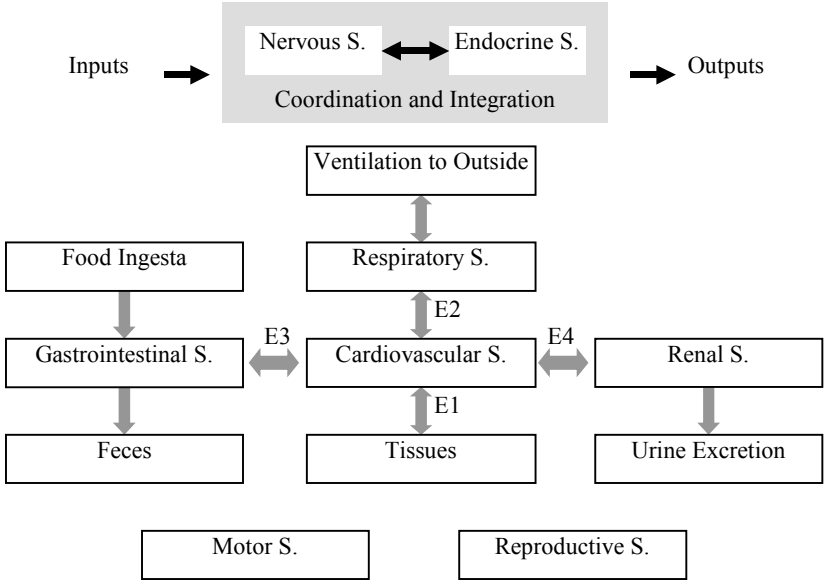


Figure 2.1. THE ORGANISM IN BLOCK DIAGRAM (see text). '*S.*' stands for 'system' and the *E*'s stand for 'exchanger'.

much more complex than the previous two, which associates the vascular net with the gastrointestinal system (GIS). In it, transport processes are both passive and active (the latter require energy expenditure and the existence of specific pumping mechanisms). Many are still under study. The GI receives the ingesta through the oral cavity, breaks down food into simpler chemical forms suitable for absorption, and excretes the residues through the anal opening. Metabolic products are cleared via a fourth and extremely complex exchanger, E4, which brings together the vascular tree and the renal system, the latter offering an outlet through the urinary tract. Exchangers E3 and E4 are composed of capillary vessels showing unique anatomic and functional characteristics. Malfunctioning of any single exchanger seriously threatens the life of the individual. Their study, including the small vessels converging to and diverging from them — the micro-circulation — constitute a large and complex chapter of physiology. It is clearly seen from this block diagram that the cardiovascular system (CVS) relates to all organs, tissues and cells, with **blood being just a carrier in constant circular movement**.

The neural and endocrine systems act as the Central Processing Unit, receiving afferent signals from the different systems, integrating them, taking decisions of all sorts, sending off efferent signals and coordinating actions. The locomotion and reproductive systems complete the view, the former to move about, either searching for food, running from danger or looking for a mate. Electrical action potentials are the basic signals within the neural system while internal secretion blood levels (hormone concentrations) excite or inhibit specific receptors distributed all over and, thus, also transmitting information.

This individual block diagram, with its physiological systems, suggests right away the existence of subsystems and organs, down to the cellular and even molecular level, thus pointing out to a hierarchical arrangement. The student must identify as soon as possible where his/her research object is located in order to better choose the proper tools, either experimental or theoretical, to attack it.

Study subject: Modify the proposed block diagram in order to have that of a pregnant mammal or woman. Is there another exchanger? Explain.

2.2. Cardiovascular System

> *...it is absolutely necessary to conclude that the blood in the animal body is impelled in a circle, and is in a state of ceaseless motion; that this is the act or function which the heart performs by means of its pulse; and that it is the sole and only end of the motion and contraction of the heart. William Harvey, De Motu Cordis, 1628.*

With rhythmic pulses running down its road	(electrical activity)
the heart moves always restlessly abroad.	(mechanical activity, blood reach)
The magic of its numbers is astounding	(measurements, heart rate ...)
no matter how they're handled or transformed.	(signal processing)
Its reddish elixir is wonder	(blood)
that flows in those palpitant channels and yonder	(hemodynamics)
against oscillating resistance	(peripheral resistance)
and generously crossing fine barrier.	(capillaries)
We read all its messages, hesitatingly	(recording)
we think we understand them, expectantly	(diagnosis, interpretation)
while taking decisions, vacillatingly.	(therapeutics)
Oh! Circuit revered and much feared,	(still there are unknowns)
by probes and by catheters inspected,	(more studies are needed)
from inside our destiny thou markest!	(but in the end, we all die)

2.2.1. Cardiac Mechanical Activity

The previous historical and poetic accounts, almost 400 years apart in time and knowledge, briefly summarize what the system is about. The heart and vessels and their essential properties are presented, along with the variables of the CVS, as a continuation and more focused view of the first general diagram (Figure 2.1). Besides, Fick's principle is introduced as an application of the continuity equation, with specific use also in the determination of cerebral and coronary blood flow. A simple derivation of Stewart–Hamilton formula is given, as the basis of the indicator dilution technique for the determination of cardiac output.

The heart as a pump is reviewed by means of the pressure-volume loops, a clear cut engineering concept borrowed from thermodynamics, which leads into new insights of cardiac performance with significant clinical consequences. The concept of aortic impedance is also brought up along

with the still not fully answered question of optimal ventricular-arterial coupling.

2.2.1.1. The circuit

At the beginning of this chapter (Figure 2.1), it became clear that the blood acts as a carrier. Its movement is sustained by the heart, which as a muscle contracts and, thus, rhythmically propels the fluid to the overall network. In fact, it is a series hydraulic system (Figure 2.2) with a powerful chamber, *the left ventricle* LV ejecting a spurt of blood (*stroke volume*) in each contraction (in the order of 80 mL/beat at about 70 beats/min or slightly above 1 beat/s) through the aortic valve into the aorta — main arterial output — to the *systemic circulation*. The latter is a highly complex network, extended deeply into every tissue and presenting a finite, measurable and variable hindrance to the blood flow. Physiologists call it *periheral resistance*, R_p, and, quite often, they even calculate a handy numerical value making use of Poiseuille's Law (which is the hydraulic analog of Ohm's Law of electricity).

Suggested exercise: Obtain a possible value of the peripheral resistance in a normal adult. Establish the units. Sometimes, physiologists define the *R-unit* (the unit of peripheral resis-

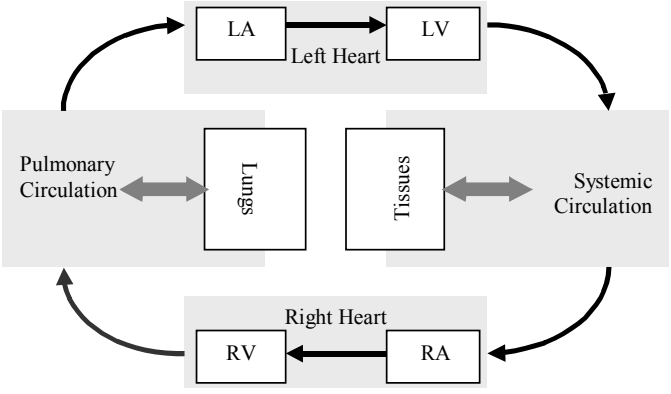


Figure 2.2. THE CIRCULATORY SYSTEM AS A HYDRAULIC SERIES CIRCUIT. LA, left atrium, connected through the mitral valve to LV, left ventricle, leading through the aortic valve into the Systemic Circulation, site of the peripheral resistance. From the latter, blood enters into the right atrium RA and, via the tricuspid valve, proceeds to the right ventricle RV. Finally, through the pulmonary valve, it fills the pulmonary circulation returning to LA. Exchanges to/from tissues and to/from lungs are also depicted, respectively, for the Systemic and Pulmonary Circulations.

tance) as 1 mmHg divided into 1 mL of blood/s. Try to find the analog variables, as for example, hydraulic pressure and electrical potential. Have you ever heard of Hopkinson's Law? Find it out. Hint: It is another analog of the two laws mentioned above. To summarize, write down the three laws and clearly establish the analogies.

After exchange (gases, nutrients and metabolites) in the capillaries (which are part of the peripheral resistance), the return pathway is via the venous system back to the heart, entering it through the vena cava (a large distensible vein) into a small contractile chamber called the *right atrium*, RA. Blood goes from here to the *right ventricle* traversing the *tricuspid valve* (which connects the RA with the RV) and, from the latter chamber (also contractile but less powerful than its left side companion) proceeds through the *pulmonary valve* into the pulmonary artery. Thereafter, there is a shorter branching to the lungs, filling all its capillaries and, thus, permitting another exchange (this time it is only a gas exchange: oxygen is taken up and carbon dioxide is downloaded). From them, little venules converge into larger and larger veins as they approach the four pulmonary veins, which dump its oxygenated content into the *left atrium*, LA, another small and contractile chamber. The final step to complete the circuit is from the LA via the *mitral valve* back into the LV, already set to start all over again.

Suggested exercise: Again making use of Poiseuille's Law, calculate the pulmonary resistance (R_L). Is it lower or higher than the systemic peripheral resistance? Why is it so? Try first to come up with an explanation of your own. Later, check in any good physiology textbook or discuss it with a classmate or with an instructor.

Study subject: Besides carrying substances, are there other secondary but yet important functions of the circulation? Suggest possible block diagrams for these functions. Explain briefly.

2.2.1.2. The customers

An urban water distribution system has its customers hooked in parallel to a main supplying pipe. Something similar is found in the mammalian blood distribution system. To begin with, from the left heart to the right heart, we have lumped all the peripheral resistance in one block (Figure 2.1). However, this is in fact extremely complex. The arteries, starting with the main outflow (the aorta), branch off into smaller ducts supplying blood to the different *vascular beds.* These regional subsystems vascularize the tissues that, in the end, are the "customers" or the final users of the blood contents. Each bed, by and large, has a relatively well defined main inflow, as for example the kidneys, with one unmistakable *renal*

artery per kidney (one to the right and the other to the left), or the ventricles themselves, with the *right and left coronary arteries* (coronary means "crown"; if you take a look at the heart, especially from above, it gives the impression of a crowned head).

Suggested exercise: Come up with a definition of *vascular bed*. It is an important concept because, even though all beds share certain common properties, they also differ significantly in others. Search for the main arteries supplying blood to the brain. Is there a single blood input to the skin? Find out what the physiologic pulmonary shunt is.

Figure 2.3 depicts the principal users: the coronary circulation, the brain, skeletal muscle, bone (yes, bone also requires some blood), the gastrointestinal and hepatic especially coupled systems, the kidneys, the skin, and other tissues. Out of the total flow expelled by the left ventricle, the figure indicates the respective approximate percentages derived to each regional section. However, depending on the physiological condition (say, exercise, rest, postprandial or preprandial state, shock) these values may change significantly. Some regions (as the brain and the heart) are favored in emergency conditions as compared to other lower hierarchical beds (as, for example, the skin and the guts).

Suggested exercise: Why a person with hypotensive shock is pale? Please, think in terms of the "big print" without trying to develop a clinical history. Hint: Does he/she need blood supply to this particular bed? Give a second thought to the concept of *regional hierarchies* from the hemodynamic viewpoint: Why should the brain and the heart be favored?

2.2.1.3. Variables of the Cardiovascular System (CVS)

Figure 2.1 and Figure 2.2 are models and, as such, already contain the basic and important *variables* any cardiovascular physiologist, hemodynamist or cardiologist have to deal with.

– *Blood volume* (V_b)

It is measured in units of length to the 3rd power, as for example in cubic centimeter (cm^3 = mL) or in liters (L). It is a physical almost intuitive concept closely related to *mass* and *density* (the latter is defined as *mass per unit volume*; recall that the density of blood is 1.055 g/cm^3, that is, it is slightly higher than that of water). Volume quantifies how much space a given material occupies. In the normal adult, V_b is in the order of 5 to 6 L. Life is not viable without blood, thus, it is essential and indispensable. By a similar token, an empty hydraulic network is useless.

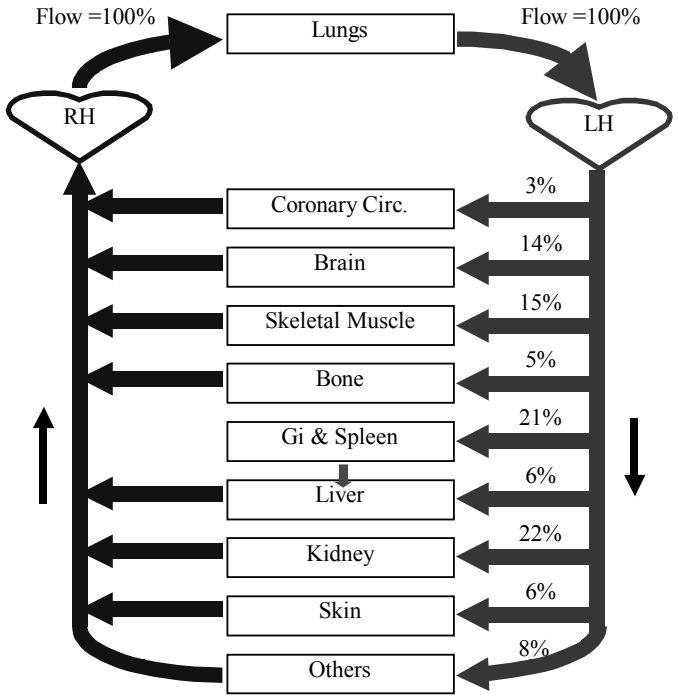


Figure 2.3. THE CIRCULATORY SYSTEM AND ITS PRINCIPAL BEDS. Observe the particular arrangement of the gastrointestinal, spleen and hepatic system. It has a double inflow of blood. Percentages by each branch correspond to the relative amounts of blood flow to the respective beds.

– *Blood flow* (F_b)

It is measured in units of volume per unit time (say, in mL/s or in L/min), has to be guaranteed through every single piece of tissue. Sometimes, this concept is referred to as *perfusion*. The total resting outflow, either from the left or from the right ventricle, has an average value of about 5 to 6 L/min (or about 100 mL/s). It is also called *cardiac output*, *CO*. However, as clearly seen in Figure 2.3, each bed takes only a portion of this total value. If one considers the pulsating action of the heart, since each ventricle ejects during each contraction the so called *stroke volume*, *SV* — about 70 to 80 mL/beat — at a frequency of 70/min, it leads to 4.9–5.6 L/min, which is roughly the figure given above. Remember clearly: **Blood supply to the tissues is absolutely necessary**. Without it tissues die. *Ischemia (ischein*, to supress, and *haima*, blood, from Greek)

is the word to describe the condition characterized by a diminished perfusion. If it happens in the brain or in the ventricles, it may end up in serious cerebral or cardiac injury. It may take place also in any tissue, as for example in the gut. The existence of pressure does not necessarily assure a good passage of blood. An obstructed vessel usually has good hydraulic pressure at the arterial side.

Suggested exercise: Search for at least three pathologies (other than the three already mentioned above), which recognize ischemia as its source.
Suggested experiment: Pick up a weight (for example, 5 kg) with your hand and start to flex the arm as if trying to build up a good-looking biceps. Do it 30 or 40 times. Thereafter, tie your arm around below the shoulder with an elastic band. Repeat the exercise. How many times can you lift it up? Explain.

– *Blood pressure* (P_b)

It is measured in mmHg or in Pa, at any place all over the circulatory system. It is also essential and is a clear physical concept, defined as force per unit area (1 Pa = 1 Newton/m^2; 1 kPa = 1,000 Pa; 1 hPa = 100 Pa). Thus, as a rule of thumb, to pass from mmHg to hPa multiply by 4/3, and vice-versa, to go from hPa to mmHg, obviously multiply by 3/4.
Maintenance of a head pressure is what keeps the blood moving. Except for friction losses, arterial blood pressure at the entrance of any bed is the same. Its mean value is somewhat lower than 100 mmHg (or 13.33 kPa = 133.3 hPa).
Application of Poiseuille's Law leads to the definition of a fourth (and this time compound) variable, the already mentioned *peripheral resistance*, R_p. However, if applied to a specific region, the concept of *regional resistance*, R_r , can also be defined.

Suggested exercise: Calculate the relative resistances of the beds outlined in Figure 2.3 assuming a constant arterial head pressure and a zero pressure on the venous side.
Suggested exercise: After the model of Figure 2.3 and using the concepts of regional resistances and different variables of the cardiovascular system, identify three types of possible collapse (*shock*, in medical language). Find analogies in an urban water system, say, during one of those unbearable summer days when, for example, everybody in town decides to water the yard, clean the swimming pool and wash the car.

2.2.1.4. The heart

By now, it is we hope quite clear to the student that the heart is a muscle divided in two sections, left and right, each with an atrium and a ventricle. Hence, it has four chambers: two are mainly collecting ones (with a

relatively minor propelling function) while the other two are essentially ejecting structures. The heart *in toto* acts as two hydraulic pumps connected in series. The unidirectionality of the circuit is undoubtedly manifested by the four cardiac valves (kind of hydraulic diodes). Moreover, the heart can be considered as a periodic pulse generator that, on the average, maintains an *equivalent constant flow.* Such flow, according to the requirements, can be adjusted (up to a maximum of about 30 L/min during heavy exercise, in a healthy young adult). However, this continuous equivalent concept is merely a practical simplification and it must be emphasized that, by its nature, **the circulatory system is pulsatile**. Even more, there is evidence indicating that life would not be viable, were the pump of the true constant type. Thus, apparently, pulsations are necessary, although some people question such stand. Besides, it is now well recognized that the system shows reflecting waves and a nonlinear behavior.

The heart is simultaneously a pump and a muscle and, as any muscle, it has contractile elements (*sarcomeres*) that change their length. The net result is a change in the cross-sectional horizontal ventricular diameter, relatively much larger than the change in the longitudinal (or *basal-apical*) diameter. Wall thickness also changes during contraction, and all this occurs in synchronism, both in the left and in the right ventricles. The increment in intraventricular pressures opens in the end the arterial output valves as these pressures become higher than the arterial blood pressures, so starting the ejection phase. Hence, valve opening is a *passive* phenomenon, with a pressure gradient being the *only* driving mechanism (no muscle or tiny motor is involved). In the aortic root, pressure reaches a maximum of about 120 mmHg (or 160 hPa, *systolic pressure*) and, because of its pulsatile characteristic, shows a minimum in the order of 80 mmHg (or 106 hPa, *diastolic pressure*). In the pulmonary artery the maximum and minimum are, respectively, 25 and 10 mmHg (or 33.3 and 13.3 hPa) with a mean value of about 15 (or 20 hPa). The difference between the two is called the *differential systemic arterial pressure* or the *differential pulmonary pressure*, respectively, both with significant clinical implications.

Within the ventricles the numbers, instead, are rather different. In both ventricles the minimum is always zero (which offers a splendid criterion to know when a catheter is within the chamber), going up to a maximum of 120 and 25 mmHg, respectively, for the left and for the right side, obviously coincident with the maxima given above (remember that the ven-

tricular and the nearby arterial cavities become temporarily a single one when the valves are opened during the ejection phase).

A unique characteristic of any muscle is its *contractility*, somewhat elusive in its definition for researchers still argue about which is the best. Nonetheless, since intraventricular pressures rise from zero to a maximum because of the contraction, the velocity of doing it, i.e., the temporal derivative of the intraventricular pressure provides an acceptable means for estimating how good the muscle contracts. Hence, contractility, either of the left or of the right ventricle, can be defined as (dP_{iv}/dt), where P_{iv} stands for the corresponding intraventricular pressure. Usually, the maximum of this derivative, $(dP_{iv}/dt)_{max}$, is chosen as the proper descriptor.

Cardiac contractility is an extremely important physiological and clinical parameter. It can be defined for any of the four cardiac chambers. However, the left ventricle's is what people are mostly concerned about for in heart disease, very commonly, ventricular contraction decreased strength leads to cardiac insufficiency and failure. Nervous or hormonal influences can deeply affect cardiac contractility (and, therefore, the ability to develop pressure). For example, sympathetic activity will increase it (the so called *positive inotropic effect*), producing stronger contractions and higher pressure. The Greek roots *inos* and *tropic*, respectively, mean "force" and "tendency". Other agents may elicit *negative inotropic responses* (that is, weaker contractions).

Study subject: Find the difference between *tropism* and *trophism*. Search for a few examples in physiology. They are important concepts.

2.2.1.5. Laws of the heart: Starling's and Laplace's

But not only contractility plays a role in the force of contraction (and, thus, in the capacity to develop pressure which finally leads to the ejection of blood). There are two beautiful laws contributing to the mechanisms associated with intraventricular pressure build up. Both are significant in cardiac physiology. The first one is physiological, born around 1914 after a series of experiments carried out in London by Ernest Henry Starling (1866–1927) and his collaborators, while the second is physical, to be found in a monumental opera, *La Mecanique Celeste*, authored by the French mathematician and astronomer Pierre Simon Laplace (1749–1827) and published in several volumes between 1790 and 1825:

– *Starling's Law* (of the heart)

The force of contraction increases as the initial fiber tension goes up (the nowadays so called *preload*).

– *Laplace's Law* (adapted to the heart)

The intracavitary pressure (P_{ic}) is directly proportional to the wall stress (W_s) and inversely proportional to the equivalent radius (R_{eq}) or

$$P_{ic} = kW_s h / R_{eq} \tag{2.1}$$

where k is a constant and h is the wall thickness.

1- Starling's Law, although first described for the heart, is also valid for skeletal muscle. Figure 2.4 is a record obtained from a naturally contracting turtle ventricle while, after each beat, the initial tension (it could be called "mechanical bias") was manually increased by a small amount using a specially designed spring mechanism attached to a force transducer. The inset (right of the same figure) displays the relationship between the force of contraction (measured as the difference between the maximum and the minimum per beat) and the corresponding stretch applied to the whole ventricle, both in grams (force), measured from the

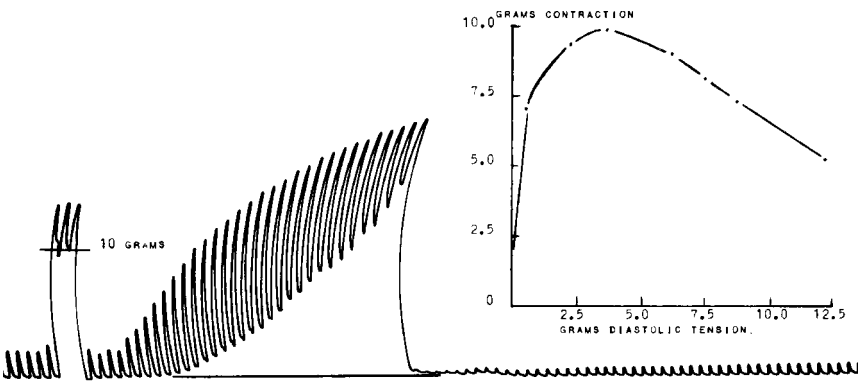


Figure 2.4. STARLING'S LAW OF THE HEART. Ventricular spontaneous contractions of a turtle's heart as the organ was gradually stretched by means of a specially designed mechanical arrangement which included the myographic transducer. Notice the increase in contraction with stretch up to a maximum value. Thereafter, because of overstretch, there was a decrease in contraction. The graph to the right shows clearly such maximum. Experimental data obtained from a turtle at the Physiology Department of Baylor College of Medicine, Houston, TX, 1966.

base line to the minima. Both, the temporal record and the plot, clearly show the important increase in the contraction force. However, as the stretch proceeds, the force goes down, indicating entrance into a dangerous operating region where, probably, the overstretch may cause damage to the fibers. This is an easy and highly demonstrative experiment, which was routinely carried out by medical and physiology students alike at Baylor College of Medicine, in Houston, TX, during the professorship years of Hebbel E. Hoff and Leslie A. Geddes, roughly from 1957 to 1975. There used to be an *Experimental Physiology Manual*, by these two authors and now probably almost impossible to find (a real jewel for those who may still keep a copy), where practical details of the exercise were clearly explained. Its foundation was presented by Starling in the Linacre Lecture at Cambridge, England, in 1915 and published a few years later in 1918. His latter phrasing was: "*The law of the heart is thus the same as the law of muscular tissue generally*, *the energy of contraction*, *however measured*, *is a function of the length of the muscle fibre*" (Patterson and Starling, 1914; Starling, 1918).

2- Laplace's Law was described for spherical bubbles (typically, soap bubbles) with a wall of negligible thickness; in other words, it refers to a film under tension, the so called *surface tension* (*T*), which is defined as the *tangential force* (*tangent to the surface*) *applied perpendicularly to a unit length lying on the surface.* The theorem of Laplace applies to a membrane separating two spaces of any shape with tension in it. Any surface — as for example that of a Bunny Rabbit or Teddy Bear balloon — can always be decomposed into many "caps" each defined by two mutually perpendicular circumferences of respective radii, R_1 and R_2, which are called the *principal radii of curvature.* Within the balloon, according to the theorem, the pressure P (actually it is the difference between the inside and the outside pressure, which most frequently is the atmospheric pressure) is given by,

$$P = T\left[\frac{1}{R_1} + \frac{1}{R_2}\right] \tag{2.2}$$

where T is the already mentioned surface tension, for example in dynes/cm, and the radii, R's, say, expressed in cm. It is easily seen that pressure comes out in dynes/cm^2. When radii are equal, we have a sphere

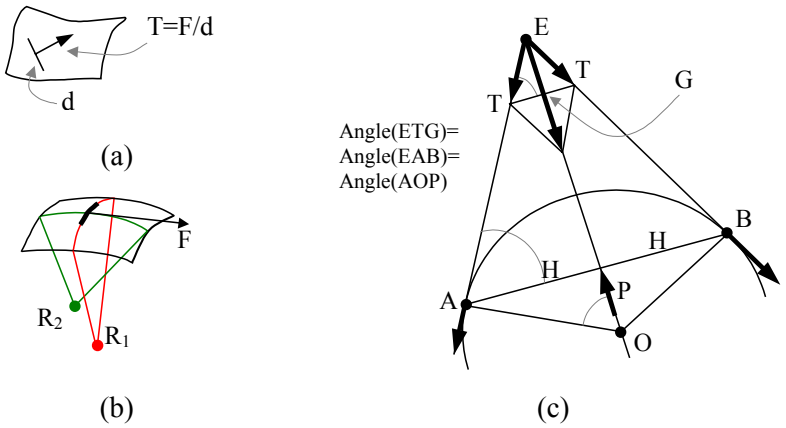


Figure 2.5. LAPLACE's LAW. (a) Definition of surface tensión. (b) Cap element showing its two principal radii and the force over the surface unit length. Pressure is applied from underneath. (c) The arc AB represents one of the two principal arcs on (b), for example, the red one, thus, it defines a plane. Notice that chord AB is divided in two halves; each can be called H, from A to its middle (where the arrow P sets) and from there to point B. It is underlined that point G corresponds to the middlepoint of the segment TT. See text.

(such as a bubble), and the equation simplifies into $P = 2T/R$ (Figure 2.5).

Suggested exercise: Obtain the equation for a cylinder. This case is applicable to the vessels in general.

If the latter equation for the sphere is compared with the above eq. (2.1) for the cardiac case, it comes out that,

$$kW_s h = 2T$$

or

$$W_s = 2T / kh = KT / h = KF /(dh) \qquad (2.3)$$

clearly showing that wall stress is proportional to the surface tension distributed over the wall thickness. The constant, $K = 2/k$, can eventually be equated to 1, while d is the unit length over the surface to which the force F acts perpendicularly and tangentially to the same surface. Wall stress is an extremely important concept in cardiology. Excessive stress causes injury to the cardiac tissue and the latter, in turn, may end up in a lower contractility and consequent insufficiency. One clinical strategy aims at reducing wall stress, which acts as an *internal load* to the heart.

Laplace's Law is quoted over and over, however, its demonstration is rarely found. Woods applied it in 1892 to hollow organs (such as the uterus, the bladder and the ventricles). K. De Snoo, a Dutch researcher in 1936, offered a simple geometrical derivation that was reviewed by Máximo Valentinuzzi (Sr), in Argentina in 1950, both in studies of the uterine contraction. The law was rephrased by the latter author as follows: The tension on any given point of the uterine wall is equal to the product of the intrauterine pressure and the curvature radius of the uterus at that point, or T = P R. Clearly, this is valid for the heart, too. These papers are difficult to find and, besides, are written in languages other than English.

Burton (1957) produced an editorial article where an excellent discussion is provided, even reporting actual numerical results. Nontheless, no mathematical demonstration was given. Hence, we will develop herein De Snoo's, even though it has a weak point which the curious and mathematically oriented student may take up for further thinking.

As already stated above, the surface of a balloon of any shape, with pressure in it, can always be broken down into a number of caps, as many as necessary. To give a visual idea, these caps would look like parachutes when quietly falling down to Earth. Let us take a section of an osculating circumference in one of these caps, coincident with one of the characterizing radii (Figure 2.5). **From now on**, **everything in this demonstration is considered on the plane of the circle defined by such circumference**. The pressure inside acts on the Chord (AB) subtending the arc, thus, the arc between points A and B is under the tension,

$$T = P\,Chord(\mathrm{AB}) \tag{2.4}$$

where P is in [dynes/cm^2] and (AB) in [cm], signaling that we are indeed getting units of tension, or [dynes/cm]. Besides, the chord AB is divided in two halves H, or *Chord* (AB) = 2H, indicating that the tension can be rewritten as $T = P$ (2H). This tension is counterbalanced by an opposing tension D, divided in two equal components tangent to points A and B, respectively. The balancing tension is graphically obtained by, first, sliding the two T's over their linear projections until they intersect at point E, and second, by application of the parallelogram rule. Thus,

$$Segment(\mathrm{EG}) = D/2 = T\sin(\alpha) \tag{2.5}$$

where α is the *angle* (ETG) = *angle* (EAB), because their sides are parallel, and is also equal to *angle* (EOP), because their sides are perpendicular. Solving eq. (2.5) for T and replacing D, yields,

$$T = D/2\sin(\alpha) = PH/\sin(\alpha) \tag{2.6}$$

but inspection of Figure 2.5 easily shows that sin $\alpha = H/R$ which, after substitution in eq. (2.6), ends up in

$$T = P \times R \tag{2.7}$$

which describes **in the plane** the statement in italics given above. Now we face an arguable step:

For one principal plane containing the principal radius R_1, we have $T_1 = P \times R_1$, which is eq. (2.7); for the other, perpendicular to the former, and following exactly the same rationale as developed above, we obtain, $T_2 = P \times R_2$. Solving each for P and adding them up leads to,

$$P = (T_1/2)/R_1 + (T_2/2)/R_2 \tag{2.8}$$

because pressure, by Pascal's Law of hydrostatics, has to be the same. Moreover, if it is accepted that $T_1 = T_2 = T_p$ and defining $T = T_p/2$, we end up with

$$P = T\left(\frac{1}{R_1} + \frac{1}{R_2}\right) \tag{2.9}$$

The latter equation is fully coincident with eq. (2.2). The weakness in the last part of the derivation is pointed out and, therefore, left to a more inquisitive (and powerful) young mind as a possible "little project". Tensors might supply a good tool for it.

If the balloon is a sphere, the two radii are equal to R and eq. (2.9) reduces to the well known $T = PR/2$. If now the concept of wall stress is introduced in a balloon or hollow organ with wall thickness, h, as defined above in (2.3), we get,

$$W_s = k_1 (P \times R)/h \tag{2.10}$$

always keeping in mind the existence of a proportionality constant. Its numerical value should be determined for each particular case without worrying about the come and go of a 2 in it. In the heart, intraventricular hypertension and ventricular dilatation increase wall stress (a risky condition); with time, such hypertension produces hypertrophy, meaning a thickening (larger h) of the ventricular wall, obviously tending to relief parietal tension. Hence, hypertrophy is a compensatory mechanism.

Suggested exercise: Search in the published literature a possible value for left ventricular volume in an adult healthy man. Thereafter, calculate the equivalent diameter (as if it were a sphere).

Suggested exercise: Applying Laplace's Law, estimate the ratio of left ventricular wall thickness, h_L, to right ventricular wall thickness, h_R. Explain the result.

2.2.1.6. Vessels: arteries and veins

By definition, artery is a vessel carrying blood from the heart to the periphery and vein is a vessel carrying it back to the heart, irrespective of whether it is or not well oxygenated. Figure 2.2 displays the arterial section on the right and the venous section on the left. Well embedded in the different tissues is the *microcirculation*, which includes (i) the last and smallest in diameter (the *arterioles*) portion of the arterial side; (ii) the capillaries, as the **only** exchange section through their highly permeable thin walls; and (iii) the first and smallest in diameter (the *venules*) segment of the venous side. The microcirculation constitutes a separate chapter of utmost importance in vascular physiology. One of the most remarkable properties, especially in beds like the brain, heart or kidneys, is the ability to control the amount of blood reaching the network. This is called *autoregulation*, independent of mechanisms controlled by higher centers.

Unlike any technological hydraulic system, all blood vessels are *elastic*, a property bestowing upon them a powerful regulatory tool, both as passive mechanism and also as an active one via smooth muscles covering the external side of their walls. One way to evaluate *passive elasticity* is by application of the concept of *compliance*, *C*, defined as,

$$C = \mathrm{d}V/\mathrm{d}P \tag{2.11}$$

that is, the differential change in volume V per differential change in pressure P. Its units, as expected, are for example [cm^3/mmHg]. The inverse of C is called *elastance*, E. One of the several reported values for the latter in normal men puts it in about 1,500 dynes/cm^5. The interested student will find more details in the current literature, as for example, other definitions to describe the elastic properties that, interestingly enough, have been kept essentially the same over the years (Valentinuzzi, Ghista & Nichols, 1979). Mostly, newer reports refer to the action of drugs on these parameters or to the relative contributions of elastin and collagen components (Armentano, Cabrera Fischer, Levenson *et al*., 1990; Armentano, Levenson, Barra *et al*., 1991; Cabrera Fischer, Levenson, Barra *et al*., 1993).

A vessel with high compliance increases greatly its volume with a small increment in pressure. Veins are much more compliant than arteries and, as a consequence, most of the blood volume (about 8% of body weight) lodges dynamically within the venous system.

Suggested exercise: A person is placed horizontal in a centrifuge with his legs pointing to the center and his head pointing outwardly. After rotation during a while, is he still alive? If not, could you explain why? (It sounds weird and awful, but it is descriptive).
Study subject: A reference is given above (Armentano, Cabrera Fischer, Levenson *et al*., 1990). In its title, it mentions "the aortic elastic response to epinephrine". What is such response? Think in terms of an athlete who is getting ready to act.
Do you remember Hooke's Law of elasticity? Try to correlate it with eq. (2.11) above.

Another characteristic of the circulatory vessels is their smooth musculature. These muscles are controlled by the autonomic nervous system and by hormonal secretions of different kind and origin. These muscles modify the vessels' caliber or lumen (see the suggested study subject above). Hence, the compliance of a vessel in active state is different than its compliance when inactive. Moreover, in some vascular disease conditions (such as arterosclerosis, flebitis, varicose veins) compliance of the vessels is significantly modified. In arterosclerosis, compliance decreases (arterial hardening) while in varicose veins, vessels tend to be overcompliant.

Study subject: Compare the musculature of arteries and veins. What part of the vascular system is the most heavily covered with smooth muscle? Where is peripheral resistance mostly concentrated on?

2.2.1.7. Cardiac output or the total flow, coronary and cerebral flows

– *The direct Fick method*

In a previous section above, blood flow was presented as one of the variables in the CVS. Here, we want to be more specific. The total flow or cardiac output — $CO = F_t$ — exits the left ventricle at high pressure, enters the right heart via the vena cava at very low pressure, is also expelled by the right ventricle at moderately low pressure, and finally returns to the right atrium at very low pressure again (Figure 2.2 and Figure 2.3). If the lungs are considered as a node (Figure 2.6), the *Continuity Principle* applied to blood (as carrier, in mLblood/min) and oxygen (as transported substance, in mLO_2/mLblood) establishes in the steady state condition that,

$$F_t[V] + F_{ox} = F_t[A] \qquad [mLO_2/min] \quad (2.12)$$

where $[V]$ and $[A]$, respectively, stand for the concentration of oxygen in venous and in arterial blood, while F_{ox} represents the net oxygen uptake in [mLO_2/min] via the respiratory system. Solving for F_t, results in,

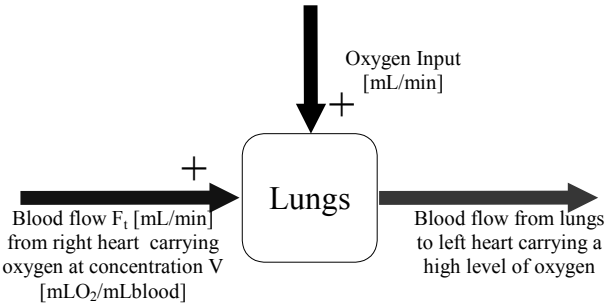


Figure 2.6. FICK's PRINCIPLE. In the steady state, everything that goes in (per unit time) must also go out. See eq. (2.12) and text.

$$F_t = \frac{F_{ox}}{\{[A]-[V]\}} \qquad \text{[mLblood/min]} \quad (2.13)$$

which is the famous and well-known Fick's formula (Hoff and Scott, 1948). Those familiar with electric circuits will find this similar to the total current converging to and diverging from a node. After all, recall that current is nothing but amount of electric charge per unit time (1 amp = 1 coulomb/s) and in eq. (2.12) we have amount of oxygen also per unit time. The numerator in eq. (2.13) is usually obtained from a metabolimeter (an easy measurement). A normal adult at rest may take about 250 mLO_2/min. A sample of blood from any artery (the method, thus, requires arterial puncture) and, subsequent determination in the biochemistry lab, gives the arterial concentration of oxygen. The venous concentration of oxygen is not easy. Samples from a peripheral vein are not acceptable because the oxygen consumption varies from tissue to tissue. *A representative sample has to be a mixture coming from all tissues.* Only the right atrium, or better, the right ventricle, or the best, the pulmonary artery, carry venous blood meeting such requirement. Hence, a probing catheter must be introduced to any of these vascular places in order to withdraw a few milliliters of blood to be tested in the lab for oxygen content. Typical expected normal values are 20 mLO_2/100 mLblood, for the oxygenated blood, and about 15 mLO_2/100 mLblood, for the mixture of venous blood. Thus, the arteriovenous difference is about 5. These units are many times referred to by physiologists as 15 volume percent. When the above figures are replaced in eq. (2.13), the result is a steady state value of 5 L/min.

Adolph Fick described the method in 1870 in a very short communication to the Society of Physics and Medicine of the city of Würzburg, in Germany (Hoff and Scott, 1948, see pp 26–31). However, he never actually put it into practice because he lacked the means to do it. Human cardiac catheterization was still many years ahead, even though Claude Bernard, on one hand, and Jean Baptiste Auguste Chauveau and Etienne Jules Marey, on the other, all three in France, performed it almost routinely in animals. Fick's idea was first tested in dogs by H. Gréhant and C.E. Quinquaud, in 1886, who reported values of 591 to 2,614 mL/min for body weights ranging from 7 to 18 kg. Zuntz and Hageman, in 1898, did it in the horse. It was Werner Theodor Otto Forssmann, in 1929, the first to introduce a catheter in his own right heart via the brachial vein, so demonstrating the feasibility of the procedure in the human being. He got later on, in 1956, shared with A. Cournand and D.W. Richards, the Nobel Prize, although at the moment he was severely reprimanded by his superior medical chief for breaking hospital rules. Finally Klein, in 1930, measured cardiac output in man by the *direct Fick method* (this is the way it is now called) obtaining venous samples with a cardiac catheter. It took 60 years (1870–1930) to reach the human application. Insufficient technology was obviously a factor against, but not enough basic knowledge was undoubtedly another (the human heart was perhaps viewed as some kind of "untouchable"). A third one may have been the relatively slowness of communications (as compared to those we enjoy nowadays). A nice epistemological little research project.

– *The indicator dilution method*

Even though it is a reference considered for many years a Gold Standard, the direct Fick method is not practical. Using an exogenous indicator (such as dye, radioactive substance, saline or heat), suddenly injected at a given site in the circulatory stream (peripheral vein, right ventricle or left ventricle), a *dilution curve* can be detected in a peripheral artery. Such curve, continuously recorded, contains information from which cardiac output can be obtained (Geddes and Baker, 1989). A basic condition is that the indicator bolus must traverse at least once the central pump or duct, either the left or the right.

Figure 2.7 represents a hydraulic and simplified model (which can actually be built and tested). It is composed of a pump, a main duct carrying the total flow F_t, a number of branches (1, 2, ..., j, ... n), and a return pipe back to the pump with the same flow. Thus, it is a closed leakless circuit. By injecting a known weight m_i of indicator in one branch and by detecting the passage of the indicator in another branch, after passage through the pump, the total flow can be determined. The calibrated detector, with its associated recording equipment, provides the dilution curve from which the necessary data are extracted.

Under ideal conditions, the indicator will be uniformly mixed after injection in a small liquid cylinder of length x_k and cross sectional area A_k.

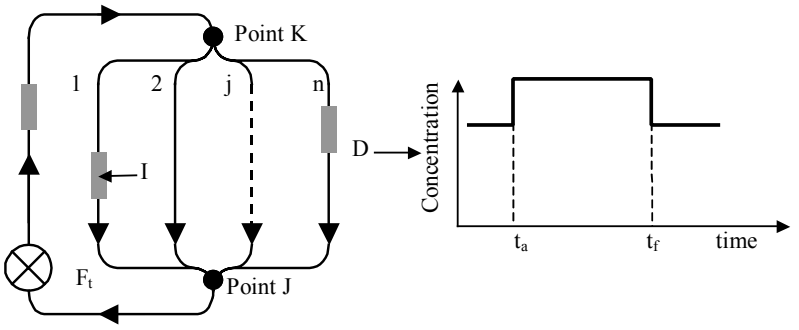


Figure 2.7. HYDRAULIC AND SIMPLIFIED MODEL. It is composed of a propelling pump, a main outflow, *n* branches and a return line. I and D are, respectively, the injection and detection sites. An adequate transducer and recording system produce an ideal concentration versus time output.

The cylinder will move along the branch and appear in the main return conduit at, say, point J, where the amount of indicator will be now uniformly dissolved in a new cylinder of length x and cross-sectional area A. This cylinder of volume $V = xA$ will advance at a velocity x/t , where t is the time required for the cylinder to progress a distance equal to its own length. Multiplying the latter by A, leads to $F_t = Ax/t = V/t$. On the other hand, the concentration of indicator within the cylinder is $C = m_i/V$ which, combined with the former and solved for flow, results in,

$$F_t = m_i / Ct \tag{2.14}$$

so offering a first expression relating total flow with indicator mass, concentration and time, the two latter derived from a highly inconvenient place in the circulation. Our little ideal cylinder proceeds its travel, gets through the pump, reappears at the main outflow and, after reaching the multiple branching off site at point K, it breaks off too into as many new cylinders as branches in the circuit. Each branch cylinder will take a portion m_j of the injected mass in such a way that the total amount emerging at point J should satisfy the *Conservation of Mass Principle*, that is,

$$m_i = \sum m_j \tag{2.15}$$

where j stands for any branch of the system of n branches. Besides, the total flow will be equal to the sum of all branch flows (once more the *Continuity Principle*). If the detector is located in branch $j = n$, at a time t_a (*appearance time*), it will begin to detect the passage of a cylinder of length x_j and cross-sectional area A_j , which moves at velocity x_j / t_j , car-

rying an amount m_i of indicator dissolved in its volume V_j . The time t_j is the time required for the cylinder to move a distance equal to its length. If the velocity equation is multiplied by the cross-sectional area of the branch, the flow in branch *j* is obtained, i.e.,

$$F_j = A_j x_j / t_j \tag{2.16}$$

where $V_j = A_j\, x_j$ is the volume of the *j*-th cylinder carrying a concentration $C_j = m_j / V_j$. **It must be emphasized that the latter branch concentration is what the recording system supplies**. Since uniform distribution was assumed and since sharp borders of the cylinder are supposed, the ideal dilution curve will be a perfect rectangle, ranging from a minimum to a maximum, the difference C_j being the concentration in the cylinder (Figure 2.7). The time difference, $t_f - t_a$, is known as the *passage time* t_j .Combining eq. (2.16) with the above branch concentration leads to the branch flow, or

$$F_j = m_j / C_j t_j \tag{2.17}$$

neither equal nor to be confused with the total flow F_t but obviously similar to eq. (2.14).

If now eqs. (2.14) and (2.17) are considered, recalling also the total flow as the sum of the partial branch flows, we can easily write,

$$m_j / C\,t = \sum m_j / C_j\, T_j \tag{2.18}$$

or, recalling eq. (2.15),

$$\sum m_j / C\,t = \sum m_j / C_j\, t_j \tag{2.19}$$

Each side of eq. (2.19) is a polynomial in m_j and, by using a corollary of a theorem of algebra, it can be seen that this equation is true only if the corresponding coefficients of m_j are equal, so that,

$$C\,t = C_1\, t_1 = C_2\, t_2 = \cdots = C_j\, t_j \tag{2.20}$$

meaning that in eq. (2.14) the denominator can be replaced by $C_j\ t_j$ and, thus, resulting in

$$F_t = m_i / C_j\, t_j \tag{2.21}$$

The latter equation is called the Stewart–Hamilton formula, widely used in the determination of the average cardiac output. It is based on the property described by the previous eq. (2.20): **the area under the dilution curve recorded at any peripheral artery is always equal to the area recorded under the main outflow** (Valentinuzzi, Geddes & Baker, 1968).

In actual situations, the indicator bolus is obviously not a cylinder. It has a drop-like shape, with a head and a long tail, tending to become longer as it advances within the conduit. Someone likened it to a huge spermatozoid. The concentration shows a maximum slightly behind the front edge decreasing steadily as one explores further back in it. This describes concentration as a function of the longitudinal axis of the bolus. Consequently, a detector, fixed in a given place, will see the bolus as it passes under producing an experimental concentration versus time dilution curve with a clear rise time, a maximum and a long tail (Figure 2.8). The area under the curve is expressed mathematically as the integral of the concentration curve as a function of time, from the beginning of the curve (*appearance time*) up to the end (*final time*). The latter should be at infinity, but for practical reasons is usually defined as the time when the concentration drops to 1% of the maximum level ($t_f = t_{1\%}$). By dividing the area (graphically obtained) into the time difference or base of the dilution curve, an average concentration results and, thus, an equivalent ideal rectangle is produced (Figure 2.8). In such a way, we can have a

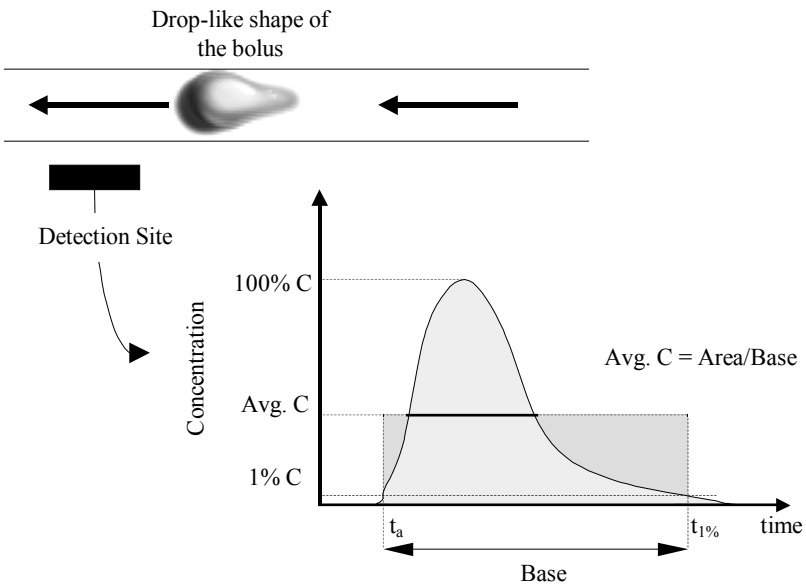


Figure 2.8. ACTUAL SHAPE OF THE DILUTION CURVE SHOWN AS A FUNCTION OF TIME. Usually, the end of the curve is taken at the 1% level with respect to its maximum value.

good graphic correlate of eq. (2.21). More properly, however, it is to rewrite the latter as,

$$F_t = \frac{m_i}{\int_{t_a}^{t_f} c(t)dt} \tag{2.22}$$

where $c(t)$ stands for the concentration function whose explicit mathematical form is not known, except for empiric approximations (Valentinuzzi, Valentinuzzi & Posey, 1972), and is only obtained graphically after performing an experiment. The difference, $t_f - t_a$, is the *passage time*, encompassing the dilution curve.

The concept of an indicator diluted in blood was already present in Fick's report. His indicator was a physiological substance, either oxygen or carbon dioxide, both part of the tissue metabolic processes and, hence, always present at expected levels in arterial and venous blood. It was G.N. Stewart who, late in the 1890's, introduced the concept of constant infusion of a foreign substance (saline solution in his case) to calculate the output of the heart. The substance is infused at a rate of I [mg/s] until saturation is detected at the outflow, i.e., when maximum concentration C_{max} is reached. Obviously, the curve will show a sigmoid shape and flow will be easily given by $F = I/C_{max}$ (Figure 2.9). Observe that Fick method is a constant infusion method (for oxygen or carbon dioxide always enter into the circulation).

Constant infusion is conceptually similar to applying a *step function* to the system under study, while the output concentration curve is the *step response* of it. Many years later, the same Stewart came up with the sudden or single injection method, thereafter improved by W.F. Hamilton and his group, when they used a dye as indicator. This is equivalent to applying an *impulse or Dirac function*. Hence, the dilution concentration curve represents the *impulse response* of the system. As a consequence, the engineering student should not be surprised if the **concentration output after constant infusion is proportional to the time integral of the concentration response after a single injection** (Figure 2.9). One important requirement is a really fast injection of the indicator bolus, which obviously faces practical limitations. Quite interesting, too, is the fact that two engineers, C.M. Allen and E.A. Taylor, back in 1924, used

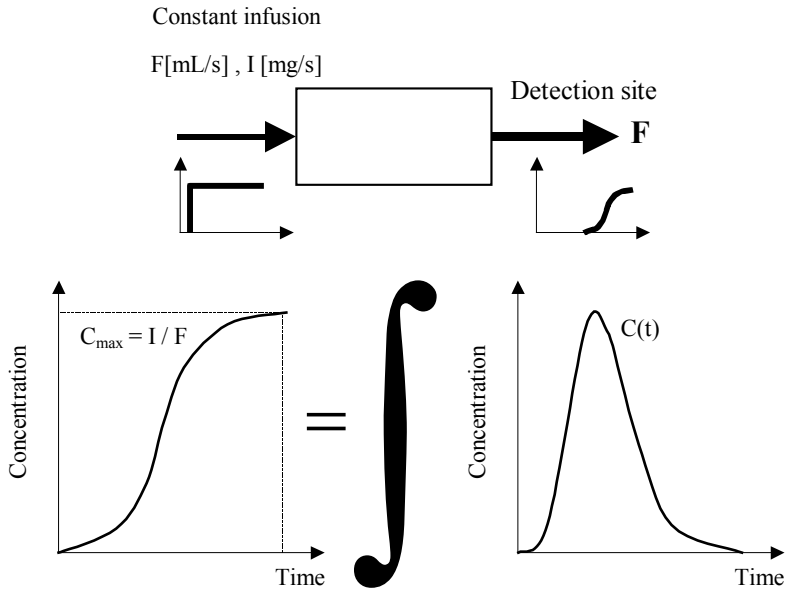


Figure 2.9. CONSTANT INFUSION METHOD TO OBTAIN FLOW. Saturation is reached at C_{max}.

salt as a marker for water measurements in hydraulic systems (Grodins, 1962; Zierler, 1962, 1963).

Suggested exercises: Make a list of the possible limitations referred to above. Remember that the sudden single injection has the Dirac function as its ideal. Make also a list of the requirements needed for a good indicator. There is a compromise of two opposing and conflicting requirements. Is there a perfect indicator?

The indicator-dilution curve, as the single injection response of the hydraulic system under study, can be viewed from another reference frame. Say, now, that we are studying the time it takes to the particles of a population to traverse a hydraulic system when carried by a flow *F*, from the injection to the detection site (Figure 2.7). Even though they are transported by the same average flow, some will move faster than others. They are running some sort of race and the analogy is quite valid: All runners start at the same time, but those that run at higher speed will reach the detecting site sooner while most of them will get there later, while a smaller slower group will be last. A recording camera would

show this phenomenon very clearly. Well, the dilution curve at the detecting site is some kind of "recording camera": the early arrival times are obviously made by a few, then there is a sharp increase until a maximum is reached (most of the particles arrive at time t_{max}), and thereafter and **asymmetrically**, less and less particles get to the detecting site (these are the slow ones). Thus, **the dilution curve can be interpreted also as the only continuously recorded event that yields the distribution of arrival times**, **at a given site**, **after introduction of a known number of particles in an upstream injecting place**. It can be demonstrated that it follows the so called *Poisson distribution* (Figure 2.8).

Cardiac output measurement, in general, and the overall theory and practice of the indicator-dilution method is a wonderful chapter of hemodynamics and hydraulics, highly attractive to the engineering student, with aspects still deserving attention and calling for further development.

– *Cerebral and coronary flows*

Knowledge of the overall and average blood flow to the brain and myocardial mass is obviously very valuable. They are the most important beds. One drawback, though, is that this value does not supply information regarding the *distribution* of blood throughout the region, for some areas may not be well perfused leading to tissue damage.

Blood gets to the brain via the two *carotid arteries* (easily felt by finger palpation on the neck, at both sides of the trachea) and, to a lesser degree, via the two *vertebral arteries*, on the back of the neck. Return takes place through veins that run parallel to the above-mentioned arteries. Kety and Schmidt developed a method, in 1948, for the measurement of the *average blood flow to the brain per unit of tissue mass.*

Let us draw a flow diagram (Figure 2.10) with three nodes: the lungs, L, the left ventricle, LV, and the cerebral mass, M. We also have a patient breathing nitrous oxide, N_2O, at a rate of N [mL/min]. The patient will fall into sleep because this gas is anesthetic, traversing very easily the tissue barriers. There is a total flow of blood, F, to the lungs from the right heart and from the lungs to the left ventricle. From the latter, the greatest portion proceeds to the general circulation, while a small fraction, F_{cer} [grams blood/min], supplies the brain. This is precisely the regional flow we are interested in. It carries a high concentration, $A(t)$ [mL gas/gram blood] and, as it perfuses the brain, it diffuses through the capillary walls, the interstitial space and the nervous cell membranes getting into the intracellular space. Thus, *the brain tissue takes up nitrous oxide*

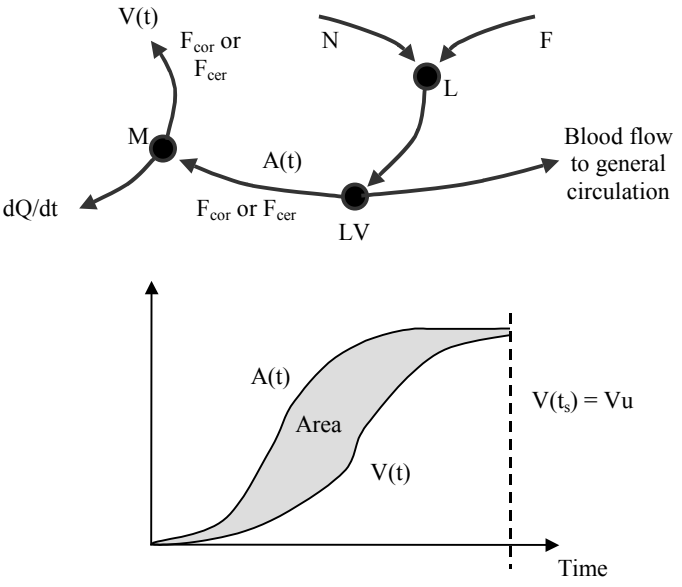


Figure 2.10. NODE SYSTEM TO OBTAIN THE CORONARY OR THE CEREBRAL BLOOD FLOW. It must be underlined that only an average value per unit mass of tissue is obtained.

at a rate dQ_b/dt, measured in [mL gas/min]. Evidently, the venous blood outflow will carry a low gas concentration, $V(t)$, because there was *tissue extraction*. **At the node M**, **the continuity principle states that**, **at dynamic equilibrium conditions**, **the amount of gas that goes in must be equal to the amount that goes out**. In mathematical terms, it writes,

$$F_{cer}\,A(t) = dQ_b/dt + F_{cer}\,V(t) \qquad \text{[g gas/min]} \quad (2.23)$$

where, we repeat, there are flow of blood in [g bl/min], flow of N_2O in [mL gas/min], and concentration of gas in [mL gas/g bl]. The student should practice with the units to fully grasp the concept. This is why there is some repetition, precisely to underline its importance.
From eq. (2.23), it is seen that the nitrous oxide uptake rate appears as,

$$dQ_b / dt = F_{cer}\left[A(t) - V(t)\right] \quad (2.24)$$

where A and V are functions of time obtained experimentally (Figure 2.10) sampling and testing blood for gas content at regular intervals. Both curves tend approximately to the same saturation level, with venous blood concentration staying always slightly below arterial concentration.

The total gas amount taken up by the brain is the integral, between the initial and the saturation time, T_s, of the latter equation, so that, after solving for F_{cer}, it becomes,

$$F_{cer} = \frac{Qb}{\int_0^{Ts}(A - V)\,dt} \tag{2.25}$$

where the denominator describes the shaded area depicted in Figure 2.10. Tracking down the units once more, it is seen that Q_b has to be measured in mL of gas, thus well checking with the concentration and time units below in order to yield grams of blood per minute. The area is graphically calculated from the experimental records since arterial and venous samples (the latter from the jugular) are relatively easy to obtain, but there is no way of measuring or even estimating the total amount of gas absorbed by the brain. However, if eq. (2.25) is divided through by the weight W of the brain (of course, unknown), we get,

$$F_{cer}/W = Q_b/W\left[AREA(A-V)\right] \qquad \text{[g blood / min] / [g tissue]} \tag{2.26}$$

an expression still looking unusable. Hence, some more assumptions are needed.

Once the patient has breathed for a while (T_s = 10 to 15 s), the arterial and venous curves reach saturation and it is said that *the gas concentration in brain tissue is in equilibrium with the gas concentration in venous blood*, or

$$V(T_s)\left[\text{mL gas/g bl}\right] = (Q_b/W) \qquad \text{[mL gas/g tissue]} \tag{2.27}$$

so describing a usual situation in exchangers in general. Moreover, it is also assumed that 1g blood is approximately equal to 1g tissue, which does not sound to bad because the density of lean tissue is about 1.10, that of fat tissue is around 1.03, and that of blood is between 1.05 and 1.06. Hence, considering (2.27), eq. (2.26) becomes simply,

$$BF = V(Ts)/\left[AREA(A-V)\right] \tag{2.28}$$

where BF represents the brain flow expressed in g or mL of blood per min per g of tissue. Sometimes, a factor α (called the partition coefficient) is included in eq. (2.28) to account for the imperfect passage of substance across the tissue barrier. This factor is about 0.98–0.99 and has to be experimentally estimated. With this ingenuous procedure, the authors mentioned above reported, from 14 patients, an average flow of 54 mL/min for each 100 g of brain tissue, with a standard deviation of 12.

Exercise: First, compare eq. (2.22) with (2.25). Do you notice any similarity?
Derive now the expression for the coronary flow using the same approach as for the cerebral blood flow. Search for the necessary information. Where would you take the blood samples from? Average value is 78 g blood/min (or 73 mL bl/min) per 100 g of myocardium. Find out the approximate myocardial weight of an adult normal heart and calculate its total coronary flow. What percentage is it of the total flow? Brain and heart show also a remarkable autoregulation: search for more informartion about it.

There are techniques (as for example with radioactive microspheres or with technecium^{99} or tallium^{201}, using the so called Single Photon Emission Computed Tomography, or SPECT) to estimate regional flows, both in the brain and the heart, so permitting a better evaluation of possible injuries and also of therapeutic strategies. It is, indeed, a highly specialized subject. The student is encouraged to search in the literature.

2.2.1.8. The pressure-volume loops

– *Origins*

Nicolas Léonard Sadi Carnot (1796–1832) was a French physicist and engineer who, in 1824, published a little and yet seminal contribution to science and technology: *Reflexions sur la puissance motrice du feu et sur les machines propres a developer cette puissance* (Thoughts about the power of fire and about the proper engines to develop such power). It was the foundational hallmark of *thermodynamics*, indeed, essential for the understanding and design of the *internal combustion machines* (such as the steam, gasoline and Diesel engines). In them, the pressure and volume handled by the cylinders with movable pistons which, in the end, transmit the force to produce the rotation of a shaft, give rise to the so called *pressure-volume diagrams* or *loops*.

Theoretical and practical thermodynamics developed rapidly all through the XIXth Century. Cardiac physiology, in turn, made significant progress. It was Otto Frank, in Germany in 1899, the first to bring the concept of pressure-volume loop into physiology, when with very rudimentary means he produced such a diagram from a frog's ventricle. However, in the 1970's and 1980's, the group of the late Kiichi Sagawa, at Johns Hopkins University, after a careful and long series of studies, put this engineering concept in its adequate place and interpreted it within the overall frame of cardiac physiology, opening with it a complementary new and fresh vision (Sagawa, Maughan, Suga *et al*., 1988).

– *Generation and description of a PV-loop*

Electronics Engineering students are familiar with the Lissajous figures. They are predictable nice looking composite patterns obtained by combining two sinusoidal signals injected, simultaneously and respectively, to the *x*-horizontal and *y*-vertical plates of an oscilloscope after disconnecting the horizontal internal sweep. The pressure-volume diagrams are some sort of Lissajous patterns except that the component signals, while still periodic, are not sinusoidal. One is the intraventricular pressure, P_{iv}, and the other, the intraventricular volume, V_{iv}, both as time course events experimentally obtained. In fact, any of the four cardiac chambers can produce such a loop.

Figure 2.11 displays ten beats recorded from an experimental dog. The upper channel is the left intraventricular volume, calibrated in milliliters and detected by a set of electrodes which measure blood conductance within the chamber (conductance is proportional to volume), while the lower channel is the left intraventricular pressure, calibrated in mmHg and detected by a miniature transducer, also placed within the chamber.

For each beat, volume, as time proceeds, shows a downstroke from a maximum value (*end diastolic volume*) to a minimum level (*end systolic volume)*, thus, describing *ejection*. Thereafter, it is followed by a rising limb back to a peak value and, hence, marking the *filling phase* of the left ventricle. Observe that the ventricle never empties completely. Pressure, instead, always starts from zero (or almost zero) abruptly climbing to the maximum value and displaying a rectangular-like shape.

As an exercise, the student should take a careful look at these records trying to correlate them in time.

Figure 2.12 combines beats 2 and 3 shown in the preceding Figure 2.11

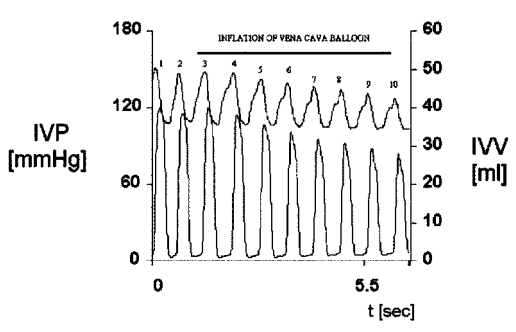


Figure 2.11. INTRAVENTRICULAR VOLUME AND PRESSURE RECORDS. They are displayed as time course events during a preload maneuver. Volume above (calibration on the right side) and pressure below (calibration on the left). Obtained at the Departement of Bioengineering, UNT, 1995.

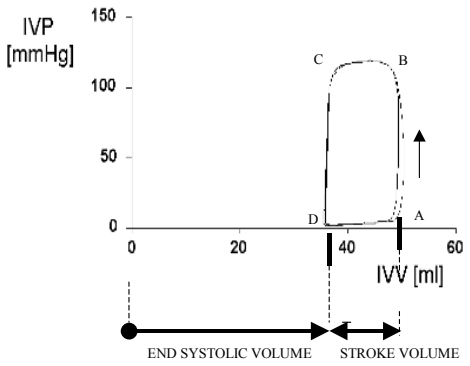


Figure 2.12. INTRAVENTRICULAR VOLUME-PRESSURE DIAGRAMS. Beats 2 & 3 of the previous Figure 2.11 were combined as orthogonal signals producing a Lissajous-like pattern, that is, two almost coincident PV-loops. The two isovolumetric phases (relaxation, on the left, and contraction, on the right) take place along the levels, respectively, of the end-systolic volume, and of the end-systolic volume plus stroke volume = end-diastolic volume. Points A and C mark the beginning and end of the mechanical systole (subdivided in the isometric contraction and ejection phases). The vertical arrow on the right shows the counter-clockwise rotation of each beat. On its way back to A, the cycle enters into its diastolic portion (subdivided, in turn, into relaxation and mainly passive ventricular filling, the latter from D to A). The four corners correspond to the openings and closures of both left cardiac valves.

to yield the Lissajous-like shape of two typical PV loops, both almost coincident. Roughly, the corners of this quasi-quadrangular figure identify the four characteristic points of the *cardiac cycle*. They signal the openings and closures of the cardiac valves: A(V_{ed}, P_{ed}), defined by the *end-diastolic volume* and *pressure*, respectively; B(V_{ed}, P_{op}), defined by the same *end-diastolic volume* and a much higher pressure than the preceding one, which can be called the *opening pressure* because it is larger than the aortic pressure and, thus, it opens the aortic valve, so starting the *ejection phase.* Ventricular contraction started at A, which coincides with the closure of the *mitral valve* (for the intraventricular pressure becomes higher than left intra-atrial pressure), intraventricular pressure rises to P_{op} and that contraction keeps on going sustaining pressure during all ejection until the pressure inside the chamber falls below the aortic level triggering the *closure of the aortic valve*. That is point C(V_{es}, P_{es}), defined by the *end-systolic volume* and *pressure*, respectively. The same point marks the end of ventricular mechanical systole, as point A characterizes its beginning.

Thereafter, the myocardium relaxes *isovolumetrically* (at constant volume) at a fast decreasing pressure, until it reaches zero or almost zero, when the low but higher left intra-atrial pressure opens the *mitral atrioventricular valve*. It is point D(V_{es}, P_o), defined by the previous *end-systolic volume* and an almost zero intraventricular pressure, and signaling the beginning of ventricular *filling phase*, which ends at A, to start all over again in the next beat. Hence, the whole cardiac cycle is divided in two halves: systole, between A and C (closure of the mitral valve and of the aortic valve), and diastole, between C and A (closure of the aortic valve and of the mitral valve). Each semicycle is, in turn, subdivided in two phases: First, the *isovolumetric contraction* of the ventricle, between A and B, that is, when both valves are closed and, thus, the chamber cannot change its volume (blood is essentially incompressible). Most of the energy consumption takes place during that short phase. Second, the already introduced *ejection phase*, between B and C, when the aortic valve is fully opened and blood is propelled into the elastic aortic reservoir. Finally, at diastole, we find the *isovolumetric relaxation*, between C and D, and the last *filling phase*.

Exercise: Compute the area encompassed by one of the two loops shown in Figure 2.12 using any means you may have at hand. How would you interpret it? Give a thought to the units.

Question: By what mechanism cardiac valves open and close?

– *The FSSS line*

Figure 2.13 displays eight loops after composing eight pressure and volume beats, both as time course events, during inflation of an intra-caval balloon, as indicated in Figure 2.11 (upper continuous bar). This is called in the hemodynamics jargon a *preload maneuver*. Transient occlusion of that great vessel (10 to 30 s, at the most) significantly reduces the venous return to the left atrium, in turn reducing the amount of blood to the left ventricle. According to Starling's Law, a reduction both in pressure and stroke volume must be expected, as clearly shown in the temporal records. The manifestation of such phenomenon in the PV-plane is a sliding of the loops to the left, with the end-systolic point (C point) describing a straight line which is called the *Frank–Starling–Suga–Sagawa* (or *FSSS*) *line.* It has been demonstrated that that line (essentially its slope) can be used as an estimator of the contractility of the myocardium, or, in other words, as a measure of the myocardium inotropic condition. A lin-

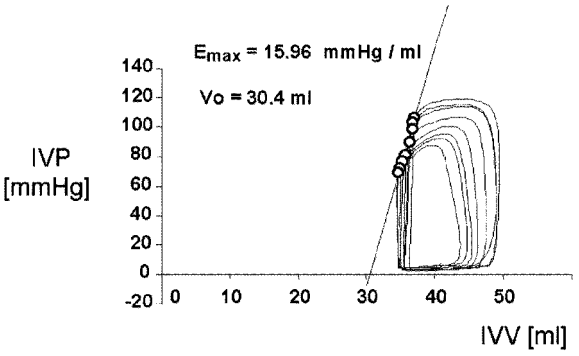


Figure 2.13. END-SYSTOLIC LINE (or Frank–Starling–Suga–Sagawa Line). It was obtained with a preload maneuver, that is, a partial occlusion of the venous return to the right heart, producing a shift to the left and downward of the PV loops. The latter behaved as "if hanging from the line and sliding down". See text for details. Obtained from a patient at the Institute of Cardiology of Tucumán with the technical assistance of the Department of Bioengineering, UNT, 1996.

ear regression of the end-systolic points obtained by a preload maneuver produces an equation of the type,

$$P_{es} = E_{\max}\left(V_o - V_{es}\right) \tag{2.29}$$

where E_{max} stands for its slope, and V_o , for the horizontal axis intercept. There has been considerable discussion over the detailed definition of the C points on the loops and over the meaning of the volume for zero pressure. One relatively well accepted stance is that the end-systolic point occurs where the absolute slope of the PV-loop is maximum, i.e., when P/V reaches a maximum. However, the E_{max} is a statistical value obtained from the collection of perhaps eight to ten such maximum slope data points in a single maneuver. When pressure is zero, there is a residual volume which Sagawa and collaborators named the *dead volume*. Much has been written regarding its interpretation. It shows wide spread and its true meaning still remains controversial, especially when now and then some negative values come up.

The filling phase of the cardiac cycle corresponds to the passive increase in pressure as volume is steadily increased in a balloon, an experiment easily carried out in any lab. If one measures the increments ΔV, with a calibrated syringe, and simultaneously one reads the resulting increments ΔP, with a manometer, the overall passive PV-relationship can be plot-

ted, so graphically describing the elastic properties of the material. Under ideal conditions, it is linear, following the well-known Hooke's Law, i.e.,

$$P = a + kV \tag{2.30}$$

from which, by differentiation, one gets $dP/dV = k$, the *elastance* of the material, obviously the inverse of the compliance, as already introduced above.

Suggested exercise: First, check in any physics textbook for the regular statement of Hooke's Law, as applied, for example, to a string under tension. Play mathematically with it to obtain eq. (2.30). Equation (2.29) has the same mathematical form, with its slope being also an elastance. However, does it describe a passive or an active parameter?

Let us underline that during a preload maneuver, the PV-loops shift progressively to the left, with the end-systolic points sliding down the FSSS line, as if "hanging" from a cable. The base, that is, the filling phase DA, also slides to the left and slightly downward, following the passive elastic PV-relationship. Such relationship, however, is not linear, being usually approximated by an exponential function. Thus, the latter acts as some sort of lower rail. An inflatable balloon can also be inserted in the thoracic aorta via, say, one of the femoral arteries. During inflation for a few seconds, it tends to hinder the outflow of blood from the left ventricle leading to an increase in intraventriculer pressure and a transient accumulation of blood in it. This is called an *afterload maneuver*. The PV-loops shift upward and to the right, but always "hanging" from the FSSS line and sliding over the passive non-linear PV-relationship (Valentinuzzi & Spinelli, 1989; Feldman, Erikson, Mao *et al*., 2000; White, Brookes, Ravn *et al*., 2001; de Vroomen, Steendijk, Lopes Cardozo *et al*., 2001). In normal physiological conditions, the heart moves constantly in this fashion, adjusting itself almost beat by beat to the demands. Even more, certain substances (as for example, epinephrine, in the case of exercising) can modify the position of the FSSS line (steeper slope) and, hence, offering more room for the left ventricle to operate. Everything that has been said so far for the left ventricle is also applicable to the right, keeping in mind the lower pressure it works with. An ischemic ventricle, instead, will contract with less force and the line would drop (lower slope). See the references given above in this same paragraph to illustrate.

Review exercise: Using published records found in the literature, draw by hand a left ventricular PV-diagram marking and identifying the four characteristic points. Do the same for the right ventricle. Identify the valves you are dealing with.

Jules Antoine Lissajous (1822–1880) was a French physicist who studied at the *École Normale Supérieure*, in Paris. His doctoral dissertation, in 1850, was devoted to the recordings of vibrations. In 1873, he was awarded the Lacaze Prize. The theoretical analysis of his famous patterns can be found in any physics textbook (for example, Page, 1965) under the heading of simple armonic motion.

Sadi Carnot graduated from the *École Polytechnique*, in Paris in 1814. He worked on the mathematical theory of heat and helped start the modern theory of thermodynamics. In 1824, he published the only work during his lifetime that includes his description of the ideal cycle. This work became well known after Clapeyron published an analytic reformulation in 1834. It was incorporated into the thermodynamic theory of Clausius and Thomson. It is not known whether Otto Frank was directly influenced by these physics and engineering concepts. In his papers there is no reference to them but, as a well-informed high-level scientist, he must have been aware of that knowledge.

2.2.1.9. Arterial input impedance

– *Definition*: *First component*

Two terminals of a given circuit will tend to oppose to a change of state when an excitation is applied. If the latter is a sinusoidal voltage, V, the system will present some hindrance to the establishment of a sinusoidal current, I. The relationship of V to I, in the s-domain, **and only in the *s*-domain**, as a generalization of Ohms' Law, gives the simplest mathematical definition of the *electrical impedance* of the system. All electrical engineering students and practitioners are well familiar with this concept. Thus,

$$Z(s) = V(s)/I(s) \tag{2.31}$$

where the impedance Z is, obviously, a complex function of the complex frequency s.

A similar rationale can be used in the vascular tree: Vascular impedance expresses the relation of the forces acting in the bloodstream to the resulting motion of blood (Milnor, 1982). The physical properties of the blood and of the blood vessels determine vascular impedance. Therefore, it contains information of paramount importance. The concept was first introduced by J.R. Womersley, in 1955, furthered by his close associate D.A. McDonald, in 1955 and subsequent years and, later on, more elaborated by other authors, like Patel, Greenfield & Fry (1964). Milnor's classic book referred to above lists all the significant literature up to that year. The concept is valid for the aorta, the pulmonary artery, arteries supplying specific beds (like the kidneys, the liver, the legs, or any other), or just any vessel, including veins.

Two basic variables in the cardiovascular system (see above) are pressure and flow, respectively analogous to voltage and current. Both are periodic and mathematically unknown functions of time t. However, **they are not sinusoidal**. Figure 2.14 displays aortic pressure (upper channel), picked up by a miniature transducer placed almost at the root of the vessel with a catheter inserted via a femoral artery, and aortic flow (lower channel), detected by means of an electromagnetic flowmeter embracing the artery after the arch. The animal was an anesthetized dog. The foot of each beat in the pressure record (diastolic pressure) marks the opening of the aortic valve, hence, ejection starts, and the record below shows a rapid upstroke. The indentation after the maximum value (systolic pressure) marks the closure of the same valve. It is called the *dicrotic notch*, DN, a handy flag present in any good quality arterial pressure record. As a consequence, flow drops to zero after having shown a maximum. The two vertical bars clearly bound *ejection time*, ET, which in this particular case is in the order of 300 ms.

After having obtained experimental digital records of pressure, $P = p(t)$, and of flow, $F = q(t)$, they are subjected to spectral analysis, usually by means of the Fast Fourier Transform (FFT). The latter algorithm is readily available in many commercial softwares, for which the student should be familiar with Fourier series to better understand it. In other words, each signal is decomposed in its dc (P_o, Q_o), and sinusoidal components (p_n, q_n), in which the fundamental frequency, f_1, coincides with the heart rate, HR. Thus,

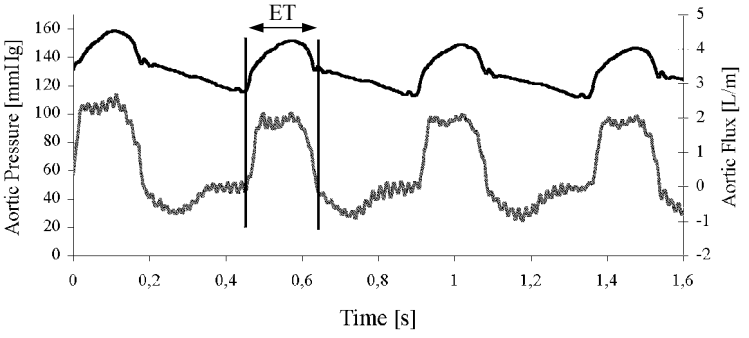


Figure 2.14. AORTIC PRESSURE (upper curve) AND AORTIC FLOW (lower curve). Experimental time course records from which the aortic input impedance can be calculated as explained in the text. The dicrotic notch in the upper channel marks the end of ejection. ET = ejection time, between the opening and the closure of the aortic valve. Obtained from an experimental dog at the Department of Bioengineering, UNT, 1994.

$$p(t) = P_o + p_1 + p_2 + \ldots + p_n \tag{2.32}$$

$$q(t) = Q_o + q_1 + q_2 + \ldots + q_n \tag{2.33}$$

where $p_n = P_n \cos(n\omega t + \theta_n)$ and $q_n = Q_n \cos(n\omega t + \psi_n)$. Phases or angles are represented by θ_n and ψ_n. In practice, it is enough to keep up to the tenth harmonic ($n = 10$), as it is well documented in the literature (Geddes, 1970).

It is now the time to calculate the impedance, both in modulus and in angle, for each harmonic component by extension of the definition used in electrical circuits:

$$Z_o = R_o = P_o / Q_o \tag{2.34}$$

$$Z_l = P_l / Q_l \tag{2.35}$$

$$\vdots$$

$$Z_n = P_n / Q_n \tag{2.36}$$

for the moduli, and 0 and $\phi_n = \theta_n - \psi_n$, for the phases. Observe that the dc component of the arterial impedance component has a zero angle (thus, pressure and flow are in phase) and its modulus Z_o coincides with its real part R_o, which, in turn, coincides with the peripheral resistance Rp, introduced earlier above. Obviously, the *hydraulic input arterial impedance concept* is more general than the simpler Poiseuille's definition,

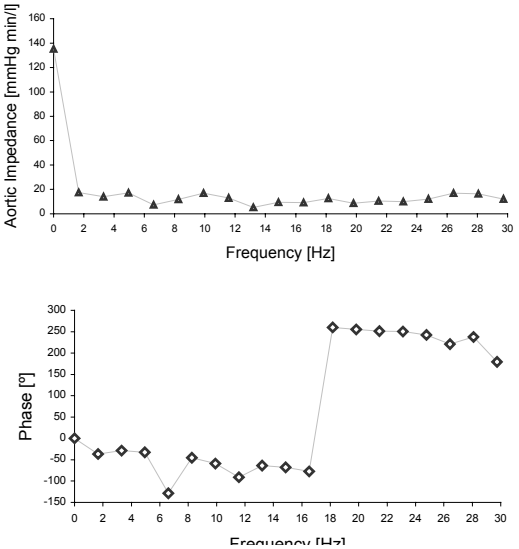


Figure 2.15. AORTIC IMPEDANCE. The calculations made with the data obtained from Figure 2.14 led to these two graphs: impedance modulus above and phase below, both as functions of the frequency expressed in Hertz.

the latter being the first component contained in the former.
With 20 beats — of which those shown in Figure 2.14 are part — Figure 2.15 was composed and drawn by means of special software. The upper part plots the moduli while the lower portion is the phase, both as functions of frequency. **The whole represents the input aortic impedance**. Each point is the average of twenty beats, with a heart rate of 159 beats/min and a peripheral resistance (read it at zero Hz) of 138 mmHg.min/L.

– *Two other components*

Two other components are present in the hydraulic input impedance: a capacitive-like reactance, X_C , and an inductive-like one, X_L , both functions of the frequency, the former describing the elastic properties of the vessels and the latter related to the mass of blood in movement. They are also scaled to unit length. Analogous to electric relationships, they are written as,

$$X_C = (1/\omega C) \underline{/-90°} \tag{2.37}$$

$$X_L = \omega M \underline{/\ 90°} \tag{2.38}$$

and are polar equations which clearly show their moduli, one inversely proportional to the angular frequency, $\omega = 2\pi f$, and the other directly proportional to it, one with a 90° lead and the other with a 90° lag, as expected in any well-behaved purely capacitive or purely inductive circuit. C is the *compliance* already introduced before, and M is the mass of blood or *inertance*, both expressed per unit length of vessel. At lower frequencies, the phase diagram of Figure 2.15 indicates a capacitive dominance, while at higher frequencies the behavior tends to be inductive.
The moduli of the two reactances defined by eqs. (2.37 and 2.38) lead to another parameter, the *characteristic impedance* of the system, or

$$Z_C = (X_L \times X_C)^{1/2} \tag{2.39}$$

which is easily read in the impedance modulus plot by averaging out the points beyond the second harmonic.

– *Reflections and ventricular-vascular coupling*

In an electric line, when there is no *reflected wave*, the generator internal impedance is equal to the impedance offered by the line (the so called *characteristic impedance*) and also to the load impedance. It is thus said

that there is *impedance matching*, with maximum transfer of energy from the generator to the load.

The pulsatile pressure and flow within the arterial tree **show significant reflections** clearly indicating that, at least from the electrical engineering point of view, no impedance matching is met. It is also obvious that the cardiovascular system is not designed for maximum transfer of energy. In fact, the actual recorded waves — as for example those displayed in Figure 2.14 — are the composition of the forward and of the backward (or reflected) waves (Yin, 1987).

If the heart is the generator and the arterial system is the line connecting it to a load, the question arises as what the optimal or the best coupling between the two is. The subject is still without a definite answer. Authors like O'Rourke, Avolio and Nichols (in Yin's 1987 book) state "Arterial function is optimal when the fluctuation around mean pressure is minimal. Ideal ventricular-vascular coupling entails as low a mean pressure as practicable for adequate organ flow, with as low a mean systolic pressure and as high a mean diastolic pressure as possible. Low mean systolic pressure allows adequate ventricular ejection with low oxygen demands by myocardium and little stimulus to cardiac hypertrophy. High mean diastolic pressure allows adequate coronary perfusion."

The student is invited to carefully peruse this previous paragraph trying to define and understand each of the concepts mentioned (such as mean arterial pressure, mean systolic pressure and mean diastolic pressure) and the reasons given in each of its sentences. What does cardiac hypertrophy mean? What can stimulate its appearance?

Summarizing: Arterial input impedance is a complex concept graphically presented as modulus and phase, both as functions of frequency. It includes peripheral resistance, elastic vessel properties and the mass of blood. Wave reflection and the coupling between ventricles and their outflow arteries are strongly related to this concept.

To think about and eventually to search in the literature: Where is the major reflection site in the arterial tree? Would an abdominal aortic aneurism or an abdominal coarctation of the aorta affect the hydraulic aortic input impedance? Explain why.

The student is encouraged to review the theory of electric transmission lines in order to find possible analogies with the arterial tree. Recall also that technological hydraulic systems **do not show elastic properties** in their conduits.

2.2.1.10. Body fluids

> "*...all are of the dust, and all turn to the dust again* (*Ecclesiastes, 3:20*) ...
> *just thirty percent, for the remaining seventy ... it evaporates*!"

Blood is indispensable. No wonder that the antiques called it the "elixir of life" and that nowadays encompasses a specialty of its own: *hematology*. It is part of the body fluids, also and obviously essential for the maintenance of viability, so much that many medical practitioners devote their talents to better handle *internal environment* derangements (as for example in kidney failure, or severe dehydration). Thus, we will offer here a brief account (McArdle *et al*., 1991).

Body weight (*BW*) — a stressing and sometimes pathological concern in the current occidental culture — is composed of two clear-cut parts: solid materials (*SM*) and total body water (*TBW*). If referred to body weight, the following simple equation can be written,

$$100 = \left[SM / BW + TBW / BW \right] 100 \tag{2.40}$$

The first term lies between 30 and 40% while the second one is in the order of 70 to 60%. Hence, most of our body is water. Since 1L of water is equivalent to 1 kg, a person after heavy exercise can easily lose 1, 2 or 3 kg of weight ... which he/she readily recovers when, driven by sheer physiological thirst, he/she drinks the same amount or maybe more. Beware: drink it with some ionic content, otherwise you may run into trouble.

The solid material, in turn, is essentially formed by proteins, *P*, mostly located in the muscular mass, by minerals, *M*, mostly held in the bone structure, and by fat, *F*, distributed all over the body anatomy (and many times showing a prominent concentration in the abdominal volume). Hence, the 30 to 40% of solid material gives way to,

$$\left(SM/BW \right) = \left(P/BW \right) + \left(M/BW \right) + \left(F/BW \right) \tag{2.41}$$

leaving out the percental 100 for the sake of simplicity. Each term in the right side of the equation above amounts, respectively, to about 18, 7 and 15% in a normal individual. In obese people the amount of fat is larger and that of water smaller.

For those swimming fans: Who floats easier, a lean guy or a rounded fellow? Explain why.

Figure 2.16 summarizes the relationships of the body fluid compartments, including the approximate percentages normally accepted. The upper one states that the extracellular fluid (*ECF*) and the intracellular fluid (*ICF*) collect all the body water, or

$$1 = (ECF/TBW) + (ICF/TBW) \tag{2.42}$$

Both compartments are separated by the cell membrane. The percentages with respect to *TBW* are, respectively, about 30 and 70%. Hence, all cells are immersed in an internal sea called the *internal environment* (the famous concept brought up by the French physiologist Claude Bernard in the second half of the XIXth Century). The cells, where all metabolic processes take place, hold most of the fluid. The *ECF* feeds, cleans and protects the cells. It acts as some sort of buffer. The latter, in turn, is described by the second diagram, where

$$1 = (IF/ECF) + (PV/ECF) \tag{2.43}$$

The interstitial fluid (*IF*), i.e., that fluid between cells and outside the capillaries and lymphatic vessels, and the plasma volume (*PV*) are the components of the *ECF*. Only a quarter of the extracellular fluid is plasma confined within the cardiovascular system. Plasma volume communicates to the interstitial fluid through the highly permeable capillary

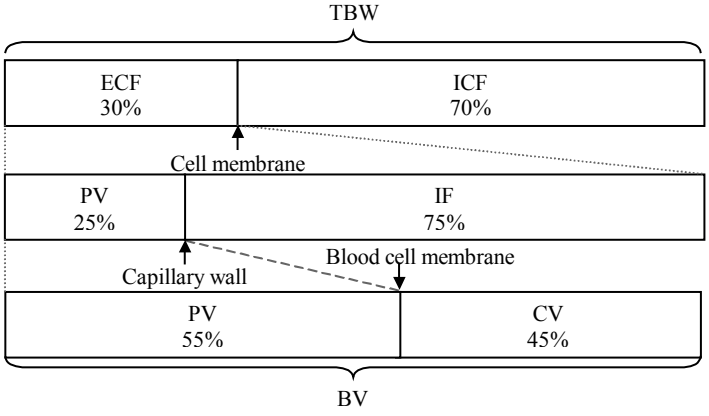


Figure 2.16. BODY COMPARTMENTS. *TBW* = Total Body Water; *ECF* = Extracellular Fluid; *ICF* = Intracellular Fluid; *PV* = Plasma Volume; *IF* = Interstitial Fluid; *CV* = Cell Volume in blood. Percentages are referred to *TBW*, *ECF* and *BV*, respectively, from the upper to the lower diagram. Remember that fluid in blood *CV* is part of *ICF*. See text for details.

walls
The volume of plasma is approximately 55% of the total volume of blood, the remaining part are cells (white and mostly red cells). In symbols,

$$BV = PV + CV \tag{2.44}$$

or, in relative terms, dividing through by BV and multiplying by 100,

$$100 = 100 \times (PV/BV) + 100 \times (CV/BV) \tag{2.45}$$

from which two clinically important relationships and concepts come out, the *plasmacrit*,

$$Pc = 100(PV/BV) \tag{2.46}$$

and the *hematocrit*,

$$Hc = 100(CV/BV) \tag{2.47}$$

both expressed in percent relative to blood volume. For the latter, values below 40% are considered as abnormal. Centrifugation of a blood sample leaves the cell volume packed at the bottom of the tube and plasma on its upper part. It is a simple way of assessing a possible *anemic* condition.
The *Hc* includes the fluid contained in the blood cells. If this material is desiccated to remove the water content, then the blood dry weight is obtained.

Exercise: Using the relationships of above, estimate all the presented parameters for a normal male adult of 70 kg. If needed, complement with data taken from any good textbook.

2.2.1.11. Closing remarks

By now, the student should have a fairly good idea of the variables, parameters, organs, and laws involved in the so-called **mechanical activity of the cardiovascular system**. Some simple exercises were suggested in order to spur curiosity along with a few historical remarks to show how concepts developed with time and knowledge. The text was also interspersed with clinical insights marking, in addition, possible avenues of study and the associated relevant bibliography. A useful practice is to keep a notebook at hand (or in the computer) where those subjects still to be disclosed or clarified are taken down. With time, the student will collect a list that may be helpful when a research project, say, for a doctoral dissertation, has to be decided.

2.2.2. Cardiac Electrical Activity

In this part, the electrical activity of the heart is dealt with, starting with the essential concepts of electrophysiology, continuing with the origin of the heart impulse, its propagation throughout the whole myocardial mass until it is detected from outside by means of surface electrodes to produce the well-known and traditionally used electrocardiogram. Finally, an introduction to changes in the normally rhythmic activity is given. Comments, suggestions and some insights into newer knowledge from molecular biology are also included plus short historical tips. Stimulating the student to put his/her inner drive into action looking back and forward is extremely important.

2.2.2.1. Essentials of electrophysiology

– *Model*

The electrical phenomena associated with tissue functions, in particular the so-called *excitable tissues*, fall within the domain of *Electrophysiology*. Nerves, skeletal, cardiac, and smooth muscles — all excitable tissues — are characterized by a permanent resting and stable electrical state, E_1 (Figure 2.17). In fact, such resting condition is so important that, when it disappears, it flags the death of the cell, offering an unmistakable criterion for the experimental researcher to decide whether his/her preparation is still viable (for example, when working with microelectrodes as recording elements). When an *adequate stimulus* is applied to the cell, the response is a change in the electrical state to another level, E_2, re-

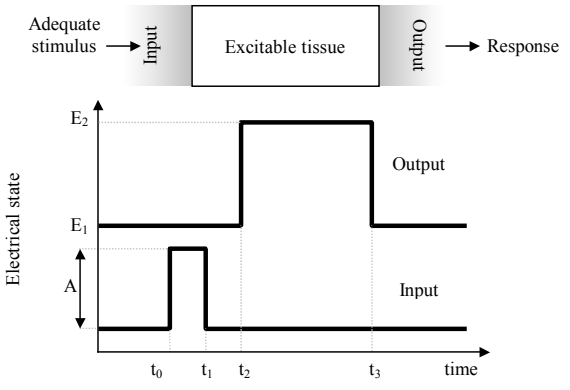


Figure 2.17. MODEL OF AN EXCITABLE TISSUE. The output has an electric stable resting state, E_1, and a metastable state, E_2, after an adequate stimulus is applied as input. The time the second state is kept, $t_d = (t_3 - t_2)$, depends on the characteristics of the tissue. Thus, the behavior reminds of an electronic monostable flip-flop.

maining at it for a very specific length of time, $t_d = (t_3 - t_2)$, *which exclusively depends on the characteristics of the tissue.* It reminds the behavior of an electronic monostable multivibrator (mono = one). The time difference between t_2 and t_0 is called the *latency*, that is, it measures how long it takes the cell to react to the stimulus and is also a characteristic of the tissue. The amplitude of the cell response, the *action potential*, is obviously the difference of the two electrical levels, E_2 and E_1.
In the laboratory environment, the stimulus is produced and controlled by an external equipment (biological stimulator) and usually it is of rectangular shape, defined by an amplitude, A, either in volts or amps, and width $(t_1 - t_0)$, in ms, both adjustable at will. In the physiological normal situation, action potentials act as stimuli, although their actual shape may deviate considerably from the ideal rectangular waveform.

Small research subject: Stimuli other than the electric type can elicit a response in excitable tissues. Find out which. Mention three more. Hint: Think in terms of types of energy. Did you accidentally hit your elbow experimenting a nasty electric sensation? Explain.

– *Resting membrane potential*

Most of the cells of excitable tissues are long and cylindrical in shape surrounded by a *membrane*, about 100 Å thick (1 Å = 10^{-8} cm), which separates the intracellular fluid (*ICF*) and its contents from the extracellular fluid (*ECF*). When a very small diameter electrode (a *microelectrode*, of about 1μm) — connected to one of the inputs of a special high impedance amplifier — is introduced into the *ICF* puncturing the thin cell membrane while the other amplifier terminal is hooked to a return electrode immersed in the *ECF*, the recording instrument (a dc coupled oscilloscope) shows a displacement of the base line from the zero level to about –80 to –90 mV, assuming that the excitable cell (say, a skeletal muscle one) is alive. This is the *resting membrane potential* or the stable electric state, E_1, already introduced above. The internal side of the membrane is negative with respect to its external counterpart. This highly summarized description is an experimental fact that can be demonstrated in any laboratory of electrophysiology. The *ECF* contains a high concentration of sodium ions and a low level of potassium ions. Conversely, the *ICF* shows a low level of sodium and a high concentration of potassium. Both ions on both sides of the membrane are also accompanied by chloride ions. The membrane is relatively permeable to all these charge carriers; however, it does not permit the passage of large proteic anions, which abound within the cell. Besides,

the membrane is a good insulator constituted by oriented proteins and phospholipids, roughly containing one ion per 5,000 water molecules. *ECF* and *ICF*, instead, have in the order of one ion per 175 water molecules, meaning that these fluids are by far better electrical conductors.

Study subject: The student should check in any physiology textbook the values reported for sodium, potassium, chloride and proteic ions in *ECF* and *ICF*, in nerve and skeletal muscle cells. Via INTERNET, we suggest DEVELOPMENT OF TRANSMENBRANE RESTING POTENTIAL, by David L. Atkins, Professor of Biology at George Washington University, atkins@qwis2.circ.qwu.edu, 1998. Calculate also the electric field, in volts/meter, stressing the cell membrane. Compare it with porcelain. Review also the fluid compartments. Notice that the *ECF* faced by the excitable cell membrane is its interstitial fluid part *IF*, with no proteins. Plasma, instead, the other portion of *ECF*, *contains a large amount of proteins* and is exclusively restricted to the cardiovascular system.

– *Resting potential by the Ionic Theory*

On December 20, 1998, Sir Alan L. Hodgkin died in his home residence of Cambridge, England, at the age of 84. He was 1963 Nobelist in Physiology or Medicine for his analysis of the ionic basis of the action potential, former Master of Trinity College in Cambridge University and President of the Royal Society. Co-recipients of the prize were also Andrew Fielding Huxley and John Carew Eccles, the latter from Camberra, Australia. However, many others have also contributed to the ionic theory and a few of their names will be mentioned here (such as Kenneth S. Cole and Bernard Katz). Some classical and enlightening references for the interested student are Hodgkin (1964), Katz (1966), Cole (1982) and Gardner (1992). Even though there are still a few who question the validity of the theory, most of the scientific community has accepted it and continues to work in its improvement.

Figure 2.18 represents the cell membrane with two vertical lines. On the left, there is the *ICF*, i.e., inside the cell, and on the right we have the *ECF*, which coincides here with the interstitial fluid. The former is essentially a compartment high in positive potassium and negative proteic ions while the latter is characterized mainly by a high concentration of positive sodium ions. Potassium and sodium cannot exist as such just by themselves because they appear from electrolytic dissociation of chemical compounds, in this case, of KCl and NaCl. Hence, both compartments also have their chloride ionic counterparts as the proteic anions in the *ICF* have theirs, in such a way that, on **both sides**, the *neutrality principle* must be met, that is to say,

Sum of all positive charges = sum of all negative charges

or

Algebraic sum of all charges = zero (2.48)

The first postulate of the Ionic Theory of Excitation states that, within the cell membrane, there is an *active ionic pump* that extrudes sodium from the cell and carries potassium into it (Figure 2.18). For the time being, let us assume that the ionic exchange is 1:1, or, for each sodium out, one potassium is brought in. It is *active* because the pump consumes energy supplied by the cell itself. Its slowing down or inactivation tends to level concentrations off and, in the end, kills the cell. The amount of *evidence* collected in favor of the $Na^+ - K^+$ pump over the years in countless laboratories all over the world is enormous and overwhelming, indeed; however, perhaps there has not been yet a **final conclusive** *demonstration*, so offering one relatively weak side to those questioning the theory. Notice the semantic difference between *evidence* and *demonstration*. In fact, other ionic pumps have been described having become a widely

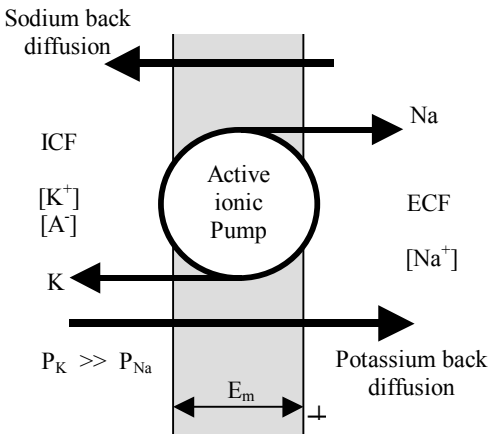


Figure 2.18. MEMBRANE RESTING POTENTIAL. The two vertical lines represent, respectively, the internal (left) and external (right) sides of the cell membrane, hence delimiting its thickness. The active ionic pump maintains the ionic concentration gradients (potassium is higher in the *ICF* than in the *ECF* and sodium is just the opposite). Besides, in the resting condition, the *permeability of the membrane to potassium is much higher than its permeability to sodium*, thus, potassium back-diffuses easier to the *ECF* than sodium does it to the *ICF*. Since both carry positive electric charges, the net result is an accumulation of positive charge on the *ECF* side, breaking neutrality and so giving rise to the resting membrane potential, E_m, positive outside and negative inside.

used concept in physiology.

The second postulate of the Ionic Theory says that, in the resting condition, the *permeability*, P_K, *of the membrane to potassium ion is higher than its permeability*, P_{Na}, *to sodium*. This is a demonstrated fact with a numerical ratio between the two of about 50 to 75. Due to the concentration gradients, there is a passive (with no energy expenditure) back diffusion of both ions; one tends to return sodium to he *ICF* and the other pushes potassium to the *ECF*. However, because of the large permeability difference, for each Na^+ getting back inside the cell, there are 50 to 75 K^+ leaving it, so accumulating positive charges on the *ECF* side, depleting the *ICF* side of them and, hence, creating a charge imbalance on both sides with the outer face positive to the inner. Thus, here we have the membrane resting potential, E_m (Figure 2.18), which obviously breaks neutrality.

The membrane is impermeable to the large proteic anions and they stay within the cell, but chloride can pass through it passively *following the electric field* sustained by the resting potential and reaching finally an equilibrium state. Due to osmotic forces that depend on the concentrations of the osmotically active particles (largely the small ions), water also moves across the membrane so determining cell volume.

Exercise: Suppose the permeability ratio potassium to sodium is one half of the stated value, would the resulting membrane potential increase, stay the same or decrease? Suppose both permeabilities were equal, what resting potential would result? Suppose the pumping rate drops to one half of its normal value, would it affect the membrane potential? Suppose a toxin stops the pump, what would E_m be? Does the pump directly contribute to the resting membrane potential? For the latter question consider a 1:1 pump ratio. Note: do not confuse membrane ionic permeability ratio with ionic pumping ratio. This exercise is important for the full understanding of changes in membrane potential due to permeability changes.

Exercise: Suppose the pump ratio changed to $2Na^+/3K^+$ and both permeabilities were equal, would the pump contribute to the membrane potential? If the answer were affirmative, which side would be positive to the other? Suppose now the ratio changed to $3Na^+/2K^+$ keeping always both permeabilities equal, which side would be positive?

In real life, very rarely if ever, the pump is electrically neutral (1:1 ratio). The ionic exchange ratio varies from tissue to tissue and, even within the same tissue, histologically different fibers may show also different ratios. A typical example is found in the heart. In fact, the determination of the sodium-potassium pump exchange ratios is still a subject of research.

Therefore, one portion (usually less than one third) of E_m is contributed by the pump. Ionic pumps, which sustain potentials, are called *electrogenic.* In other words, **the resting membrane potential recognizes two sources**, **one** (**the largest**), **comes from the large difference in permeabilities via back diffusion**, **in turn dependent on concentration gradients sustained by the sodium-potassium pump**, **and the other** (**much smaller**), **directly produced by the pump due to a not neutral pumping sodium-potassium ratio**, **different from 1**:**1**.

– *Action potential*

It was said above that an *adequate stimulus* applied to the cell triggers a response characterized by a change in the electrical stable state to another level, E_2, remaining at the second one for a very specific length of time, t_d; thus, it can be called *metastable.* Let us describe the phenomenon with more detail: The external side of the membrane is electrically positive. If some of these charges are removed, i.e., if the membrane is *depolarized*,

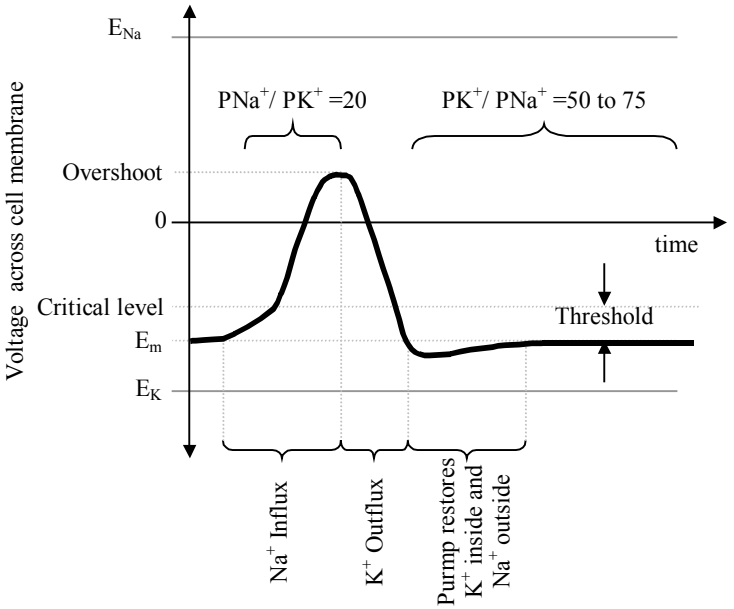


Figure 2.19. ACTION POTENTIAL AND ITS BOUNDARIES. The action potential is always bounded by the potassium potential, near the membrane potential, and the sodium potential above, which is a limiting unreachable value. See text.

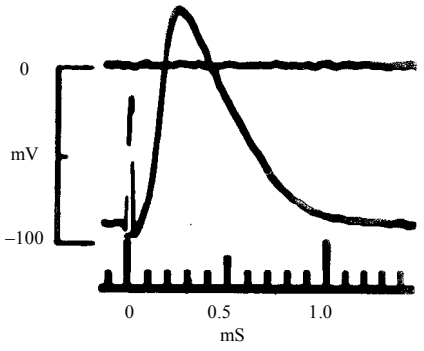


Figure 2.20. ACTUAL ACTION POTENTIAL. Recorded with microelectrode from a cat dorsal root nerve fiber. Observe the stimulus artifact just before the depolarization upstroke, where the zero time is placed. The overall duration of the event is shorter than 1 ms. The positive overshoot appears clearly depicted. The resting potential lies around –90 mV. Reproduced from Ruch and Patton, Physiology and Biophysics, Copyrigth 1966, Chapter 2 by Walter Wodbury, figure 2C, page 30, with permission from Elsevier.

the resting potential shifts towards zero (Figure 2.19). If depolarization continues, the decrease in E_m (less negative inside) eventually reaches a level called the *critical* or the *firing* or the *threshold potential*. At the latter level, the membrane takes command of its own depolarization which rapidly proceeds until reaching a maximum, E_2, to, thereafter and somewhat slower, returning to the initial resting state (Figure 2.19). Beyond the critical or firing level, the applied stimulus loses control completely. Figure 2.20 is an actual nerve action potential photographed on the oscilloscope screen after sweep synchronization with the stimulus (seen as a small spike preceding the fast depolarization upstroke). By comparison of both Figure 2.19 and Figure 2.20, the student can easily identify potential levels and durations. Notice the conceptual difference, many times mixed up in the daily jargon, between *critical* (or *firing*) *threshold potential* and *threshold stimulus*: the first one is just the negative potential measured from the zero line (in Figure 2.20, it could be in the order of –70 mV), above the resting potential E_m (about –90 mV), while the latter is *defined as the threshold potential minus the membrane resting potential*, or

$$S_{th} = E_{th} - E_m \tag{2.49}$$

which, using the numerical values stated above to illustrate, would yield (–70) – (–90) = 20 mV. This equation clearly says that the applied stimulating voltage must overcome the difference separating the resting level from the threshold potential and, since it is applied via an internal microelectrode, it has to be a positive pulse.

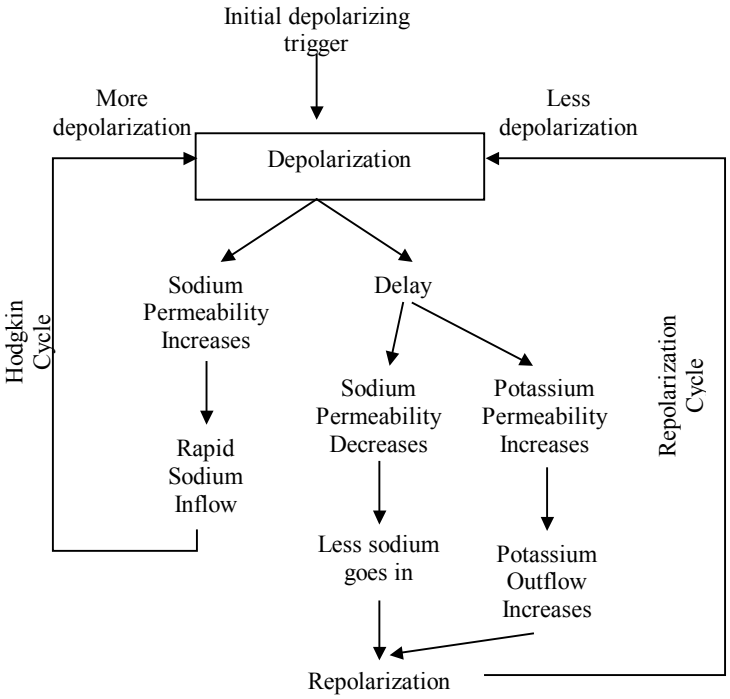


Figure 2.21. HODGKIN ACTIVATION CYCLE. Depolarization produces an increase of the membrane permeability to sodium leading to a positive feedback loop. With some delay, depolarization starts an increase of the permeability to potassium so initiating repolarization. See text.

Thinking exercise: An *adequate* stimulus was mentioned above, implying some conditions to be met by the stimulating rectangular pulse. Two conditions were already introduced. Search for them in the preceding text. Think of a third, which is easily fulfilled by any electronic stimulator. Hint: Do practical rectangular pulses change instantly from one level to the other?

Study subject: Find out what the strength-duration curve is. From it, the important concepts of utilization time, rheobase, and chronaxy are derived. This experimental relationship was introduced in 1909 by the French physiologist Louis Lapicque (1866 – 1952). Old beautiful theoretical and still valid in many respects descriptions were given by Nicholas Rashevsky in his classic book (Rashevsky, 1960). The student is encouraged to at least take a look at some of them, as for example, Blair's Theory of Excitation. Notice that these theories were developed long **before** Hodgkin's.

So far, we merely described the action potential. Let us continue with Hodgkin's group contributions. A **third postulate of the Ionic Theory** is that *depolarization produces an increase in the membrane permeability to sodium ions* which, due to the outside-inside concentration gradient, leads to an increase in sodium influx with more positive charges toward the internal face and more depolarization to the membrane (recall that depolarization means either negative charges on the external side or positive on the inside). Beyond the critical level, however, this process is greatly and irreversibly accelerated so that the external stimulating front is no longer needed because more depolarization brings about higher permeability to sodium with a rapid positive ionic entrance and even more depolarization. A positive feedback loop, called *Hodgkin activation cycle*, is established. It is characterized by a *fast inward sodium current* (Figure 2.21). However, after a certain delay, the effect of depolarization on sodium permeability slows down and eventually there is a reversal while, simultaneously, the permeability of the membrane to potassium ions increases. Thus, less sodium gets into the cell and a quite rapid and steady potassium outflow takes place, both leading to depolarization slowing down and reversal to repolarization. The latter inactivation cycle is characterized by potassium extrusion.

Ionic membrane permeabilities to both ions are, thus, main actors in the resting state and during excitation. Changes are rapid and dramatic, from a stable potassium to sodium ratio of 50 to 75, depending on the tissue, to a full reversal of sodium to potassium in the order of 20, at the metastable state or overshoot. **Permeabilities act**, **indeed**, **as true controllers**. Associated with these changes, there are ionic currents, a very fast sodium one (depolarization) and a somewhat slower but also fast potassium efflux (repolarization). Ionic exchange has become a tremendously important physiological tool. Now the specialist talks in terms of *ionic channels*, because the mechanisms involved may be different even for the same ionic species, thus, they refer to, say, potassium channel 1 and potassium channel 2. Whole scientific conferences are being devoted to the subject with participation of electrophysiologists, cell physiologists, molecular biologists, geneticists and, why not, biomedical engineers. They discuss concepts such as channel gating, channel pore, channel associated proteins, assembly of channels, channel regulation, and ionic channel dysfunction associated with disease. The student is encouraged

to enter the American Physiological Society INTERNET Web Site to search for these and related subjects.

Study subject: There is a famous figure, after Hodgkin and collaborators, reprinted over and over in papers and textbooks, which clearly shows an action potential and both ionic membrane permeabilities, the two latter depicted as electrical conductances. In the figure, the delay is very well seen. Find this figure; inspect units and times taken for each phase (depolarization and repolarization). If you cannot find it, go to the library and check in Hodgkin and Huxley (1952) original paper (p 530, its figure 17). It would be a good exercise at least to glance it over.

A previous experimental finding associated the action potential with a temporal membrane impedance change. It was a superb biomedical engineering accomplishment carried out by Cole and his collaborator, H.J. Curtis, in 1939 in the USA, with the cell membrane forming one arm of an impedance bridge and rather rudimentary but well designed vacuum tube electronics. We urge the student to search also for this frequently reprinted oscilloscope photograph. Hodgkin *et al*. broke up the impedance change in its two sodium and potassium ionic components.

– *Nernst equation*

Figure 2.19 displays two horizontal lines marking two levels: one, positive and above the overshoot, called the *sodium equilibrium potential*, E_{Na}, and another, negative and below the resting level, called the *potassium equilibrium potential*, E_K. They represent theoretical potentials delimiting precisely the band within which the action potential occurs. It never can go beyond such band, neither above nor below it.

We need the concept of Nernst–Gibbs–Donnan passive equilibrium potential. Hermann Walther Nernst (1864–1941) was a Prussian born physicist who developed methods for measuring dielectric constants. He was awarded the Nobel Prize (Chemistry, 1920) for his theoretical work on the Third Law of Thermodynamics. Probably, he is best known to electrochemists for his elucidation of the now called Nernst equation. A simple, illustrative, and perhaps not too rigorous demonstration is given below, encompassing both, electrolytic and solid semiconductor solutions.

Let us assume two compartments, 1 and 2, separated by a semipermeable interface, holding respectively concentrations C_1 and C_2 of electrically charged carriers, which can be either ions in electrolytic solutions or electrons/holes in doped semiconductor materials (such as silicium or germanium or the like). If $C_1 > C_2$, a *diffusion current*, i_{DIF}, appears, proportional to the concentration gradient,

$$i_{DIF} = q \times D(\partial C / \partial x) \tag{2.50}$$

where q stands for the electric charge of the carrier, D represents the diffusion constant, and $(\partial C/\partial x)$ is the concentration gradient in the x-direction. In fact, the gradient should be written in space, but to simplify the mathematics we will stick to the one-variable case. It does not affect the concept.

The diffusion current across the semipermeable interface (a membrane in electrolytic solutions or the *pn*-junction in semiconductors) creates an electric field because of accumulation and depletion of charges on the sides, thus producing an *opposing current* called a *drift current*, i_{DRI}, or

$$i_{DRI} = q \times \mu \times D(\partial U / \partial x) \tag{2.51}$$

where q is the carrier charge in coulombs, as above, μ represents the *mobility of the carriers*, C stands for the concentration of carriers, and $(\partial U/\partial x)$ is the linear electrical potential gradient created by the diffusion current. The same comment as above is valid, i.e., a three dimensional gradient should be considered. Carrier mobility is defined as the velocity per unit of electric field, obviously measured in [m/s]/[V/m]. Ions have a much lower mobility than holes and electrons. Observe that the charge transfer across the interface breaks neutrality on each side.

Both currents (diffusion and drift) soon reach an equilibrium when they become equal, fully opposed, and thus stopping any further exchange, $i_{DIF} = i_{DRI}$, or after equating eqs. (2.50 and 2.51),

$$D(\partial C/\partial x) = \mu C(\partial U/\partial x) \tag{2.52}$$

which, after separation of variables and integration between C_1 and C_2, on one side of the equation, and between U_1 and U_2, on the other, yields,

$$\Delta U = U_2 - U_1 = (D/\mu)\ln(C_2/C_1) \tag{2.53}$$

Eq. (2.53) describes the equilibrium difference of potential, passively developed, when two compartments containing different concentrations of electrically charged carriers are separated by a semipermeable interface. This is already an expression of Nernst equation for electrolytic solutions or the *junction* potential well-known in semiconductor *pn*-junctions. Such potential difference, however, **is unable to sustain a current and**, **thus**, **cannot act as a generator**.

The student is encouraged to make a little drawing showing both compartments, their respective concentrations, marking the above-mentioned currents and the resulting voltage

difference. Review carefully the rationale so far developed trying to understand its essentials.

Now, Nernst equation will be rewritten in two other mathematical forms using well-known relations in physics. One of Einstein's equations relates several physical constants and parameters,

$$D/\mu = kT/q \tag{2.54}$$

where k = Boltzmann constant, T = absolute temperature, and q = carrier charge. Besides, Boltzmann constant is also given by R/N, with R = constant of gases, and N = Avogadro's number. The student should search for the numerical values and units associated with these classical parameters. After replacement of the above in eq. (2.53), it obtains,

$$\Delta U = (RT/Nq)\ln(C_2/C_1) = (RT/FZ)\ln(C_2/C_1) \tag{2.55}$$

where $Nq = FZ$ represents the electric charge of one mole of substance, with F being Faraday's constant and Z standing for the chemical valence of the element species. Eq. (2.55) is the form usually found in electrochemistry books. When the ion species is monovalent, $Z = 1$, further simplifying the right hand expression. However, semiconductor physicists prefer to write the equation as

$$\Delta U = kT/q \times \ln(C_2/C_1) \tag{2.56a}$$

easily rewritten as

$$C_2 = C_1 e^{[q\Delta U/kT]} \tag{2.56b}$$

The latter equation obviously is nothing but the well-known relationship between hole and electron concentrations in a *pn*-slab. An interesting consequence of eq. (2.55) is that the product of the concentrations on one side equals the product on the other side, leading to the so-called Donnan equilibrium (Donnan, 1925-6).

When Nernst eq. (2.55) is applied to potassium and sodium concentrations, respectively, found in the *ECF* and *ICF* compartments separated by the semipermeable and excitable biological membrane, the following typical values can be calculated.

$$E_{K^+} = RT/F \times \ln[K_o]/[K_i] = 61.5\log_{10} 31 = -93\,\text{mV}$$

$$E_{Na^+} = RT/F \times \ln[Na_o]/[Na_i] = 61.5\log_{10}(1/24) = 84\,\text{mV}$$

The latter are, respectively, the potassium and sodium equilibrium Nernst–Gibbs–Donnan potentials, if the system were a passive one dominated either by the first or by the second ion. Hence, they are only

theoretical potentials impossible to coexist. Obviously, they are different and of different sign (Figure 2.19). In the resting state, the membrane potential, E_m, however, is rather close to the potassium potential, E_K, say, –90 mV versus –92 mV, or thereabouts, depending on the specific tissue. Thus, it is said that the resting membrane potential is dominated by potassium. On the other hand, the sodium potential is way up (about +84 mV). The positive overshoot (in the order of +20 mV) tends to the sodium potential but it falls rather short of reaching it. As stated at the beginning of this section, **both theoretical equilibrium potentials mark the lower and upper limits of the action potential operating band**.

The student should check the numerical values of the equilibrium potentials applying concentrations quoted in the textbooks for nerve and muscle tissues. The temperature is, by and large, body temperature, but other physiological values are valid as well (as the case may be in amphibians or *in vitro* preparations).

Thinking exercise. Many experiments have been performed and reported modifying sodium and potassium concentrations and measuring the possible influence on the resting potential and on the overall amplitude of the action potential. With the preceding information, the student should be able to answer the following questions:

If the concentration of potassium ions were increased in the *ECF*, (a) would E_m change? (b) if the answer is yes, in what direction? (c) Is the amplitude of the action potential affected?

If the concentration of potassium were decreased in the *ECF*, (a) would the resting membrane potential change? (b) Would the stimulus threshold be different? (c) If the answer is yes, would a larger or smaller stimulus be needed?

If the concentration of sodium were decreased in the *ECF*, (a) would the resting potential be affected? (b) Would the action potential overshoot change? (c) If the answer is yes, would it be smaller or larger?

Draw an equivalent electric circuit modeling the membrane in the resting and in the maximum activity conditions. Permeabilities to sodium and potassium are represented by variable resistors and the equilibrium potentials by ... (complete the text). This circuit can be constructed and tested.

When the stimulus amplitude does not reach the firing level the action potential (is)(is not) triggered and the response, if any, is called subliminal. Select one of the answers between parentheses.

Demonstrate the Donnan equilibrium product in a two compartment NaCl system separated by a semipermeable membrane. Hint: equate the electrochemical potentials.

From the above, it is clearly seen that Nernst equation cannot describe neither the resting nor the maximum activity membrane potential. A modified empirical version was proposed by Hodgkin's group, called the

Goldman–Hodgkin–Katz or *GHK equation*, which takes into account both permeabilities and ionic concentrations and, thus, can be applied at any moment of the action potential,

$$E_m = -61.5 \log_{10} \frac{\left[K^+\right]_i + P_{Na}/P_K \left[Na^+\right]_i}{\left[K^+\right]_o + P_{Na}/P_K \left[Na^+\right]_o} \tag{2.57}$$

It is illustrative to insert actual numbers into the equation in order to test it and see how the different components affect the membrane potential. A better exercise is to plot the membrane potential as a function of each parameter in the equation keeping the others at fixed values. Use any of the available programs (such as MathCad, Excel, or the like).

An excellent and more detailed account of the preceding paragraphs is to be found in the classic textbook by Ruch and Patton (1966). Through INTERNET, the student can have access to ELECTROPHYSIOLOGY OF THE NEURON (http://tonto.stanford.edu/eotn), a rather recent nice interactive tutorial, by John Huguenard (John.Huguenard@stanford.edu) and David A. McCormick (David.McCormick@Yale.edu). The authors are, respectively, with the Department of Neurology and Neurological Sciences at Stanford University School of Medicine, Stanford, CA, and with the Section of Neurobiology at Yale University School of Medicine, CT. More elaborate and rigorous demonstrations of Nernst equation can be found elsewhere in the literature, as for example in Plonsey's classic book on bioelectric events (1969).

Two significant contributors to the subject were Frederick George Donnan (1870–1956), from the Physical Chemistry Laboratory at University College, London (Donnan, 1925–6), and Josiah Willard Gibbs (1839–1903), born in New Haven, Connecticut. The latter received from Yale University, in 1863, the first doctorate of engineering to be conferred in the United States. During a three-year stay in Europe, he received influence of high caliber scientists, such as Kirchhoff and Helmholtz (considered perhaps as the first biomedical engineer). He returned more a European than an American scientist in spirit, one of the reasons why general recognition in his native country came so slowly. In 1876 Gibbs published the first part of the work for which he is most famous, *On the Equilibrium of Heterogeneous Substances*, publishing the second part of this work in 1878. It is universally recognized that its publication was an event of the first importance in the history of chemistry.

– *Propagation of the action potential*

Excitability *per se*, as an electrical change, is just a local phenomenon, quite interesting and amazing, but of limited use for a living organism. The second characteristic that really enhances bioelectricity raising it to its full power is the propagation or conduction of the electrical change — the action potential — to adjacent parts of the excitable tissue. It is so

that it can travel at high speeds to areas distant from the original stimulus, traversing complex networks, and eliciting all kind of responses essential for life.

Once a given portion of an excitable fiber is fully depolarized to maximum activity (sign reversal), small potential gradients appear between the depolarized region and the yet resting neighbor areas to each direction, creating local current loops which depolarize them and, thus, generate new action potentials, one to the left and another to the right (Figure 2.22, upper part). This is the basic *smooth conduction* mechanism in nerves and muscles.

A discontinuous sheath of an electrically insulating fatlike substance called myelin, however, covers an important part of nerve fibers. It forms sleeves approximately 1 to 2 mm long and about 2 μm thick. In-between,

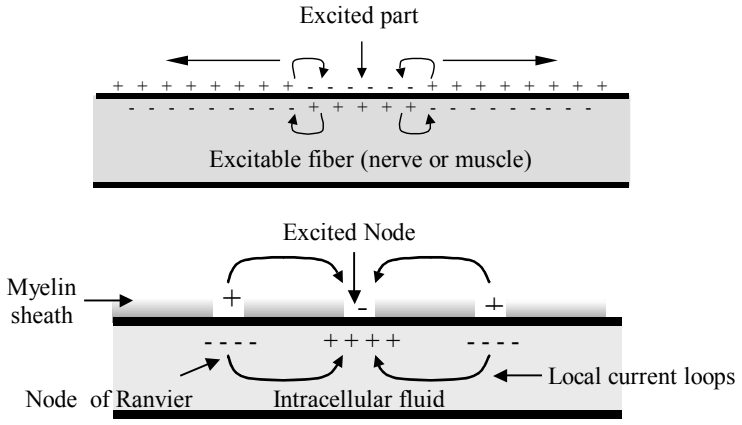


Figure 2.22. PROPAGATION OF THE ACTION POTENTIAL. Schematic representing a longitudinal section. Upper section - An excited portion acts as the stimulus to its adjacent right and left sections via the generation of depolarizing local current loops from positive to negative (curved arrows). This mechanism repeats itself to both sides resulting in the propagation of an action potential to the right and another to the left. By the time the two pulses reach their respective ends, the originally excited part is already recovered and ready to accept a second stimulus, if any. Lower section - It shows an important type of nerve fibers, covered by discontinuous sleeves of an insulating material called myelin. The cell membrane shows up only in-between, at 1μ gaps, the nodes of Ranvier. The local current loops only can be established between the excited one and the two immediate non-depolarized neighbors. Hence, there is some kind of jump (saltatory conduction).

the cell membrane shows up, at the nodes of Ranvier (after the French pathologist Louis Antoine Ranvier, 1835–1932). They are only 1μm wide. Hence, electrical charges concentrate around them, both outside and inside the cell. The electrical resistance across myelin is much higher than across the naked membrane (about 200 times). Conversely, the electrical capacitance across the membrane is about 200 times that across the sheath. Figure 2.22 (lower part) illustrates the conduction mechanism, similar to the one explained before, except that the current local loops "jump" or "leap" from node to node, for which reason it is named *saltatory conduction*. It has been demonstrated that a myelinated nerve can elicit a response only when the stimulus is applied to a node. If a node is blocked (as with cold or with an anesthetic), the local depolarizing loop can reach the next adjacent one and still propagate the action potential, but if two nodes are blocked, then propagation cannot proceed. The velocity of propagation greatly increases by the saltatory (jumping) trick. Quite an ingenious design!

Suggested exercise: There are several INTERNET Web Sites, entering for example with the word "Ranvier", where more detailed explanations, even with animated images, very didactically illustrate both mechanisms of action potential conduction (smooth and saltatory). The student can also find names and places all over the world still active in the subject.

The velocity of conduction of the action potential, in any excitable tissue, is one aspect of concern with definite patho-physiological and clinical implications. Its measurement is based on the estimation of the time required for the action potential to traverse a known distance. Historically, it was first Hermann von Helmholtz (1821–1894) who, in 1850, in frog's sciatic nerve and using a rudimentary technology (a rheotome, a transformer and a galvanometer, very ingeniously arranged), reported a value of 30 m/s (Geddes and Hoff, 1968; Grüsser, 1997). Later on, in 1928, Erlanger and Gasser, both Nobel laureates, confirmed the value within 10% of error introducing in physiology with these measurements the oscilloscope and the electronic amplifier (Erlanger and Gasser, 1937). In whole muscles, it is now possible to estimate such velocity using certain properties of specific types of signal processing (Spinelli, Felice, Mayosky *et al.*, 2001).

Suggested study subjects: Find out what a rheotome was. Analyze the etymology of the word. Search in the literature for the arrangement used by Helmholtz to measure conduction velocity.

Analyze the direct approach used first by Helmholtz and, thereafter, by Erlanger and Gasser to measure the action potential propagation velocity. Study any of the indirect techniques based on signal processing and currently applied in electromyography to obtain the same parameter.

When reading this, think in terms of the evolution and development of concepts and in the influence of technology in physiology. The application of a new technology does not necessarily mean a new physiological concept.

After going over the preceding introductory electrophysiology concepts, it is suggested to read a series of four didactic articles in the subject written by Nassir H. Sabah (1999, 2000a, b, c).

– *Electrophysiology of the heart*

The myocardium is composed of bundles of *myocytes* (cardiac cells). Under the microscope, it shows *striae* (lines, bands) somehow reminding those seen in skeletal muscle, but really not quite because the bands are not as well organized and arranged as in the latter. Cardiac cells differ from skeletal myofibers in that cardiac myocytes have as a rule only one

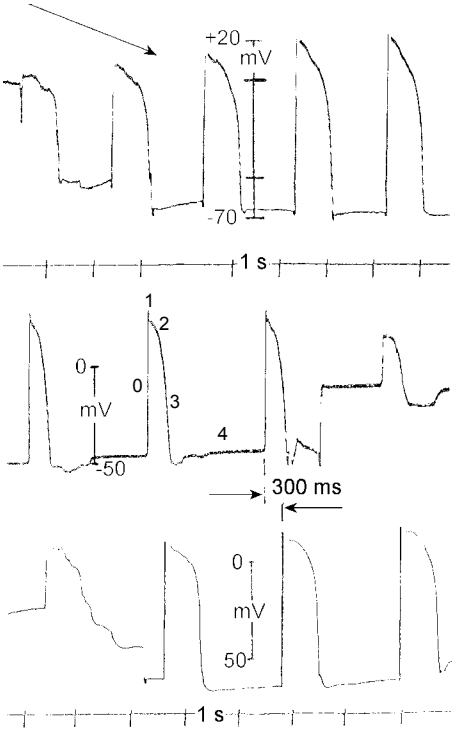


Figure 2.23. CARDIAC ACTION POTENTIALS. Upper record: A floating glass microelectrode pierced the cell membrane of a ventricular frog ventricle. Arrow above on the left indicates the piercing process, with a first beat when the electrode was on the surface, and a second beat with the electrode almost inside. The three last beats displays full action potentials. Diastolic potential was approximately – 70 mV with an overshoot of + 20 mV; thus, the total amplitude was 90 mV.

Lower two records: Two floating glass microelectrodes recorded simultaneously from the left atrium (above) and from the ventricle (below). The time difference of about 300 ms measures essentially the A-V delay. After the third beat (atrial channel) the electrode came off. The lowest record shows also penetration of the electrode on the left side. Records obtained by the author in 1967 at the Department of Physiology of Baylor College of Medicine.

nucleus per cell. Myocardial fibers are shorter than the fibers of skeletal muscle and rarely exceed 100μm in length, with a diameter in the order of 30μm–50μm. The shape is pretty irregular, showing projections to contact other cells in such a way that can be described as an anastomosing network. Such arrangement, very likely, tends to facilitate the propagation of the electric impulse.

The electrical phenomena associated with cardiac activity find their background in the general concepts of electrophysiology, but they show a number of peculiarities that have established *cardiac electrophysiology* as a specialty on its own. Any physiologist or biomedical engineer or physician can make a life in this important, attractive and fascinating field.

As any excitable tissue, all myocytes have a resting membrane potential and, when properly stimulated, trigger an action potential. Figure 2.23, shows atrial and ventricular action potentials recorded with microelectrodes. Hoffman and Cranefield (1960) published a superb and well-known set of action potentials recorded from several cardiac fibers.

Roughly, these action potentials look like distorted rectangles and, in principle, can be (and have been) modeled by a rectangular pulse. The two nodes (S-A and A-V), however, deviate most from that model, displaying smooth rounded shapes. All show a depolarization upstroke (phase 0, a standardized denomination in cardiac electrophysiology, see Figure 2.23, atrial channel), from a minimum (the resting or diastolic membrane potential) to a maximum (the overshoot). The overshoot is not present in nodal fibers. The first and short repolarization return is usually called phase 1 (Figure 2.23, small spike). Many papers have been published to explain it; its study well belongs to the specialist and is beyond the boundaries of this textbook. The overshoot is followed by a plateau (phase 2), well marked in Purkinje and ventricular fibers and slowly falling off in atrial and His bundle fibers. It is as if the repolarization process were somewhat held on for a given time. Thereafter, repolarization speeds up (phase 3) until the membrane returns to the initial resting state (phase 4). The plateau is not shown by nodal fibers.

For the interested student, we suggest to visit the WEB entering with the words *overshoot of the action potential*. For example, Joe Patlak and Ray Gibbons (Department of Physiology, University of Vermont), in 1998 and 1999, offered simple and short accounts of the

phenomenon without going into much details. The first microelectrode recordings from false tendons of dog hearts demonstrating the overshoot of the action potential upstroke were obtained by Coraboeuf and Weidman (1949). These are classic authors who produced substantial contributions to cardiac electrophysiology. This paper is written in French. We understand the difficulty, but the student should be aware that there is imporntant scientific literature written in languages other than English and that it would not hurt to learn how to read at least one foreign language (French, German, Italian, Russian, Portuguese or Spanish). Besides, it is fun. One extra language means another opened door.

Nerve and skeletal muscle action potentials are always short in duration, usually less than 1 ms. Cardiac action potentials, conversely, last between 200 and 400 ms, depending on the species and on the type of fiber. In other words, it is a very long metastable period. The duration of the contraction elicited by this action potential is of the same order of magnitude. In skeletal muscle, instead, a very short action potential triggers a longer contraction.

Study subject: Why cannot the heart enter into tetanic contraction? Do you know what tet-

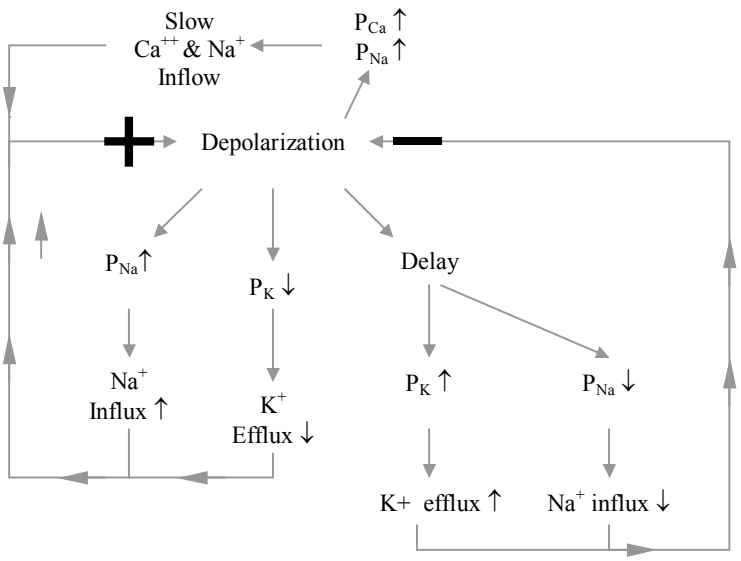


Figure 2.24. NOBLE's ACTIVATION CYCLE. Compare this figure with Hodgkin's activation cycle (Figure 2.21) for nerve and skeletal muscle. The long duration of the cardiac action potential is due to the effects of Ca^{++} and Na^{+} getting into the myocyte, but mainly to the former.

anic contraction is in skeletal muscle? Search in the literature for actual numerical values, say, comparing durations of action potentials and contractions in biceps or triceps with their counterparts in ventricular and atrial fibers of mammalian hearts. As a hint, check the following website: Medical Curriculum Objectives Project© 2001, The American Physiological Society (APS), Bethesda, MD.

Noble's cycle

The activation process in heart cells is somewhat different than in either nerve or skeletal muscle. Figure 2.24 modifies Hodgkin's cycle and summarizes the chain of events (Noble, 1962). Depolarization, after it is started, increases the permeability to sodium ions and simultaneously decreases the permeability to potassium. Due to the existing concentration gradients maintained by the sodium/potassium pump, an increment in Na^+ influx and a decrement in K^+ efflux take place, both events so reinforcing depolarization. Thus, a positive feedback loop is established leading to the fast depolarizing upstroke (phase 0). Slower Ca^{++} and Na^+ channels due to increases in their respective permeabilities allow entrance of these ions into the cell and, hence, since they carry positive charges, depolarization is held on (plateau or phase 2). After some delay, the effect of depolarization on the permeabilities is reversed leading to the repolarization process in a braking negative loop. Explanation and interpretation of phase 1 are relatively controversial and, thus, should be left for more advanced courses in electrophysiology.

Noble did not consider these latter calcium and sodium channels in his first 1962 paper. They were reported years later, by McAllister, Noble and Tsien who developed a Purkinje fiber model in 1975 and which, thereafter, led to the Beeler–Reuter mammalian ventricular model in 1977. Many experiments were performed to provide a greater insight into the working of the ion channels in cardiac tissue. Finally, Di Francesco and Noble (1985) constructed a new and now well-recognized model of cardiac electrical activity. It is recommended to search in INTERNET entering with the words *calcium and cardiac action potential*. There are several didactic and even interactive proposals that the student can enjoy and profit from. Send a requesting message, for example, to info@cellml.org. The Department of Physiology of Loyola University Chicago (Maywood, IL 60153), in turn, has developed another model that incorporates many findings on cardiac calcium regulation and ionic currents (LabHEART, by J.L. Puglisi and D.M. Bers, 2001).

Figure 2.25 graphically describes the time course development of cardiac action potentials and the various permeability changes of the membrane associated with them (both, in the case of ventricular and S-A node fi-

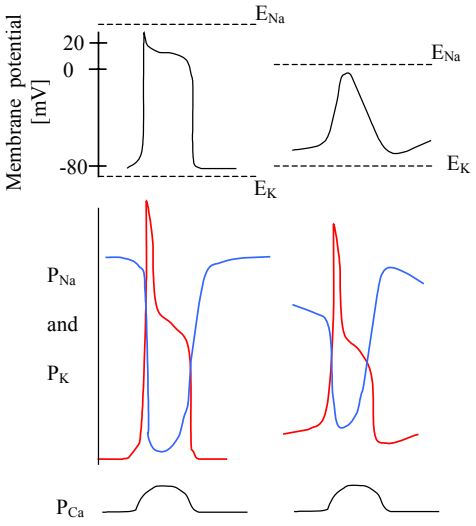


Figure 2.25 PERMEABILITIES AS TIME COURSE EVENTS. The upper portion displays action potentials of a ventricular fiber (left) and an S-A fiber (rigth), while the two lower lines show permeabilities to sodium (red), potasium (blue) and calcium (black), as they temporally correlate to each other. Sodium permeability opposes potassium's, so reminding the behavior of a multivibrator collector potentials. Mainly after a course of Electrophysiology offered by Dr. Robert Vick, Department of Physiology, Baylor College of Medicine, Houston, TX, 1967.

bers). Action potentials are also bounded, above and below, by the potassium and sodium potentials (see Figure 2.19). Observe that, tipically, the potassium potential of a nodal fiber is always less negative than that of any other cardiac fiber. Such characteristic partially accounts for the instability of the nodal diastolic potential. The permeabilities to sodium and potassium change in opposite directions, reminding the behavior of a bistable multivibrator.

Partially summarizing: There are several types of cardiac fibers, showing histological and electrophysiological differences. In all likelihood, there are also differences in ionic concentrations. There might be differences too in their permeability rates of change. The four most significant ionic channels activated during the course of an action potential are, a rapid Na^+ influx during phase 0, slow Ca^{++} and Na^+ influx during the plateau phase 2, and slow K^+ efflux during repolarization.

How did all this get started? Luigi Galvani, an Italian obstetrician, conducted in the city of Bologna a long series of experiments at the end of the 18th Century. They can be divided essentially in three kinds: With the first, he rediscovered the old phenomenon of electrostatic induction. Thus, there was no news. With the second, and by sheer serendipity, he found that the junction of a metal and an electrolytic solution could generate an electric potential to stimulate a frog's muscle, but he never realized nor understood that and mistakenly thought he had found animal electricity. Unfortunately, he published a paper that was destroyed by Alessandro Volta, who developed further the idea and invented the elec-

tric pile, the first DC generator. In the midst of humiliation, poverty and family sorrow (for his wife had died), Galvani came up with a third experiment in frogs finally demonstrating the existence of animal electricity (now known as *injury potential*): When one portion of a sciatic nerve (still innervating its muscle, let us call it muscle 2) touched the injured part of another muscle (called, say, muscle 1) while a distant portion of the same nerve was held in contact with a normal uninjured portion of muscle 1, muscle 2 contracted. Obviously, the injury potential (negative inside and positive on the surface) acted as a stimulus during the "on" action. It was a manifestation of the membrane potential, recorded with microelectrodes 150 years later (in 1948), and the birth of modern electrophysiology. On the other hand, the strong controversy he maintained with Volta gave rise to the electric pile and, with it, to the beginning of current electricity (Hoff, 1936; Geddes and Hoff, 1971). Quite a couple of outstanding accomplishments.

2.2.2.2. Origin of the heart beat: the pacemaker

In herptiles (amphibia and reptilia), the heart has three stages and four chambers: the sinus venosus SV, two atria and one ventricle with the left and right side partially connected through the opening of an incomplete septum. Yes, there is some mixture of blood, but not too much due to the particular helicoidal configuration of this wall. Each stage has an electrical activity and the associated mechanical contraction. Hence, blood moves from the SV to the atria to the ventricles.

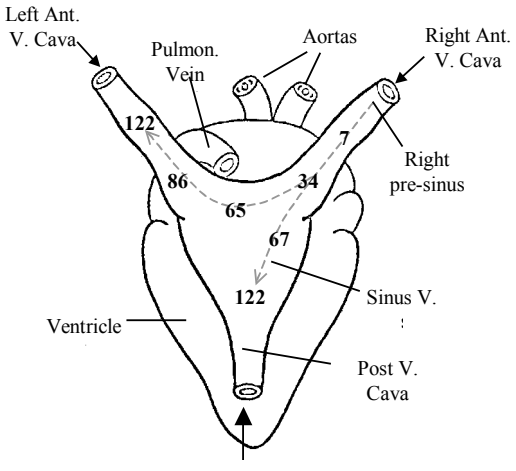


Figure 2.26. FROG's HEART. Posterior view. The sinus venosus SV has a triangular shape. It receives blood from the posterior vena cava (which brings blood from the lower body) and from the left and right venae cavae (returning blood from the upper portion of the body). The pacemaker is usually placed either at the right or at the left side of the upper SV, at the presinus area. From there, the electric impulse propagates to the other side and to the lower SV, to the atria atria and to the ventricles. Numbers show propagation times in ms measured from the origin. The frequency was 45/min.

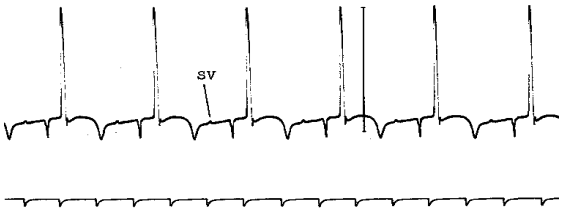


Figure 2.27. SINUS VENOSUS ELECTRICAL COMPLEX. Surface electrocardiogram recorded from an anesthetized snake (*Elaphe obsoleta*, the commonly called chicken snake in Texas) showing the SV activity followed by the atrial P wave, the ventricular QRS complex and an inverted T wave. Time marks are 1 s apart. The vertical bar by the third beat corresponds to a calibration of 1 mV. Record obtained by the author.

The sinus venosus SV has a triangular shape (Figure 2.26); it receives blood from the posterior vena cava (which brings it from the lower body circulatory system) and from the left and right venae cavae (returning it from the upper portion of the body). The *pacemaker* is usually placed either at the right or at the left side of the upper SV, at the so-called pre-sinus area. It is a natural oscillator (marking the heart pace or rhythm), with an unstable resting potential (called *diastolic potential*). From there, the electric impulse propagates to the other side and to the lower SV, to the atria and to the ventricles.

The snake is a particular good model to record the electrical and mechanical activities of its heart (Valentinuzzi, 1969). Figure 2.27 displays

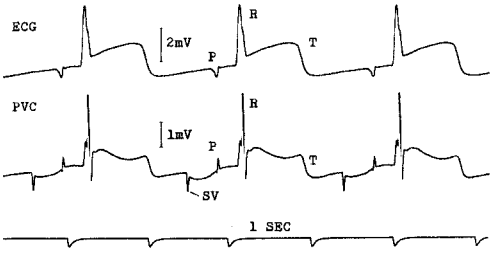


Figure 2.28. SINUS VENOSUS ELECTRICAL COMPLEX AND OTHER COMPONENTS. From an anesthetized snake (*Elaphe obsoleta*). The upper channel is a surface ECG while the lower channel was obtained with a bipolar catheter in the sinus venosus introduced via the posterior vena cava. The sequence is clear: SV, P, R and T. The elevation of the S-T segment in both channels indicates some ventricular damage. After Valentinuzzi and Hoff (1970).

a surface electrocardiogram of one of these specimens where the sequence mentioned above is easily seen: a small upward spike SV (sinus venosus electrical activity), a negative pulse signaling the atrial activity called P, a large positive ventricular spike R and the last (negative in this case) smoother T wave (Valentinuzzi, Hoff & Geddes, 1969). The meaning of all these components is explained in detail further down in the text.

Figure 2.28 demonstrates the same electrical cardiac components, also in the snake, this time recording also with needle skin electrodes and with a catheter inserted in the sinus venosus (Valentinuzzi & Hoff, 1970).

Figure 2.29 shows clearly the full electromechanical correlation of events. The animal was an anesthetized snake. The upper channel is the surface electrocardiogram while the second and third channels were recorded with a catheter introduced in the sinus venosus carrying electrodes and tubing connected to a sensitive pressure transducer. The amplitude of the SV signal in channel 2 is very small and it could be improved by moving the catheter, but at the expense of losing signal from the sinus pressure, thus, a compromise had to be reached. SV (the pacemaker) was followed by the sinus contraction (channel 3) reaching ap-

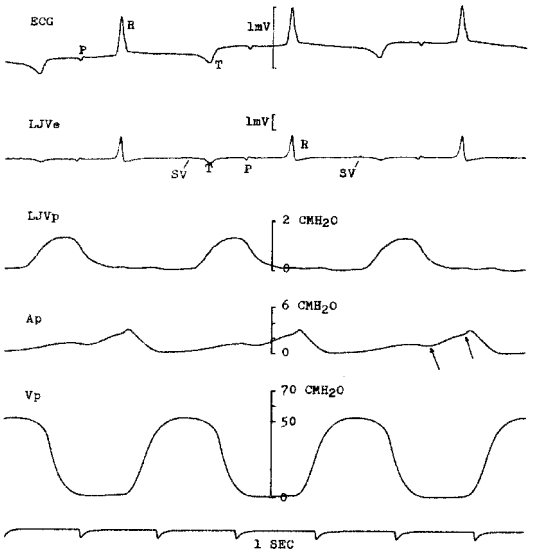


Figure 2.29. FULL ELECTROMECHANICAL CORRELATION OF CARDIAC EVENTS. The animal was an anesthetized snake *Constrictor constrictor*, a Brazilian boa. After Valentinuzzi (1969). See text for explanation.

proximately a maximum of 1.5 cmH_2O. Atrial activity P is seen in the first two channels just preceding atrial contraction, which is shown in channel 4 (atrial pressure recorded with a tubing inserted through the atrial wall). The lower channel is intraventricular pressure obtained with a catheter and another sensor. There is a clear build up right after the R wave. The first arrow in channel A_p indicates the beginning of atrial contraction. The second arrow probably corresponds to the closure of the A-V valve, while the peak of atrial pressure reflects the isometric ventricular contraction through the bulging back of the A-V valve. The wave before atrial contraction is an indication of the sinus venosus contraction (Valentinuzzi, 1969; Valentinuzzi & Hoff, 1970). Briefly stated, the electrical depolarization of each stage is the mandatory triggering signal for the mechanical contraction, which, in the end, builds the pressure up and propels blood.

2.2.2.3. Spread of excitation: the conduction system

As in any excitable tissue, the action potential (often referred to as the cardiac impulse) propagates to all the heart: from the pacemaker site throughout the whole SV, from the latter to the right atrium and just slightly later to the left atrium, it traverses the atrioventricular junction to proceed to the ventricular base, ventricular apex and all the ventricular mass. In each stage, we underline, once the whole mass is fully depolarized, the mechanical contraction takes place. Frogs, turtles and snakes, because of their low frequencies (between 20 and 50 per minute, depending on the temperature), permit an easy visual inspection in experimental open chest preparations so that the sequential sinus, atrial and ventricular contractions can be followed when, by picking up the ventricle with a pair of tweezers, the heart is lifted to look at its posterior face. The SV is dark blue, the atria tend to be dark bluish-red and the ventricle has a lighter red color (well, perception of colors may vary from observer to observer). Filling and emptying of the chambers change the colors, too.
Crocodiles and caimans are the highest herptiles in the zoological scale. The last ones to have sinus venosus and the first with a full septum and, hence, separated right and left ventricular chambers. Birds follow them and, as in mammals including man, the sinus venosus disappears as a contractile chamber. However, a residual piece of specialized tissue remains only showing spontaneous electrical activity. This is the *cardiac pacemaker* located at the *sinoatrial node* (or *SA node*), in mammals usually found at the junction of the superior vena cava with the right atrium. The electrical impulse propagates to a second node — the *atrioventricu-*

lar node (or *AV node*) — via three internodal tracts (anterior, middle and posterior) proceeding, thereafter, through the conduction system formed by the bundle of His, its right branch, the common left branch, the left posterior and right anterior branches, the Purkinje fibers and, finally to reach the regular ventricular fibers which make up the main myocardial mass. The conduction system is a transmission line showing very little mechanical activity, almost none. Its main function is to provide a fast and smooth path for the depolarization wave to reach the contracting fibers.

The atrioventricular (AV) node lies on the right side of the partition that divides the atria, near the bottom of the right atrium. It is somewhat similar to the SA node because, under certain conditions, it may produce spontaneous electrical activity. We should add that any cardiac fiber is potentially able to develop spontaneous activity.

The velocity of propagation of the cardiac action potential varies enormously all throughout the heart, being as slow as 0.02 m/s at the AV node and as high as 4 to 5 m/s at fibers of the Purkinje type or of the so called false tendon. Definitely, the AV node small region introduces a significant delay in the overall conduction process that has important clinical consequences when it reaches certain levels.

Thus, the heart displays four important properties: *excitability* (ability to change its electrical condition), *automaticity* (ability to generate and maintain the electrical changes), *propagability* (ability to propagate the electrical change) and *contractility* (elicitited by the electrical change, ability to produce mechanical work).

Study subject: Concept of refractory period, in nerve, skeletal and cardiac muscle. Check in any textbook of physiology or search in INTERNET.

2.2.2.4. Concept of block

Figure 2.30 is a simplified schematic of the cardiac conduction system, as described above. Similar to the traffic in a road, the depolarization wave can be detained at any place in its transit through the system, as for example at the bundle of His (see mark) so producing a bundle block, common and worrisome condition. The block can also occur somewhat further down, say at the common left bundle branch (CLBB), which is also rather frequent, or at either the posterior or the anterior left branches, the two latter usually referred to as *hemiblocks* (see two other marks in Figure 2.30). Blocks may take place between the two nodes and, most likely, localized at the very AV node. However, they can be seen within

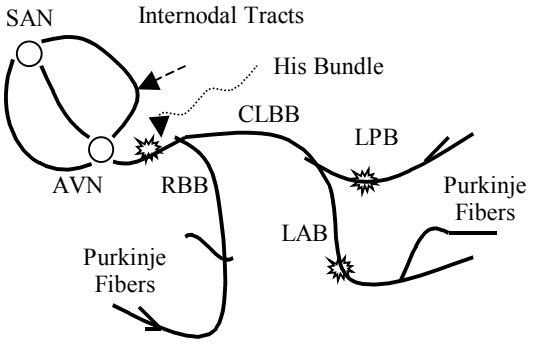


Figure 2.30. CONDUCTION SYSTEM AND CONCEPT OF BLOCK. This is a schematic of the conduction system. SAN and AVN represent, respectively, the sino atrial and the atrioventricular nodes. Cardiac depolarization spreads throughout the atrial mass to elicit its contraction but it has rapid highways, the three interatrial or internodal tracts, to reach AVN from SAN. The bundle of His stems at AVN, is short, and soon divides into the right bundle branch (RBB) and the common left bundle branch (CLBB) that, in turn, subdivides into the left posterior (LPB) and left anterior branches (LAB). Both bundle branches give off the Purkinje terminal fibers that penetrate finally into the ventricular mass, the real recipient of the electrical signal. Depolarization can be blocked at any place in the system, as for example the marked locations.

the myocardial mass, too (many times under the general name of intraventricular blocks). As a rule, the higher the level of the block, going from peripheral ventricular sites to the AV node, the more serious the clinical condition generated by the block.

Study subjects: Find out who His and Purkinje were. Find out what Stokes–Adams disease is. The latter was probably the strongest motivation, circa 1958–1960, for the development of an electronic implantable device, now a commonplace that has saved and prolonged millions of lives. What device is it?

2.2.2.5. The surface electrocardiogram (or ECG)

When electrodes, connected to an adequate amplifier and the latter to a recorder, are hooked to any place on the skin, a typically well-defined signal called electrocardiogram (ECG) is registered. Willem Einthoven, in The Netherlands, was the first (in 1903) to obtain a good, reliable and reproducible recording using, after his invention, the famous string galvanometer. He founded modern electrocardiography getting for his contributions the Nobel Prize in 1924. In fact, there were antecedents to this

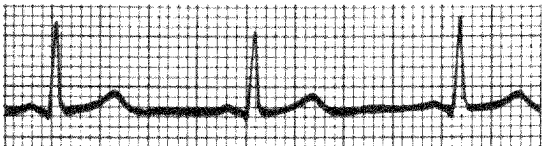


Figure 2.31. THE NORMAL SURFACE ECG. Human surface record at the standardized recording speed of 25 mm/s with the usual amplitude calibration of 10 mm/mV. Each little square represents 1mmx1mm. The three beats show a prominent spike, the QRS complex, followed by the smaller and smoother T-wave and preceded by the very small atrial P-wave. In this case, the rate was 68/min.

outstanding and longstanding apparatus, for it took decades to be displaced by the electronic electrocardiograph (after 1950). The capillary electrometer, for example, was one of its ingenious but not well-suited ancestors. It made up a nice piece of history in the development of cardiac electrical activity recording (Geddes and Hoff, 1961).

Figure 2.31 displays a human ECG recorded with surface electrodes. The first beat shows only a spike, the QRS complex, followed by the smaller and smoother T-wave. The other two beats are complete, with the short and also curved P-wave preceding the QRS complex. R is the tallest and steepest spike, Q is just a very small negative spike at the beginning of R and practically non-existent in this record, and S is the small negative projection at the end of R. Einthoven named these waves with these letters and they were accepted ever after.

Obviously, the origin of this triphasic signal (three components or waves, no relationship whatsoever with the triphasic system of electrical engineers) must be searched in the myocytes themselves. However, each cardiac cell produces an action potential with a waveform completely different than that shown in Figure 2.31. How can this be explained?

– *Relationship to the action potential*

A differential signal from two monophasic action potentials

One first and simple explanation makes use of the subtraction of two cardiac action potentials, slightly shifted in time, as if they were recorded with microelectrodes. Take for example any of the recordings shown in Figure 2.23 (say, the ventricular signal), redraw it on a piece of paper, redraw it once more below it but shifted a little to the right (as if delayed); thereafter, subtract point by point the latter from the former

choosing an appropriate sampling interval, and a waveform resembling an ECG will be obtained. If the width of one of the action potentials is shortened or lengthened, the last ECG component (or T-wave) can be made either positive or negative. This would be the combination obtained from just two fibers (Geddes, 1972).

The two electrodes of a surface ECG produce the differential signal between one electrode with respect to the other. One of the electrodes (placed at point A) can be thought of as recording a compound monophasic signal similar in shape to the cardiac action potential but the result of many fibers, say *m*, while the other electrode (placed at point B) would be recording the result of *n* cells and delayed due to the propagation time between the two points. Hence, we can write,

$$ECG = \sum_{j=1}^{m} V_{Bj} - \sum_{k=1}^{r} V_{Ak} \tag{2.58}$$

where the first summation reflects the result from one of the electrodes, say B, and the second summation represents the compound monophasic signal of the second electrode, A, while *m* and *r* are the number of fibers "seen" by the electrodes, respectively. Still, there are other fibers whose action potentials are not detected by any electrode due to unfavorable relative locations.

The student is urged to actually do the graphic exercise described above in order to get the feeling of how both, delay and width of two monophasic signals determine the shape of the resulting wave. When there is no delay and widths are exactly the same, the result is zero or a flat line. It must be underlined that each monophasic component is not a "pure" action potential of full fiber-like amplitude but the composition of many, *m* and *r*, respectively, slightly different signals (in amplitude, width and may be in delay too). This rationale is applicable to the sinus venosus in lower animals as it is also to the atrial and ventricular electrical activities. We learn immediately that always the upstroke (or depolarization) of the monophasic wave is associated with a short spike in the surface recording, while the return (or repolarization) of the former roughly coincides in time with a smaller and smoother wave of the latter.

A biphasic signal from an excitation wave

A single excitable fiber in resting state is electrically negative inside and positive outside (Figure 2.32). Two external electrodes located at points A and B, respectively, are connected to a recording system. Always, the

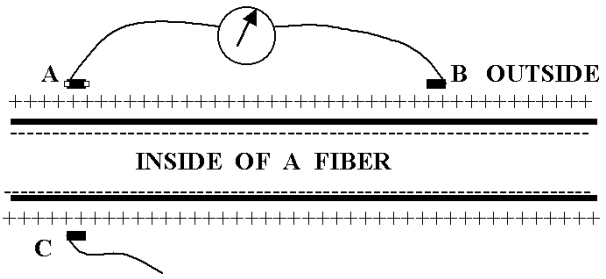


Figure 2.32. GENESIS OF THE BIPHASIC ACTION POTENTIAL. A single excitable fiber in resting state is electrically negative inside and positive outside. Two external electrodes located at points A and B, respectively, are connected to a measuring system. Always, the result is the difference of potential between A and B. Since at rest the potential of A is equal to the potential of B, the difference is zero and the detector will not show a displacement of its indicator. Point C indicates another electrode position which, when paired with A, will always yield a zero signal, irrespective of the fiber resting or excitatory state.

result is the difference of potential between A and B. Since at rest the potential of A is equal to the potential of B, the difference is zero and the detector will not show a displacement of its indicator (needle, oscilloscope beam or whatever). When a depolarizing front moves from the left to the right, as it passes under electrode A, there is a reversal of polarity and, say, that an upward difference is recorded by the recording system. If the refractory period of the tissue is short and the distance AB is long enough, the excited front will soon be between the two electrodes so that again the difference of potential will be zero. Hence, the display will show a positive spike signaling full passage of depolarization under A. Thereafter, the front reaches electrode B and the system will detect a potential difference similar to the first one but downward, because now B is negative to A (and before A was negative to B) while the arrangement of the system was kept constant. As the front proceeds to the right, the potential difference between A and B returns back to zero for the fiber is again at rest. Our record shows then two spikes: the first is positive and marks the passage of excitation under A while the second is negative signaling passage under B.

The refractory period of cardiac tissue is very long (easily, 200 to 400 ms). Thus, and repeating the reasoning, when the excitation front reaches A, an upward deflection will be recorded, as in the previous case, but as

the front proceeds to the right, since the refractory period is long, it will touch simultaneously *both* electrodes which will see only the negative side and, thus, will show a zero difference. Depolarization stretches from A to B so that you can draw a shaded area between both points in order to visualize the concept. Thus, *the recorder will have written a positive spike representing depolarization.* Repolarization is now to be expected, and there are two possibilities depending on the metabolic conditions of the tissue: either repolarization starts at A or at B. If it starts at A (first region to depolarize is the first region to repolarize), a second negative spike will be recorded. If it starts at B (first region to depolarize is the last to repolarize), a second positive spike will show up. The student is advised to work out this reasoning by drawing sequentially the different situations and thinking carefully about the resulting polarities and differences of potential. A more detailed step-by-step description of these events can be found in the traditional textbook of Leslie A. Geddes (1972).

Summarizing: When recording the electrical activity of cardiac tissue with a pair of surface electrodes, two waves will be displayed, the first one (always positive) indicating depolarization, the second (either positive or negative) signaling repolarization.

The last important concept that we can learn from Figure 2.32 is when the position of the electrodes is changed so that one is placed at A and the other at C. In other words, the electrodes axis is perpendicular to the fiber axis, either in the resting or in the excited state, the potential difference recorded by the system will always be zero. As corollary, when the electrodes axis, instead, is parallel to the propagation pathway, the signal is a maximum in amplitude.

The student is advised to carefully read the previous paragraphs (the two explanations relating cardiac action potentials with surface ECG) and think over their contents. They have, indeed, significant practical consequences in the understanding of the electrical heart activity.

The ECG as the second derivative of the membrane potential

Figure 2.33 depicts a simplified electric model of the fiber schematic presented in Figure 2.32. It accounts for the extracellular fluid resistance r_2, the intracellular fluid resistance r_1, and the membrane impedance Z_m, all expressed per unit length. The junction (or node, in the electric circuit meaning), with a negative sign, is at potential V_2 and represents the excited or depolarized region. The difference between V_2 and V_1 is the membrane potential at any time, V_m. Longitudinal currents i_1 and i_2,

through the intracellular and extracellular fluid, respectively, are driven by the small potential gradients generated by the excited region, so that,

$$r_1 i_1 = \partial V_1 / \partial x \tag{2.59}$$

$$r_2 i_2 = \partial V_2 / \partial x \tag{2.60}$$

where x stands for distance along the fiber axis. In fact and more exactly, the three dimensional space should be considered, but for the sake of mathematical simplicity and without losing concept, we will stick to the single dimension x.

Taken the derivative on both sides of the above equations leads to,

$$\begin{aligned} r_1(\partial i_1 / \partial x) &= \partial^2 V_1 / \partial x^2 \\ r_2(\partial i_2 / \partial x) &= \partial^2 V_2 / \partial x^2 \end{aligned} \tag{2.61}$$

Inspection of the electric analog easily shows that the current, i_m, across the membrane appears because of gradients of current along the outside and the inside. In mathematical terms,

$$i_m = -\partial i_1 / \partial x = \partial i_2 / \partial x \tag{2.62}$$

The negative sign simply considers the opposite directions of the two longitudinal currents. Replacement of the latter in eqs. (2.61) yields,

$$-r_1 i_m = \partial^2 V_1 \big/ \partial x^2 \tag{2.63}$$

$$r_2 i_m = \partial^2 V_2 \big/ \partial x^2 \tag{2.64}$$

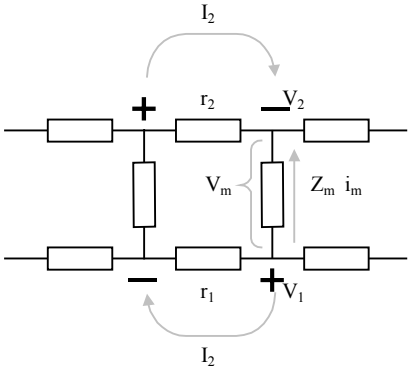


Figure 2.33. ELECTRIC ANALOG OF THE CARDIAC FIBER (see also Figure 2.32). Extracellular fluid resistance r_2, intracellular fluid resistance r_1, and membrane impedance Z_m are expressed per unit length. The outside junction with a negative sign, at potential V_2, represents the excited or depolarized region. The difference between V_2 and V_1 is the membrane V_m potential at any time. Currents i_1 and i_2 through the intracellular and extracellular fluid, respectively, are driven by the small potential gradients generated by the excited region. The lower side of the picture, where i_1, r_1, and V_1, are located, stands for the inner part of the fiber.

Subtracting the above eq. (2.64) from its mate (2.63) produces,

$$(r_1 + r_2) i_m = \partial^2 V_m / \partial x^2 \tag{2.65}$$

It should be recalled that the membrane potential V_m at any instant, either at rest or during excitation, is the difference between V_2 and V_1, as defined above.

This result is *per se* quite interesting, for it relates in a direct and simple way the current through the membrane with its difference of potential across. As parameters, only the two extracellular and intracellular fluid resistances are involved, with low numerical values relative to the membrane impedance. The membrane potential intervenes with its second derivative with respect to space, but it also changes with time when the action potential is triggered and, very important and significant, is its propagation along the fiber. Thus, we ask assistance from electromagnetic theory borrowing its wave equation,

$$V_m = f\left[(t - x/c)\right] \tag{2.66}$$

In other words, the action potential is looked upon as a mathematically unknown function of time and space taken place at the excitable tissue and moving at speed *c*. The student should perhaps review the subject to remind that the general solution of the wave eq. (2.66) is well-known and usually written as,

$$\partial^2 V_m / \partial x^2 = \left(1/c^2\right)\left(\partial^2 V_m / \partial t^2\right) = (r_1 + r_2) i_m \tag{2.67}$$

Expression (2.67) is the differential form of the wave equation. Its last right portion only applies to the particular excitable fiber herein discussed. The mathematical description of the membrane potential change is not explicitly known, but it can be experimentally recorded and there are numerical approximations of different types. In fact, the latter still is a matter of study undertaken by many concerned electrophysiologists.

If the two right hand parts of eq. (2.67) are now considered, solving for the membrane current leads to,

$$i_m = \frac{1}{c^2 (r_1 + r_2)} \frac{\partial^2 V_m}{\partial t^2} = K \frac{\partial^2 V_m}{\partial t^2} \tag{2.68}$$

and this reads as the membrane current being directly proportional to the second time derivative of the membrane potential, i.e., the action potential. A rather attractive concept because we can proceed one step further by saying that the surface ECG could be considered as proportional to the i_m and, hence,

$$ECG = k \times i_m = K(V_m)'' \tag{2.69}$$

In the latter, the double prime over V_m stands for the second derivative with respect to time. This theoretical result has been experimentally tested by recording monophasic cardiac action potentials, electronically deriving such signal, and comparing the latter with the recorded surface ECG. The agreement was satisfactory (Geddes, 1972).

In summary: In any cardiac chamber, depolarization of the monophasic action potential is associated with the upward spike of a surface recording and repolarization is associated either with an upward or a downward smooth deflection also of the surface record. Hence, the duration of the monophasic action potential is also measured by the time distance between the sharp and smooth spikes of the ECG.

– *Mathematical relationships of the surface ECG*

Any physiology or electrocardiology textbook describes the standard leads located on the frontal plane of an individual. Let us refer to Figure 2.34 (left), which represents the differences of potential as recorded by an ECG machine from leads I, II and III. They form a triangular circuit where the First Kirchhoff's Law can be applied, as in a loop, going from the negative to the positive or R to L, in branch I, from L to F, in branch III, to equate the potential difference found between F and R, that is, the

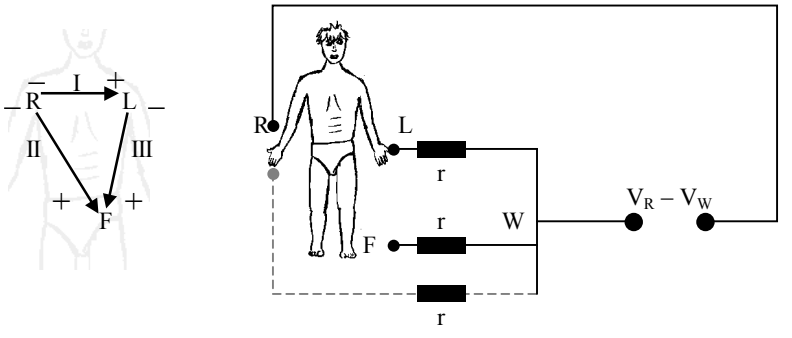


Figure 2.34. EINTHOVEN'S ELECTROCARDIOGRAPHIC LEADS. Leads I, II and III, respectively, measure the differences of potential between the left and right arm, the left foot and right arm, and the left foot and left arm, forming a triangle. Polarities are conventional, but the student should understand them better if thinking of cardiac depolarization and repolarization in terms of an equivalent dipole. See text. Isn't he a nice guy? It's Federico's drawing (the ACKNOWLEDGMENTS section will tell you who Federico is).

algebraic sum of all the electrical potential differences must be zero, or

$$V_I + V_{III} = V_{II} \tag{2.70}$$

where each term V stands, respectively, for the potential difference between LR, FL and FR. Thus, electrical engineers are using an old well-established knowledge lending a hand to the cardiologists. Equation (2.70) is called by cardiologists Einthoven's First Law. However, there is more, because the Second Kirchhof's Law can be applied to the node W (Figure 2.34, right), where the sum of the currents should equal zero, or

$$\frac{V_R - V_W}{r} + \frac{V_F - V_W}{r} + \frac{V_L - V_W}{r} = 0 \tag{2.71}$$

Since the r cancels out, working algebraically a little bit with the equation leads to,

$$V_R + V_F + V_L = 3V_W \tag{2.72}$$

where each V represents the electrical potential (not difference) of the right arm, left arm and left foot, respectively. The last potential is called Wilson's terminal and works as a virtual reference for the arrangement. Resistors r act as equalizers and may take practical values between 5 to 10 kΩ, at the most. Three cardiac electrical signals can be obtained by using as outputs the differences ($V_R - V_W$), ($V_L - V_W$) and ($V_F - V_W$). However, Goldberger, in 1942, introduced a modification to get bigger amplitudes: one of the resistors is removed (as marked by the dashed line in Figure 2.34) and the junction of the two resistors is called now W′ (modified Wilson's terminal). When this is done, the difference between R and W′ is called aV_R, so that,

$$aV_R = V_R - V_{W'} \tag{2.73}$$

Applying Second Kirchhoff's Law to the new node and solving for $V_{W'}$, leads to,

$$\frac{V_F + V_L}{2} = V_{W'} = \frac{3V_W - V_R}{2} \tag{2.74}$$

which, after replacement in eq. (2.73) above, yields,

$$aV_R = (3/2)(V_R - V_W) \tag{2.75}$$

In eq. (2.74), the last two terms are obtained from eq. (2.73). Similarly, two other relationships can be gotten for aV_L and aV_F, respectively. The student should work them out as an exercise remembering that the two resistors should be shifted to the corresponding limbs as the left arm and

left foot are considered. The "*a*" stands for "augmented" because the potential these leads produce is 1.5 bigger than that obtained with respect to the original Wilson's terminal. Thus, they are called the *augmented leads of Goldberger*.

Tips for the interested reader: Gustav Kirchhoff (1824–1887) was a student of Gauss. He taught at Berlin and at Breslau. In 1854 he was appointed professor of physics at Heidelberg, where he collaborated with Bunsen (inventor of the common laboratory gas burner). He was a physicist who made important contributions to the theory of circuits using topology and also to elasticity. Kirchhoff's laws, announced in 1854, allow calculation of currents, voltages and resistances of electrical circuits extending the work of George Ohm. His work on black body radiation was fundamental in the development of Quantum Theory.

By joining the wires from the right arm, left arm and left foot with 5000 ohm resistors, Frank N. Wilson defined an 'indifferent electrode' later called the 'Wilson Central Terminal'. The combined lead acts as a virtual earth and is attached to the negative terminal of the ECG recorder. An electrode attached to the positive terminal then becomes 'unipolar' and can be placed anywhere on the body. Wilson defined the unipolar limb leads *VR*, *VL* and *VF* where 'V' stands for voltage (Wilson, Johnston, Macleod *et al.*, 1934). Eight years later, in 1942, Emanuel Goldberger increased the voltage of Wilson's unipolar leads by 50% and created the augmented limb leads *aVR*, *aVL* and *aVF*, as explained above (Goldberger, 1942). When added to the original Einthoven's three limb leads and the six chest leads, we arrive at the 12-lead electrocardiogram that is used today (Goldman, 1970).

Suggested exercise: Derive Einthoven's Second Law of the ECG by adding up aV_R, aV_L, and aV_F, and taking into account eq. (2.73). The sum of the three-recorded differences of potential should be zero.
There are several websites to visit where plenty of information can be found in order to expand or complement what is presented here.

– *The heart signal as a vector*

Let us consider only ventricular depolarization. If we go back to Figure 2.32 and its associated part in the text, depolarization can be also viewed as a dipole moving from left to right with the positive head to the right and the tail being negative. Disregard the positive and negative charges outside and inside (which are real but not necessary for this kind of reasoning). We will work with a fictional model that came out to be very

useful in practice. The electrodes in this case will detect an upward deflection: A sees the negative tail and B sees the positive head. A similar situation takes place during repolarization: the dipole moves to the right but backwards, because its negative tail looks to the right (when recovery starts at the left side). The opposite occurs when recovery starts at the right side: the dipole runs backwards to the left.

Actually, the whole cardiac electrical activity can be looked at as if it were a set of dipoles, one for ventricular depolarization and another for ventricular repolarization, a third for atrial depolarization and a fourth for atrial repolarization. However, each dipole, in turn, changes with time and would be the resultant of a very large number of small component dipoles, perhaps located at the fiber level.

Let us take the ventricular depolarizing dipole, for the others follow a similar rationale. It is in space, with the positive head looking downwards, slightly forward and somewhat to the left. Hence, it is a vector (with an amplitude and a direction). It produces three projections; one is on the frontal plane (already introduced), another on the horizontal plane (that sectioning the body through the thorax and dividing it in a upper and lower portion), and a third, the sagittal, sectioning the body through the middle and leaving a left and right side (two hemi-lateral bodies). Einthoven's and Goldberger's ECG leads only explore the frontal plane and from any two of them we can get the frontal projection of the cardiac vector.

Quite interesting is the fact that the values recorded by the ECG leads are scalar magnitudes so that Einthoven's laws are exclusively applied as such. They are placed in an angular axial system, which is *not orthogonal* and, therefore, does not comply to the rules of the parallelogram, as vectors do. But the cardiac dipoles, either in space or on any of the three mentioned planes, are true vectors. For years, the apparent discontinuity between cardiac vectors and cardiac electrical potentials was painful for physicists and cardiologists alike, even a source of controversy, until Burger, in 1968, solved the problem with the introduction of the concept of *lead vector.* Einthoven–Goldberger hexaxial system is a graphic algorithm that transforms two scalars into a vector or, viceversa, a vector into scalar components (Valentinuzzi, Geddes, Hoff *et al.*, 1970). Going into the details of the demonstration is beyond the scopes of this textbook and, thus, the interested student could check the given references or perhaps can wait until he/she advances more in the courses. Suffice it to say that, by and large, for any electrode location, the difference of potential

recorded is given by the scalar product of the heart vector and the lead vector. The lead vector fully describes the lead configuration.

2.2.2.6. Derangements of the cardiac rhythm

This subject is clinical, but the biomedical engineering candidate needs at least an idea, hence, we will give here a few basic and simple concepts. Whenever each P wave (atrial depolarization) is followed by a QRS (ventricular depolarization), people speak of *normal sinus rhythm*. Normal P waves show a frequency range between 60 and 100 bpm, with a variation smaller than 10%. If the rate is below 60, it is called *sinus bradycardia*, if it is higher than 100, clinicians call it *sinus tachycardia*, and if the variation surpasses 10%, the proper name is *sinus arrhythmia*.

The latter is marked in infants and children, it decreases with age until it becomes barely detectable in the very old; however, it is always present. A fully and perfectly stable cardiac frequency has a prognosis of certain death within 15 to 40 days, for it indicates irreversible lesion at the central nervous level, typically found in patients after serious cervical trauma. Some species, as sheep, dolphins, and sharks, show a marked sinus arrhythmia. It can be enhanced in experimental animals, such as dogs, with anesthetic agents that potentiate the parasympathetic system.

The PR interval measures the time taken for the depolarization wave to traverse essentially from the pacemaker site to the ventricles. Most of the delay, however, occurs at the AV node, which is a true bottleneck for conduction (see the delay in Figure 2.23). The normal range lies between 120 and 200 ms. Intervals shorter than 120 ms usually reflect the existence of Wolff–Parkinson–White syndrome, indicative of an extra electrical pathway between the atria and the ventricles. This shortcut, known as "accessory pathway", may at times encourage a rapid rhythm. When the PR interval is longer than 200 ms, it is termed *first-degree block*. Usually, it is asymptomatic, but should be medically followed if discovered in regular check ups.

The normal QRS complex lasts typically 115–125 ms, if longer, it may mean right or left bundle branch block. The amplitude should be around 1 to 1.5 mV, if larger, there is a possible ventricular hypertrophy. However, 2 mV is most of the time taken as the limit cutting value.

The QT interval is also an important parameter regularly measured by electrocardiologists. It is directly related to the duration of the monophasic action potential because it stretches from the Q wave to the end of the T wave, that is, from the beginning of depolarization to the end of repolarization (see paragraphs above). There is an inverse non-linear relation-

ship with cardiac frequency, the higher the latter, the shorter the former. Bazett, in 1918–20, gave an empirical equation of the form $QT = K/(HR)^{1/2}$, which is still very much used. Other approximations and theoretical derivations have been tried, in man and in different species as well (Valentinuzzi, 1971a). Prolongation of this interval may recognize a number of different causes such as changes in calcium concentration.

The ST segment is a long fraction of the QT interval. In the normal ECG, it has zero amplitude and coincides with the base line. Elevation or depression has profound clinical meanings, such as myocardial ischemia or infarction. Its detailed interpretation has become one of the hottest and always actual subjects in cardiology.

Very generally speaking, arrhythmias can be classified in two large groups, rhythmic and arrhythmic ("a" means lacking something, that is lack of rhythm).

– *Rhythmic arrhythmias*

As the name implies, they are derangements that maintain periodicity although it is not the physiologically required periodicity. Blocks are perhaps the best examples. A 2:1 atrioventricular (AV) block will show a P wave getting through and, thus, eliciting the associated QRS complex with its final T wave, followed by the blocked P that does not have a

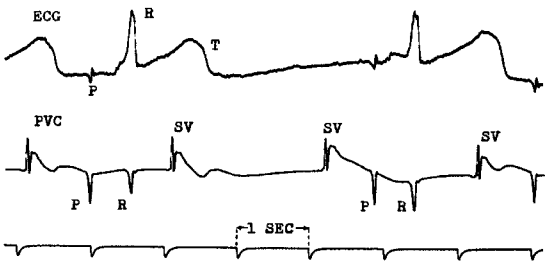


Figure 2.35. SINUS VENOSUS-ATRIAL BLOCK (OR SVA BLOCK). The animal was a snake. The upper channel was the surface ECG while the lower channel was obtained with a catheter inserted via the posterior vena cava up to the sinus venosus. The SV complex is clearly seen followed by the P wave and, thereafter, by the R wave. The second SV, however, is alone and, after a period equal to the previous one, another SV complex is seen with nothing in between. Hence, we have a 2:1 spontaneous SVA block. A third SV period shows again the normal full sequence. Observe the magnitude of times: 2 s for the SV-SV period (30/min), 800 ms the SV-P interval and about 500 ms the P-R interval. Reproduced after Valentinuzzi (1969).

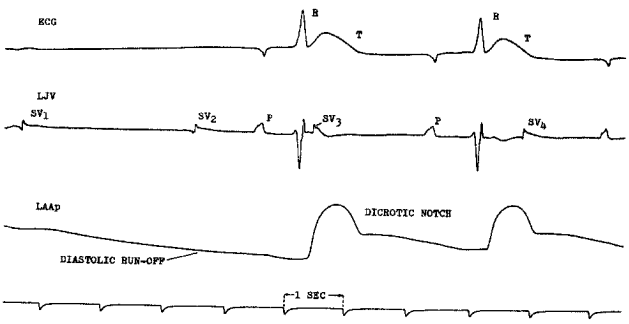


Figure 2.36. SV-A BLOCK. Records obtained from a snake. Channel 1: Surface ECG. Channel 2: Electrogram recorded with a catheter in the region of the sinus venosus. Channel 3: Aortic pressure recorded with a catheter and an external pressure transducer. After Valentinuzzi (1969).

QRS and, thereafter, another P that does get through and is followed by its QRS. The net result is a sequence P, QRS, P, block, P, QRS, P, block, P, QRS, P, block … which, notwithstanding its condition of cardiac arrhythmia and clearly pathological, is rhythmic. There are missing ventricular beats affecting the hemodynamic condition of the individual. The heart acts as a frequency divider because the ventricular rate is exactly one half the atrial rate. Such pathological rhythm may remain for a long time or may take place just for one or two beats. Figure 2.35 illustrates a similar case but at the level of the sinus venosus and atria, in a snake. The concept is exactly the same with a sequence SV, P, R, SV, block, SV, P, R, SV, block.

Figure 2.36 shows a clear spontaneous SV-A block. Between SV1 and SV2 there is a block and its immediate consequence appears as the diastolic run off of the blood pressure. The second sinus venosus activity

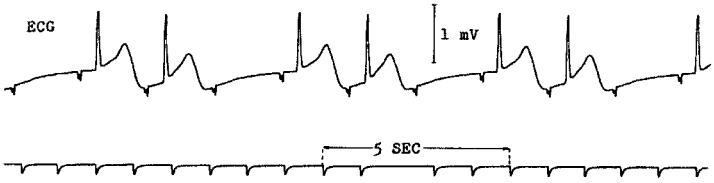


Figure 2.37. SNAKE's ECG. It displays a long series of 3:2 AV blocks showing a rather constant periodicity. Afer Valentinuzzi (1969).

SV2 was conducted to the atria (P-wave) and, thereafter, to the ventricles (R-wave). The latter triggered a contraction and the sudden increase in pressure. SV3 was a premature sinus venosus beat followed by a compensatory pause, thus, the SV2-SV4 interval is approximately twice as long as the SV1-SV2 interval.

A sequence of spontaneous 3:2 AV blocks is shown in Figure 2.37 in a single subcutaneous ECG also recorded from a snake. It is possible to find SV-A and A-V blocks, as illustrated in Figure 2.38, where there are two simultaneous channels, with a catheter, from the sinus venosus (upper channel), and right to left atrial electrogram (lower channel). Two large spikes on the right mark ventricular activity elicited by the previous SV, P series, respectively; however, there is a missing R in between. On the left side, instead, a series SV-P, SV-P, SV, clearly shows SA blockade and a missing P wave after the third SV complex. In other words, a prolonged AV block occurred with an SV-A block in between. After the S-A block, a 2:1 AV block followed. The following SV-P intervals can be measured: 0.88 s, 0.91 s, block; 0.65 s, 0.79 s, 0.82 s. Thus, there was a *progressive* SV-P first-degree block that ended up in second degree. This type of block is relatively common in man at the level of the atria and ventricles and it is called Wenckebach block. Observe the constancy of the SV period (the pacemaker), in the order of 1.6 s.

These unusual records are quite illustrative because indicate the variety

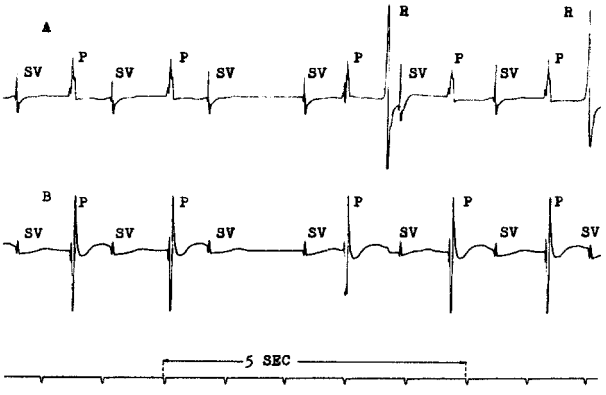


Figure 2.38. SNAKE's ELECTROGRAM. Trace A: bipolar electrogram from the sinus venosus. Trace B: right to left ventricular electrogram. After Valentinuzzi (1969).

of blockades that are possible in the cardiac system. The student is encouraged to search in the literature for other more common examples in man and mammals.

Study subject: Search for bigeminal and trigeminal rhythms, which are to be concerned about when present, and for electrical alternans, too.

– *Arrhythmic arrhythmias*

Periodicity is lost. Fibrillation is the classical and more dramatic example. A high cardiac frequency (tachycardia) may lead into it, which is some kind of chaotic activity (Figure 2.39). The coordinated contraction of all myocardial fibers is replaced by a disorganized electrical and mechanical activity. All three cardiac stages can suffer this arrhythmia. If

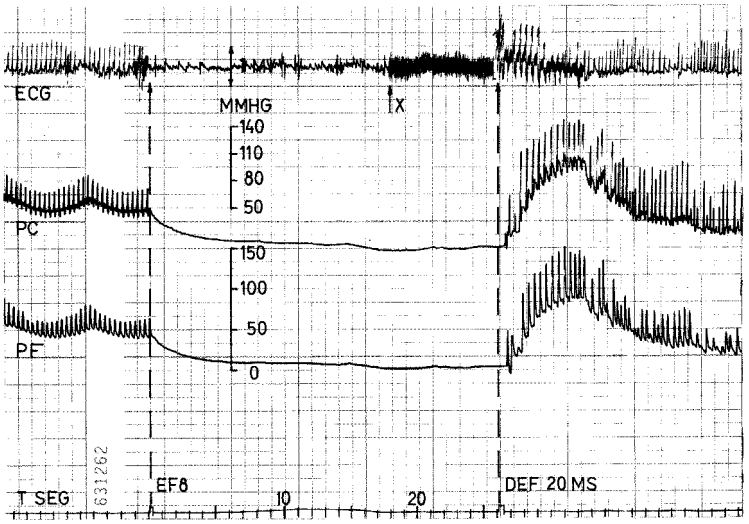


Figure 2.39. FIBRILLATION–DEFIBRILLATION. An episode of ventricular fibrillation triggered by electrical stimulation at time EF8. An electrical shock of 20 ms was delivered about 21 s later to defibrillate. The animal was a dog and the defibrillating electrodes were two paddles laterally and directly placed to the heart. The chest was open. Channel 1 is ECG, channel 2 shows the carotid pressure and channel 3 is a record of the blood pressure at the femoral artery. The chaotic fibrillatory waves are well seen on the first trace while blood pressure fell rapidly to zero level. Point X signals the moment of insertion of the electrodes manifested by noise on the record until defibrillation took place. Records obtained at the Bioengineering Department, Universidad Nacional de Tucuman.

established in the sinus venosus or in the atria, life can proceed without much interference. In the ventricles, instead, the effects include lack of adequate pressure build up and ejection of blood. In mammals and man, death is the outcome unless an emergency treatment is instituted within 3 to 5 minutes of having started.
Arrhythmic arrhythmias include the appearance of ectopic beats (that is, beats out of the expected location in time), rhythmic pieces mixed with tachycardic runs, wandering ectopic beats, even including blockades. At times, if the extrasystole (another word for ectopic) is premature, it may not elicit a good efficient contraction and ejection is either poor or non-existent at all. In other words, the ECG is unstable even though it may show pieces of apparent rhythmicity.

2.2.2.7. Closing remarks

The electrical activity of the heart was the main theme. We started with the essentials of electrophysiology, as viewed by the Ionic Theory, proceeding thereafter to the concept of cardiac pacemaker and the spread of excitation via the conduction system. The concept of block was easily introduced, almost as a consequence of the latter system. Usually, the ECG is recorded either from surface or needle electrodes, showing a triphasic pattern not readily associable with the cardiac action potential (the actual generator). Three possible explanations were offered: as a differential signal from two monophasic action potentials, as a biphasic signal from an excitation wave, and as the second derivative of the membrane potential. In summary: In any cardiac chamber, depolarization of the monophasic action potential is associated with the upward spike of a surface recording and repolarization is associated either with an upward or a downward smooth deflection also of the surface record. Hence, the time distance between the sharp and smooth spikes of the ECG also measures the duration of the monophasic action potential.
Several simple mathematical relationships, including the first and second Einthoven's laws, were derived from basic circuit theory, to briefly introduce the concept of heart vector in space and its projections on the frontal, horizontal and sagittal planes. A bare hint is also given to the conciliation problem between the electrocardiographic differences of potential, as scalar magnitudes, and the cardiac vectors on the three planes. The relationship is found in the scalar product of two vectors, the lead

vector and the heart vector. The final subject refers to the derangements of the cardiac signal. The biomedical engineers, especially those headed to the medical environment or to cardiologic research, should get a good grasp of these concepts. We gave here a bird's eye view, pointing mostly to the big print. Arrhythmias were tentatively and grossly divided in rhythmic and arrhythmic, that is, originated in more or less regular blocks and those related to instabilities.

2.3. Respiratory System

Is Earth so unique as to be alone in the Universe, coated with its feeble oxygen film? Can you imagine the challenge for a human being to survive, say, in Mars? And man is getting closer and closer to it…

2.3.1. Introduction

The respiratory system is physically related to the cardiovascular system (Figure 2.1) through the exchanger E_2, where the pulmonary capillaries are located. On one side, the airways via the mouth and nose connect the organism to the external atmosphere. On the inner side, there is the pulmonary circulation and its blood stream. In simple words, respiration can be described as the optimal adjustment of the depth and rate of breathing to meet the metabolic demands of the body (input of O_2, output of CO_2) with minimal work of the respiratory muscles. This is valid for any mammal, including the human being.

However, the respiratory system takes care of other secondary functions: it cooperates in the maintenance of the acid–base equilibrium, it helps in the regulation of water loss (in some species, as the dog, this function is extremely important because this animal does not have sweating glands), it helps in the heat loss process (especially when the environmental temperature goes up), it collaborates in the circulation of blood through the pulmonary vessels with its mechanical pumping action, it plays a role in emotional manifestations such as crying, sobbing and laughing, it is essential in protecting reflexes (coughing, sneezing, vomiting), and it is central to other reflexes of imprecise and yet not well understood nature (yawning and hiccupping).

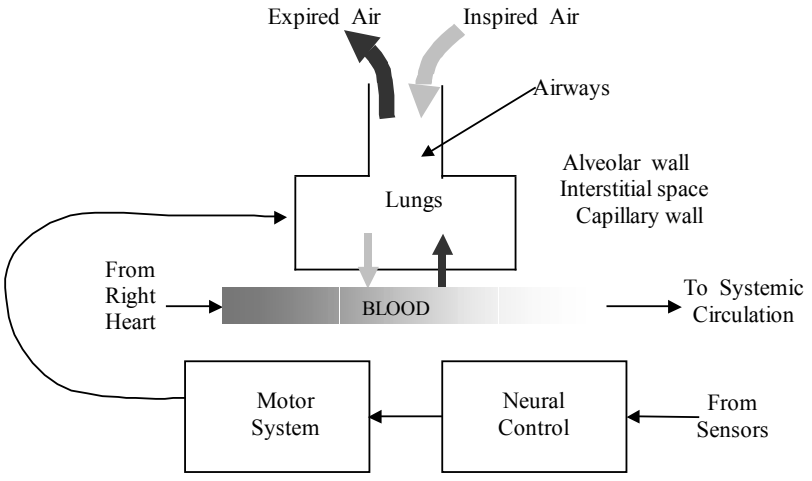


Figure 2.40. RESPIRATORY SYSTEM. The lungs, mechanically operated by the motor system (respiratory muscles), in turn driven by a nervous control system which also receives signals from specific sensors, on its external side moves air in and out, and on its internal interface exchanges oxygen and carbon dioxide with the pulmonary circulation through and exchanger formed by the alveolar membrane, the interstitial space and the capillary wall. Blood low in oxygen and high in carbon dioxide enters the pulmonary circulation (left side of figure, shaded area) and leaves it with these concentrations reversed (right side of figure, white area). See text for details.

The inspired air contains O_2 at 158 mmHg and CO_2 at 0.3 mmHg while partial pressures in the expired mixture are, respectively, 116 and 32 mmHg (Figure 2.40). The normal concentration of these two gases in alveolar air is 100 and 40, respectively, always expressed in mmHg, clearly marking an inward gradient of O_2 and an outward one for CO_2 Blood from the right heart carries oxygen at 40 mmHg and carbon dioxide at 46 mmHg while, after equilibration at the outflow heading to the left heart, the respective levels are 100 and 40 mmHg.

Problem: Based on a sea level atmospheric pressure, calculate the values given above for inspired air. What physics law are you applying?

2.3.2. Mechanics of Respiration

2.3.2.1. Basic mechanisms and principal muscles

Air inflow to the pulmonary chambers requires their expansion and return to the original volume means outflow of air. Two basic mechanisms are responsible for such events:

1. A downward movement of the diaphragm (during inspiration) and an upward movement of the same muscle (during expiration). Thus, this upwardly convex or domed muscle behaves as some kind of piston increasing and decreasing the thoracic cage volume. Approximately 70–75% of the normal respiration takes place by this mechanism. The phrenic nerves (one on each side) command, respectively, the two halves of the diaphragm (which is a skeletal muscle).
2. Rising and lowering of the frontal face of the ribs, which pivot on the vertebral column. Hence, the net effect is an increase and decrease of the anteroposterior diameter of the thorax and almost no change on the lateral diameter (Figure 2.41). The external intercostal muscles act as reins or straps to lift up the ribs. This mechanism accounts for the other 25–30% of the respiratory act.

Study subject: What other muscles act during forced respiration? Think of these muscles in paraplegic patients with spinal lesions. If the lesion affects the respiratory center, what muscle would you stimulate to obtain a relatively normal respiratory action?

Figure 2.42 illustrates the action of the intercostal muscles after section-

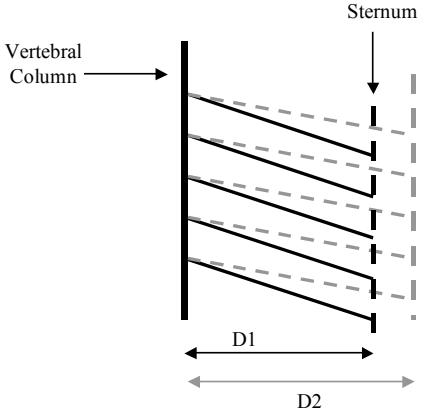


Figure 2.41. ANTEROPOSTERIOR DIAMETER CHANGE. Lateral view. The arrow indicates pulling up of the external intercostal muscles. As the ribs rotate, the diameter increases from D1 to D2. No change occurs in the perpendicular direction (lateral). Only during forced expiration the internal intercostal muscles pull down the ribs to decrease the diameter, otherwise this occurs passively.

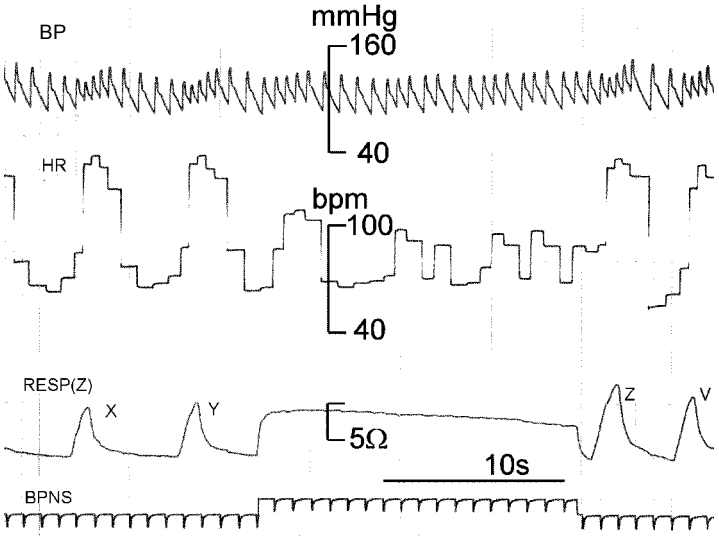


Figure 2.42. INTERCOSTAL AND DIAPHRAGMATIC COMPONENTS OF THE RESPIRATORY MOVEMENTS. Records obtained from an experimental dog whose two phrenic nerves had been sectioned. Thus, the animal was breathing only with its intercostal muscles. First channel: Carotid artery blood pressure. Second channel: Heart rate in beats/min. Third channel: Respiration detected with two transthoracic lateral electrodes connected to an impedance meter. Inspiration is shown by an increase in impedance and expiration by a decrease. Calibration of the third channel was 22.5 ohms per 500 ml of respired air. Records obtained by the author at the Department of Physiology, Baylor College of Medicine, Houston, TX, 1970.

ing of the two phrenic nerves in an experimental animal. Respiration was detected with two electrodes connected to an impedance meter. During inspiration there was an increase in impedance and expiration corresponds to a decrease (fourth channel). Breaths X and Y were produced only by the action of the intercostal muscles. Electrical stimulation of the two phrenic nerves at BPNS (during approximately 21 s) produced a good diaphragmatic contraction and consequent long inspiratory phase, as manifested by the plateau. Hence, even though the nerves had been cut, they were still responsive. Breaths Z and W were also due to the intercostal muscle contractions.

Problem: Using the calibration given in Figure 2.42 for the respiratory channel, calculate the amount of air moved by the intercostal and the diaphragmatic contractions, respectively.

2.3.2.2. Pulmonary capacities and volumes

A physiological air volume detecting apparatus (as for example, a spirometer) may show changes with time that shift from one level to another, either upward or downward, as the subject breathes following instructions. As illustrated in Figure 2.43, there are four important levels to define: Maximal Inspiratory Level (MIL), reached when the subject is asked to take air in all the way up; Normal Inspiratory Level (NIL), which is a maximum value during normal respiration and, thus, it may be either higher or lower; the resting Normal Expiratory Level (NEL), a minimum obtained also during a normal respiratory act; and the Maximum Expiratory Level (MEL), only reached when the subject voluntarily expels all the air he/she is able to.

The difference between NIL and NEL is called *tidal volume* TV (about 400 to 500 mL of air/breath in the normal adult). The gas remaining in the lung at the end tidal position NEL is called Functional Reserve Capacity (FRC). If one keeps in mind that "capacities" are comprised of one or more "volumes", the lung volume subdivisions are easy to deduce. Say, a maximal expiration measured from NEL is called Expiratory Reserve Volume (or NEL–MEL). The air remaining in the lungs at the end of this maneuver is termed Residual Volume (RV, about 1.2 liters). Therefore, the sum of the two volumes, RV and ERV, results in FRC (about 2.2 liters). Similarly, the sum of tidal volume (TV) and Inspiratory Reserve Volume (IRV) equals Inspiratory Capacity (IC, in the order of

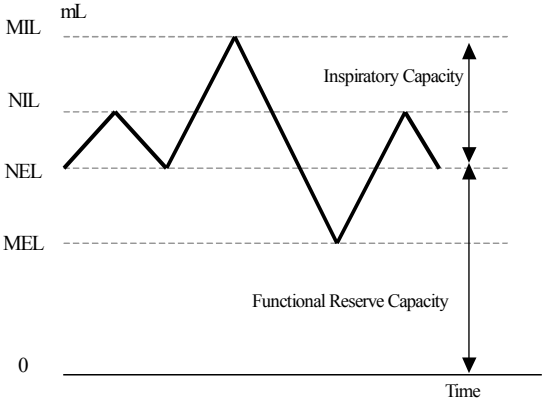


Figure 2.43. PULMONARY CAPACITIES AND VOLUMES. A physiological air volume detecting apparatus (as for example a spirometer) may show changes with time that shift from one level to another as the subject breathes following instructions.

3.8 liters), remembering that IRV (about 3.3 liters) is obtained after (MIL–NIL). In turn, the sum of IC and FRC equals Total Lung Capacity (TLC, 5 to 6 liters), i.e., from zero level up to MIL. The subject can never expel the RV, even after full expiratory exertion

Study subject: If a newborn mammal is found dead, how can it be determined whether it died before or after birth?
Exercise: Write down, as a summary, the simple equations that describe ERV, FRC, IC, IRV and TLC. They are very handy in respiratory volumetry. Find out what a spirometer is. Subject to be refreshed: Review the laws of gases, especially the general law expressed as PV=nRT, where P and V stand, respectively, for pressure and volume, and n represents the number of moles. What is R, what is T? These concepts are extremely important when the numerical experimental values need to be standardized. Hint: R is a universal constant and T is expressed in a unit honoring a famous physicist. Search in INTERNET entering with words like "gases and respiratory volumes". You will find in the WEB a concise hand out by Dr. Robert A. Furrilla (rfurilla@prtc.net), currently with the University of Puerto Rico.

2.3.2.3. Respiratory variables

– *Rate and ventilation*

Identification and clear definition of the variables is a basic requirement for any measurement. We have referred to this concept when discussing the cardiovascular system in Section 2. Respiratory rate *RR* or respiratory frequency f_r is perhaps the easiest and most salient. In any emergency situation, it should be checked right away by the attending person (physician, nurse or even a responsible passerby). Typical values in a normal adult range between 12 and 16 breaths/min, however, such rate is not constant for it shows variations. When tidal volume is multiplied by rate, a new variable is defined, *ventilation*, or

$$Q' = RR \times TV \qquad \text{(breaths/min)x(mL/breath = mL/min)} \qquad (2.76)$$

In other words, it is flow of air, conceptually similar to flow of blood or of any fluid, either gas or liquid. Accepting 15 breaths/min and 500 mL of air/breath as average normal resting values, Q' turns out to be 7.5 liters of air/min. During exercise, these values change considerably depending on the physical fitness of the individual (McArdle, Katch & Katch, 1991)

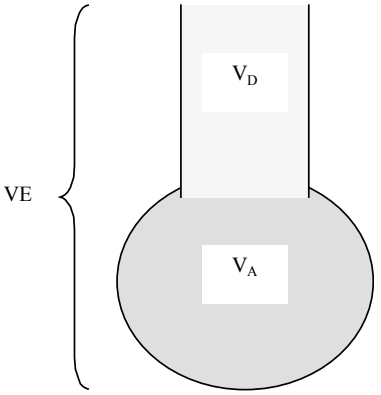


Figure 2.44. BASES FOR THE DETERMINATION OF THE DEAD SPACE. The lower shaded area represents the alveolar useful space and the vertical tubing stands for the dead volume (not participating in the exchange process). V_E is just the sum of the two former.

– *Dead space*

However, an apparent normal ventilation value does not necessarily mean that the subject is well ventilated. It is required that air reaches the alveolar space in order to actually accomplish the gaseous exchange. A fraction of the air in the respiratory compartments (airways like mouth, nose, throat, trachea, bronchi, bronchioles) doest not get in contact with the diffusion alveolar surface and, as a consequence, does not participate in the respiratory act. All this non-operative volume is called the *dead space*, V_D. As rule of thumb and for a normal adult, either the body weight directly in pounds, or the same body weight in kilograms multiplied by two, both yield a number equal to the dead space in mL. As an example, a person weighing 83 kgs (≈166 lbs) would have a V_D = 166 mL.

Thus, assuming a normal tidal volume of 500 ml, about 30% of this air is useless in the sense that it does not participate in gas exchange. Physiologic dead space includes all the previously mentioned non-respiratory parts of the anatomic dead space plus alveoli which are well-ventilated but poorly perfused (or viceversa) and are, therefore, less efficient in the gas exchange process.

Let us work with a simple model (Figure 2.44) to obtain a method of measurement of the dead space. In it, the total volume V_E is the sum of the other two volumes V_D (dead volume) and V_A (alveolar volume). Besides, we recall that the concentration of a gas in any volume multiplied

by the respective volume is the amount of that gas in it. Thus, the following mass equation for any specific gas holds true,

$$V_E C_E = V_D C_D + V_A C_A \tag{2.77}$$

where the V's represent volumes (Figure 2.44) and the C's stand for concentrations of a trace gas within the respective volumes. Solving the equation for V_D, and considering the sum of the two volumes, leads to,

$$V_D = V_E \frac{C_E - C_A}{C_D - C_A} \tag{2.78}$$

Since the atmospheric carbon dioxide content is negligible, all the CO_2 expired in a breath comes essentially from the alveoli and none from the dead space, thus, $C_D = 0$, and the equation becomes,

$$V_D = V_E \frac{C_A - C_E}{C_A} \tag{2.79}$$

because always the concentration in the expired mixture, C_E, is smaller than C_A.

Hence, by measuring the CO_2 partial pressure in the alveolar region, which is the same as that in arterial blood (by definition, at the outflow of an exchanger the concentration equals the concentration of the supplying side) and the carbon dioxide partial pressure in the expired air, one can determine the dead space. Remember that the partial pressure of a gas in a mixture is directly related to its concentration. Expression (2.79) is called Bohr's Equation.

In healthy individuals, the anatomic and physiologic dead spaces are roughly equivalent, since all areas of the lung are well perfused. However, in disease states where portions of the lung are poorly perfused, the physiologic dead space may be considerably larger than the anatomic dead space. Hence, physiologic dead space is a more clinically useful concept than is anatomic dead space. The dead space is significantly augmented in emphysema and silicosis.

Christian Bohr (1855–1911), father to the Nobel laureate physicist Niels Bohr, was one of the founders of modern respiratory physiology. In his laboratory at the University of Copenhagen, students like August and Marie Krogh bloomed, heavily contributing to the field (Schmidt–Nielsen, 1984). August Krogh got the Nobel Prize for Physiology in 1920.

– *Alveolar ventilation*

From the functional point of view, this is the important variable, for it removes the undesirable effect of the dead space. It is defined as,

$$Q_A' = RR \times (TV - V_D) \tag{2.80}$$

and so clearly establishing that only a fraction of the tidal volume (about 70%) is actually involved in the gas exchange. For 15 breaths/min, TV = 500 mL/breath and V_D = 150 mL/breath, the alveolar ventilation is only 5.25 L/min and not the 7.5 predicted by the overall ventilation.

Numerical exercise: If you breath with shallow breaths of only 200 mL and at a high rate of 30 breaths/min, do you hypo- or hyper-ventilate yourself? If you do the opposite, say at deep breaths of 600 mL/breath but at a low frequency of 10 breaths/min, do you hypo- or hyperventilate yourself? Based on this information, assume you are training a group of kids in a swimming club and you want them to dive and swim underwater for the whole pool length, what type of the two previously described respirations would you recommend them to do just before diving, say for a few breaths?

Thinking exercise: Kids like to play tricks on girls when they play in a pool or, better, at the beach, as for example swimming underwater and grabbing their legs. One smarty gets hold of a long plastic tubing with a small short curve at one end in order to place it in his mouth, dive in and breathe through it while the other pipe end just vertically sticks out of the water surface, as in a submarine. Now he can swim underneath undisturbed as long as he likes while he has the heck of a time creating a nice mess among the poor defenseless scary gals above! Do you think he can really stay down there as long as he likes? Could he go as deep as he wants?

– *Pressures*

Three absolute pressures play a role during the respiratory act: *alveolar pressure* P_A, *pressure in the pleural cavity* P_{PL}, and the *atmospheric or barometric pressure* P_B. With this in mind, three differences can be defined that are of daily use,

$$P_A - P_B = P_{tw} \text{ or } \textit{transairway pressure} \tag{2.81}$$

$$P_{PL} - P_A = P_{tp} \text{ or } \textit{transpleural pressure} \tag{2.82}$$

$$P_{PL} - P_B = P_{tt} \text{ or } \textit{transthoracic pressure} \tag{2.83}$$

From the three latter equations, it is easily seen that $P_{tt} = P_{tp} + P_{tw}$ or, in words, the *transthoracic pressure is the sum of the transpleural and the transairway pressures*. These pressures vary over the respiratory cycle

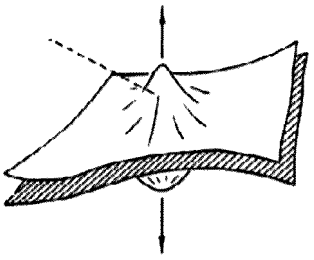


Figure 2.45. GENERATION OF A SUBATMOSPHERIC PRESSURE. Two wet rubber sheet tend to stick together when one is set over the other. A *virtual space* or *cavity* is generated.

and the most important and interesting is P_{tp}. The pressure in the pleural cavity is always lower than the atmospheric pressure, thus, P_{tp} is always negative (slightly subatmospheric). Let us describe its mechanism of generation.

Two wet rubber sheets tend to stick together when one is set over the other (Figure 2.45). If we try to separate them off by pulling out perpendicularly in opposite directions, a partial vacuum volume will be created by simple forced expansion. The inner thoracic wall and the outer pulmonary wall are in apposition, filled with a small amount of fluid so that they can easily slide over each other. Due to their intrinsic elastic properties, the former tends to recoil outwardly and the latter tends to do it inwardly. The net result is a phenomenon similar to the subatmospheric pressure effect illustrated in the picture. In fact, there is no space between the two sheet or between the lungs and the thorax, thus, it is called a *virtual space* or *cavity.*

Under resting expiratory conditions, intrapleural pressure is about 2 mmHg lower than atmospheric pressure because of the basic passive mechanism depicted in Figure 2.45. Due to the action of the inspiratory muscles, the thorax expands and intrapulmonary pressure decreases. Meanwhile, the added pulling out of the chest wall decreases intrapleural pressure even more. The net result is air flowing in with a concomitant increase in intrapulmonary pressure ending up inspiration when lung recoil counteracts chest recoil. Thereafter, intrapleural pressure goes up, but always below atmospheric, because of passive exhalation (not forced expiration) due to temporary predominance of lung recoil and everything

returns to the initial resting situation. During maximal normal inspiration, intrapleural pressure can be 6 to 7 mmHg below atmospheric.
The records of Figure 2.46 were obtained from an experimental dog. Intrapleural pressure (third channel) was recorded with a cannula inserted through the ribs into the pleural cavity and connected to a sensitive transducer. Its base line lies below the zero pressure level (which coincides with the atmospheric pressure). The dog has only one pleural space because it has no mediastinum, as man has. The decrease during inspiration is clearly associated with the increase in impedance (inspiration) shown by the fourth channel, recorded with two lateral electrodes and an impedance meter. Besides, there is also a good respiratory heart rate response manifested by the varying rate in the ECG and in the blood pressure channels, respectively. The anesthesia given to this animal favored

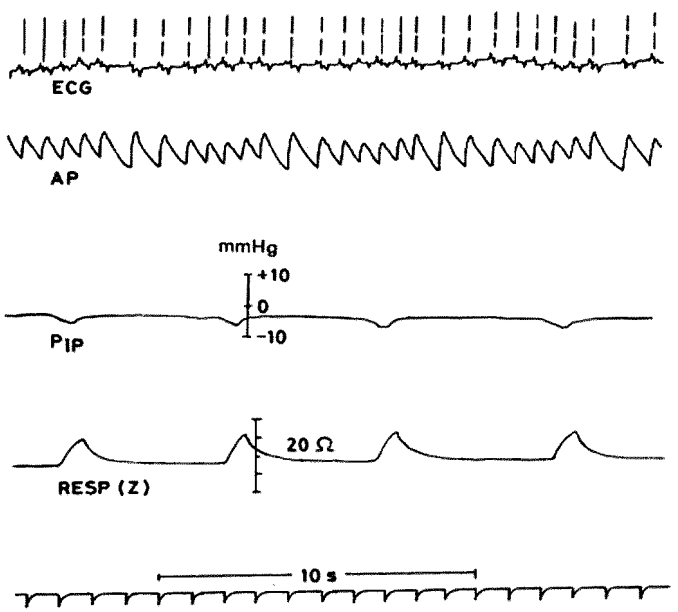


Figure 2.46. INTRAPLEURAL PRESSURE. These records were obtained from an experimental dog. Intrapleural pressure (third channel) was recorded with a cannula inserted through the ribs in the pleural cavity and connected to a sensitive transducer. Its base line lies below the zero pressure level (which coincides with the atmospheric pressure). Records obtained by the author in the Department of Physiolgy, Baylor College of Medicine, Houston, TX, 1970.

such response: inspiration associated with higher RR.

If air gets into the pleural cavity, as the case is in chest stabbing or in some car accidents, it is called a pneumothorax (*pneumo* = air). Due to the intrinsic elastic properties of the lung walls, they recoil and collapse while the thoracic wall pulls in the opposite direction. The person is unable to breathe, it is painful and, obviously, it means an emergency. If it is a dog, since it has a single pleural cavity (no mediastinum dividing the thorax in two), artificial respiration must be instituted and the air accumulated in the cavity has to be removed using well-known and not too difficult techniques. In man, instead, with a mediastinum and two pleural cavities (left and right), the situation is somewhat easier because frequently only one side gets air in it (it depends on the accident), and the person, although painfully, can continue breathing with one lung. However, it is also an emergency that requires immediate attention.

Question: Artificial respiration can be instituted in two ways, with positive pressure and with negative pressure. How would you basically implement them? Hint: In the old days of poliomyelitis (before the vaccine was developed in the early 50's), many patients (especially children) suffered respiratory paralysis. Sad but true, the negative pressure respirator used to become their enclosure. It was commonly termed the "lung machine". The positive pressure type is easier to implement.

Practice to demonstrate the negative pressure effect: Select a short (15 to 20 cm) plastic or glass bottle, about 8 to 12 cm in diameter, with a wide neck (3 to 5 cm). Remove the bottom and replace it with a piece of soft rubber, as a diaphragm. You can hold it with an elastic band if the piece of rubber is bigger than the bottle diameter. Drill a hole in a cork to be inserted later on in the bottleneck. Pass through the hole a glass or plastic tube about 3 to 5 mm in diameter and fasten a little rubber balloon on one end (lower end) so that when you blow in the tubing the balloon is inflated. Deflated balloon and tube inserted in the cork are to be placed within the bottle, the cork closing its upper mouth and showing the tubing sticking out while the balloon is in the central region inside the bottle. Thus, you can see through the transparent bottle wall the deflated balloon. There should be no leaks. When the diaphragm is pulled down with the fingertips (say, thumb and index), the balloon should inflate. Why?

2.3.2.4. Compliance

The elastic properties of the lungs and thorax play a significant role in respiratory mechanics. The concept is similar to that already introduced in the cardiovascular system, i.e., a relationship between volume change and pressure change. Since one structure (thorax) embraces the other

(lungs), they behave as a series coupling so that the resulting compliance is always smaller than the smaller of the two. In other words, the hardest bag to inflate will offer the limiting value no matter how soft or compliant the second one may be. Mathematically, such behavior is described by,

$$\frac{1}{C_{TOT}} = \frac{1}{C_{TH}} + \frac{1}{C_L} \tag{2.84}$$

where C_{TOT}, C_{TH}, and C_L stand, respectively, for the total resulting compliance (thorax and lungs), thoracic compliance and lung compliance. The lung compliance and the thoracic compliance are each in the order of 200 mL/cmH_2O, leading to about 100 mL/cmH_2O for the total or overall compliance.

Pulmonary compliance is relatively high (that means easier to inflate) because of a substance, called *surfactant* (a lipoprotein), covering the alveolar surface and secreted by specific small glands. If the amount of surfactant is low, the lungs become less compliant and difficult to inflate and the inspiratory muscular effort increases (respiratory distress syndrome). This is precisely the situation in cases of *hyaline membrane*, mostly found in premature newborn babies. Although reduced, the incidence and severity of complications of this syndrome continue to present significant morbidities.

An excess of oxygen may also lead to the condition, but it is reversible. An interesting fact is that saline solution further reduces pulmonary surface tension to nearly zero. John A. Clements (1957), a famous respiratory physiologist, demonstrated that surface tension in the lung fluid is high when the lungs are inflated and very low when they are deflated. It may be thought of as an intrinsic passive protective mechanism.

2.3.2.5. Work of respiration

The pulmonary system is essentially an expansible chamber holding at any instant of the respiratory cycle a given variable volume under a definite variable pressure, thus, as in the cardiac chambers, a pressure-volume curve can be obtained experimentally by recording simultaneously, as an *x*-*y* plot, pulmonary volume and intrapulmonary pressure (both with suitable detectors, respectively). In such case, the data will refer to the combined double bag system as outlined in the inset of Figure

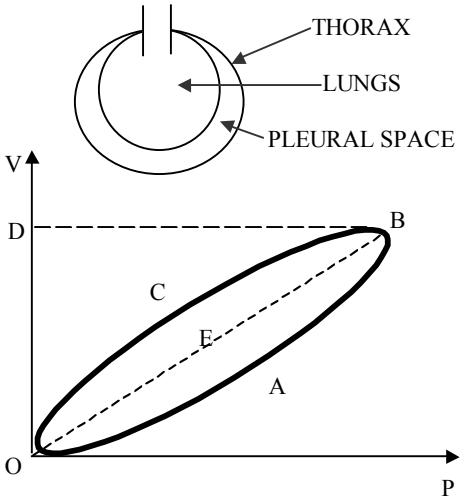


Figure 2.47. RESPIRATORY PRESSURE-VOLUME DIAGRAM. Experimentally, it can be obtained by recording simultaneously as an *x*-*y* plot pulmonary volume and intrapulmonary pressure. In such case, the data will refer to the combined double bag system as outlined in the inset. If intrapleural pressure is used instead, then the data will yield information regarding the lungs alone. The two types of data collection must be used when one wants to get pulmonary compliance and overall compliance (see text).

2.47. If intrapleural pressure is used instead, then the data will yield information regarding the lungs alone. The two types of data collection must be used when one wants to get pulmonary compliance and overall compliance. Chest compliance is, thereafter, calculated from eq. (2.84).
In the diagram there are several areas with different physiological meanings: The area contained within the boundaries marked by OABDO represents the work during inspiration. A smaller area within the former is OEBDO, which considers the work performed to stretch the lungs. The area below the straight line E and limited by the curve A, that is, OABEO, is work to overcome airway and tissue resistances. The area on the upper left side, BCODB, is work returned mostly passively during expiration. Finally, the area encompassed by the respiratory loop, or OABCO, represents the work expended during the respiratory cycle.

Exercise: Work out the units of the pressure-volume diagram of Figure 2.47, where volume is usually given in mL of air and pressure is in cmH_2O. Their product must produce units of work. If the cycle length is considered, then the power consumed in one breath can be cal-

culated. A normal value should be between 0.3 and 0.8 kgm/min. Search in the literature for the appropriate numerical values.

The work of respiration is negligible in the normal animal or person, however, in conditions like asthma, emphysema or congestive heart failure, say, in respiratory distress generally speaking, it may reach significant clinical levels that call for assist procedures.

2.3.3. Pulmonary Circulation

In many respects, this minor or lesser circulation is similar to the systemic or major or greater circulation, however, there are some important differences: it is a distensible low pressure and low vascular resistance system. Systolic pressure is about 25 mmHg and diastolic just about 10 or 12 mmHg with a mean of 15-16 mmHg. Inserting a special catheter carrying a very small pressure transducer through the subclavian vein, the pressure records shown in Figure 2.48 can be recorded as the catheter tip crosses the right atrium, right ventricle, pulmonary artery and, beyond the latter and as far as it can go, the so called pulmonary artery *wedge pressure*. The *pulmonary pressure gradient* is given by the difference

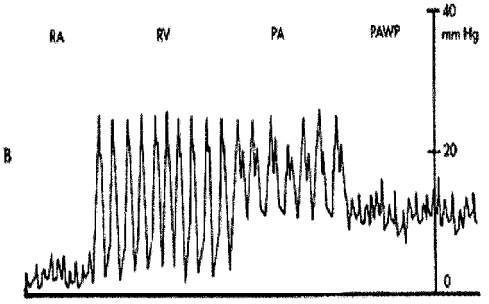


Figure 2.48. PULMONARY WEDGE PRESSURE. A special catheter was inserted via one of the subclavian veins. It held at its tip a very small pressure transducer. As the catheter was pushed in, it traversed the right atrium, right ventricle, pulmonary artery and, beyond the latter and as far as it could go, acting as a wedge in that artery. Thus, the respective blood pressure records correspond to the right atrium (low pressure of about 4–5 mmHg), right ventricle (26/1 mmHg), pulmonary artery pressure (25/10 mmHg), and wedge pressure (about 13/8 mmHg). Taken from www.mtsinai.org/pulmonary/books/physiology/chap8_2.htm. Proper recognition is acknowledged to the authors at Mount Sinai Hospital who generously make this material freely available without even giving their names. Students are invited to visit the site.

between the mean pulmonary arterial pressure (15 or 16 mmHg) and the venous pressure (about 7 mmHg) looking towards the left heart, i.e., about 8 to 9 mmHg. This gradient is responsible for the **indispensable** capillary blood flow through the lungs. It takes approximately 750 ms for a red cell to fully traverse the capillary network.

Its principal function is taking venous blood returning from all the peripheral system to the alveolar capillaries (*pulmonary perfusion*) to permit the gaseous exchange. As a consequence, a salient physiological parameter is represented by the *alveolar ventilation-perfusion ratio*, or

$$\frac{\dot{Q_A}}{F_t} = ratio \approx 0.8 \tag{2.85}$$

where the numerator stands for the alveolar ventilation (about 4.5 L of air/min) and the denominator represents cardiac output or total blood flow (about 5.5 L of blood/min). Thus, a typical numerical normal ratio is as indicated above. In other words, for each liter of blood getting through the lungs, only 0.8 liters of air actually interact with the former. Non the less, do not be fooled by an apparent normal ratio. Let us imagine a hypothetical situation in which one lung gets normal blood flow and no air at all (obstruction of the bronchus) while the other is well ventilated but there is full obstruction of its blood inflow. The ratio (2.85) would give a normal value but the subject will doubtless die.

Pulmonary vessels contain almost one liter of blood and, out of that amount, more than one half is held within the venous side. This blood volume is about 1/5 to 1/4 of the total blood volume in a normal adult, in fact much more than the lung blood requirement for only about 100 mL can be captured within the capillaries at any instant. This is why the lungs behave as a dynamic blood reservoir from which the left heart takes blood when, for a few beats, the right heart ejects less than the normal stroke volume.

Body position affects significantly the pulmonary blood content. Lying down increases it by as much as 300 or even 400 mL more than normal, while standing up brings that amount back to the general circulation. Under conditions of cardiac failure, the individual may be led to *orthopnea*, that is, the patient can breath normally only when in the vertical position

(*orthos* = vertical) because, when in bed, his/her lungs are congested making more difficult the respiratory act.

The discovery of the pulmonary circulation is an interesting and debated subject. It is commonly believed that the discovery of the pulmonary circulation had its inception in Europe, in the sixteenth century, by Servetus, Vesalius, Colombo, and Harvey. However, in view of the discovery of ancient manuscripts, the real credit for the discovery of the pulmonary circulation seems to belong to an eminent physician of the thirteenth century: Ibn Nafis. Ibn Nafis, or Abu-Alhassan Alauldin Ali Bin Abi-Hazem Al-Quarashi, was born about 1210 in Damascus. On December 17, 1288 he died at the age of 78 after an unknown illness. Downloaded from INTERNET, after *The Discovery of the Pulmonary Circulation Revisited* (*1994*), *by* Ayman O. Soubani and Faroque A. Khan, from New England Medical Center, Tufts University School of Medicine (Dr. Soubani), Boston, and the Department of Medicine (Dr. Khan), Nassau County Medical Center, SUNY at Stony Brook, East Meadow, New York. Address reprint requests and correspondence to Dr. Soubani, 208 Gerry Road, Brookline, MA 02146.

2.3.4. Gas Exchange

We reach now the core function of the system, the *gas exchange process*, during which the *diffusion capacity*, D, plays a fundamental role. It is defined, either for oxygen or for carbon dioxide, as the amount of gas G (in mL) crossing from the alveolar space to the capillary blood compartment, or viceversa, per unit of area A (in square cm), per unit time t (in min or s) and per unit of differential partial pressure ΔP (in mmHg), that is,

$$D = \frac{G}{A\,t\,\Delta P} \tag{2.86}$$

For all the alveolar exchange surface area, in the order of 70 m^2, the total oxygen diffusion capacity is 20 mL/min×mmHg. Under exercise, it increases up to 65 or even more. For carbon dioxide, instead, the diffusion capacity is much larger and extremely difficult to measure.

In other words and from the latter equation, the flow of gas per unit time is proportional to the partial pressure difference. This is usually referred to as Fick's Law and describes an entirely *passive process*. Under normal resting conditions, blood is in contact with the alveoli for about 0.75s and O_2 reaches equilibrium between capillary blood and alveolar air in about 0.25 s. If the O_2 pressure gradient decreases or the diffusion resistance increases, the movement of O_2 from alveoli to blood will slow. From

Table 2.1. Common numerical blood values

Parameter	Arterial Blood	Venous Blood
P_{02} (mmHg)	95	45
O_2 (mL/100 mLblood = V%)	19.5	15
P_{CO2} (mmHg)	40	46
CO_2 (V%)	47	52
H_b (% saturation)	97	75

Fick's law: decreasing the pressure gradient (hypoxia) will reduce the rate of oxygen transfer into blood, and increasing resistance (low diffusing capacity) will decrease the rate of oxygen uptake by the blood.

Oxygen and carbon dioxide dissolve in blood, however, hemoglobin increases by about 70 fold the transport capacity of the former and the different processes involved in the latter produce a 17 times increment in its transport capacity. Each gram of saturated hemoglobin can carry 1.3 mL of oxygen and 100 mL of blood contain about 15 g of hemoglobin. Table 2.1 below summarizes some common numerical blood values. Note for the first and third rows: Partial pressure of a gas in a liquid is that pressure in the gaseous phase which, in equilibrium with a liquid, would produce the concentration of gas molecules found in the liquid. Note for the second and fourth rows: Frequently, the number of mL of a substance found in 100 mL of blood is called volume percent (V%).

Let us explain the concept of "saturated hemoglobin" by comparing the molecule of this substance to a bus that is to transport passengers. The bus has, say, 40 seats, and if all seats are occupied, it is said to be full or, in our terms, "100% saturated with passengers". However, at times, the bus could carry only 10 or 20 passengers, in which case it would only be saturated to 25% or to 50% or simply not fully saturated. The hemoglobin molecule, in a complex biochemical dynamics, can bind to more or less oxygen and, as already stated, at maximum capacity it is said to be saturated (see Table 2.1). In fact, it never reaches full saturation although it gets pretty close (97%).

Any perturbation that affects the passage of gases affects also the individual's life. The abovementioned parameters indicate the normal expected values that are used as reference in daily practice. Factors such as air pollution, anesthetic agents and general metabolic conditions of the

subject tend to deeply influence them. Thus, detectors to implement monitoring systems are of the utmost importance.

Study subject: Search in the literature or in INTERNET for information about the oxygen-hemoglobin dissociation curve (which relates percent of O_2 in hemoglobin with partial pressure of O_2 in blood) and factors that influence the curve (such as temperature and blood pH).

2.3.5. Respiratory Control

It is a major subject not fully explained yet and, thus, opened to research. Figure 2.49 is a schematic of the neural control of the respiratory act. The circuit formed by the pneumotaxic center (located in the pons, a structure within the brain stem) and the inspiratory center (at the level of the me-

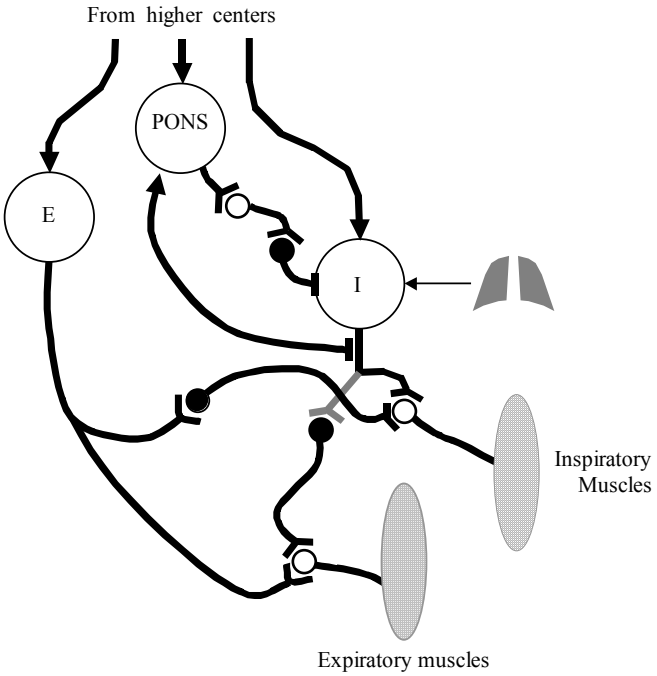


Figure 2.49. NEURAL CONTROL OF THE RESPIRATORY ACT. The circuit formed by the pneumotaxic center (located in the pons) and the inspiratory center (at the level of the medulla) constitutes an oscillator that determines the respiratory frequency. Neurons in white are excitatory and in black are inhibitory.

dulla) constitutes an oscillator that determines the respiratory frequency. The inspiratory center tends to permanently send activating signals to inspiratory muscles. Signals from the pons activate facilitatory neurons (in white) and, thereafter, inhibitory neurons (in black) to stop the inspiratory signal from I. The latter center sends signals to the motor neurons of the inspiratory muscles (such as the diaphragm and the external muscles) while inhibit the action of the expiratory agonist ones. The expiratory center E, only stimulated by voluntary action, distributes its signals (action potentials) to the expiratory muscles motor neurons inhibiting simultaneously the inspiratory agonist ones. In the lungs L, there are stretch receptors which, according to it stretching state and via afferent vagal fibers, send action potentials to the inspiratory center I. The tighter they become due to inflation, the larger the number of action potentials so that they signal the inflation level to the inspiratory center which, in turn, reduces its activity in order to decrease and even stop the inspiratory muscular action. This is called the Hering–Breuer reflex that can be thought of as a protective negative feedback mechanism.

Besides, there are central and peripheral chemoreceptors to detect oxygen, carbon dioxide and acidity levels in the blood stream from which information is sent back to the central nervous system where, after processing (called integration in the physiology jargon), a quick decision is reached to compensate for the initial deviation by changing rate and depth of respiration. For example, a carbon dioxide increase in blood is a powerful drive to triggering a deeper and more rapid respiration so that this gas is washed out faster. However, other controlling sensors are distributed in different regions of the respiratory system. Everybody is familiar with the effect of tickling inside the nose with a very fine brush (leads to sneezing) or of getting a small crumb of bread in the throat (leads to coughing or, if the stimulus is strong and deep enough, even to vomiting). Obviously, these are some kind of mechanical sensors. In this respect, all the respiratory system is extremely sensitive. Small foreign particles in the bronchi or in the bronchioles can cause major noxious effects. Temperature and emotions are also strong stimuli to respiration.

Josef Breuer (1842–1925), Austrian physiologist and pioneer of psychoanalysis. After successfully treating hysteria in one of his patients, Breuer collaborated with Sigmund Freud in writing *Studies in Hysteria*, in 1895. Earlier, working with Ewald Hering (1834–1918), Breuer

described the Hering–Breuer reflex involved in the nervous control of breathing movements (The Macmillan Encyclopedia 2001, Market House Books Ltd). See also the INTERNET to search for more information regarding this reflex and, in general, about the nervous control of respiration.

Derangements of respiration, usually involving frequency, depth and pattern, indicate many times abnormal function of its control system and may recognize metabolic and/or neural origins. There are three classic types that have been clinically described in the literature and that should be briefly presented here as examples of mostly unsuccessful efforts of the system to restore normal action: Kussmaul–Kien, Cheyne–Stokes and Biot respirations.

- *Kussmaul–Kien respiration or breathing* was described by Adolf Kussmaul and Alphonse Kien in 1874. It is also referred to as air-hunger syndrome. Characterized by rhythmic gasping and very deep type of respiration with normal or reduced frequency, associated with severe diabetic or renal acidosis or coma (Kussmaul's coma).

- *Cheyne–Stokes respiration* is an abnormal breathing pattern which commonly occurs in patients with decompensated congestive heart failure and neurologic diseases, in whom periods of tachypnea and hyperpnea alternate with periods of apnea. In the majority of these patients, the ventilatory patterns may not be recognized and, eventually, may resemble a slow modulation of the depth of respiration. The clinical features are generally dominated by the underlying disease process. Cheyne–Stokes respiration may, however, have profound effects on the cardiopulmonary system, causing oxygen desaturation, cardiac arrhythmias, and changes in mental status. Treatment of Cheyne–Stokes respiration in congestive heart failure with supplemental oxygen or nasal continuous positive airway pressure, in addition to conventional therapy, may improve the overall cardiac function and perhaps the patient's prognosis. John Cheyne was a Scottish physician (1777–1836) and William Stokes was an Irish physician (1804–1878).

- *Biot's respiration* originates mainly in cerebral lesions (e.g., hemorrhage, the so called cerebrovascular accident or CVA), when intraneal pressure goes up. Biot originally described it in patients with meningitis in 1876, in France. There is intense hyperventilation that is sometimes noisy and stertorous. Occasionally, irregular periods of apnea alternate with periods in which four or five breaths of similar depth are taken. The

pattern may resemble the on-off keying modulation (that is, a few constant depth breaths followed by a short period of apnea). Hyperventilation is frequently seen after head injury, with a consequent decreased in carbon dioxide concentration that causes reflex central nervous system (CNS) vasoconstriction with reduced cerebral perfusion leading, in turn, to a beneficial secondary decrease in intracranial pressure.
The interested student may find more details in the INTERNET entering with the appropriate words via, say, the searching machine GOOGLE.

2.3.6. Closing Remarks of Section 2.3

Summarizing, the respiratory act of mammalian animals and man is a complex result of neural, humoral and perceptual factors, which, in turn, are influenced by changes in the respiratory mechanics. Besides, there is also a close relationship with the cardiovascular system.
The lungs are particularly prone to malignant tumors, especially and unfortunately because smoking is still rather widespread among people. Air contaminants of different origin — smog in daily language — appear also as biasing factors. Prevention is always the best therapy but, even so, there are other oncogenic factors. Molecular biology is one of the big hopes of humanities in this respect. Gene-based therapies for cancer are based on the augmentation of the host's antitumor immunity or the augmentation of sensitivity to antineoplatic drugs.

– *Numerical illustrative exercises*

1. A volume V_t of 24 liters of air was collected from a subject in $t = 3$ minutes within a water filled spirometer at 25°C. The barometric pressure was 765 mmHg and the respiratory rate *RR* was 10 breaths/min. Find the average tidal volume *TV* and express the value in ATPS (*ambient temperature and pressure*, *saturated with water vapor*), BTPS (*body temperature and pressure*, *saturated with water*) and STPD (*standard temperature = 0℃ and pressure =760 mmHg*, *dry*) conditions. These are the usual standardized forms encountered in Respiratory Physiology.
Solution 1:

$$TV\,(ATPS) = \frac{V}{t \times RR} = \frac{24}{3 \times 10} = 800\,\text{mL/breath}$$

Solution 2:
The general gas equation $P_1 \times V_1/T_1 = P_2 \times V_2/T_2$ is applied, where the condition 1 stands, say, for ATPS, and the condition 2 refers to BTPS. Thus,

$$TV(BTPS) = \left(\frac{P_1 \times V_1}{T_1} \right) \times \frac{T_2}{P_2} = \frac{800 \times (765 - 23.8) \times (273 + 37)}{(273 + 25)(765 - 47)} = 859.7\,\text{mL}$$

From tables that can be found in the literature, 23.8 and 47 are, respectively, the water vapor pressures at 25 and 37°C. To the centigrade degrees, 273 must be added in order to obtain the absolute temperatures in Kelvin degrees. Expired air is always saturated with water.
Solution 3:
Applying the same equation and with the appropriate interpretation of the subindeces, it becomes,

$$TV(STPD) = \frac{800 \times (765 - 23.8) \times 273}{760 \times (273 + 25)} = 714.5\,\text{mL}$$

Notice how important is to express numerical values within the same standardized frame of reference if comparisons are to be made.

2. Obtain the ventilation of a subject breathing at a rate of 15 breaths/min with a tidal volume of 0.4 L and, thereafter, his/her alveolar ventilation assuming 70 kg of body weight. What is his/her alveolar ventilation if the respiratory frequency goes up to 30 breaths/min and the tidal volume is reduced to 0.2 L? Now, calculate the alveolar ventilation but with f_r = 12 breaths/min and TV = 0.5 L/breath. Finally, find an adequate combination of TV and f_r when the subject is at rest under the water, at a depth of 2 m, breathing through a tube of a cross-sectional area of 1 cm^2. If the maximum TV is 0.6 L, what would the maximum tube length be irrespective of RR?

$$Q' = RR \times TV = 15 \times 0.4 = 6\,\text{L/min}$$

$$Q_A' = RR \times (TV - V_D) = (0.4 - 0.14) \times 15 = 3.9\,\text{L/mir}$$

because the dead space in a 70 kg subject is roughly equal 2×70 = 140 mL of air. Application of the same equation leads to 1.8 L/min with 30 breaths/min and 0.2 L as tidal volume, and to 4.3 L/min when the RR is 12 and TV is 500 mL/breath. It clearly illustrates the better results with the second combination.

With the tube, the dead space is artificially increased, so that the equation needs a term *VT* (the tube volume) to be added to *VD*, and thus, becomes,

$$Q_A' = RR \times (TV - V_D - V_T) = 3.9 = (TV - 0.34) \times RR$$

assuming the first alveolar ventilation calculated above. If the subject breaths with *TV* = 0.5 L, he/she should increase the rate quite a bit, up to about 24 breaths/min.

The maximum theoretical tube length results for a zero alveolar ventilation, that is 0.6 – 0.14 = 0.46 L or a length of 4.6 m when the cross-sectional area of the tube is 1 cm^2, as stated above. The subject would merely move up and down a column of air without getting any oxygen. Of course, he/she would strongly pant long before that length would be reached because the carbon dioxide sensors would signal a high level in blood of that gas.

3. The FRC of a subject yielded 3,000 mL in a Physiology Laboratory. His intrapleural pressure at that FRC was – 5mmHg. After inspiring 1 L of air, intrapleural pressure decreased to – 10 mmHg. Calculate the pulmonary compliance.

Solution:

$$C = \frac{\Delta V}{\Delta P} = \frac{(3{,}000 - 4{,}000)}{-5 - (-10)} = 0.2\,\text{L/mmHg}$$

disregarding the negative sign for a compliance cannot be negative. It is simply the volume change what matters. Since 1 mmHg = 13.6 mmH_2O, the above value can also be expressed as 0.147 L/cmH_2O.

2.4. Renal System

Its evolution truly describes the passage from early fish to man, —the Philosopher— from watery confinement to freedom.

2.4.1. Introduction

This section deals with the Renal System and, following the general philosophy of the book, it is an overview predominantly presented from an engineering conception, leaving somewhat aside some biological and

mostly biochemical aspects. More specific and detailed information can be found in classic but still valid books or reviews (Pitts, 1966; Thurau, 1974, 1976; Wright and Briggs, 1979).

The objectives of the renal function are essential for life as all have a fundamental regulatory action over the different concentrations of metabolites in blood, osmotic pressure, fluid volumes, and electrolyte levels. The exchanger associating the kidneys to the cardiovascular system is a highly complex arrangement, delicate in its adjustment mechanisms and of the utmost importance for a free life, as expressed by Claude Bernard (1813–1878), French physician and scientist who, in the XIXth Century, introduced the concept of *internal environment* (the extracellular fluid or plasma plus interstitial fluid). Smith (1959), in his profound book, says: "The story of how the kidneys operate and how they came to function in the way they do is the vertebrate story, from which man is the most notable and clever actor and, besides, is the only philosopher." Briefly stated, the renal system centers its *homeostatic* activity on the total body water and, more specifically, on the extracellular fluid compartment, where all the cells are immersed, as in a small and personalized sea. *Homeostasis* (from the Greek words for "same, *homeo*" and "steady, *stasis*") refers to the dynamic regulation and readjustment processes of the physiological variables sustaining life.

Walter B. Cannon (1871–1945) devised the term *homeostasis* in 1930. His book, *The Wisdom of the Body*, published in 1932, describes how the human body maintains steady levels of temperature and other vital conditions such as the water, salt, sugar, protein, fat, calcium and oxygen contents of the blood.

Hence, and being somewhat repetitious to emphasize the concept, the kidneys regulate the chemical composition of body fluids by removing metabolic wastes and retaining the proper amounts of water, salts, and nutrients. Waste is removed from the body by the kidneys in the form of urine. The kidneys produce approximately 1 mL of urine per min (≈ 1.5 L/day), and maintain an average extracellular fluid (ECF) osmolarity of 300 mOsmoles/L.

2.4.2. Anatomical Features

All the blood volume must pass through the kidneys and it does so approximately 30 times per day or about 160 L/day (assuming a blood vol-

ume of 5 to 6 L). It enters via the two renal arteries (one per kidney) getting out through the renal veins to return to the general circulatory stream. Between inflow and outflow there is a highly complex structure of vessels and tubules. Besides, each kidney has an exit duct, the ureters, both connected to the bladder (a temporary urine reservoir) ending in the urethra, the final path for urinary excretion.

Each renal artery branches off until it reaches the level of very many and small caliber *afferent arterioles* that get into a minute capillary network called *glomerulus*, in turn contained into the capsule of Bowman, some

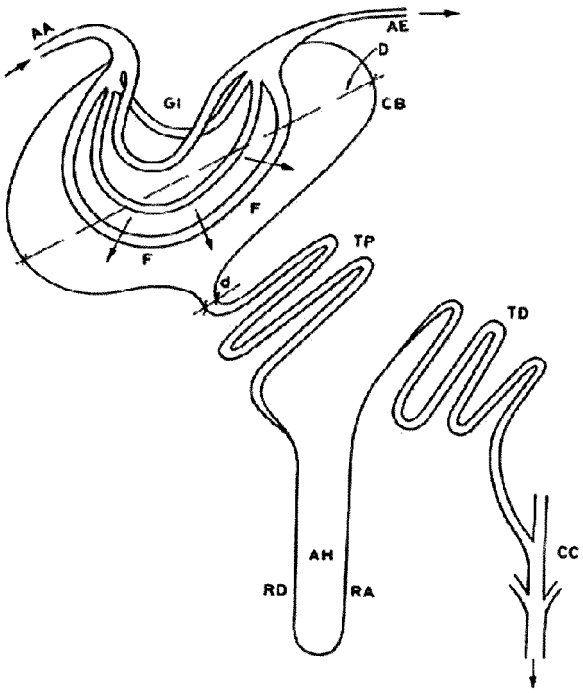


Figure 2.50. FUNCTIONAL RENAL UNIT: THE NEPHRON. AA: afferent arteriole. AE: efferent arteriole. Gl: glomerulus or glomerular capillaries. The total area of these capillary walls, in man, is in the order of 1.5 square meters. CB: capsule of Bowman, with a diameter D ≈ 200 μm. TP: proximal tubule. Its output from the capsule has a diameter d ≈ 55 μm. AH represents the loop of Henle, with its descending and ascending limbs RD and RA, respectively. As a continuation, the distal convoluted tubule TD is depicted ending at the colleting duct CC. Arrows F within the capsule show the direction of filtration.

kind of basket-like structure, from which the *efferent arteriole* comes out (Figure 2.50). The *proximal convoluted tubule* originates in this capsule continuing, thereafter, with the *loop of Henle* (descending and ascending legs), the *distal convoluted tubule* and the *collecting duct*). The glomerulus and its associated tubular-loop system constitute the renal unit called *nephron*. Each human kidney has approximately 1,000,000 nephrons while in the dog, instead, the number of nephrons is about 400,000, always per kidney. Up to a point, the kidney size depends on the number of nephrons. All collecting ducts converge to the minor, major calyces and the renal pelvis that, finally, end up in the ureter. Calyces act as some sort of funnels to direct the urine to its destination (the bladder).

Blood, after traversing the glomerular capillaries, exits this minute system via the efferent arteriole, which, in turn, branches off to form the *peritubular capillaries*. They surround and wrap the tubular system in such a way as to establish the *renal exchanger* which permits the back and forth shift of substances between blood and the intratubular fluid. These capillaries successively converge (or fan in) into venules and larger diameter vessels until reaching the renal vein to get back to the general circulation. The *renal parenchyma* is the soft and well-wet tissue found between the peritubular capillaries and the tubular system.

2.4.3. Techniques Employed

To study renal function several techniques have been applied. Their description is not a relevant subject to this text although they make a magnificent piece of ingenuity, scientific accomplishment and technological development; thus, we will only mention them so that the interested student may access to other suitable sources, as for example the INTERNET (say, Renal Physiology, by D. C. Mikulecky, Virginia Commonwealth University, mikulecky@gems.vcu.edu; http://views.vcu.edu/~mikuleck/)

2.4.3.1. Micropuncture

Wearn and Richards, in 1924, successfully inserted a minute needle, a *micropipette*, into the glomerulus of a single amphibian nephron to collect and analyze fluid samples. Their results provided conclusive evidence that ultrafiltration of the blood occurred at this site. Walker, Bott, Oliver and MacDowell, in 1941, extended the technique to the proximal and distal convoluted tubules. In 1958, Gottschalk and Mylle, micro-

punctured the loop of Henle, and Sakai, Jamison and Berliner, in 1965, did the same at the collecting duct. It is certainly no understatement to say that the technique of micropuncture has been the single most important event in the history of our understanding of renal physiology.

2.4.3.2. Microperfusion

Another technique, first developed by Windhager and Schatzmann in 1953 and later on improved by Gertz in 1963, was *stationary microperfusion*, sometimes called the "split-drop" method. It is based on micropuncture and, as such, it can be considered as an extension of the latter. A droplet of oil was injected into the tubule, followed by a test solution that divides the oil drop. Time-lapse photography was used to measure the changes in the droplet length and, thus, demonstrate concentration changes along discrete tubule lengths. Burg and Orloff perfected this technique in 1966.

2.4.3.3. Stop flow technique

It was devised in the 1950's by Malvin, Sullivan and Wilde. One ureter of an anesthetized animal was catheterized while a given substance was infused intravenously. When the urine flow stabilized, the ureter was clamped. After a few minutes, the clamp was released and urine gushing out was collected in a series of samples that were subsequently analyzed. The first samples came from the pelvis and distal parts of the nephrons while the final samples derived from more proximal sections of the tubules (that is, closer to the glomerular region). This technique has grossly provided information as to the loci of secretory and reabsorptive processes.

2.4.3.4. Slices of medulla and freezing point

The freezing points of distilled water and salt solutions depend on the concentration of solute in the solution. An important property of solutions states that the greater the ionic concentration, the lower the freezing point (at constant pressure), a result of vapor pressure lowering caused by the presence of ions in the solution. The change in freezing point can be represented by the equation

$$\Delta T_f = K_f C_{solute} \tag{2.87}$$

where, ΔT_f is the difference between the freezing point of the solution

and that of the pure solvent, K_f is a constant that is characteristic of the solvent and is called the *freezing point depression constant*, and C stands for the concentration of the solute in the solution.
Slices of the renal medulla removed from known regions are subjected to the freezing point technique in order to derive the concentration values by application of the above equation. The freezing point is found by detecting the temperature at which a small quantity of the sample under study freezes in a cuvette placed within a cooling device. Also urine samples can be used with this technique.

2.4.4. Renal Processes

All blood getting into the glomeruli is *ultrafiltrated*, that is, only particles of molecular weight smaller than 75,000 are able to pass through the glomerular capillary thin walls and appear in the capsule's volume in order to continue their trip along the tubules. Thus, ultrafiltration is a special very fine filtration. Cellular somas and proteins, which are larger than the size stated above, cannot pass the capillary walls and are never supposed to appear in normal urine. Filtration does not involve local expenditure of metabolic energy and depends only on the hydrostatic pressure imparted to the blood by the heartbeat.
The glomerular filtrate, loaded with substances, undergoes highly significant and physiologically important modifications as it moves within the tubular system until it reaches the ureter. There are processes of *secretion* and *reabsorption*, that is, by way of passive and mostly active mechanisms, substances pass from the blood in the peritubular capillaries to the renal parenchyma to the intratubular fluid and, viceversa, substances traverse an inverse pathway, i.e., from the intratubular fluid to the blood.
A simple linear model will be presented to obtain the equations that describe filtration, reabsorption and secretion (Valentinuzzi, Geddes, Baker *et al*., 1968). All the glomeruli are represented by a single node *Gl*, all the peritubular capillaries by another node *Ca*, and all the tubules by a third node *Tu* (Figure 2.51). Since this can be considered as a hydraulic system, let us apply the continuity principle to node *Tu*, which is equivalent to one of Kirchhoff's law in electric networks. In other words, in steady state conditions, what gets into a node must be equal to what gets out of it, or the algebraic sum of the flows to a node is zero. Flow is in-

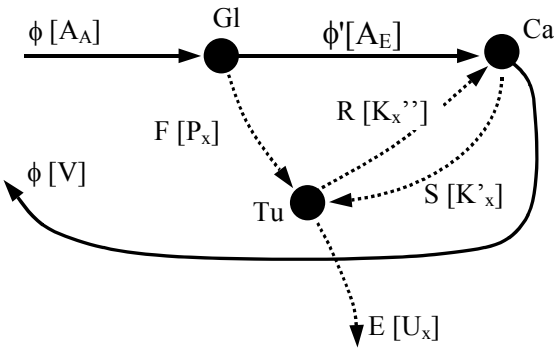


Figure 2.51. FLOW DIAGRAM OF THE RENAL SYSTEM. The renal equations for the three basic renal processes are obtained after application of the Continuity Principle to the nodes *Gl*, *Tu* and *Ca* (see text).

terpreted as the amount of fluid per unit time (say, mL/s), or the amount of a disolved substance carried by the fluid (say, mg/s), or charge per unit time (say, coulombs/s, which mean current in amperes in an electric circuit). In the case of Figure 2.51, there are four branches converging to node T_u, thus,

$$F[P_x] + S[K_x^{'}] - R[K_x^{''}] - E[U_x] = 0 \tag{2.88}$$

where the subindex x represents the substance disolved in the fluid, F, S, R and E are the rates of filtration, secretion, reabsorption and excretion, respectively, expressed in mL/min, and the variables between brackets stand for the respective concentrations of the substance x within the indicated branches of the system. Obviously, the net units of the equation are mg of substance/min. Equation (2.88) describes a dynamic equilibrium of the flow of substance with respect to node *Tu*.

Study subject: Search in the literature for the concept of the continuity principle or equation, usually found in physics textbooks. Search for the two laws of Kirchhoff in electric circuits. It is basic and necessary knowledge for a bioengineer.

Historical note: Gustav Robert Kirchhoff (1824–1887) graduated from the University of Koenigsberg in 1847. After a few years of lecturing in Berlin and Breslau, he took on a position at the University of Heidelberg in 1852, where he remained until 1875. After moving to Berlin, his failing health forced him to prematurely retire in 1886. Still a student, Kirchhoff made important contributions to electrical circuit theory, and in 1857, he produced a theoretical calculation demonstrating that an alternating electrical current flowing in a zero-

resistance conductor would flow at the speed of light, which provided an important stepping stone towards the electromagnetic theory of light formulated in 1880 by James Clerk Maxwell.

2.4.4.1. Filtration

Let us assume a substance that is only filtrated (as, for example, *inulin*, frequently used in the determination of glomerular filtration). In such case, secretion and reabsorption are absent and the equation reduces to,

$$F[P_x] - E[U_x] = 0 \tag{2.89}$$

which corresponds to a straight line when the *excreted load* $E[U_x]$ is represented as a function of the plasmatic concentration of substance P_x (Figure 2.52, left upper panel, line *a*). The slope F is precisely a measure of the *glomerular filtration rate* (also called GFR). By definition of ultrafiltrated fluid, the concentration of x in plasma is equal to the concentration in the capsule fluid. Experimentally, it has been found that the inulin filtration rate is about 120 to 130 mL/min, or in the order of 180 L/day (which is abou 30 times the blood volume).

If equation (2.89) is divided through by the plasmatic concentration $[P_x]$, we obtain,

$$C_x = F = GFR = \frac{E[U_x]}{[P_x]} \tag{2.90}$$

which is the glomerular filtration rate and, by definition, it is *the excreted load over the plasma concentration*, called also the *clearance* of substance x. In other words, clearance of x is the excreted load of that substance per unit concentration of the same substance in plasma. The numerator is measured in mg/min and the denominator in mg/mL meaning that clearance is expressed in mL/min, thus, another common definition in medical practice and in renal physiology states that clearance is the volume of plasma that in the unit time (say, one minute) is completely cleared of the substance x. For inulin or for any substance that only is being filtered by the kidneys, C_x is constant with respect to P_x (Figure 2.52, right upper panel, horizontal line *a*).

Study subject: Find the definition and meaning of the terms *extraction* and *filtration fraction*. They are rather common in the specialist's jargon. They are important concepts mainly applicable in renal diseases.

2.4.4.2. Secretion

Essentially, all substances are filtered by the kidneys, and some are also secreted into the tubular fluid but not reabsorbed. The general equation (2.88) becomes,

$$F[P_x] + S[K'_x] - E[U_x] = 0 \tag{2.91}$$

because $R = 0$. If it is accepted that $[K'_x] = [P_x]$, we can define

$$T'_x = S[P_x] \tag{2.92}$$

as the *secreted load* Figure 2.52, lower panel, line *a*). That straight line, as the plasma concentration increases, reaches a plateau or saturation value T'_{mx}. Experimental curves show that the breaking point is not

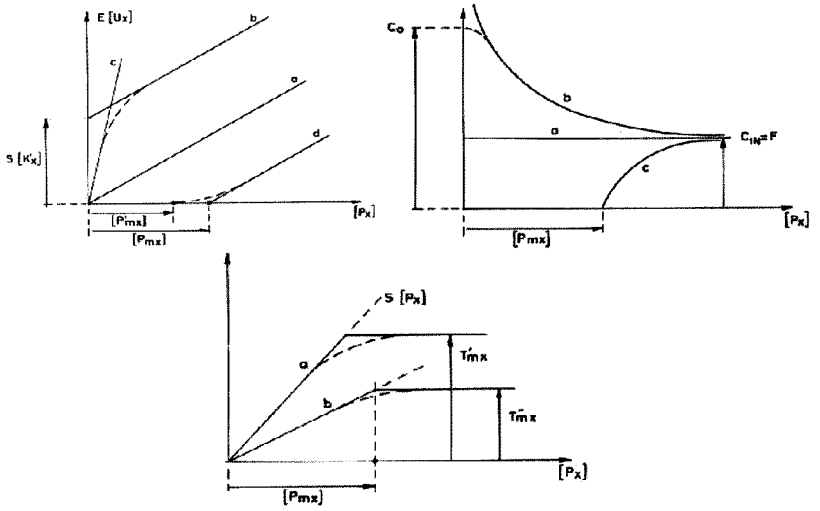


Figure 2.52. THE THREE BASIC RENAL PROCESSES. Left upper panel: Excreted load as a function of the plasmatic concentration. For pure filtration (a), for secretion first (c) and filtration predominance thereafter (b), and reabsorption followed by filtration predominance (d). Rigth upper panel: Clearance as a function of the plasmatic concentration. Filtration (a), secretion (b) and reabsorption (c). At high plasmatic concentrations, all three processes tend to the same constant glomerular filtration rate, which equals the inulin filtration value. Lower panel: Secreted and reabsorbed loads as functions of the plasmatic concentration. Curve (a), secretion, and curve (b), reabsorption. Notice that, in the case of reabsorption, the maximum plasmatic concentration $[P_{mx}]$ determines the breaking points of (d), excreted load, (c), clearance, and (b), reabsorbed load.

abrupt but rather there is a smooth bending to reach the maximum level (dashed line in Figure 2.52). This phenomenon is due to the spread out or splays of the individual nephron behaviors.

Clearance is also defined from equation (2.91) dividing it through by the plasma concentration and solving for the quotient excreted load to the same concentration, i.e.,

$$C_x = \frac{E[U_x]}{[P_x]} = F + \frac{T'_x}{[P_x]} \tag{2.93}$$

which is represented in Figure 2.52 (right upper panel, curve *b*). When the plasmatic concentration is low, the secretory process dominates and the clearance curve starts at a high value C_o (not infinite, as incorrectly predicted by the equation, meaning that the model is far from being perfect). The curve falls rather sharply with increasing plasmatic concentrations tending to the inulin clearance value at high concentrations, when there is a net filtration predominance. This is clearly seen in equation (2.93) when its mathematical limit is taken for $[P_x] \to \infty$.

In turn, the excreted load displays a steep slope first, ($F + S$), as shown in Figure 2.52 (upper left panel, line *c*) to become equal to the GFR (same figure, line *b*) as the plasmatic concentration goes up. Thus, line *b* runs parallel to line *a*, and with an upward shift. The passage from the first slope to the second one in actual experimental curves is smooth and the breaking point represents only a theoretical limiting value.

A typical substance frequently used in renal secretion studies is para-aminohippurate (PAH), with a maximum secreted load of 80 mg/min and a clearance of 650 mL/min at low plasmatic concentrations.

2.4.4.3. Reabsorption

The third process does not have secretion and the equation, with S = 0, becomes now,

$$F[P_x] - R[K''_x] = E[U_x] \tag{2.94}$$

because the excreted load was moved to the right hand side. This is represented by the straight line *d* in Figure 2.52, shifted to the right with respect to line *a*, and crossing the horizontal axis at $[P_{mx}]$. If $[K''_x] = [P_x]$, it is possible to define the *reabsorbed load* as,

$$T_x'' = R[P_x] \tag{2.95}$$

which is represented by line *b* in Figure 2.52 (lower panel). Breaking points, as mentioned before, are really defined by the respective projections of the straight lines and experimental curves display a smooth transition from one to the other, as shown in the figure. The reabsorbed load reaches a maximum value (saturation) when the tubules are no longer able to take more substance and that happens beyond a given plasmatic concentration specific for each substance. At lower concentrations, instead, there is full reabsorption. As we did before, clearance is obtained by dividing equation (2.94) by $[P_x]$ leading to,

$$C_x = \frac{E[U_x]}{[P_x]} = F - \frac{T_x''}{[P_x]} \tag{2.96}$$

The latter describes a hyperbola (Figure 2.52, upper rigth) that tends to the constant value given by inulin as the plasmatic concentration increases. For low concentrations, the reabsorptive process dominates and the substance in full is returned to the circulation. This is the case for substances of high physiological value, as glucose is. The maximum reabsorbed load is in the order of 380 mg/min at a plasmatic threshold level of about 300 mg/100 mL that, in practice because of the splay phenomenon, lies between 180 and 200 mg/100 mL (Figure 2.52 A*d*, B*c* and C*b*, where A, B and C stand, respectively, for the upper left, upper right and lower panels). A diabetic person will have a high level of glucose in blood (*hyperglucemia*), usually beyond the renal plasma threshold. The reabsorptive capacity is saturated and glucose appears in urine (*glucosuria*). In the old days, the physician used to taste the urine for its sweetness in order to determine whether a patient was diabetic. Do not laugh, it is true! Fortunately, techniques are more advanced and sophisticated nowadays.

All substances are filtered by the kidneys, including exogenous substances as drugs are. Almost common knowledge is the fact of a person taking vitamin B who, a few hours later, produces reddish urine, typical of that substance. Many times the attending physician will warn the patient not to be scared in such case. Some substances are reabsorbed, others are secreted and there is a group that is both absorbed and secreted (as urea and creatinine, with a predominance of secretion, the latter much

more than the former). Sodium, instead, is also secreted and reabsorbed but approximately in equal amounts. In a quantitative sense, the filtration and reabsortion of ions and water are by far the most significant operations of the mammalian kidney. The body cannot afford, for example, to lose potassium and, thus, about 93% is reabsorbed and retained. The interplay of these processes, in the end, is responsible for the proper electrolyte balance.

2.4.4.4. Osmosis

When a membrane, which is permeable to solvent but not to solute, separates a solution and pure solvent, the solvent passes into the solution by *osmosis* (from Greek, meaning "push"). The *osmotic pressure* is that particular hydrostatic pressure which must be applied to the solution to prevent the entry of solvent. In other more general words, water passes from a solution with lower concentration to a solution with higher concentration so that both concentrations tend to equilibrate.

Osmotic pressure is one of the properties of solutions and depends on the number of particles per unit volume of solvent, not on their chemical characteristics. Similar to the law of gases which relates pressure, volume and temperature, van't Hoff's equation states that,

$$\Pi = C_m \times R \times T \tag{2.97}$$

where Π represents the osmotic pressure or "water attraction" generated by the solution, C_m stands for the molar concentration of the solution, R is the gas constant (= 0.08 atm×L/mole×°K) and T is the absolute temperature. Since the molar concentration $C_m = n/V$, n being the number of moles and V the volume in liters, thus expressing it in moles/L, equation (2.97) can be also written as,

$$\Pi \times V = n \times R \times T \tag{2.98}$$

which well reminds the law of gases because pressure P, in the latter, is replaced by osmotic pressure Π, in the former. Equation (2.98) defines 1 osmol, a unit, as the osmotic strength generated by a concentration of 1 mole per liter at the measured temperature T in Kelvin degrees, or

$$1 osmol = 1\frac{mole}{liter} \times 0.08\frac{atm \times liter}{mole \times {}^{o}K} \times 310^{o}K = 25.4 atm = 19{,}304 mmHg \tag{2.99}$$

Above, we take the temperature as equal to the body temperature, 37°C, and adding 273 to obtain the Kelvin degrees. If the temperature instead were 0°C, that is, $T = 273$°K, 1 osmol becomes equivalent to a pull of 22.4 atm. The osmotic strength of a solution can be found by summing the molar concentrations of all the ions and non-dissociating molecules.

Some numerical examples will help in the understanding of the subject. On purpose, we are using several times the whole word and the abbreviation for a given unit in order to be clearer in its meaning. For any substance, one gram-molecular weight (or one mole = 1 M) contains $N = 6.02\times10^{23}$ molecules (*Avogadro's number*). If the substance is glucose, sucrose or any non-dissociating compound, the osmotic strength is 1 osmol. Now, let us consider a substance that dissociates when in solution. One gram-molecular weight of NaCl consisting of 6.02×10^{23} molecules dissociate into twice this number of ions in solution. Thus, 1 M of NaCl exerts an osmotic effect of nearly two osmols (we say "nearly" because the degree of dissociation depends on the concentration, the lower the concentration, the higher the dissociation). If it were a substance dissociating in three ions (such as Na_2SO_4), the effect would approach 3 osmols. The osmotic concentration of plasma, interstitial fluid and intracellular fluid are all kept in man within the band (283 ± 11) milliosmoles/L (mOsm/L), basically by the ingestion and excretion of water. Gain of water induces prompt water diuresis while loss of water induces thirst and antidiuresis. Such value, roughly equal to 300 mOsm/L, produces at body temperature an osmotic effect or pull equivalent to 7.62 atmospheres or 5,791 mmHg, which is quite an impressive value (check the calculation with the relationship given above). In other words, this is the pressure needed to counteract the tendency of plasma (or to any solution similar to plasma) to draw water.

The osmotically active solutes are to a large extent (around 90% or more) the electrolytes such as sodium, chloride, and bicarbonate. Glucose, amino acids and urea contribute not more than 10%. Hence, the osmolar concentration is a way of measuring the "total concentration" of a fluid. Besides, and as practical information, the weight concentration of a substance, in g/L, is given by the molecular weight MW (a number) multiplied by the molar concentration, in moles/L, or $C_w = MW\times C_m$. A 5.4% solution of glucose means a weight concentration of 54 g in 1 liter which,

with a molecular weight of 180 yields the molar concentration C_m = 300 mM/L at 37°C. By the same token, a 0.9% NaCl solution means a weight concentration C_w = 9 g/L which, with MW = 58, produces a molar concentration C_m = 155 mM/L and the same osmolar strength of plasma. Verify these calculations using the relationships given above.

Another piece of useful information refers to the freezing technique already and briefly described above: 1 mole/L of ideal solute will depress the freezing point by 1.86°C, from which the freezing point depression (in °C) experimentally obtained from a given sample divided by 1.86×10^{-3}°C yields directly the number of mOsm/L of the sample solution. "Osmolarity" refers to the number of osmols per liter of solution (the solution contains the solute) while "osmolality" is the number of osmols per kilogram of solvent (the solvent does not contain the solute). These terms are common in the texts as the terms "molarity" and "molality" are. The student can easily deduce the definitions for the two latter.

Amadeo Avogadro (1776–1856), Italian chemist, stated in 1811 that "equal volumes of all gases under the same conditions of temperature and pressure contain the same number of molecules" (Avogadro's Hypothesis), but he never tested it.

The name "Avogadro's Number" is just an honorary name attached to the calculated value of the number of atoms or molecules in a gram-mole of any chemical substance. If we used some other mass unit for the mole, such as "pound-mole", the "number" would be different than 6.022 x 10^{23}. The first person to have actually calculated the number of molecules in any mass of substance was Josef Loschmidt (1821–1895), an Austrian high school teacher, who in 1865, using the new Kinetic Molecular Theory, obtained the number of molecules in one cubic centimeter of gaseous substance under ordinary conditions of temperature and pressure to be around **2.6 x 10^{19} molecules**. This is usually known as "**Loschmidt's Constant.**" Check INTERNET to get more details regarding this subject.

Maintenance of osmolarity is essential for life. It is basically kept by water ingestion and excretion. An individual can survive many days without food but not too long without water. If a large amount of water is drunk, a diuresis of diluted, uncolored, and almost odorless urine reestablishes normal osmolarity. Conversely, if there is dehydration (as in diarrhea, or heavy sweating during a sunny day), oliguria restricts urine outflow and the reduced amount excreted is characterized by high concentration (high osmolarity), amber color (yellowish to brownish), and penetrating offensive odor. Besides, the mechanisms of thirst are activated so that the sub-

ject is driven to drink water. The renal system is much more efficient in defending the organism against dilution (water excess) than against dehydration. Children and old persons are particularly sensitive to the latter and sportspeople must be always warned of the risks faced when heavy exercise is practiced under high temperature conditions, especially during summer time. In the end, the thirst mechanism leads to compensating for the water deficit. The kidneys cannot do this by themselves for they only are able to either remove or conserve water. There are special sensors, the osmoreceptors, located in the central nervous system, that constantly check osmolarity to effect thirst and the proper renal actions.

The kidneys maintain a strong osmolar concentration gradient from their deep medullar region, where osmolarity is 1,200 or even 1,400 mOsm/L, to gradually decrease down to 300 mOsm/L at the cortical region. There is a group of nephrons with their glomeruli placed at the level of the renal cortex and the loops of Henle and collecting ducts penetrating deep into the renal medulla. Besides, their peritubular capillaries run almost parallel to Henle's limbs, which show a U-like shape (Figure 2.50). When a subject is dehydrated, central osmoreceptors located in the hypothalamus order the secretion of antidiuretic hormone (ADH), also called *vasopressin*, from the *posterior hypofisis*. ADH acts on the distal convoluted tubules and the collecting ducts increasing their permeability to water which, due to the concentration gradient across the tubular wall, as fluid moves along the tubules and duct, water goes easily from the intratubular fluid to the interstitium by simple passive osmotic gradient and from there to the blood in the peritubular capillaries. Water is, thus, retained. An opposite situation is when the subject drinks too much water. The osmoreceptors suppress the secretion of ADH and the permeability to water of the distal tubules and collecting ducts decreases. Hence, in spite of the concentration gradient, water cannot traverse the walls and there is diuresis.

In these two extreme situations, the total urine excretion in 24 hours can be from 0.5 L in dehydration at a concentration of 1,200 or even 1,400 mOsm/L meaning an excreted load of 600 to 700 mOsm, and up to about 20 or 24 L in dilution because of excess water at a concentration of only 30 mOsm/L, meaning an excreted load also of 600 or 720 mOsm. In

other words, the excreted load is kept essentially constant in both situations.

Jacobus Henricus van 't Hoff was born in The Netherlands, in 1852, and died in 1911, in Germany. In 1885, he came up with a study on the chemical equilibrium in gaseous systems and strongly diluted solutions. Here, he demonstrated that the “osmotic pressure” in solutions, which are sufficiently dilute, is proportionate to the concentration and the absolute temperature so that this pressure can be represented by a formula that only deviates from the formula for gas pressure by a coefficient. Thus, van 't Hoff was able to prove that thermodynamic laws are not only valid for gases, but also for dilute solutions. His pressure laws, given general validity by the electrolytic dissociation theory, are considered comprehensive and seminal in the realm of natural sciences. He can be regarded as one of the founders of physical chemistry. In 1901, van’t Hoff received the first Nobel Prize in Chemistry.

Friedrich Gustav Jacob Henle (1809–1885), German anatomist and pathologist. His name is best known today for the loop-shaped portion of the nephron. It consists of a thin descending limb and a thicker ascending limb. His observation of it in 1862, supported by isolation preparations, was correct in itself but the interpretation was wrong. Nevertheless, his study resulted in a new series of investigations on the kidneys through which, between 1863 and 1865, their structure was definitely determined.

2.4.5. Countercurrent Mechanisms

We have seen in the previous paragraphs that the kidneys require a medullo-cortical osmolar gradient to regulate the osmolar-excreted load and, with it, to keep the extracellular fluid osmolarity. Let us explain now how this gradient is built up and maintained. Each operation makes use of a countercurrent system, a principle well known by chemical engineers in many types of industrial exchangers to improve the exchanging efficiency.

2.4.5.1. Countercurrent multiplication

There is a basic mechanism between the ascending and the descending limbs of the loop of Henle, at any level, that by active transport of sodium ions from the ascending branch into the descending one generates a constant transversal difference of 200 mOsm/L. This is a physiological renal property. The difference includes the interstitial fluid, too (Figure 2.53). The ascending limb walls, we underline, are impermeable to water while the descending branch as the collecting duct are not, implying that water cannot get into that portion of the tubules and dilute its content.

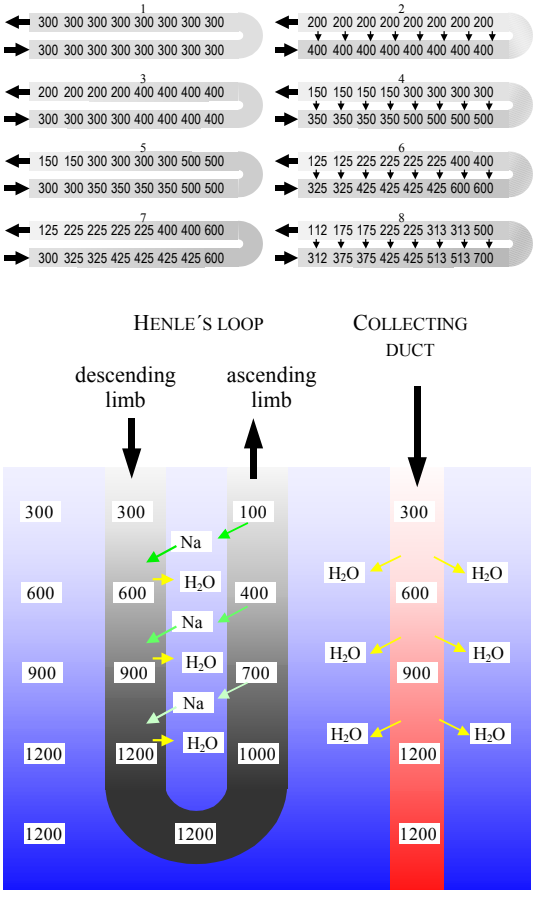


Figure 2.53. MECHANISM OF COUNTERCURRENT MULTIPLICATION. Upper panel: Stage 1 depicts the loop of Henle full of fluid at 300 mOsm/L. At stage 2, due to a basic active process, a transverse osmolar difference of 200 mOsm/L is established between the descending and ascending limbs.
Lower panel: It shows the countercurrent exchange relationship between Henle's loop and collecting duct. See text for further details. Redrawn after Pitts (1966).

Because of the U-like shape of the loop, the fluid in both limbs flow in a countercurrent way (opposite directions) so that there is a multiplicatory effect that, in the end, becomes a longitudinal gradient.
(Figure 2.53, upper panel, drawn horizontally) explains the build-up process step by step: At stage 1, say that fluid gets into the loop filling it fully at the same concentration, entering via the descending limb and exiting along the ascending portion. Thereafter, at stage 2, due to the basic active transport of sodium, a 200 mOsm/L is transversely generated all along the loop (say that the descending side rises its concentration to 400 mOsm/L while the other side lowers it to 200 mOsm/L). However, flow continues and a moment later, stage 3 depicts the situation when fresh

fluid at 300 mOsm/L gets in shifting the whole column by a certain length. Again the active mechanism restores the osmolar difference (stage 4), but observe that the region of the bent (right side in the figure, right upper panel) begins to increase its concentration creating a difference with respect to the entrance. Successive fluid shifts followed always by the basic transversal osmolar gradient build-up plus the countercurrent flow give rise to a longitudinal much larger gradient between the entrance (at 300 mOsm/L) and the medullar region (reaching there 1,200–1,400 mOsm/L). The mechanism is most interesting and ingenious, so much that it can be qualified as outstanding. A neat design of the Great Engineer, indeed! Rash (1984) simulated with a microprocessor this multiplying phenomenon, which may even have technological applications. With the present technology, such simulation should be easier and a real challenge for biomedical engineering students.

2.4.5.2. Countercurrent exchanger

The medullo-cortical concentration gradient, however, tends to equilibrate, as a ball placed on a ramp tends to roll down unless something is done either to prevent it or at least to partially brake it. Another mechanism is needed to conserve such gradient. The vasa recta of the peritubular capillaries that run parallel to Henle's limbs take care of that function (Figure 2.53). Notice the specialization of the medullary nephrons. Blood gets into the vasa recta at 300 mOsm/L and, as it moves down in their descending branches, water traverses the capillary walls because of the osmotic pull from outside while active osmotic particles get into the blood, also due to a small transverse gradient. As blood goes up following the ascending limbs of the vasa recta the opposite shifts take place, that is, water gets into blood and solutes go out because at any level the ascending side has a slightly higher concentration than the descending one. Thus, the countercurrent exchanger reduces excessive loss of osmotically active solutes from the inner medulla. Blood in the vasa recta remove sodium and water. Loss of the medullocortical osmotic gradient would be disastrous for animal or human life.

This mechanism is similar to the countercurrent heater exchangers widely used in industrial plants, it has been amply studied both theoretically and experimentally and its better efficiency has been fully demonstrated as compared to heater exchangers with parallel streams in the

same direction. Penguins have in their legs a circulatory arrangement of the same kind that helps them in keeping a temperature gradient from the trunk to the feet. Does it make the feet warmer? No, but the upper legs, the abdomen and the upper body better conserve body temperature. The extremities of the sloth also have the same circulatory arrangement, however, its function is unknown, especially if we recall that they live in tropical regions, like NE Brasil. Another stimulating subject for the inquisitive mind.

2.4.5.3. Osmotic exchanger

The distal convoluted tubules and the collecting ducts (more the latter than the former) regulate the final osmotic urine adjustment by the action or not of the antidiuretic hormone (ADH), which, as mentioned before, controls their wall permeability to water. Most of the water is recovered and only a minor amount is excreted, either as hypotonic or hypertonic urine (measured with respect to the plasmatic 300 mOsm/L), and depending on the hydration degree of the subject. Sodium is also recovered along this final pathway along with other ions (such as phosphate and bicarbonate). Thus, the final urine equilibrates with the hypertonic interstitium of the renal medulla and papillae.

The major osmotically active constituents of urine are sodium and chloride ions and urea. The osmolar clearance may be calculated from a formula derived from the clearance definition given above, that is,

$$C_{osm} = \frac{[U_{osm}]V}{[P_{osm}]} \tag{2.100}$$

In this expression, $[U_{osm}]$ represents the collected urine osmotic concentration, V the collected volume in a given period of time and the denominator stands for the plasma osmolarity. In words, it is defined as the volume of plasma per unit time completely cleared of osmotically active solutes or, also, as the osmolar excreted load per unit of plasmatic osmotic concentration. The interested student can find details and mechanisms in the literature.

2.4.6. Renal Blood Flow

Let us go back to Figure 2.51 writing now the continuity equations for nodes *Gl* and *Ca*,

$$\Phi[A_A] - F[P_x] - \Phi'[A_E] = 0 \tag{2.101}$$

$$\Phi'[A_E] - S[K_x'] + R[K_x''] - \Phi[V] = 0 \tag{2.102}$$

Solving for $\phi'[A_E]$ the two previous equations, equating, and considering equation (2.88), the following is easily obtained,

$$\Phi[A_A] - \Phi[V] = E[U_x] \tag{2.103}$$

where ϕ stands for the renal plasma flow, so that,

$$\Phi = \frac{E[U_x]}{[A_A - V]} \tag{2.104}$$

or in words: The excreted load of a given substance divided by the renal arteriovenous concentration difference yields the renal plasma flow. If the test substance is fully extracted from blood in its passage through the kidneys, then, the venous concentration becomes zero. Besides, if the other tissues do not extract that substance from blood, the renal artery concentration will be equal to the concentration in any systemic artery or vein, that is, it will be equal to $[P_x]$. With all this in mind, equation (2.104) simplifies to,

$$\Phi = \frac{E[U_x]}{[P_x]} \tag{2.105}$$

the latter coinciding with the plasmatic clearance of the substance (see clearance definition above). Para-aminohippurate is one substance that approximately meets the described conditions. As a consequence, the measurement of the plasmatic renal flow is relatively simple: After administration of PAH, a volume of urine is collected for a specific time span to have E. Immediately thereafter, the PAH concentration in urine is determined. From a venous blood sample, its PAH concentration is also determined. Finally, equation (2.105) is applied to obtain the renal plasma flow. Normal adult human kidneys yield a value of about 650 mL/min. Taking the hematocrit into account (arount 0.45), the renal blood flow becomes in the order of 1,200 mL/min, which is 20% of cardiac output.

Thinking exercise: Review the definition of hematocrit and the calculation that leads from plasma flow to blood flow. Do the kidneys really need such a large amount of blood? Check the percentage it represents relative to cardiac output.

2.4.7. Closing Remarks of Section 4

The Renal System is essentially the complex exchanger shown as E4 in the block diagram of an organism shown earlier in the book (Figure 2.1), connecting it on one side with the Cardiovascular System and on the other with the external world to excrete the waste products. The metabolic needs of the kidneys do not justify the large perfusion they receive, rather this comes about because several times per day all the blood volume must pass through them in order to guarantee the adequate homeostasis of the internal environment. Most remarkable is the urine concentration (or dilution) process by which water and other substances are conserved: Mainly, the collecting ducts use a large medullo-cortical osmolar gradient. At every level of Henle's loops ascending branches, a basic transversal difference of 200 mOsm/L is established by active sodium transport as compared to the descending limb. The ascending branch is impermeable to water, so preventing the immediate loss of that gradient. The basic gradient is longitudinally multiplied by the countercurrent flow of the intratubular fluid. The peritubular vasa recta also working in a countercurrent arrangement reduce losses and tend to conserve the medullocortical osmotic difference. Finally, the collecting ducts act as true osmolar exchangers controlled by the antidiuretic hormone. In this way, only a small volume of fluid is lost per day. Such fluid carries in solution different substances and electrolytes.

The mechanics and control of micturition have not been touched in this section. For the time being, they can be ignored. The interested reader will find good descriptions in any physiology textbook.

A person can survive and have a normal life with one kidney but life is incompatible without renal function. Thus, when we think in terms of renal failure, immediately we think of transplantation, dializers and eventually total kidney replacement by an artificial one. From the perspective of Biomedical Engineering, there is a lot to offer to the problems posed by the subject, from theoretical, physiological to technological aspects. It is an endless avenue.

2.5. Gastrointestinal System

Man is the only species using the GI system for pleasure!

2.5.1. Introduction

As indicated in Figure 2.1, the Gastrointestinal System (GIS) can be considered as an input-output physiological unit designed to supply "combustible" to the organism. Its objective is to transform (digest) the ingested food into simple nutritive substances that are to be transferred (absorbed) to the Cardiovascular System (CVS), which, as already described, distributes and delivers them to the tissues. Chewing and swallowing, digestion, absorption and defecation are all processes controlled

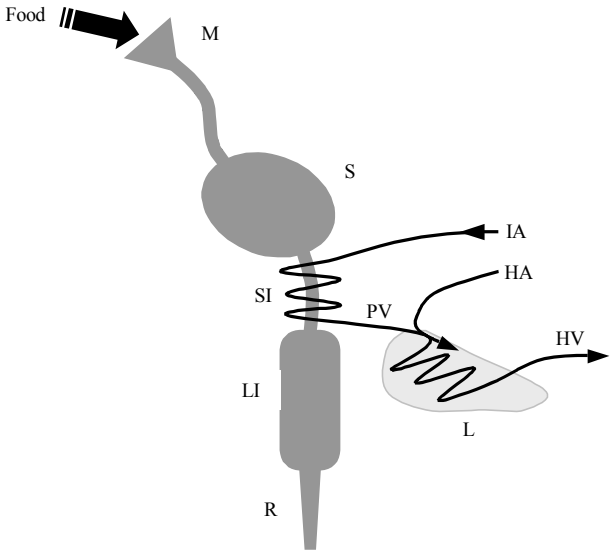


Figure 2.54. SCHEMATIC OF THE ALIMENTARY CANAL. From top to bottom, left side: M, mouth; S, stomach; SI, small intestine; LI, large intestine; R, rectum. The intestinal artery IA vascularizes the SI in a complex network of capillaries that constitutes the first section of the GI system exchanger. Blood loaded with nutritive substances comes out of the intestinal capillaries via the portal vein PV, which enters the liver L after a short pathway. The liver receives also blood from the hepatic arteries HA. The hepatic capillaries or sinusoids form the second section of the GI system exchanger. Blood returns to the general circulation via the hepatic veins.

by autonomic, nervous and hormonal mechanisms. Digestive glands act to provide moisture, lubrication, emulsification and enzymes for digestion of proteins, carbohydrates and lipids. In a simplified manner, the gastrointestinal tract (Figure 2.54) is a canal that,

a. receives food through the mouth;
b. receives secretions from the salivary glands, liver, pancreas, stomach and intestines to digest food;
c. absorbs the digested nutritive substances, especially at the level of the small intestine;
d. supplies the absorbed nutritive substances to the liver via the portal vein;
e. elaborates and stores other substances in the liver; and
f. excretes the residues.

The entire tract has an associated muscular mechanics, a vasculature and a complex regulatory system. The classic book by Davenport (1977) is a must for any interested student of the subject. Ganong's textbook of physiology (1981) is also recommended in any of its many editions. Besides, there are several websites to visit, as for example
http://www.e-gastrointestinal.com.

2.5.2. Mechanics

Movements can be felt and detected all along the alimentary canal. The experience of the hunger pangs just before dinnertime or the bowel noises, especially after having eaten certain gas producing foods (such as milk or soja beans), is well known by everybody. They are GI mechanical manifestations. Such movements are complex in nature and rather irregular and they are generated by the smooth muscles that cover the different GI structures. This musculature is controlled by the autonomic nervous system, by hormones, by local mechanisms (which, by and large, involve stretching due to the entrance of material into the gastric cavity or the intestinal lumen) and, also, originates in intrinsic automatism (that is, there are temporary pacemakers).

Basically, motility takes the form of two types of movements, that is, as **segmental contractions** and as **peristaltic waves** (peri and *stalsis* = contraction, from *Greek*). The former, mix and churn the intestinal content (**chyme**) while the latter essentially propel the chyme along the intestine. Segmental contractions are ring-like contractions that appear at fairly

regular intervals along the gut, then disappear and are replaced by another set of contractions in the segments between the previous contractions. When the intestinal wall is stretched, a deep circular contraction or peristaltic wave forms behind the point of stimulation and passes along the intestine toward the rectum at rates varying from 2 to 25 cm/s. Such response to stretch is called the **myenteric reflex**. Movement normally takes place in the aboral or oral-caudal direction; occasionally, and in pathological conditions, antiperistalsis (antidromic propulsion) can be seen.

Figure 2.55 illustrates the spontaneous mechanical activity of a piece of rabbit small intestine, placed in a thermic physiological bath, and attached to a sensitive isotonic transducer. These contractions, as shown in the figure, are usually regular in amplitude, frequency and pattern, allowing the recording of rather long strips. In this particular case, there was a pacemaker activity of about 20 periods/min. Such intrinsic automaticity may vanish or may be enhanced (for example, by pharmacological stimulation) decreasing in the oral-caudal direction (Alvarez Law). The latter means that always a sample from either the duodenum or the jejunum or the ileum will show a higher spontaneous frequency than a sample from lower regions, as, say, the large intestine (ascending, transverse or descending colon).

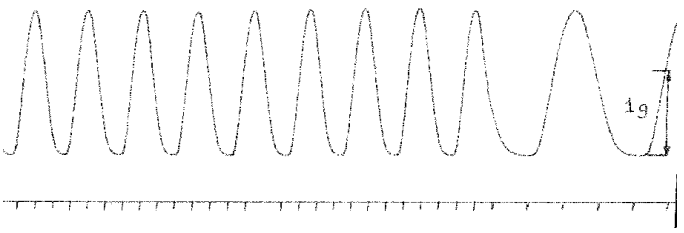


Figure 2.55. RABBIT ILEUM ACTIVITY. Spontaneous contractions of a small intestine sample (ileum) immersed in an adequate physiological solution, temperature controlled, and with gentle air bubbling. The maximum force of contraction was about 1.75 g and very regular. The pattern was also regular and quasi-sinusoidal. Lower marks are 1 s apart. To the right of the record, paper speed was increased to better visualize the waveform. An isotonic transducer attached to one end of the sample while the other end was anchored to the holder longitudinally detected these contractions. Thus, they cannot be identified either as segmental or peristaltic. Records obtained by the author at the Department of Bioengineering, UNT, 1980.

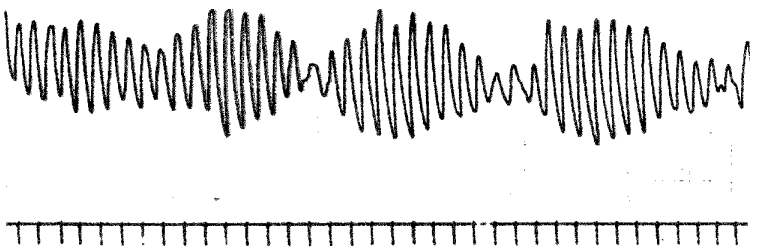


Figure 2.56. BEATING EFFECT FROM A RABBIT ILEUM SAMPLE: The sample was about 2 cm in length; thus, it probably caught two pacemakers with slightly different intrinsic frequencies (recall Alvarez Law: frequency decreases towards the tail, hence, the caudal portion of the segment should show a slightly lower activity than the upper one). The net result was the beating effect as manifested by the slow modulation. Time marks are 5 s apart. The carrier had a frequency of about 15 cycles/min while the beat had a very low 1.5 cycles/min. Records obtained by the author at the Department of Bioengineering, UNT, 1980.

Observe the sinusoidal-like waveform of the record shown in Figure 2.55, a rather interesting feature that can be explored a little further. When the sample is too short (say, less than 1 cm), no activity is detected because no pacemaker site is in it; conversely, if the sample is longer than 1 cm and about 2 cm, a waveform resembling the beating effect of two near sinusoidal frequencies can be produced (Figure 2.56). This phenomenon is typical of coupled systems, which can be modeled with a second order differential equation. If the sample is even longer, then more than two pacemakers may be picked up and a complex waveform will be the result.

Historical tip: Bayliss and Starling, in I899–I901, formulated the "Law of the Intestine" (not to be confused with Alvarez Law) to provide an explanation for peristalsis. They found that the response of the small intestine to a local stimulus consisted of contraction of the *muscularis externa* layer immediately above, and relaxation immediately below the point of stimulation. It was attributed to a reflex that involved the myenteric plexus and was independent of the external innervations of the intestine. Cannon, in I911, called it a "myenteric" reflex, as referred to above; the biphasic wave of relaxation and contraction was found to pass over the muscular layer in an aboral direction for short distances from the point of stimulation. Later workers, Alvarez among them, in I940, and Bozler, in I949, improved these studies.

There are longitudinal and circular smooth muscles all along the intestinal walls. Contractions of the former and of the latter musculature during

peristalsis are out of phase by 90 degrees, according to Davenport. On distension of the lumen the longitudinal muscle contracts, followed by progressive contraction of the circular layer. The circular layer begins contraction when contraction of the longitudinal layer is half complete; contraction of the circular layer is complete when relaxation of the longitudinal is half complete. Peristalsis is defined rather superficially in many scientific and medical dictionaries. The consensus appears to be that it is a vermiform or a progressive, wave-like movement in tubular organs, consisting of alternating waves of relaxation and contraction in the muscular coat by means of which the contents are propelled. More details can be found in the website med.plig.org/index.html (look for *The Pyloric Sphincteric Cylinder in Health and Disease*, by A.D. Keet).

Contraction of a smooth muscle cell is associated with change in potential of the cell membrane; the potential depends on the distribution of electrolytes between the cell and the extracellular space, much like in skeletal muscle. Two main types of spontaneously arising changes in the membrane potential may be detected, i.e., slow potential variations and spike potentials. The latter are superimposed on the former and are associated with mechanical activity. Thus, the electrophysiology of the gastrointestinal tract appears as another research field still with many unresolved problems and unknowns.

2.5.3. Secretions

Many substances are secreted into the alimentary canal. Some are *hormones* (which are internal secretions, that is, they get into the circulatory stream). The first to be found is saliva. The salivary glands in the mouth produce about 1,200 mL/day of saliva (an external secretion) at a pH of 6 to 7, that is, from acid to neutral.

The stomach secretes a hormone, *gastrin*, whose principal effect is to stimulate the production of pepsine (an *enzime*) and of gastric juice, both from the stomach itself. They are exocrine secretions, i.e., they do not get into blood. Gastric juice contains a variety of substances, it shows a very low pH (from 1 to 3.5, which means high acidity, due to the predominance of hydrochloric acid, HCl), and is secreted by parietal cells at a rate of about 3,000 mL/day. Such high acid level should damage the gastric wall, however, the surface membrane of the mucosal cells and the tight junction between cells seem to act as a protective barrier. Sub-

stances that tend to break the barrier (such as aspirin or vinegar) may lead to gastric irritation, from mild to severe. Vagal activity also stimulates acid secretion from the stomach cells. This perhaps explains the rather common and so-called *heartburns* (upper gastric burning sensation) reported by worried or overstressed persons who may show an increased tone of the vagi.

As many as 33% to 44% of Americans experience heartburn at least once a month and up to 13% have heartburn each day. The likelihood of having heartburn increases with age and among women who are pregnant. Having heartburn every once in a while is something almost everyone experiences, but if it occurs 2 or more days per week, it can be a sign of a more serious problem called *gastroesophageal reflux disease*.

The small intestine secretes two hormones, *secretin* and *cholecystokinin-pancreozymin* (CCK-PZ). They both act on the exocrine pancreatic function stimulating the secretion of pancreatic juice (1,200 mL/day, pH = 8.0 to 8.3, rich in enzymes to break up carbohydrates, proteins and lipids). Besides, CCK-PZ stimulates the contraction of the gall bladder to inject bile (produced by the liver) into the duodenum. Bile is temporarily stored in the gall bladder. There is a production of about 700 mL of bile per day at a constant pH of 7.8, that is, it is on the alcaline side.

Secretin was the first hormone ever found, discovered by Bayliss and Starling, in 1902. It causes the secretion of a watery, alcaline pancreatic juice. Its action on the duct cells of the pancreas is mediated by cyclic AMP. This hormone is produced by cells located deep in the glands of the mucosa of the upper portion of the small intestine.

Suggested reading exercise: Search for information about cyclic AMP (adenosine monophosphate). Find the definition of *enzyme*. The are extremely important substances.

A stimulating factor was isolated in crude form from hog intestines by William Maddock Bayliss and Ernest Henry Starling (the latter is the same of the law of the heart), very early in the XXth Century. These scientists, realizing the messenger role of a chemical substance, coined a new word — hormone (which means “I move” in Greek) — and called it secretin. It turned out, however, to be rather difficult to achieve its isolation from gut mucosa in pure form. The pure hormone was finally obtained about fifty years later by Jorpes and Mutt of the Karolinska Institute, in Stockholm in 1961. The same group was also successful in determining the sequence of the twenty-seven amino acids constituting the peptide chain of secretin.

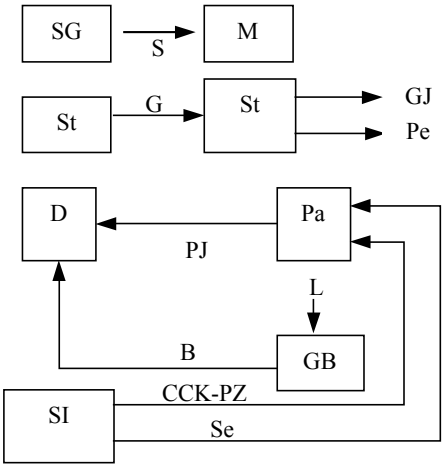


Figure 2.57. MAIN SECRETIONS OF THE GIS. SG = salivary glands; Sa = saliva; M = mouth. The stomach, St, secretes gastrin, G, which in the same stomach stimulates the production of gastric juice, GJ, and pepsin, Pe. SI = small intestine secretes cholecystokinin-pancreozymin, CCK-PZ, and secretin, Se, to stimulate pancreatic juice secretion and the gall bladder contraction, respectively. The liver, L, sends bile into the gall bladder, GB, and the latter into the duodenum, D.

Figure 2.57 summarizes in a block diagram the main secretions (hormones and juices) of the gastrointestinal system. There are several websites that can be visited to fetch detailed information about this subject. One good piece is

http://www.kutchai/medphys/handouts/gi-secr.doc,

www.hsc.virginia.edu/med-ed/phys/pdf/Gi_secr.pdf .

2.5.4. Perfusion

Mesenteric circulation and splanchnic circulation are terms many times used as synonims, however, the former actually refers specifically to the intestinal vasculature while the latter encompasses the blood flow to all the viscera within the abdominal cavity.

Figure 2.58 summarizes the main avenues giving also some numerical values. It can be observed that the liver receives blood from two important inputs: from the hepatic artery, to take care of its tissue needs, and from the portal vein, to supply via its exchanger the hepatic parenchyma with the substances to be metabolized and stored in it. The first one, supplying in the order of 700 mL of blood/min, feeds the hepatic artery, HA, which carries into the liver 500 mL/min to satisfy the tissue needs of this organ, and distributes also blood to the stomach, St, the spleen, Sp, and the pancreas, Pa. The SMA carries another 700 mL/min with ramifications to the pancreas, the small intestine, SI, and the colon, Co. Finally, the IMA, with 400 mL/min, completes the blood supply to the colon. The

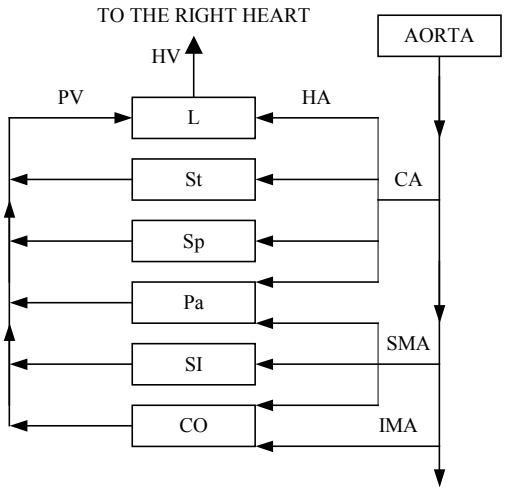


Figure 2.58. SIMPLIFIED DIAGRAM OF THE SPLANCHNIC CIRCULATION. The descending abdominal aorta gives off three main branches: the celiac (CA), the superior mesenteric (SMA) and the inferior mesenteric (IMA) arteries.

portal vein, PV, high in nutritive contents loaded mainly through the small intestinal capillaries, returns blood to the liver at a rate of 1,300 mL/min, by far much more than its tissue needs because this blood goes into the delicate hepatic sinusoids. The liver parenchyma is the place of essential metabolic and storing processes. The hepatic veins, HV, return blood to the vena cava and, from there, on to the right heart.

The liver is an essential station for the sustenance of life. Claude Bernard, around 1860, already recognized its importance. Any substance, good or bad, that we ingest is metabolized in the liver. Nutrients, among them —such as glucose— are metabolized, elaborated and stored in the liver to be delivered later on to satisfy body needs. The liver acts, for example, as a glucostat to regulate the blood sugar level.

To think about: Many times, say at mid-afternoon, we feel some dizziness and headache, especially when already tired after some demanding work. However, by sheer will we keep on going and soon the headache and weakness vanish. It was hypoglycemia, that is, our sugar level went down. We could have quickly solved the discomfort by taking a cup of tea with a couple of cookies to restore the normal glycemia, but we did not, and the liver sensing the deviation released glucose into the blood stream from its glycogen reserves. Thus, it was an “internal cup of tea” served by the liver. The student is advised to dig further into the meanings of glucose, glycemia, glycogen, and glucostat.

No wonder then that knowledge of the hepatic blood supply is of paramount significance. Based on a simple linear model, similar to that de-

veloped previously for the kidneys, a method for its determination will be outlined (Valentinuzzi, 1971b). Let us remind the definition of a portal system, schematically represented in Figure 2.54, as one that connects two capillary exchangers, in this case the intestinal network with the hepatic sinusoids.

Figure 2.59 depicts succinctly the procedure: The liver is considered a node where the continuity principle can be applied under steady state conditions, that is,

$$\phi_a C_a + \phi_p C_p = \phi_h C_h + Q_o \tag{2.106}$$

where ϕ_a, ϕ_p, and ϕ_h are, respectively, the average blood flows of the hepatic artery, portal vein and hepatic veins, while C_a, C_p, and C_h stand for the corresponding concentrations of the indicator. A fraction Q_o, expressed in mg/min, is the amount of indicator coming out of the biliary duct.

The indicator is constantly infused via a peripheral vein at a rate of Q_i mg/min until the stationary condition is reached when $Q_i = Q_o$, that is,

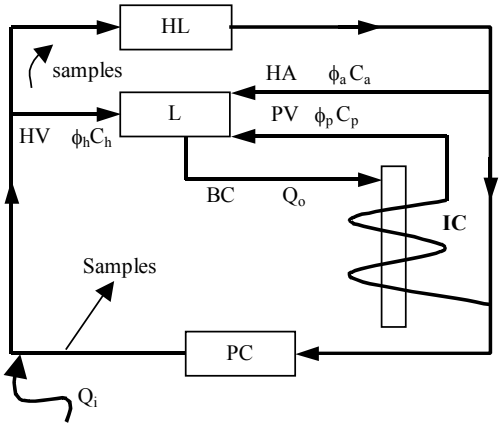


Figure 2.59. HEPATIC BLOOD FLOW DETERMINATION. The heart-lung HL supplies blood through the aorta (above). One of its branches is the hepatic artery HA going into the liver. It carries a flow ϕ_a, in mL/min, which has a concentration C_a of an indicator substance, in mg/mL. The indicator is constantly infused via a peripheral vein at a rate of Q_i mg/min. The intestinal capillaries IC are represented by the loop on the lower right, receiving blood from the arterial supply and converging into the portal vein PV, with flow ϕ_p carrying a concentration C_p of indicator. From the liver exits also the biliary canal BC which dumps bile into the intestine at a rate of Q_o mL/min.

such substance (as sulphobromophtalein sodium) must not be taken up by any tissue and only has to be excreted through the liver to the intestine through the biliary duct. Besides, and for the same reason, the arterial and portal concentrations are constant and equal, that is, $C_a = C_p = C$. As a consequence, equation (2.106) simplifies to

$$(\phi_a + \phi_p)C = \phi_h C_h + Q_i \tag{2.107}$$

from which the hepatic blood flow ϕ_h can be solved for as

$$\phi_h = \frac{Q_i}{[C - C_h]} \tag{2.108}$$

because $\phi_a + \phi_p = \phi_h$. Notice that the latter equation is nothing but the dilution principle already described in the Cardiovascular Section. The numerical value of Q_i is the rate of indicator administration, C is obtained by analyzing blood samples removed from a peripheral vein and C_h requires catheterization of one of the hepatic veins or a surgical procedure. In the dog, a typical value is 35 mL/min per kg of body weight, i.e., a 30 kg animal is expected to have an hepatic flow of about 1 L/min. Since cardiac output is in the order of 8%–9% of body weight (or ≅2.5 L/min, at rest), hepatic flow would represent 40% of the total cardiac outflow. In man, hepatic flow is about 25 to 35% of cardiac output in resting adults.
More detailed information on the liver, especially on its pathologies, can be obtained from Howard J. Worman, Division of Digestive and Liver Diseases, Departments of Medicine and of Anatomy and Cell Biology, College of Physicians & Surgeons, Columbia University, New York, NY 10032, E-mail: hjw14@columbia.edu. He has an excellent INTERNET site.

2.5.5. Exchangers

The gastrointestinal exchanger is highly complex and, actually, is divided in two sections: the intestinal exchanger, with its capillary network, and the hepatic sinusoids or hepatic capillaries, both connected by the portal vein. The portal blood, especially after a meal, is well loaded with nutritive substances (carbohydrates, aminoacids and lipids). Transport from the intestinal lumen to the capillary blood is essential for life. This is the absorptive function of the intestine.
Such function takes place mainly at the level of the jejunum and ileum, with a total length of about 6.5 m in the adult. The caliber goes from 3

cm in the duodenal-jejunal angle, decreasing gradually to 2 cm at the beginning of the large intestine. The ileocecal valve marks the limit between both intestinal sections. After that, the large intestine becomes much larger in diameter. Several hundred grams of carbohydrates are being absorbed per day plus 100 or more grams of fatty acids, including monoglycerides and cholesterol, 50 to 100 g of aminoacids, 50 to 100 g of different ions (such as Na, K, Mg and the like) and 8 to 10 liters of water. However, the absorptive capacity of the small intestine is much greater than these values.

The intestinal mucosa acts as an amplifier. Its surface area is approximately 100 times the skin body surface, that is, it is in the order of 200 m^2 (an adult's body surface area ranges from 1.5 to 2 m^2, typically 1.75). Such enormous contact area is attained by successive convolutions: intestinal loops, mucosal convolutions, intestinal cilia or villi, and epithelial microcilia. Each villus has a capillary network with a small arterial input, a small output venula and a central exit lymphatic vessel, everything in a countercurrent arrangement to improve the exchange efficiency (see the Renal System). Intestinal lymphatics play a significant role in the absorption of fatty acids. Absorptive mechanisms are passive and active and many are still not well understood.

Using the GOOGLE searching engine and the words "intestinal absorption", an interactive computer-simulation of experiments that may be performed on one of the classical *in vitro* preparations — the isolated, everted intestinal sac of the rat — can be found. The authors are David Dewhurst, Jake Broadhurst, Peter and Jacqueline Hardcastle, who are part of the 2000 Sheffield Bioscience Programs, in England (David Dewhurst, d.dewhurst@lmu.ac.uk).

2.5.6. Control and Regulation

Control and regulation of the GI system recognize different origins without existing at all a unique center to collect and integrate the information. There are autonomic nervous factors as well as hormonal and local ones. Also, the central nervous system plays an important role.

The autonomic system — both parasympathetic (with the vagus nerves as major representatives) and sympathetic branches — innervates the stomach and the intestine. Besides, there are two nervous plexuses — Auerbach and Meissner — intrinsic to the gastrointestinal tract. The

plexuses are interconnected and contain nerve cells with processes that originate in receptors in the wall of the gut or the mucosa. They are responsible for peristaltic contractions while coordinated motor activity occurs in the total absence of extrinsic innervation. Generally speaking, cholinergic stimulation (typically of vagal origin) increases intestinal activity while the adrenergic discharge (typically sympathetic) has an opposite effect.

Reflexes play a definite role in GI control and regulation. Let us briefly review the most important. Details can be found in specialized bibliography or in INTERNET:

Deglutition or swallowing. It is controlled via the vagus nerves and a center in the medulla oblongata. Initiation is voluntary after collection of the oral content. Thereafter, a wave of involuntary pharyngeal muscle contractions is triggered that pushes the bolus into the esophagus. Inhibition of respiration and glottic closure are part of the reflex response.

Enterogastric reflex. The entry of chyme into the duodenum stimulates duodenal chemo and mechanoreceptors. This inhibits gastric motility by inhibiting the vagal nuclei in the medulla, thus slowing the transport of material from the stomach. It also activates sympathetic fibers that cause the pyloric sphincter to tighten. The intrinsic plexuses mediate the inhibitory effects, by long autonomic pathways and by the release of several duodenal hormones (such as secretin and other inhibitory hormones known collectively as *enterogastrones*). In other words, since the duodenum is full and has a small capacity, the reflex provides a protective mechanism to prevent further food entry into the small intestine.

Gastroileal reflex. Enhanced activity of the stomach initiates the gastroileal reflex, which is a long reflex that enhances the force of segmentation in the ileum. When the gastric content leaves the stomach, the ileocecal valve relaxes letting chyme get from the ileum — last portion of the small intestine — into the ascending colon. Presumably, this is a reflex mediated by vagal activity.

Gastrocolic reflex. This reflex is the colon's equivalent to the gastroileal reflex in the small intestine, and is responsible for initiating mass movements. Distension of the stomach trigger contractions of the colon and rectum and, frequently, it leads to defecation. Every mother knows well the quick baby's response after nursing or feeding it the bottle. Babies do

not yet have the social inhibitions of the adults. Apparently, the reflex is not neurally mediated and may be due o the action of gastrin on the colon.

Defecation. It is a complex action requiring coordination and sequential activation of a large number of muscles. It is controlled by the autonomic nervous system, but is also under voluntary control. Defecation is initiated by distension of the rectum by feces arriving from the sigmoid colon. This sensation leads to a chain of events that ends in expulsion of feces from the anus. The act of defecation is voluntarily controlled in healthy, normally functioning people.

A mechanism of intestinal blood flow regulation. Blood flow to the intestinal bed is essential for normal activity but highly variable depending on the activity of the individual. After a heavy meal, when there is predominance of the parasympathetic vasodilating discharge, it is highly increased and playing a football game is not advisable nor the subject feels

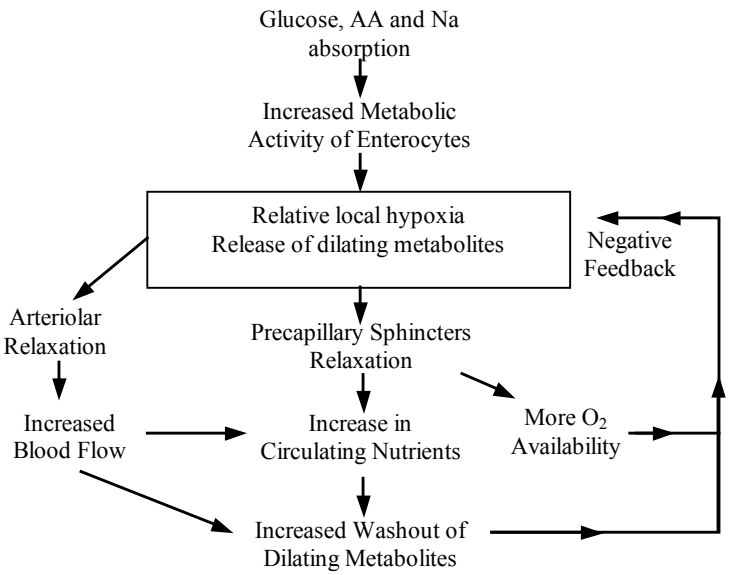


Figure 2.60. MESENTERIC BLOOD FLOW REGULATION. This interesting mechanism links the intestinal absorption of glucose, aminoacids (AA) and sodium to the amount of blood supplied to the region. The relative local hypoxia generated by the metabolic activity of the ciliary enterocytes plus the release of dilating metabolites are central actors in the process.

like doing it. Conversely, when ready to perform a heavy exercise or during it, the sympathetic activity is enhanced with a consequent vasoconstrictor action on the splanchnic region and all the reflexes described above are inhibited.

There is a neat feedback coupling between the absorptive intestinal activity and the amount of blood to the region (Jacobson, 1982). The activity of the ciliary enterocytes during absorption of glucose, aminoacids and sodium generates a relative local hypoxia and the release of dilating metabolites (Figure 2.60). The latter induce arteriolar and capillary presphincter relaxation, which, in turn, increase the blood flow to the intestinal bed. This results in a better circulation of nutrients, more oxygen availability plus the washout of the same initial dilating metabolites. The end result is the compensatory effect of the triggering local hypoxia. Thus, the system keeps balancing from hypoxia to normoxia according to the enterocytes absorptive activity while the arterioles and capillary sphincters change their calibers to regulate the blood supply to the region.

2.5.7. Closing Remarks of Section 5

All the GIS activity aims at the exchange of substances that in the end must sustain viable cells and tissues. The exchanger is actually divided in two separated sections: an absorptive one — intestinal — and a metabolic exchanger — hepatic — the latter storing in its parenchymal cells a number of essential compounds. The portal vein connects both sections. It is easy then to understand the harmful and even lethal effects of the so-called malabsorption syndromes (due, for example, to the lack of certain enzymes or to surgical partial ablation). By the same token, hepatic degenerative processes (as cirrhosis) can also lead to serious and sometimes irreversible conditions.

Since we have here introduced already the concept of *secretion*, it is a good time to better clarify the current recognized types and their respective definitions. They will be useful. They are also referred to as types of chemical communication or mediation. The suffix *–crine* derives from the *Greek* word *krinein*, to separate or elaborate products, and is attached to a prefix that modifies its meaning. Etymologically, ‘secretion’ is the Latin equivalent and has exactly the same meaning. These substances can also be considered as mediators. Thus,

Exocrine secretion or mediation: Since *exo* means 'outside', it is a secretion that goes out, usually via a duct. Saliva, pancreatic juice and bile are typical examples we have mentioned above. The student is encouraged to find other examples.
Endocrine secretion or mediation: Since *endo* means 'inside', it is substances secreted into the blood vessels to act on distant target cells. They are called collectively 'hormones' and the paragraphs above have already introduced a few of them, such as secretin, which was the first to be discovered.
Intracrine secretion or mediation: It regulates intracellular events. Examples are the second messenger systems.
Autocrine secretion or mediation: The prefix means 'self'. They are substances secreted outside the cell that influence the same cell. Thus, there are autoreceptors.
Paracrine secretion or mediation: The prefix here means 'near' or 'from the side of'. They are substances that influence adjacent cells. Neurotransmitters are typical examples.
Ectocrine secretion or mediation: The prefix means 'outside'. They are substances released into the environment to communicate with other individuals. Examples are the pheromones. Smell is usually the detecting organ (a male animal follows the track of a female in heat).
For the time being, learn well the first two concepts and just read the four latter and know they exist. In due time, they will show up again and in more detail. To complement the concepts of this section, there are in the libraries several textbooks of recent editions (Rhoades & Tanner, 1995; Guyton & Hall, 1996; Henderson, 1996; Cormack, 1998; Fox, 1999; Christensen, 2001, the latter via INTERNET).

A last historical tip: William Beaumont (1785–1853) was a pioneer investigator in the area of gastric physiology. A rather attractive and colorful piece of history frame his significant contributions. In 1809, Beaumont began "reading" under Dr. Benjamin Moore. There were few medical schools in the U.S., so it was common to be trained by reading medical subjects under the direction of an established physician. In 1812, the Medical Society of Vermont approved William to practice surgery. That same year, Beaumont enlisted as a surgeon's mate in the U.S. Army. That event in his life marked for many years to come his interests and activities.
On June 6, 1822, in the American Fur Company on Mackinac Island, a French–Canadian voyageur named Alexis Saint Martin was accidentally shot in the upper left abdomen. The

musket wound was “more than the size of the palm of a man's hand,” Beaumont wrote after examining the patient, “and affected part of a lung, two ribs, and the stomach”. Beaumont was unsuccessful in fully closing the hole. The patient was now unable to work as a voyageur, so in April 1823, **Beaumont hired him as the family's live-in handyman** (would you do a thing like this?). A voyageur's job was to paddle a canoe to pick up furs from Indian trappers to deliver to the fur company. The hole in St. Martin's side was a permanent *open gastric fistula*, large enough that Beaumont **could insert his entire forefinger into the stomach cavity**.

It was not until 1825 that Beaumont began his experiments with St. Martin, becoming the first person to observe human digestion as it occurs in the stomach. Beaumont tied pieces of food to the end of a silk string and dangled the food through the hole into St. Martin's stomach. Beaumont pulled out the string one, two, and three hours later, to observe the rate of digestion for the different foods. He even took samples of gastric juice to observe the rate of digestion of a piece of meat, while also placing the same-sized piece of meat directly into the stomach. The stomach digested the meat in two hours; the vial of gastric juice took 10 hours. The experiments shown that gastric juice has solvent properties. Saint Martin returned to Canada, so Beaumont was unable to experiment on him further at this time.

In June 1829, Alexis St. Martin returned to the Beaumonts, this time bringing his family. One set of observations was to try to determine any relation between digestion and weather. By observing St. Martin on different days and times and in varying weather conditions, Beaumont saw that dry weather increases stomach temperature, and humid weather lowers it. He also learned that gastric juice needed heat to digest (cold gastric juice has no effect on food). Beaumont found that vegetables are less digestible than other foods. St. Martin sometimes became irritable doing experiments, and Beaumont observed that being angry can hinder one's digestion. In April 1831, St. Martin and his family left for their home in Canada.

Beaumont located Alexis St. Martin in October 1832 and traveled with him to Washington, D.C., where he again tried different foods. Beaumont focused on gastric juice, but did not study the importance of saliva on digestion. Another limitation on Beaumont's work is that he could not obtain a chemical analysis of the gastric juice, as chemical analysis was severely limited in the mid-nineteenth century. Beaumont published in 1833 his results in a famous book, “*Experiments and Observations on the Gastric Juice and the Physiology of Digestion*”. Sometime later, St. Martin left for Canada; he expected to rejoin Beaumont for more experiments, but as it turned out, St. Martin and Dr. Beaumont never again saw each other.

2.6. Endocrine System

Perhaps most of our acts and responses are driven by the internal secretions.

2.6.1. Introduction

The general and bold block diagram of Figure 2.1 (bold because it is an overt oversimplification) shows in its upper portion a modest inner rectangle, linked to the Nervous System, representing the compound to be developed in this section, the Endocrine System (ES). The task is quite daring, demanding and, at the same time, fascinatingly attractive because perhaps many of the unknowns still hidden in the ES contain the answers to the flows and ebbs of our behavioral misfortunes and states of elation. And the task is aimed at the young fresh mind, just starting the long biomedical engineering road, with a core material that has to be clearly and succinctly explained. On top of it, it must motivate.

It is a system with a high "engineering content". In it, control is omnipotent, showing delicate high sensitivities all over, with extremely low concentrations of hormones in many cases. Its derangements lead always to unhappy endings. We will subdivide it in functional subsystems according to the currently accepted knowledge, defining functions, variables and feedback loops, always to the best of our possibilities. The following sections, (2.6.2 through 2.6.7), include respectively, the Hypothalamic-Hypophyseal Axis (HHA), the Catecholamine System: Adrenal Medulla, the Thyroid-Parathyroid System for Calcium Regulation, the Insulin-Glucagon System: Pancreas, the Renin-Angiotensin-Cardionatrine System, and the relatively recently revisited Pineal Gland and Biological Clock. In the closing remarks, we will try to pose some of the pragmatic and heuristic sides of this basically scientific specialty of physiology.

The Endocrine System controls, coordinates and regulates different functions in the organism, many times in conjunction with the Nervous System. Its actions are relatively slow to take place because the hormones — with their messages — are released into the blood stream by the internal secretion glands, as briefly described in the previous section. Hormones are highly specific and only trigger effects on well-determined targets

(cells or organs). The elicited response is usually proportional to the stimulating hormonal concentration. For more details, we refer to any textbook of physiology, as for example Ganong (1981) or Guyton and Hall (1996), either in the stated editions or in any of their many others, considering always that the newer, the better.

2.6.2. Hypothalamic-Hypophyseal Axis (HHA)

The hypothalamus — a functionally ubiquitous portion of the diencephalon — is the principal center of this group for it skillfully pulls the strings to keep most of the whole system under control. It receives information from higher centers in the brain, it has vascular connections with the adenohypophysis (anterior pituitary) and it has, too, neural pathways linking it to the neurohypophysis (posterior pituitary). However, both hypophyses, while anatomically attached to each other, are embriologically different. Thus, the hypothalamic-hypophyseal axis (HHA) is, in fact, split in two, the hypothalamic-adenohypophyseal (HAH) and the hypothalamic-neurohypophyseal (HNH) axes. Each axis, in turn, gives rise to a system that projects into the whole organism. Hence, from now on we will rather speak of 'systems'. Let us start with the first of them.

2.6.2.1. Hypothalamic-Adenohypophyseal System (HAHS)

It is characterized by a portal vascular arrangement transporting minute quantities of hypothalamic releasing and inhibiting hormones directly to their target cells in the anterior pituitary, that is, they are not diluted out in the systemic circulation. The distance they travel is very short. Specific hypothalamic hormones bind to receptors on specific anterior pituitary cells, modulating the release of the hormone they produce. Thus, some of the neurons within the hypothalamus — neurosecretory neurons — secrete hormones that strictly control secretion of hormones from the anterior pituitary. Hans Selye and Roger Guillemin, in Montreal, put the hypothesis of hypothalamic mediators forward in the early 1950's and that can be considered the beginning of neuroendocrinology (Guillemin, 1998).

The hypothalamic hormones are referred to as *releasing hormones* and *inhibiting hormones*, reflecting their influence on anterior pituitary hormones. In the early years of neuroendocrinology, these hormones were baptized as *releasing factors*, thus, TRH (thyrode releasing hormone)

was TRF (thyrode releasing factor), CRH was CRF, and so on. The hypophysis or pituitary gland (both sections), in turn, is often portrayed as the "master gland" of the body. Such praise is justified in the sense that the anterior and posterior pituitaries secrete a battery of hormones that collectively influence all cells and affect virtually all-physiologic processes.

Other secretions of the hypothalamus, still with unsettled functions, include the melanocyte-stimulating hormones (α-MSH and β-MSH) and the endorphins. In lower species, like amphibians and reptiles, the two former seem to be related to skin coloration. In mammals, however, their function remains uncertain for the time being. The latter, which are morphine-like substances, have analgesic effects. All hypothalamic secretions are polypeptides, proteins or glycoproteins

Exercise: Compare the hepatic portal system with the hypothalamic portal system. Try to find differences, if any. Are there other portal systems in the organism? Search in the literature.

– *Thyroid*

The hypothalamus secretes the thyrotropin-releasing hormone (TRH), which binds to receptors on anterior pituitary basophilic cells called thyrotrophs (or thyrotropes), stimulating them to secrete the thyroid-stimulating hormone (TSH) or thyrotropin. It is a glycoprotein with a molecular weight of approximately 28,000 daltons. This pituitary hormone enters the systemic circulation and binds to their receptors on other target organs. In the case of TSH, the target organ is the thyroid gland (Figure 2.61) that, in response, produces triiodothyronine (T3) and thyroxine (T4). These two hormones affect the metabolism, growth and cell differentiation of practically all the tissues. Two negative feedback loops presumably control the blood concentration of these hormones, one detecting TSH with hypothalamic sensors and another checking T4 with pituitary sensors. Higher centers always influence the hypothalamic action, implicitly meaning that external perturbations may also have an effect.

Iodine is a raw element essential for thyroid hormone synthesis. Ingested iodine (I_2) is converted to iodide (I^-) and absorbed. The minimum daily intake that will maintain normal thyroid function is 100–150 μg in the

adult. Deficiency of the thyroid hormones (hypothyroidism) leads to *cretinism*, in children, and *myxedema*, in the adult. Both are serious conditions. *Goiter* is another form of hypothyroidism triggered by low dietary content of iodine. In that case, since T3 and T4 synthesis and secretion are low, due to the feedback loop, TSH production increases to abnormally high levels and, thus, to a permanent stimulation of the gland which responds with its excessive growth (hypertrophy) in an effort to unsuccessfully make up for the deficiency. The enlargement of the thyroid gland results in bulging of the neck that may become extremely large. Occasionally, it may cause some difficulty in breathing and swallowing. This is why it is common practice to add iodide to table salt (NaCl) as a goiter prevention measure.

Conversely, excess levels in blood of T3 and T4 are generically termed hyperthyroidism. Graves' disease is its most common form. Robert Graves discovered it in 1835. It affects approximately three out of 1,000 people and is more prevalent in women and in families with a history of the disorder. Graves' Disease is an autoimmune disorder in which an as yet unknown immunological defect results in production of autoantibodies to the TSH receptor located on the surface of thyroid cells. These antibodies bind the receptor and stimulate it to overproduce thyroid hormones. This activation is not subject to the normal regulatory negative

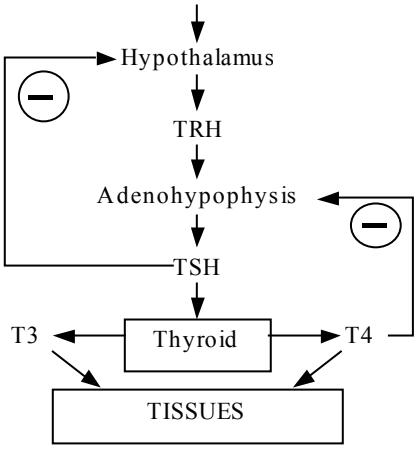


Figure 2.61. HYPOTHALAMUS-ADENOHYPOPHYSIS-THYROID RELATIONSHIP. TRH is a hypothalamic neurosecretion carried by the portal system to the adenohypophysis where it stimulates the secretion of TSH. The latter, in turn, carried also by the circulatory stream has a specific stimulating action on the thyroid gland which, in response, produces triiodothyronine (T3) and thyroxine (T4). These two hormones affect the metabolism of practically all the tissues. Two negative feedback loops control the blood concentration of these hormones, one detecting TSH with hypothalamic sensors and another checking T4 with pituitary sensors.

feedback loop even though the blood level of TSH is lower than normal (as low as 1/100 the level of euthyroid subjects). Symptoms of Graves' disease include nervousness, irritability, weight loss, increased appetite, heat intolerance, excessive sweating, rapid pulse, diarrhea, fine tremors in fingers and warm moist skin. About 50% of patients also develop exophtalmia (bulging eyes).

Just to think about. Roger Guillemin (1998) — Nobel Prize of Physiology in 1977 — states in his touching and clear autobiographic recollections: "I consider the isolation and characterization of TRF the major event in the establishment of modern neuroendocrinology, the inflection point that separated confusion and a great deal of doubt from real knowledge. Contemporary neuroendocrinology was born of that event. Isolation of LRF [the luteinizing hormone-releasing factor, now called LRH or LHRH], somatostatin, the endorphins, others later, were all extensions of that major event — the isolation of TRF — a novel molecule in hypothalamic extracts, with hypophysiotropic activity, the first so characterized. The event was the vindication of 14 years of hard work within the paradigm of a hypothalamic neurohumoral control of adenohypophyseal secretions. From observation of what has happened in neuroendocrinology since 1969, the isolation of TRF was also the vindication of my early decision, as a physiologist, that the most heuristic event in neuroendocrinology would be the isolation and characterization of the first one (any one) of the then-hypothetical hypothalamic hypophysiotropic factors." And he added as a foot note about the cost: "I once calculated that the first 1 mg of native, pure, ovine TRF made 1 kg of pure, native TRF, 2.5 times more expensive than a kilogram of moon rock brought back from the Apollo XI mission. Today the cost of synthetic TRF is a few cents per milligram". After TRF, pioneering in neuroendocrinology ceased and became the harvesting of a new expanding science, because far more interesting and revolutionary observations were to follow. TRH (or TRF) was isolated in 1968 by Guillemin and collaborators in the Department of Physiology of Baylor College of Medicine, in Houston, Texas.

– *Adrenal Cortex*

The adrenal glands, which are also called suprarenal glands, are small, triangular pinkish structures located on top of both kidneys. This gland — essential for life — is made of two parts: the outer region (or the adrenal cortex) and the inner region (or the adrenal medulla). They are quite different from a developmental point of view as they are also functionally. We will consider now the former.

The hypothalamus produces corticotropin-releasing hormone (CRH), which stimulates the adenohypophysis (or anterior pituitary gland) while the latter, in response, produces adrenocorticotropin hormone (ACTH), which in turn stimulates the adrenal cortex to secrete a set of hormones:

aldosterone (a mineralocorticoid) and glucocorticoids (corticosterone and cortisol). Figure 2.62 briefly summarizes the relationships, including three negative feedback regulatory loops. The hypothalamus senses the blood levels of ACTH and of cortisol so that an increase in them elicits a decrease in CRH and in ACTH triggering in this way an opposite compensatory change. Similarly, the adenohypophysis constantly measures the concentration of glucocorticoids (mainly cortisol) to decrease or increase its ACTH production and, hence, make up for any initial increase or decrease in cortisol. Aldosterone activates the retention of sodium and water playing, as a consequence, a role in the electrolytic balance and in blood pressure regulation. Glucocorticoids have an effect on the intermediary metabolism of carbohydrates, proteins and lipids in different tissues. They also suppress inflammatory reactions in the body and affect the immune system. Besides, the adrenal cortex secretes androgenic steroids (androgen hormones). These hormones have minimal effect on the development of male characteristics.

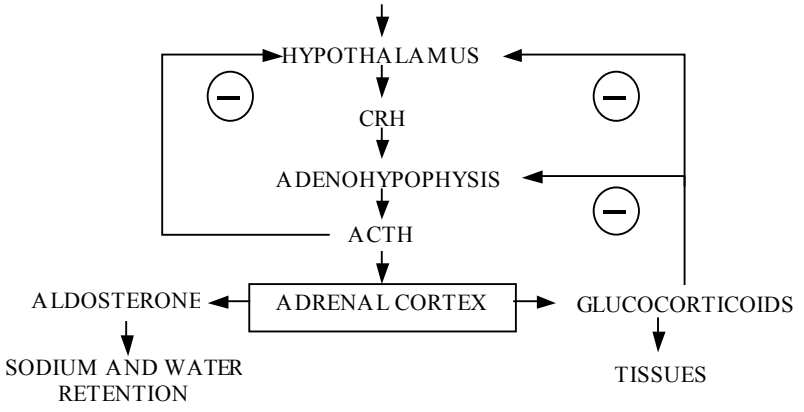


Figure 2.62. HYPOTHALAMUS-ADENOHYPOPHYSIS-ADRENAL CORTEX SYSTEM. The relationship is similar to that of the thyroid gland. Higher nervous centers always influence the hypothalamus. Corticotropin Releasing Hormone (CRH) is another hypothalamic neurosecretion carried by the portal system blood to stimulate the adenohypophysis, which, in response, produces adrenocorticotropic hormone (ACTH). The latter, in turn, via the general circulation, reaches its target organ — the adrenal cortex — to elicit secretion of two types of hormones: a mineralocorticoid (aldosterone), and glucocorticoids (corticosterone and cortisol).

Another recent historical tip: Hans Selye (Guillemin, 1998), through his stress concept as related to the adrenal cortex, had a major stimulating role in orienting the early efforts in neuroendocrinology toward the study of the hypothalamus-pituitary ACTH-adrenal cortex functional relationship. According to Guillemin's comments, strangely enough, and unwittingly on Selye's part, this is probably about the worst thing that happened to nascent neuroendocrinology. The search for CRF was to prove so complex and baffling that it was not completed until 1981 through the elegant work led by Wylie Vale, one of Guillemin's students and collaborators. They should have better put the effort on some other hypothalamic secretion. Perseverance and vision are essential characteristics of an investigator.

There are important pathologies associated with adrenal cortex malfunction: *Cushing's syndrome* (excess production of glucocorticoids), *Conn's syndrome* (excess production of aldosterone), and *Addison's disease* (deficiency of corticoids).

Cushing's syndrome (named after Harvey Williams Cushing, American surgeon, 1869–1939) occurs when the body's tissues are exposed to excessive levels of cortisol for long periods of time. Many people suffer the symptoms of Cushing's syndrome because they take glucocorticoid hormones such as prednisone for asthma, rheumatoid arthritis, lupus and other inflammatory diseases (allergies, for example), or for immunosuppression after transplantation. Because of the negative feedback loop, the higher than normal blood concentration of these hormones inhibits production of the physiological hormones and the gland may stop secretion even after treatment is stopped. It is relatively rare and most commonly affects adults' aged 20 to 50. An estimated 10 to 15 of every million people are affected each year.

The second condition affecting the adrenal gland is Conn's syndrome, also called primary aldosteronism (named after Jerome W. Conn, an American internist, 1907–1981). It is due to the presence of an adrenal tumor, usually benign. The excess aldosterone secreted in this condition increases sodium reabsorption and potassium loss by the kidneys and results in electrolyte maladjustments (Conn, 1955). Risk factors are being female and being between 30 and 50 years old. The incidence is 2 out of 100,000 people; fewer than 10 children have been reported in the literature with Conn's syndrome. Secondary aldosteronism originates in other causes, not directly related to the adrenal cortex.

The third significant pathology of the adrenal cortex is Addison's disease (named after Thomas Addison, English physician, 1793–1860). This

primary adrenal failure may be the result of congenital or acquired lesions. In the early part of the XXth century, tuberculosis was the most common cause. Other infrequent conditions may lead to adrenal cortex damage and insufficiency. However, in recent years it appeared that most cases represent the outcome of an autoimmune process.

For more details about these three pathologies check the website http://www.mc.vanderbilt.edu/peds/pidl/endocr/index.htm, copyright 1998, Vanderbilt University Medical Center, http://www.mc.vanderbilt.edu/cgi-bin/mail?webmasteror Kaplan (1982) or Hughes (1982). The textbooks of physiology usually give some information on these subjects.

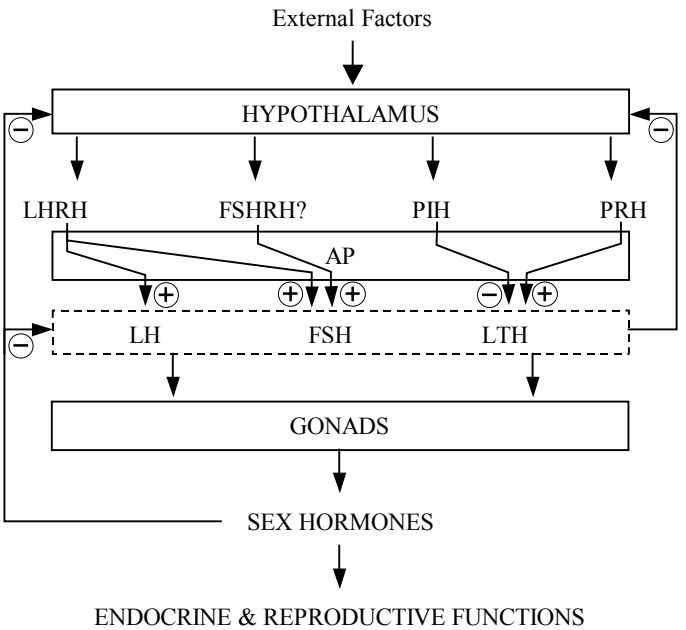


Figure 2.63. HYPOTHALAMIC-ADENOHYPOPHYSEAL-GONADS SYSTEM. Four releasing hormones are produced by the hypothalamus (one of them is still uncertain, as indicated by the interrogation mark). Three are stimulating hormones (thus, a plus sign is shown) and one has an inhibiting effect (negative sign), respectively, on the adenohypophysis or anterior pituitary gland (AP). In response, the latter produces LH, FSH and LTH, which, in turn, having the gonads as target organs, mediate the liberation of the sex hormones to act on the reproductive, developmental and other behavioral functions. The probable negative feedback loops are also shown.

Thomas Addison committed suicide. He was a sensitive, shy, unhappy man who became increasingly despondent during the last two years of his life. There were indications that he suffered from some mental illness (Cirillo, 1985). Science, medicine, is not always associated with happy events. Human beings are the principal actors and this is something a future bioengineer must learn from the very beginning: to deal with men and women and to accept them as they are.

– *The Gonads*: *Sex Glands*

This system follows a scheme similar to the two previously described; perhaps, it is somewhat more complex than them. Figure 2.63 summarizes the hypothalamic-adenohypophysis-gonads relationships. The hypothalamus produces the releasing hormones: luteinizing hormone-releasing hormone (LHRH), follicle-stimulating-hormone-releasing hormone (FSHRH) — still not fully demonstrated — prolactin inhibiting hormone (PIH), and prolactin releasing hormone (PRH). The first one, that is LHRH, stimulates the production of the luteinizing hormone (LH) and of the follicle-stimulating hormone (FSH) from the adenohypophysis or anterior pituitary. This peptide, first characterized and synthesized by Roger Guillemin and Andrew Schally (both shared the 1977 Nobel prize for this labor), has been also called the gonadotropic hormone-releasing hormone, or GnRH (Knobil, 1999). The second, FSHRH, if true, would stimulate the secretion of FSH. In turn, PIH and PRH have opposite actions on the anterior pituitary, the former inhibiting the secretion of luteotropic hormone, LTH, also called prolactin, and the latter stimulating its secretion. Thus, the three hormones from the adenohypophysis are LH, FSH and LTH, with direct actions on the gonads as target organs that, in the end, release the sex hormones. There is evidence of the existence of negative feedback loops to control these hormonal blood levels.

In women, the gonadotropins (collective name of the anterior pituitary hormones aimed at the gonads) have effects on the ovaries (the female gonads, as opposed to the testicles, the male gonads): FSH stimulates follicular growth (thus, it is a trophic hormone), while LH stimulates ovulation and luteinization of the *corpus luteum*. Luteinization is the process by which a postovulatory ovarian follicle transforms into a corpus luteum through vascularization, cell hypertrophy, and lipid accumulation, the latter giving the yellow color indicated by the term (*luteus*, yellow). Both, follicle and corpus luteum, secrete estrogens (estradiol,

estrone and estriol) while the corpus luteum secretes progesterone. All these are termed the female sex hormones, responsible for uterine endometrial modifications, development of the genitalia and secondary female characteristics. Prolactin causes the postpartum milk secretion from the breasts after estrogen and progesterone priming. Besides, prolactin inhibits the effects of gonadotropins, possibly through an action on the ovaries via negative feedback because it stimulates progesterone production. Surprisingly, it was found that TRH also stimulates the secretion of LTH (Guillemin, 1976).

A fascinating phenomenon is the pulsatile nature of GnRH secretion — demonstrated in all vertebrates studied in this regard — that is in turn responsible for LH pulses with a frequency of approximately one per hour. Moreover, it has also been shown that each LH pulse is associated with electrical activity at the mediobasal hypothalamus (Knobil, 1999).

The central event of the female cycle — ovulation — is initiated by a bolus of LH from the pituitary gland. Such surge is superimposed upon, or perhaps temporarily replaces, the pulsatile pattern of LH secretion. Interestingly enough, the pacemaker for this phenomenon is not the hypothalamus but the ovary itself, which, with a preovulatory estradiol rise, acts on the hypothalamic-hypophyseal axis to start the LH surge. It means, then, that estradiol has negative and positive feedback actions (fall in LH and FSH plasma levels and preovulatory surge), both occurring at the adenohypophysis. How this can happen still remains unknown (Knobil, 1999).

In man, LHRH is the only hypophyseal hormone with definite known stimulating action on FSH and LH, for the existence of the other three is still uncertain. FSH acts on the testicular seminiferous tubules to control spermatogenesis. These tubules would produce inhibin, which acting on the adenohypophysis, would inhibit FSH production and, thus, indicating a regulatory loop for sperm production. LH, in turn, has an action on the Leydig's cells in the testicles, main source of testosterone and the essential male sex hormone with androgenic and general anabolic effects. Besides, testosterone stimulates spermatogenesis from the seminiferous tubules and feeds back to the hypothalamus to regulate LHRH release, in another negative control loop.

Exercise and to think about: With the negative feedback loop in mind, explain the conceptual mechanism of the oral contraceptives in women. Remember that these substances are estrogens and progesterone. Most of the women who were prisoners in concentration camps during the Second World War stopped their menstrual cycles. Why? Indicate in the block diagram the signal pathway. Refer to the basic concept introduced by Hans Selye, also valid for the adrenal cortex. Draw separate block diagrams for the hormonal relationships in woman and man. The menstrual cycle is one of the several biological clocks: refer to the pineal gland and try to find links. Ovulation is another still uncertain question: How does it start?

– *Growth*

The Growth Endocrine System (GES) does not have a single target organ for its adenohypophyseal hormones — a systemic gland, as the previously described thyroid or adrenal cortex or gonads. In this case, the hypothalamus controls the adenohypophysis — its main target organ — with two secretions, the growth hormone releasing hormone (GRH) and the growth hormone inhibiting hormone (GIH) or somatostatin, with opposing effects as their names clearly identify, for the former stimulates the secretion of growth hormone (GH) and the latter inhibits its secretion from the anterior pituitary (Figure 2.64).

Growth hormone, also known as *somatotropin*, is a protein of about 190 amino acids that is synthesized and secreted by cells called *somatotrophs* in the anterior pituitary. It is a major participant in control of several complex physiologic processes, including growth and metabolism. It has direct and indirect effects. The former are the result of growth hormone binding on specific cells, such as fat cells (adipocytes), stimulating them to break down triglyceride and suppressing their ability to take up and accumulate circulating lipids. The latter, instead, are mediated primarily by an insulin-like growth factor (IGF-1), a hormone that is secreted from the liver and other tissues in response to growth hormone. IGF-1 stimulates proliferation of chondrocytes (cartilage cells), resulting in bone growth. It also appears to be the key player in muscle growth stimulating amino acid uptake and protein synthesis.

Growth hormone is one of a battery of hormones that serves to maintain blood glucose within a normal range. It has anti-insulin activity because it suppresses the abilities of insulin to stimulate uptake of glucose in peripheral tissues and enhance glucose synthesis in the liver. Somewhat

paradoxically, administration of growth hormone stimulates insulin secretion, leading to hyperinsulinemia.

This amazing system has demonstrated unexpected and not yet elucidated ramifications (Figure 2.64): Somatostatin or GIH inhibits the secretion of thyroid stimulating hormone (TSH) from the adenohypophysis, insulin and glucagon from the pancreas, gastrin from the stomach and secretin from the small intestine.

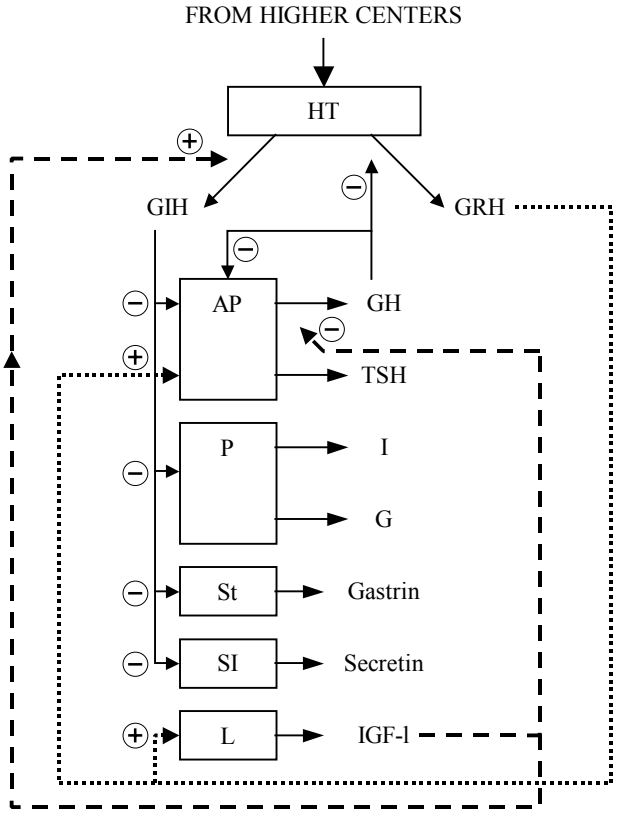


Figure 2.64. GROWTH ENDOCRINE SYSTEM (GES). The hypothalamus (HT), always influenced from higher nervous system centers which includes external stimulation, produces two releasing hormones with opposite actions on the adenohypophysis or anterior pituitary (AP): the growth hormone-releasing hormone (GRH) and the growth hormone-inhibiting hormone (GIH), also called somatostatin. The figure also depicts the relationships with the pancreas, stomach, small intestine and liver. See text for more details.

Besides, growth hormone secretion is also part of a negative feedback loop involving IGF-1. High blood levels of IGF-1 lead to decreased secretion of growth hormone not only by directly suppressing the somatotroph, but also by stimulating release of somatostatin from the hypothalamus. GH also feeds back to inhibit GHRH secretion (Figure 2.64).
States of either growth hormone deficiency or excess provide very visible evidence to the role of this hormone in normal physiology. Clinically, deficiency leads to retardation or dwarfism. Conversely, excessive secretion of GH — very dependent on the age of onset — may lead to two distinctive disorders: *giantism* (very rare, usually resulting from a tumor), the result of excessive growth hormone beginning in childhood or adolescence, and *acromegaly*, which results from excessive secretion of growth hormone in adults.

Guillemin (1998), in his beautiful account of the development of neuroendocrinology, stated that the nature of the hypothalamic-releasing factor for growth hormone (now called GRH) was not to be established until 1982 and in a totally unexpected way, triggered by studies made on a pancreatic peripheral tumor that was functioning as an ectopic source of the much searched factor. This is one example of the fascinating serendipities of science that only the alert and well-prepared mind can catch. Thus, have always ready the reception antennas.

2.6.2.2. Hypothalamic-Neurohypophyseal System (HNHS)

Figure 2.65 briefly synthesizes the pathways involved in this system. The neurohypophysis originates from neural tissue; it stores and secretes two hormones, oxytocin and vasopressin (antidiuretic hormone or ADH). These hormones are synthesized in the cell bodies of neurons located in the hypothalamus and transported along the axons to the terminals located in the neurohypophysis and are released in response to neural stimulation.
Oxytocin acts on the uterine and vaginal smooth musculature and also on the breasts to help sperm transport during sexual intercourse or during child delivery or milk expulsion during nursing. Mechanical receptors located in these anatomical structures send afferent neural information to the hypothalamus that, in a positive loop, enhance the effects by increasing the secretion of oxytocin. The antidiuretic hormone has a direct effect on the renal distal tubules and collecting ducts permeability to produce water retention and, thus, to regulate the extracellular fluid osmolarity.

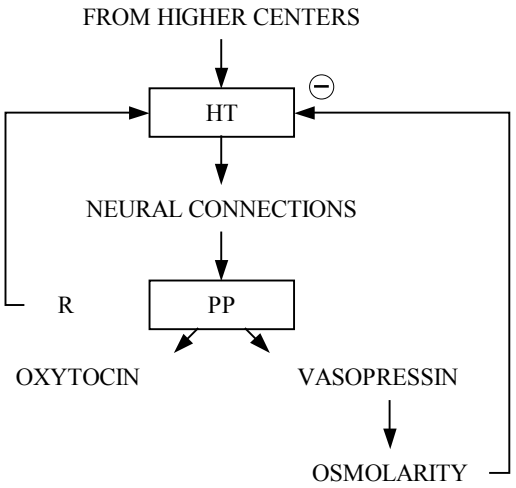


Figure 2.65. THE HYPOTHALAMIC-NEUROHYPOPHYSEAL SYSTEM (HNHS). The hypothalamus (HT), via neural connections, stimulates the secretion of oxytocin and vassopressin (or ADH) from the posterior pituitary (PP), also called neurohypophysis. Mechanical receptors R located in the genitalia, uterus, and in the breasts stimulate hypothalamic detectors to elicit contractions in the smooth musculature of these organs to favor sperm transport during sexual intercourse, birth delivery and milk secretion during nursing. ADH, in turn, has an effect on the extracellular fluid osmolarity that is constantly checked by other hypothalamic receptors, thus, establishing a negative feedback loop.

The latter is being checked by the so-called hypothalamic *osmoreceptors* so creating a regulatory feedback loop.

A typical derangement of this system is *diabetes insipidus*, which is characterized by low ADH secretion leading to *polyuria* (increased diuresis) because much less water is being retained. This urine is diluted and not sweet, as the case is with *diabetes mellitus*, that is, it does not contain glucose. There is also *polydipsia* (drinking of large amounts of fluid), when the thirst mechanism works correctly, to compensate for the renal water loss. Usually, the cause is a hypothalamic lesion. Alcohol induces diuresis because it inhibits ADH secretion.

Detailed and well-versed information can be found in the excellent edited book by Bayliss and Padfield (1985) or in any of the many INTERNET sites.

Thinking exercise: You have seen so far the several negative feedback loops that tend to keep in check different hormonal blood levels. Try to improve the block diagrams by searching more information in the literature. Try to devise ways of measuring the feedback gain, as it is done in technological systems. Probably, you will find many practical problems, but think in theoretical terms. An old paper by Sidney Roston (1959) may serve as inspirational source.

2.6.3. The Catecholamine System: Adrenal Medulla

The adrenal medulla, the inner part of the adrenal gland, is not essential to life, but helps a person in coping with physical and emotional stress. It consists of masses of neurons that are part of the sympathetic branch of the autonomic nervous system. Instead of releasing their neurotransmitters at a synapse, these neurons release them into the blood. Thus, although part of the nervous system, the adrenal medulla functions as an endocrine gland. It secretes epinephrine (also called adrenaline) and norepinephrine (also called noradrenaline). Both are derived from the amino acid tyrosine and are collectively called catecholamines, a group that includes other related substances with similar properties. The former hormone increases the heart rate and force of heart contractions, blood is shunted from the skin and viscera to the skeletal muscles, coronary arteries, liver, and brain, causes relaxation of smooth muscles, and helps with conversion of glycogen to glucose in the liver. Other effects include bronchial and pupillary dilatation, hair stands on end (de so-called "gooseflesh" in humans), clotting time of the blood is reduced, increased ACTH secretion from the anterior lobe of the pituitary. The latter (nor-

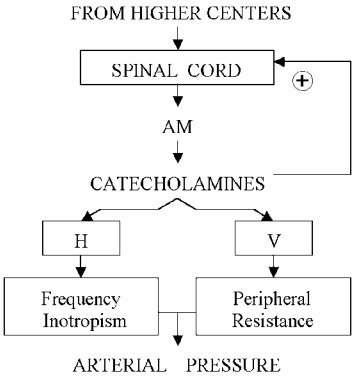


Figure 2.66. THE ADRENAL MEDULLA SYSTEM. The splanchnic nerves via the celiac ganglion innervate the adrenal medulla. These nerves originate in the spinal cord, at the thoracic levels 5 to 12. Thus, stimulation of them produces secretion of catecholamines. The positive feedback loop would be part of the alarm reaction.

epi), instead, has little effect on smooth muscle, metabolic processes, and cardiac output, but has strong vasoconstrictive effects, thus increasing blood pressure. Students are encouraged to visit
http://users.rcn.com/jkimball.ma.ultranet/BiologyPages .
This gland is innervated, via the celiac ganglion, by the splanchnic nerves, which originate in the spinal cord, at the thoracic levels 5 to 12. Thus, stimulation of them produces secretion of catecholamines. Any external stressful perturbation, such as fear or anger, may easily lead to the so-called alarm reaction perhaps enhanced by a positive feedback loop (Figure 2.66). The cardiovascular effects of catecholamines converge to at least a reversible and temporary increase in blood pressure, so explaining that common mother reproach to her children when she says "don't bring me a headache with your behavior". Even getting ready for an action, as before a physical exercise (a race, a given competition, or soldiers in combat) calls for the catecholamine discharge and its physio-

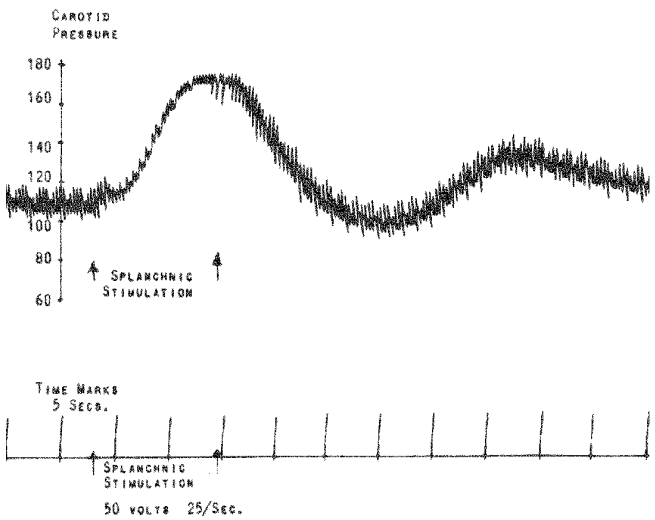


Figure 2.67. SPLANCHNIC STIMULATION in an anesthetized dog. The first increase in arterial blood pressure was due to the direct action of the train of pulses applied to the nerve. However, release of catecholamines was also triggered and their effect was manifested somewhat later, as demonstrated by the second hump of the record about 25 s after the stimulus was disconnected. Records obtained at the Department of Physiology, Baylor College of Medicine, Houston, TX (from the Laboratory Manual by H.E. Hoff and L.A. Geddes, 1965, referred to earlier in the text).

logical effects. There are also demonstrated neural tracts that descend from higher brain centers able to contribute to the phenomenon. Figure 2.67 illustrates the effects of electrical stimulation of the splanchnic nerve in an experimental animal.

2.6.4. The Thyroid-Parathyroid System for Calcium Regulation

Calcium concentration in plasma is well kept at 10 mg/100 mL (or 5 mEq/L or 2.5 mM/L). It is an ion of paramount importance because it plays a role in several essential physiological functions, such as blood coagulation, cardiac and skeletal muscle mechanics, electrical activity of excitable tissues and also in neuromuscular transmission. The largest calcium store is bone.

Three hormones are responsible for calcium homeostasis:

1) 1,25-*dihydroxycholecalciferol* (DHC), which is a steroid formed in the liver and kidneys from vitamin D;
2) *parathyroid hormone* or *parathormone* (PTH), secreted by the parathyroid glands; and
3) *calcitonin* secreted by the thyroid and with no relationship whatsoever with the hypothalamic-adenohypophyseal system.

Figure 2.68 sumarizes calcium regulation mechanisms. Vitamin D, supplied by the daily diet, is essential for DHC production in the liver and kidneys. In the intestine, an increase of DHC stimulates absortion of calcium (second row inset) and, thus, an increase in its blood level (far right inset, lines *b*). DHC and PTH cause calcium resorption (release) from bone, thus, the curves relating it to DHC and PTH increase with an increase in calcium liberation (third row, center inset) and that calcium from bone contributes to an increase in blood calcium; that is represented by the input termed *c* in the figure. In turn, blood calcium (dashed-dotted arrows) acts as input to the thyroid and parathyroid gland, the latter decreasing its PTH production when blood calcium level increase (hence, the inverse relationship shown in the third row, left hand inset) while the former increases calcitonin secretion after the same calcium increase. PTH and calcitonin levels, in a true negative feedback loop, respectively, stimulates and inhibit bone resorption. The second effect is represented by the inverse curve (upper row, center inset). PTH has also stimulating effect in the liver and kidneys to produce DHC.

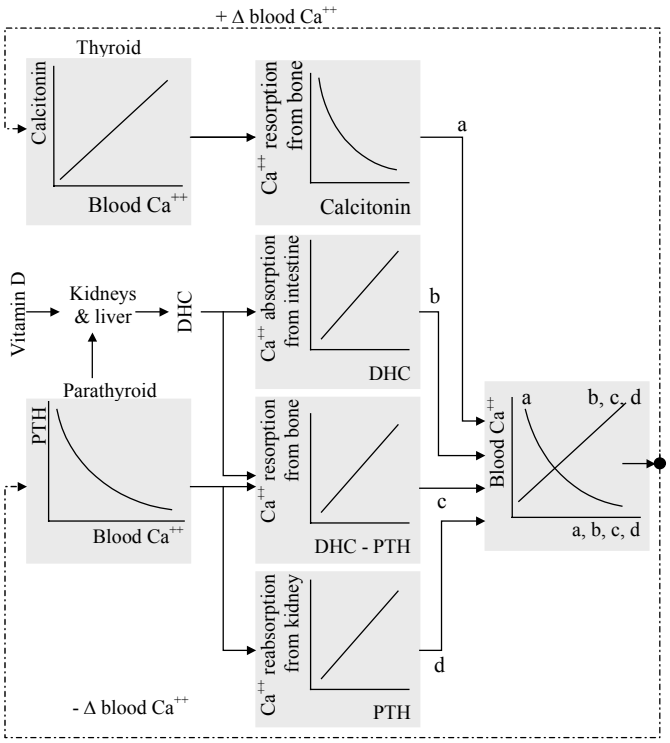


Figure 2.68. CALCIUM REGULATION SYSTEM.

Summarizing: PTH, a polypeptide hormone produced in the parathyroid gland, along with Vitamin D, is the principal regulator of calcium and also phosphorus homeostasis. The most important actions are: (1) rapid mobilization of calcium and phosphate from bone and long-term acceleration of bone resorption; (2) increase of renal tubular reabsorption of calcium; (3) increase of intestinal absorption of calcium (mediated by an action on the metabolism of vitamin D); and (4) decrease of renal tubular reabsorption of phosphate. These actions account for most of the important clinical manifestations of PTH excess or deficiency. Calcitonin, in turn, acts directly on osteoclasts (via specific receptors). Bone biopsies from patients treated with the drug show no effects on mineralization. It has a short half-life. Given as a subcutaneous injection showed significant improvements in bone density but there was a high incidence of side effects, including pain at the injection site, flushing and nausea, which

limited the use of the drug. Calcitonin now is available as a nasal spray, which has made it much more tolerable for patients. Calcitonin is a safe alternative to estrogen in women who cannot take estrogen.

Knowledge of this system is relatively recent (Bronner, Sammon, Stacey *et al*., 1967; Schweitzer, Thompson, Harness *et al*., 1979). Mathematical models have also been proposed (Powell and Valentinuzzi, 1974). One important surgical consequence is that in a thyroidectomy, the parathyroids must be preserved, otherwise derangemenst in the calcium system are to be expected. Another interesting concept is that of the liver as an internal secretion organ.

2.6.5. The Insulin-Glucagon System: Pancreas

In a previous section dealing with the Gastrointestinal System, the pancreas was described as an exocrine gland collaborating in the digestive process. The pancreas, however, is also an endocrine gland secreting four hormones that originate from specific cells located in well-differentiated structures called the *islets of Langerhans*:

a) *insulin*, produced by the *beta* or B cells (make up 60-75% of islets);
b) *glucagon*, secreted by the *alfa* or A cells (make up 20% of islets);
c) *somatostatin* (SS or GIH), secreted by the *delta* or D cells (10% of islets) ;and
d) *pancreatic polypeptide*, secreted by the F cells (just a few percent of islets).

The two first, insulin and glucagon, are essential in the normal regulation of blood glucose, which is maintained at a level of 80 mg/100 mL (or 80 V%). The third one could have an influence in the case of diseased pancreatic cells, such as tumors leading to hyperglycemia. The fourth — sometimes called the "hunger hormone" because it might stimulate the hunger center in the hypothalamus — suppresses SS secretions from gut and pancreas and pancreatic enzyme output, too. However, its function is not known yet. All four are dumped into the **portal blood** and pass through the liver before getting into the general circulation.

The curves shown in the central part of Figure 2.69 represent blood glucose (vertical axis) with insulin and glucagon concentrations, respectively, also in blood (horizontal axis). An increase in blood insulin produces a decrease in the concentration of blood glucose (curve In) while the opposite occurs when blood glucagon goes up (straight line Gl). The

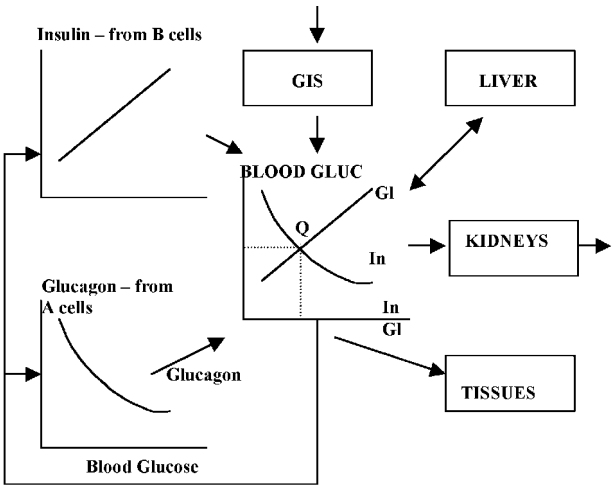


Figure 2.69. GLUCOSE REGULATION. Curves shown in the central part of the figure represent blood glucose in the vertical axis and insulin and glucagon concentrations, respectively, also in blood, in the horizontal axis. The operating point is Q. Inputs act on the horizontal axes and outputs come off the vertical axes. See text for details.

crossing point, Q, of both curves is the operating point. The horizontal dashed line intercepts the vertical axis at the normal blood glucose level. The *beta* or B-cells from the pancreas supply insulin to blood and increase their output as the glucose blood concentration goes up (left upper curve). By a similar token, the lower left relationship represents glucagon release from the *alfa* or A-cells as a function of blood glucose level: when the latter increases glucagon liberation goes down. The latter subsystem also feeds into blood. Glucose enters the system in the daily diet via the gastrointestinal tract and the liver constantly checks the blood glucose level; if needed, because of temporary hypoglycemia, it gives off an amount of glucose to compensate for the fall. Excretion of this sugar takes place through the kidneys when its blood concentration exceeds the threshold of about 180mg/100 mL. Finally, the tissues in general are the users of glucose for their metabolic processes. Thus, blood acts as a dynamic reservoir.

The most important derangement of pancreatic function leads to diabetes mellitus, highly frequent, largely genetic, caused by deficiency or lack of

insulin. Juvenile diabetes is insulin dependent as opposed to diabetes of the elder, which can be taken care of by just a well-adjusted diet. It is a disorder characterized by hyperglycemia or elevated blood glucose (blood sugar). Our bodies function best at a certain level of sugar in the bloodstream (80 to 100 V%). If the amount of sugar in our blood runs too high or too low, then we typically feel bad. **Diabetes is the name of the condition where the blood sugar level consistently runs too high**. Diabetes is the most common endocrine disorder. Sixteen million Americans have diabetes, yet many are not aware of it. Diabetes has potential long-term complications that can affect the kidneys, eyes, heart, blood vessels and nerves. The bibliography is enormous and growing because it is a field of basic and clinical research. The INTERNET sites are countless. It is a very big topic, indeed.

There are two different types of diabetes, which are similar in their elevated blood sugar, but different in many other ways: Type 1 diabetes and Type 2 diabetes. The former is due to insulin deficiency **while the latter stems in** insulin resistance. Insulin deficiency means there is not enough insulin being made by the pancreas due to a malfunction of the *beta* cells. Insulin resistance occurs when there is plenty of insulin made by the pancreas but the cells of the body (the ultimate users) are resistant to its action, which results in the blood sugar being too high.

Glucose is absorbed from the intestines into the bloodstream after a meal. Insulin is then secreted by the pancreas in response to this detected increase in blood sugar. Most cells of the body have insulin receptors that bind the insulin, which is in the circulation. When a cell has insulin attached to its surface, the cell activates other receptors designed to absorb glucose from the blood stream into the inside of the cell. Without insulin, a person can eat lots of food and actually be in a state of starvation since many of his/her cells cannot access the calories contained in the glucose. It is as having a refrigerator full of food but closed with a padlock. This is why Type 1 diabetics who do not produce insulin can become very ill without insulin shots. Insulin is an indispensable hormone. Those who develop a deficiency of insulin must have it replaced via shots or pumps.

More commonly, people will develop insulin resistance (Type 2 Diabetes) rather than a true deficiency of insulin. In this case, the levels of insulin in the blood are similar or even a little higher than in normal, non-

diabetic individuals. However, many cells of Type 2 diabetics respond sluggishly to the insulin they make and, therefore, their cells cannot absorb the sugar molecules well. This leads to blood sugar levels that run higher than normal. Occasionally Type 2 diabetics will need insulin shots but most of the time other methods of treatment will work.

A superb and dramatic piece of history. In 1920, a Canadian surgeon, Frederick G. Banting (1891–1941) visited the University of Toronto to see the newly appointed head of the department of physiology, John J.R. Macleod (1876–1935). The latter had studied glucose metabolism and diabetes, and Banting had a new idea on how to find not only the cause but a treatment for the so-called "sugar disease." Late in the nineteenth century, scientists had realized there was a connection between the pancreas and diabetes. The connection was further narrowed down to the islets of Langerhans. From 1910 to 1920, Oscar Minkowski and others tried unsuccessfully to find and extract the active ingredient from the islets. While reading a paper on the subject in 1920, Banting had an inspiration. He realized that the pancreas' digestive juice was destroying the islets of Langerhans hormone before it could be isolated. If he could stop the pancreas from working, but keep the islets of Langerhans going, he should be able to find the stuff. He presented this idea to Macleod, who at first scoffed at it but finally gave him lab space, 10 experimental dogs, and a medical student assistant, Charles Best (1899–1978). In May, 1921, as Macleod took off for a holiday in his native Scotland, Banting and Best began their experiments. By August, they had the first conclusive results: When they gave the material extracted from the islets (called "insulin," from the Latin for "island") to diabetic dogs, their high blood sugars were lowered. Macleod, back from holiday, was still skeptical of the results and asked them to repeat the experiment several more times. They did, finding the results the same, but with problems due to the varying purity of their insulin extract. Macleod assigned a chemist, James Bertram Collip (1892–1965) to the group to help with the purification. Within six weeks, the purified insulin was tested on a 14-year-old boy dying of diabetes. Insulin lowered his blood sugar and cleared his urine. Banting and Best published the first paper on their discovery a month later, in February, 1922. In 1923, the Nobel Prize was awarded to Banting and Macleod for the discovery, and each shared their portion of the prize money with the other researchers on the project, Best and Collip. Banting initially threatened to refuse the award because he felt Charles Best's work as research assistant had been vital to the project and that he should be included in the honor.

The discovery of insulin was one of the most revolutionary moments in medicine. Though it took some time to work out proper dosages and to develop manufacturing processes to make enough insulin of consistent strength and purity, the introduction of insulin seemed literally like a miracle. One year the disease was an automatic death sentence; the next, people had hopes of living full and productive lives even with the disease. Estimates show there are more than 15 million diabetics living today who would have died at an early age without insulin. Animal experimentation was essential to obtain this knowledge.

The first successful insulin preparations came from cows and later on from pigs. The pancreatic islets and the insulin protein contained within them were isolated from animals slaughtered for food in a similar but more complex fashion than was used by our doctor and medical student duo. The bovine and porcine insulin were purified, bottled, and sold. Bovine and porcine insulin worked very well for the vast majority of patients, but some could develop an allergy or other types of reactions to the foreign protein (a protein which is not native to humans). In the 1980's, biotechnology had advanced to the point where we could make human insulin. The advantage would be that human insulin would have a much lower chance of inducing a reaction because it is not a foreign protein (all humans have the exact same insulin). The technology, which made this approach possible, was the development of recombinant DNA techniques. In simple terms, the human gene, which codes for the insulin protein was cloned and then put inside of bacteria (*Escherichia coli*). A number of tricks were performed on this gene to make the bacteria want to use it to constantly make insulin. Big vats of bacteria now make tons of human insulin. From this, pharmaceutical companies can isolate pure human insulin.

Summarizing: The purpose of insulin is to regulate blood glucose. It forces many cells of the body to absorb and use glucose thereby decreasing blood sugar levels. Insulin is secreted in response to high blood glucose. Conversely, low blood glucose inhibits its secretion from the *beta* pancreatic cells. There is a type of tumor called *insulinoma*, which secretes large amounts of insulin, and thus, it produces severe hypoglycemia. Glucagon, instead, has actions opposite of insulin. It assists insulin in the regulation of blood glucose because it forces many cells of the body to release or to produce glucose. It is secreted in response to low blood glucose while its secretion is inhibited by high blood glucose. Deficiency may lead to hypoglycemia, but not always. Tumors called glucagonomas cause hyperglycemia.

Many mathematical models have been devised to predict responses to given stimulations and, thus, to better dose insulin. The group of Claudio Cobelli, in Padova, Italy, has extensive contributions to the field. Another area of development refers to automatic infusion pumps, either for the hospital environment or for ambulatory patients.

2.6.6. The Renin-Angiotensin-Cardionatrine System

Figure 2.70 outlines in a very simplified manner an accessory system to regulate blood pressure and body fluids. Many concepts are relatively recent and it still poses questions highly attractive to the physiologist, clinician and bioengineer.

Let us consider blood pressure BP as the central variable assuming a momentary decrease in its value. As a consequence, there will be less stretch of the juxtaglomerular apparatuses in the kidneys, located at the afferent arterioles (see Section 2.4). Thus, the hormone called *renin* will be secreted which, as it enters into the circulation, triggers a complex chain of events that end up with the formation of the polypeptide angiotensin II, in itself a potent vasopressor. This simple negative feedback loop tends to compensate for the initial decrease in BP; however, that is

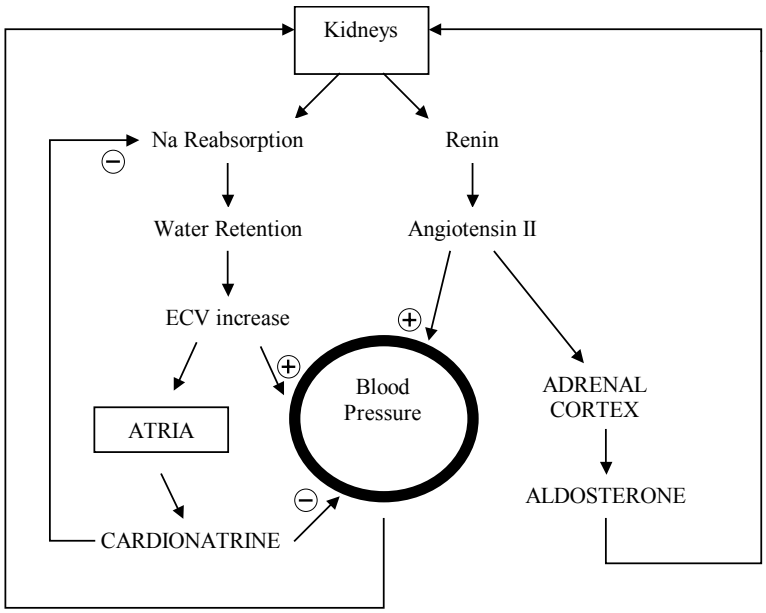


Figure 2.70. RENIN-ANGIOTENSIN-CARDIONATRINE SYSTEM. Angiotensin and cardionatrine have opposite effects. Negative and positive signs indicate activation or inhibition, respectively. It is a slow acting system. Arterial blood pressure is presented as the regulated variable that acts on the juxtaglomerulus apparatuses of the kidneys: when it goes down, they secrete renin. Production of angiotensin II occurs in the blood stream, after a complex chain of biochemical reactions. See text.

not the main pathway, for angiotensin II stimulates the adrenal cortex to secrete aldosterone. The latter acts on the kidneys enhancing sodium reabsorption, water retention and its consequent increase in extracellular volume ECV, which has also a compensatory effect on BP. This is a second negative feedback loop. The atria, in turn, have stretch volume receptors that activate the secretion of the atrial natriuretic peptide or cardionatrine. Surprisingly, the heart appears also as an endocrine organ. This relatively new substance stimulates the kidneys to excrete sodium (natriuresis), opposite to sodium reabsorption. Besides, it has a vasodilating effect, thus, tending to lower BP (Wardener and Clarkson, 1985).
Natriuretic peptides are a group of naturally occurring substances that act in the body to oppose the activity of the renin-angiotensin system. Heart failure is a leading cause of morbidity and mortality. In the United States, there are more than 5 million patients with heart failure and over 500,000 newly diagnosed cases each year. Sodium and water retention play a significant role in this disease. There are three major natriuretic peptides: atrial natriuretic peptide (ANP), synthesized in the atria; brain natriuretic peptide (BNP), synthesized in the cardiac ventricles; and C-type natriuretic peptide (CNP), synthesized in the brain. Both ANP and BNP are released in response to atrial and ventricular stretch, respectively, and cause vasorelaxation, inhibition of aldosterone secretion in the adrenal cortex, and inhibition of renin secretion in the kidneys. Both ANP and BNP cause natriuresis and a reduction in intravascular volume, effects amplified by antagonism of antidiuretic hormone (ADH). Increased blood levels of natriuretic peptides have been found in certain disease states, suggesting a role in the pathophysiology of them, including congestive heart failure (CHF), systemic hypertension, and acute myocardial infarction. For more details, see James A Hill (hillja@medmac.ufl.edu) and B.J. Strickland (stricbj@shands.ufl.edu). Produced by the Office of Medical Informatics, University of Florida College of Medicine
http://www.medinfo.ufl.edu/cme/grounds/hill/intro.html.

De Bold and co-workers, in 1981, first demonstrated that atrial extracts contain a substance that produced natriuresis and diuresis. Soon after that, in 1983-4, the ANP molecule was purified and sequenced. Some years later two other major new peptides of this family were discovered. BNP was originally isolated from porcine brain in 1988.

2.6.7. The Pineal Gland and the Biological Clock

by Veronica S. Valentinuzzi and Max E. Valentinuzzi

> *Time permeates the Universe and timing permeates animal and human life.*

Every day our serum cortisol level peaks at approximately our wake-up time, growth hormone reaches its maximum blood level at the beginning of our sleep phase, prolactin does so soon after, while thyrotropin hormone peaks in the last part of our sleep phase. In the same way, mostly every hypothalamic, hypophyseal and hypophysis-dependent hormones show a fixed temporal pattern with peaks and troughs everyday at the same time and with a stable phase-relationship between them as well as with other types of physiological and behavioral rhythmic variables. For example, blood pressure and core body temperature, subjective alertness and potassium excretion, also vary predictably day by day, with values raising during the day and falling during sleep. Figure 2.71 illustrates the

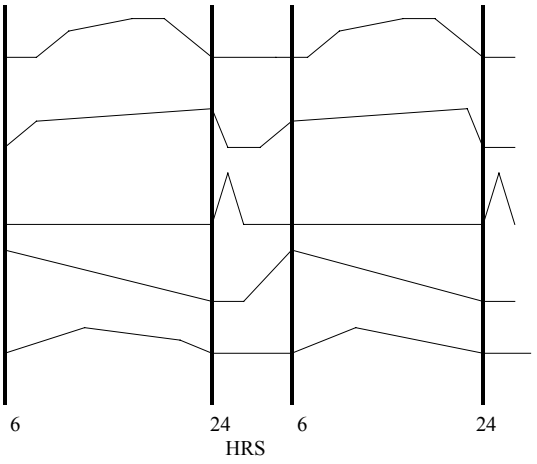


Figure 2.71. PHYSIOLOGICAL CIRCADIAN RHYTHMS of a person sleeping 6 hours per day, from midnight to 6:00 A.M. From top to bottom: Alertness (in arbitrary subjective units); Body Temperature (from about 36 to 37°C); Growth Hormone and Cortisol (relative blood levels); and Potassium Excretion (in flow units), during two day-night cycles.

approximate phase relationship among alertness, body temperature (with a range of about 1°C), growth hormone and cortisol blood levels (shown in relative terms), and potassium excretion through the kidneys in flow units, during two 24 hr cycles. This temporal organization is achieved due to an internal clock with a free-running period of about 25 hours that, in normal conditions, is entrained (i.e., it determines or modifies the phase or period) to the 24-hour solar period resulting from the Earth's rotation.

The period and phases of these physiological parameters must be synchronized with respect to each other as well as with the environmental cycles in order for them to have an optimal effect on body organs and systems. Regular alteration of peaks and troughs is now known to be essential for normal function. Evidence showing that pathological states are associated with abnormal rhythms has accumulated, and there is now substantial support for the hypothesis that disorders of temporal organization may be involved in the etiology of various diseases.

Any biological process that varies predictably with a period in the range of 20–28 hours may be described as a *circadian* rhythm (*circa*, about; *dies*, day). The current theory is that multiple independent oscillators exist in the body and that their periods are locked together with different phases by a pacemaker system resulting in coherent circadian rhythms. This circadian pacemaker system has three major components:

1) the biological clock, generator of circadian rhythms;
2) neural inputs, or entrainers, into the clock from other parts of the brain; and
3) output from the clock to target tissues which couple their internal rhythmicity to that of the clock.

The result of all this is a stable, organized and predictable expression of circadian rhythms. In mammals, the master biological clock is located in the anterior hypothalamus at each side of the third ventricle on top of the optic chiasm, the suprachiasmatic nucleus (SCN). The cells that compose these nucleuses have the capacity to generate rhythmicity in diverse cell parameters due to a mechanism of negative feedback loops of *gene expression* and protein synthesis. *Gene expression* is the process by which the information contained in genes is used to produce proteins or functional RNA (*ribonucleic acid*) molecules.

It is of interest to point out that the molecular mechanism of the clock is essentially the same in every living species, from the fruit fly to human beings. In evolutionary terms, this reveals the importance of a time keeping mechanism for the survival of an organism to an environment that is also essentially rhythmic (the regular alternation of day and night and all the other resulting environmental rhythms such as social cycles, food availability cycles, temperature cycles and the like).

Circadian rhythms are entrained by external factors termed *zeitgebers* (from German, *zeit*, time; *geber*, one that gives) leading to a synchronization or temporal adaptation of our internal clocks with the environment. The light/dark cycle seems to be the most important factor in synchronization of the endogenous circadian rhythms in animals and humans. The importance of this *zeitgeber* is reflected in the neural connections between the retina (photosensitive structure of the eye) and the SCN, the *retinohypotalamic tract* (RHT), as the most obvious afferent connections of the circadian system. Light reaching the eyes every day is transduced into a neural signal and this information travels through the RHT reaching the biological clock. Let us underline that this transduction is done by photoreceptors that are independent from the traditional visual photoreceptors (cones and rods).

A brand-new class of light-responsive cells in the mammalian retina was discovered that connect directly to the brain's clock. First, researchers found a pigment called melanopsin in a small subset of retinal ganglion cells (RGCs) in the eyes of rats. Most RGCs do not respond to light, but it turned out that the **melanopsin-containing ones do**, **making them a brand-new class of previously unknown light-responsive retinal cells**. Researchers traced their connections and found that they hook up directly to the suprachiasmatic nucleus. See, for example, D. M. Berson *et al*. (2002), Phototransduction by retinal ganglion cells that set the circadian clock, *Science*, 295:1070. Several other issues of the same journal during the year 2002 published specific articles on melanopsin, showing the interest attracted by this subject.

With input from the environment, such as the light/dark cycle, the endogenous period is entrained to 24 hours. If the external stimulus is artificially removed by creating a "constant" condition, the organism would continue to show circadian rhythms, but it would have a free-running period consistent with its own internal clock; in humans this is approximately 24.2 to 24.9 hours. Finally, the SCN conveys the endogenous temporal information (in free-running conditions) or the synchronized

temporal information (when in a cycling environment), to the rest of the organism through many neural and neurohormonal paths that guarantee adequate arrival of rhythmic information to brain structures and body organs. A hormonal path could be the rhythmic synthesis in the SCN and rhythmic local secretion of hormones directly into the cerebral spinal fluid as well as in the extra-cellular space. For example, vasopressin is liberated by the SCN only during the light phase. The SCN also sends many neural connections to different brain areas. Most of them terminate directly within the hypothalamus and a few others reach other more distant brain structures. The neurotransmitters secreted rhythmically in these nerve endings may be important time markers for the different brain structures they innervate, including the hypothalamic nucleus that control the endocrine system.

Finally, another path for conveying temporal information is through a relatively complex connection of the SCN with the pineal gland. Under the SCN's command, the pineal's main product, *melatonin* (N-acetyl-5-methoxytryptamine), is secreted rhythmically in the systemic circulation. The abundance of melatonin-binding sites in several brain areas suggests that this hormone is an important coupler of circadian rhythms and endocrine functions. Although many considered the pineal gland a functional vestige just three decades ago, relatively recent research has clearly established it as an integral and important component of the neuroendocrine system. The pineal gland and its hormonal products have been functionally related to virtually every endocrine gland in the organism.

In 1917, Carey Pratt McCord and Floyd Pierpont Allen, at Johns Hopkins University, demonstrated that when the pineal gland was crushed and added to water, tadpoles released in the water would lighten significantly in color, a phenomenon that defied explanation at the time. The pigment of the melanophores (surface pigmentation cells), would concentrate around the nucleus, thus lightening the color of the amphibians' skin. Since this pigment was known as melanine, the unkown substance was called melatonin (melanin + tonus). Melatonin was isolated and its chemical structure determined in 1958 by Aaron B. Lerner, at Yale University. One of the first experiments were to test if this hormone would lighten the skin of mammals as it did in amphibians. It was shown that skin pigmentation in mammals was not altered by this hormone.

The human pineal gland is a very small structure attached by a stalk to the posterodorsal aspect of the *diencephalon* (see Section 7). Melatonin levels are low during the day and high during the night and this rhythm is

dependent on the light portion of the light/dark cycle. Exposure to bright light (2500 lux, but not to the normal room light of about 250 lux) when melatonin levels are highest, quickly causes suppressing of further melatonin production. In addition, phase shifts in the light-dark cycle, as in the case of individuals who take a transmeridian flight, produce phase shifts in melatonin rhythm that take a few days to re-entrain to the post-flight light-dark cycle. Usually, it takes one day per out-of-phase hour to resynchronize, that is, if the passenger travels to a place where the time difference with his/her original place is, say, 5 hours, he/she would need 5 days to readapt to the new environment. Melatonin cycles in old individuals seem to adjust more slowly than in younger subjects. Since readjustment of the melatonin periodicity requires several days, a number of authors feel that *jet-lag* may, in part, be a manifestation of the lack of re-entrainment of the melatonin cycle.

The RHT and its connection to the clock play a critical role in entrainment of this melatonin circadian rhythm. This is accomplished via a long and circuitous multi-synaptic pathway by which light information that reaches the SCN through the RHT ends up reaching the pineal, leading to entrainment. In some blind human subjects, those that have retinal destruction including the melanopsin-containing RGCs, the melatonin rhythm free-runs with a period of 24.7 hours. Under such conditions, peak serum melatonin levels may be achieved at any time during the daily light or dark period.

The production of melatonin within the pinealocytes (the secretory unit of the pineal) requires the uptake of the amino acid tryptophan from the circulation, which, in turn, is obtained mainly from ingested proteins. This amino acid is converted through a series of steps, first to serotonin and then to melatonin. Two of the enzymes essential for this conversion, hydroxyindol-O-methyltransferase (HIOMT) and N-acetylserotonin (NAT), show circadian rhythmicity in their concentration and activity level. Assays that quantify NAT and HIOMT activity are often used in legal medicine to determine the approximate time of death. The activity of these two enzymes at the time of autopsy is higher in the pineal gland of humans who died during the night than in those who succumbed during the day.

Melatonin does not have a single target organ; binding sites have been found in several brain areas, mainly in the hypothalamus and hypophysis. It is clear that melatonin circadian rhythms modulate the hypothalamic-pituitary system in many species. Considerable data exist on its effects on the hypothalamic-pituitary-gonadal system; darkness, or the absence of light perception, stimulates the pineal gland to synthesize and secrete melatonin, which, in turn, has an inhibitory effect on the reproductive system. Rats housed in constant darkness or made blind suffer a delay in sexual maturation, whereas rats kept under constant light undergo premature ovulation. In humans, data are more conflicting, however, there is evidence that the circadian system interacting with the pineal has a role in determining maturation of the hypothalamic-pituitary-gonadal axis and, consequently, in affecting puberty, the stage of life when a person matures sexually. Other studies have indicated that melatonin may help regulate menstrual cycles in women and sperm production in men. A correlation with menopause has also been suggested; higher levels of melatonin can lead to early menopause. Since melatonin mediates seasonal changes in animals, it is natural that melatonin is involved in residual features in human reproduction as well. Humans are non-seasonal breeders, nevertheless, it was found in different regions of the globe that the monthly birth rate was correlated to the environmental light. Time of conception seems to be dependent on the melatonin cycle, which is influenced by daylight intensity. The duration of the night time release of melatonin is longer in winter than in summer, and it is the prolongation in the duration of the night time release of melatonin with the change of season from summer to winter which acts as the endocrine signal for inactivating the hypothalamic GnRH pulse generator (see above).

The pineal gland also plays a role in the production and release of growth hormone by the pituitary. Rats blinded before puberty exhibit a lag in bodily growth due to a decrease in growth hormone, however, if the blinded rats are pinealectomized (surgical removal of the pineal gland), growth hormone level is normal. Similarly, melatonin inhibits the synthesis and release of hypothalamic thyrotropin-releasing hormone. High-dose administration of melatonin in the light period leads to a decrease in the levels of T3, T4 and TSH, while pinealectomy restores thyroid hormones to normal.

Melatonin is readily available in drug stores and health food stores, and it has become quite popular as dietary supplement being used for everything from cancer to menstrually related migraine and Alzheimer's disease. Ingesting even modest doses of melatonin raises the melatonin level in the blood to as much as 100 times greater than normal. Apparently, it tends to promote sleep and thus relieve insomnia, and to hasten recovery from jet lag. It plays a role in many physiological and behavioral processes. Melatonin is an important mediator of external light to physiological adaptation to day and night rhythms; it significantly participates in neuroendocrine and reproduction regulation. There is also evidence of its participation in other processes such as aging, calcium homeostasis and immune regulation. Changes in the melatonin plasma rhythmicity and concentration have been correlated with many clinical symptoms. For example, in bulimia, neuralgia and in women with fibromyalgia as well as in several forms of depression. Attention in melatonin has increased significantly in the last 3 years with more than 2000 publications including books directly related with its clinical applications. It has the promise to improve the quality of life of many patients, however, much more research is necessary to assure security in long-term use treatments. Most melatonin investigators consider that this hormone should not be sold over the counter until a better understanding of its physiology has been reached.

Suggested bibliography for the interested student can be the papers by Cardinali and Pévet (1998), Rosato and Kyriacou (2001), Wright (2002) and Albrecht (2002), or just to visit the several websites dealing with the subject.

Veronica S. Valentinuzzi got the undergraduate degree in Zootechnical Engineering (1989) from the Universidad Nacional de Tucumán, Argentina, and the Master of Sciences in Biology (1993) and Ph.D. in Sciences (1999) from the *Universidade Estadual de Campinas*, São Paulo, Brazil, the latter with a dissertation carried out in Northwestern University, Evanston, IL, under the supervision of Dr. Fred Turek, in the field of chronobiology and learning, supported by fellowships from the *Conselho Nacional de Pesquisas do Brazil*. Currently, she is a Research Associate with the Department of Physiology at the *Centro de Biociências*, *Universidade Federal de Rio Grande do Norte*, Natal, Brazil. Veronica has published several papers in her specialty.

2.6.8. Closing Remarks of Section 6

The reader of this section supposedly has now an idea of the Endocrine System. We have covered all its parts showing the scientific basic facts, some historical events associated with the development of the subject plus a few practical possibilities. The understanding and treatment of endocrine diseases have benefited to unthinkable levels using knowledge obtained within the last few decades. Its future looks bright, indeed.

One of the fields where endocrinological advances are going on at fast steady pace is biotechnology, as for example, in the food industry. In China, the hypothalamic hormones are applied to breed carps, a traditional fish in the Chinese diet which used to require a significant human effort to collect it in the mountains. Now, it is conveniently harvested in well-installed breeding sites. Improvement has been attained both quantitatively and qualitatively (Guillemin, 1998).

Lam, Hwang *et al.* (2002) have developed optimal and robust control for the automated infusion of insulin to Type I and Type II diabetic individuals. Wireless technology, insulin pumps, embedded computation and emerging non-invasive glucose monitoring systems may be interconnected and a closed-loop control system realized. Ultimately, an effective control system should be able to automate over 95% of diabetic individuals' regular insulin care, freeing them from significant stress and to better handle any exception that arise. The potential for future implementation of an implanted artificial pancreas is investigated in terms of the feasibility of the interface and the trade-offs between control performance and sensor and actuator limits.

Research in circadian systems and pineal gland has also led to practical applications. For example, elite athletes, by following simple chronobiological prescriptions, can adapt themselves in just two days to as much as 12 hour *jet lags* so giving them better chances for success in high performance competitions, as already demonstrated by several authors. This procedure has been tested also in well-bred dogs, like German Shepherds, when they are transported to compete in international championships. Otherwise, the animals would not respond as expected to the master's commands.

The young bioengineer no doubt will find in endocrinology, should he/she be attracted to it, a potential area of development either in basic or applied research, in the clinical setting or in biotechnology.

2.7. Nervous System

The very information system, still a mystery, as vast as the Universe itself.

2.7.1. Introduction

In the block diagram of Figure 2.1 we introduced the nervous and endocrine systems as the Central Processing Unit, receiving afferent signals from the different regions of the body, integrating them, taking decisions of all sorts, sending off efferent signals and coordinating activities. The preceding Section 2.6 dealt with one part of it, relatively slow in its actions. This Section 2.7 will describe the fast acting portion, where the voluntary decisions take place along with other non-voluntary actions. Thus, the main pathways with their central and peripheral sections will be briefly introduced followed by the voluntary and autonomic divisions. Some anatomic and cellular characteristics will be given underlining the *synapse* as a unique functional one-way unit.

The Nervous System integrates internal and external information having the capacity of storing it. As such, it is the place of memory — a marvelous, ubiquitous and still unknown property of so many representatives of the Animal Kingdom — that may reach unpredictable and astounding levels in the human being. The electrical action potential is the messenger for transmitting information via the complex network of peripheral nerves that reach all central and distal places of the organism. Once the information reaches the integration center, it enters in a complex neural array interconnected through *synapses* (from Greek, meaning junction, connection). The number of synapses in the brain may be in the order of 10^{13} to 10^{14}.

Each specific center of the Central Nervous System (CNS) is, thus, a neural network with multiple synapses interconnecting the neurons, which are the basic cells of this system, formed by a cell body (the

soma), a long projection from it called *axon*, and many smaller projections, the *dendrites* (from Greek, *dendron*, meaning "tree"). The latter connect via synapses to other axons, somas or dendrites. The total number of neurons is about 10^{10}, and probably even more. Not much is known of these networks. Many theories and models have been developed inspired in whatever knowledge from them is available giving rise to the method called Neural Networks, widely applied in biomedical engineering in a variety of problems. Just a quick browsing of the *IEEE Proceedings of the Engineering in Medicine and Biology Society*, say of 2002 or 2003, which were the 24th and the 25th conferences, respectively, will offer the student an overview of recent possibilities already under study, such as in neuromodeling, neuro-robotics, neuroelectromagnetic imaging and source localization, fuzzy logic and neural networks for biological signal analysis, neural signals and analysis, neural prostheses, neuromechanical systems, neural microdevices and interfaces … The list seems to be endless and far reaching.

The philosophy of the section is similar to what we already developed: it is introductory, incomplete in many respects, interspersed with historical notes and a few suggested study subjects. No doubt, it represents another challenging field for the young future bioengineer.

2.7.2. Central Centers and Segmental Levels

Basically and very schematically, the CNS is composed of two main sections: the *encephalon* (located within the cranium or skull) and the *spinal cord* (located within the articulated spinal column), as a protruding and segmented long tail. The upper section receives information from the external world via the traditional five senses (vision, audition, taste, smell and touch) while the lower long portion gets its afferent inputs from the many sensors placed all over the different organs and structures (Figure 2.72).

Three major neural avenues composed of long nerve fibers interconnect both sections: the **pyramidal**, the **extrapyramidal** and the **sensory tracts**. Responses from the different segmental levels of the spinal cord constitute efferent outputs.

The pyramidal system originates in the anterior region of the cerebral cortex central fissure known as the *motor area*. It is an efferent system. Its stimulation produces discrete movements in any skeletal muscle. Vol-

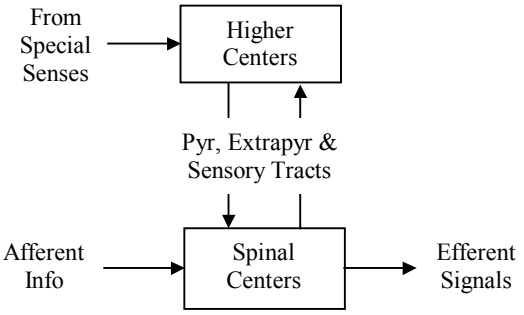


Figure 2.72. HIGHER AND SPINAL CENTERS. Three main tracts interconnect them. The pyramidal channel controls voluntary movements. See text.

untary movements follow this path and injuries in this area lead to either total or partial paralysis depending on the extension of the damage. *Hemiplegia* (paralysis of one side) and *hemiparesia* (weakness of one side) represent the two typical and rather common human conditions after a cerebrovascular stroke (CVS). Since there is a crossover from one side to the other (the *pyramidal decussation*), the left hemisphere controls the right side of the body and vice versa. Classical experiments in monkeys and during cerebral surgery in man have mapped the so-called *homunculus*, which is a motor representation on the cortex.

Study subject: Dig a little deeper in the *homunculus*. What anatomical areas are largest in it and why? How would you design the experiment to obtain the map? Search for the decussation of the pyramids. Is there a full or 100% crossover? For those mathematically inclined, think in terms of two domains, the body skeletal voluntary muscles and the motor cortical area, of fully different shapes, but functionally linked by the nervous tree in a predictable way. Would it be possible to find a mathematical relationship? Should it exist, would it be of any use? The latter question poses a research non-trivial subject with an open answer.

All other pathways that are neither pyramidal nor cerebelar related to movement are, by definition, named extrapyramidal. It is a complex system that includes several structures. Extrapyramidal activity is integrated at different levels, from the very cerebral cortex to the spinal cord. Many of its actions have not yet been elucidated. This system is involved in automatic motor movements and in gross rather than fine movement. It works with the autonomic nervous system to help with posture and mus-

cle tone and has more influence over midline structures than those in the periphery. The extrapyramidal system dampens erratic motions, maintains muscle tone and trunk stability. It is *phylogenetically* older that the pyramidal system and, thus, plays a relatively more important role in lower animals. Many of its synaptic connections are extremely complex and even today, poorly understood. Neurodegenerative disorders that affect the extrapyramidal system have yielded much of our knowledge about its normal function. Facial expression is one important communicative behavior that is mediated by the extrapyramidal tract. This is the reason that some Parkinson patients have little facial expression. In contrast to the pyramidal tract, the extrapyramidal tract is an indirect, multisynaptic avenue. Components of the extrapyramidal tract include the *basal ganglia*, the *red nucleus*, the *substantia nigra*, the *reticular formation* and the *cerebellum*. All of these structures send information to lower motor neurons.

Sensory information (cutaneous and deep visceral signals), not the special senses because they follow specific paths of their own, constitutes an afferent route with a definite cortical representation forming another homunculus. The latter lies on the posterior region of the central cerebral fissure. Peripheral stimulation, say on the big toe or on the lips, will evoke action potentials on that cortical area. The technique of the evoked potentials is widely used, both in research and in the clinical setting. When carried out with external electrodes, it calls for very specific signal processing and filtering in order to recover and isolate the desired action potential.

Study subject: Dig a little deeper in the *sensory homunculus*. What anatomical areas are largest in it and why? Appearance of the evoked potential after stimulation indicates integrity of the pathway. What kind of neurological diseases would justify the application of this technique? A mathematical study, similar to the one suggested above, might be attempted, too. Dig into the concept of *phylogeny*. What does it mean?

The spinal centers receive afferent information of different types from the different segmental levels (cervical down to coccygeal) and have efferent outputs from the same levels. The dorsal roots of the spinal cord are sensorial or afferent and the ventral roots are efferent or motor. This anatomical rule is known as Bell–Magendie law, after Charles Bell (1774–1842) and François Magendie (1783–1855), English and French

physicians respectively who described it. For years there was a controversy over who first produced it (Brazier, 1959). The typical reflex arc well exemplifies a segmental circuit: Say that while walking barefooted, we step on a sharp little stone. The painful stimulus is detected by pain receptors, which transmit the signal to the corresponding spinal segment, where it is integrated eliciting a motor efferent outflow to the leg muscles in order to quickly lift the foot up and avoid or minimize injury. This is a withdrawal protective reflex response, fully involuntary. The number of nerve fibers entering into the spinal cord is in the order of 3×10^6.

Study subject: Search for the patellar reflex, that one elicited when the proper knee site is mildly hit with a little hammer and the leg responds with a kick. Analyze the path and the components involved in it. Make a drawing to better visualize it.

Spinal segmentation is better manifested in lower animals, such as fish, amphibians or reptiles. In man, the phenomenon is less marked even though it clearly shows up in the paraplegic and in the quadriplegic subject, that is, in persons who suffered lesions (full or partial sections), low or high, in the spinal cord, respectively. Car and motorcycle accidents and wars are, unfortunately, a significant cause of this kind of patients in young people. They are candidates to participate as active members in Rehabilitation Engineering research and development groups, which have become a leading specialty of Biomedical Engineering.
Spinal nerves are connected in pairs to the spinal cord: 8 cervical (C), 12 thoracic (T), 5 lumbar (L) and 1 coccygeal (Cg). Each has dorsal and ventral roots, according to Bell–Magendie law. They distribute to the head posterior region, the neck and upper extremities (C); to the upper limbs, abdomen and back (T); to the back and lower limbs (L); to the lower limbs, genitals and glutei (Cg).

Study subject: Find out what the *dermatomas* are. Can a quadriplegic patient with a high-level injury breath by himself/herself? Explain. Find out about the concerns of Rehabilitation Engineering and important centers devoted to the specialty.

The higher centers located within the skull are usually divided in three basic and complex parts: *proencephalon* (or anterior cerebrum), *mesencephalon* (or medial cerebrum), and *rhomboencephalon* (or posterior cerebrum). They are well interconnected receiving and sending off information from and to different regions.

The first one, also called forebrain, contains the *telencephalon* (cerebral cortex, *rinhoencephalon*, *striated body* and *neopallium*) and the *diencephalon* (*thalamus*, *hypothalamus*, *metathalamus* and *epithalamus*). The former is responsible for the initiative actions, memory and conditioned reflexes, while the latter handle many locomotor reflexes and emotional responses. The hypothalamus got particular attention in the previous section, however, it is also responsible for other functions: temperature regulation, thirst and appetite control, sexual behavior, defensive reactions, circadian rhythms and the so-called motivation. The latter is not well understood, but it refers to the inner driving force that leads an individual, such as an artist or scientist, to work to exhaustion forgetting essential needs, even in the face of poverty or highly stressful obstacles, including disease.

The history of science, technology and arts is full of examples of this kind (such as Marie Curie, Michael Faraday, Vincent van Gogh; Ludwig van Beethoven). It would not hurt if the student reads about these personalities and tries to find other similar examples. They have been by far more beneficial to mankind at large than many supposedly famous generals and politicians.

The mesencephalon or mid-brain consists of, (1) a ventrolateral portion, composed of a pair of cylindrical bodies, named the *cerebral peduncles*; (2) a dorsal portion, consisting of four rounded eminences, named the *corpora quadrigemina***;** and (3) an intervening passage or tunnel, the *cerebral aqueduct*, which represents the original cavity of the mid-brain and connects the third with the fourth ventricle. Several locomotor reflexes are integrated here.

The posterior cerebrum or brainstem houses the *cerebellum*, *pontine protuberance* or *pons* and *medulla oblongata*. The former takes care of posture and coordination; the second structure participates in the arousal of the individual, some autonomic functions and relays sensory information between cerebrum and cerebellum, while the latter is the place of the respiratory and cardiovascular centers.

From the cranial cavity twelve important pairs of nerves come out carrying afferent and efferent information: the olphactory nerve (or pair I), the optic (or pair II), the common oculomotor (or pair III), the pathetic trochlear (or pair IV), the trigeminal (or V), the external oculomotor (or

VI), the facial (or VII), the auditive (or VIII), the glossopharyngeal (or IX), the vagus (or X), the spinal (or XI) and the hypoglossal (or XII).

Study subject: Identify the specific functions of the above-mentioned cranial nerves and find out which carry only afferent information, only efferent information or both. The two vagii, for example, carry both types of information.

The term *basal ganglia* refer to a group of five subcortical nuclei that play an important role in the regulation and coordination of cortically originated movement. These nuclei include: the *caudate nucleus*, *putamen*, *globus pallidus*, *nucleus accumbens septi*, and *olfactory tubercle*, the latter two are collectively called the *ventral striatum*.
The student should search for more details about all these structures, although they will be covered in other subsequent specific courses.

2.7.3. Reticular Activating System (RAS)

This is a diffuse and an ill-defined system that partially includes structures such as the diencephalon, mesencephalon, and rhomboencephalon down to the medulla oblongata. Its function is related to sleep. The lower portions have excitatory (mesencephalon) and inhibitory (pons and medulla oblongata) actions over the upper diencephalic area, leading to characteristic sleep responses when lesions or stimulation of these regions take place.
This unique system seems to be also related to the attention capacity of the individual, to the ability of associating ideas and to that attitude or behavior called *introspection*. All these apparently unrelated responses are important and sometimes determining components of human adaptation. The RAS is very capable of generating dynamic effects on the activity of the cortex, including the frontal lobes, and the motor activity centers of the brain. It plays a significant role in determining whether a person can learn and remember things well or not, on whether or not a person is impulsive or self-controlled, on whether or not a person has high or low motor activity levels, and on whether or not a person is highly motivated or bored easily. The Reticular Activating System is the center of balance for the other systems involved in learning, self-control or inhibition, and also motivation; when functioning normally, it provides the neural connections that are needed for the processing and learning of information, and the ability to pay attention to the correct task. The so-

called Attention Deficit Hyperactivity Disorder (ADHD), frequently seen in school children and also in adults, might find its roots in this system.

2.7.4. The Autonomic System

It is absolutely involuntary. We cannot constrict our skin and muscle vessels to favor blood circulation in the guts or commit suicide by stopping respiration. These actions are carefully handled irrespective of our wishes by the *autonomic system*, which conveys sensory impulses from all of the organs in the chest, abdomen and pelvis through nerves to other parts of the brain (mainly the medulla, pons and hypothalamus). These action potentials often do not reach our consciousness, but elicit largely automatic or reflex responses through the efferent autonomic nerves, thereby eliciting appropriate reactions of the organs to variations in environmental temperature, posture, food intake, stressful experiences and other external changes.

There are two major divisions, the *parasympathetic* and the *sympathetic* sections. By and large, they have opposite effects. The afferent nerves subserving them convey impulses from the body organs to the controlling centers. From these centers, efferent signals are transmitted to all parts of the body by the parasympathetic and sympathetic nerves. The impulses of the parasympathetic system reach the organs of the body through the cranial nerves 3, 7, 9, and 10, and some sacral nerves, to the eyes, the gastrointestinal system, and other organs. The sympathetic nerves, in turn, reach their end organs through more complex pathways down the spinal cord to clusters of sympathetic bodies (ganglia) alongside the spine where the messages are relayed to other neural pathways that travel to all parts of the body.

Like other nerves, those of the autonomic nervous system send their messages to the end organs by releasing transmitter substances to which the receptors of the target cells are responsive. In the parasympathetic system, *acetylcholine* is responsible for most of these transmissions. In the sympathetic division, instead, information is transmitted by the release of *norepinephrine* (*noradrenaline*). There are, nonetheless, important exceptions to this general rules.

Let us illustrate briefly with some examples. When a stimulus arises in an organ, such as a bright light into the eyes, the message is conducted through sensory fibers to the midbrain to give rise to a stimulus that trav-

els through the parasympathetic fibers of the oculomotor nerves to the pupils, resulting in automatic contraction of the pupillary muscles to constrict the aperture and so reduce the amount of light reaching the retinae. Similarly, the stimuli associated with the entry of food into the stomach are conveyed by afferent fibers of the vagus nerve to the command station of the vagus in the brain, whence messages are automatically sent through efferent vagal fibers back to the stomach. These stimulate the secretion of gastric juices and peristaltic contractions to mix the food with the secreted digestive juices and gradually shift the gastric contents into the intestines where a similar process is initiated through essentially the same parasympathetic nerve pathways. Fortunately, emptying of the rectum and of the urinary bladder is not entirely automatic, but is subject to parasympathetic impulses that are voluntarily controlled.

The sympathetic nervous system is even more automatic and only exceptionally susceptible to any voluntary control. When the environmental temperature is raised on a hot summers day, the increased temperature initiates several responses. Thermal receptors send stimuli to control centers of the brain from which inhibitory messages travel along the sympathetic pathways to the blood vessels of the skin resulting in dilatation of blood vessels, thereby greatly increasing blood flow to the surface of the body from where heat is lost by radiation. Dilatation of the blood vessels in this way tends to lower the blood pressure and to promote oozing or transudation of fluid from the capillaries, which may result in swelling of the dependent limbs. Thus, fine adjustment in sympathetic control of vascular contraction and "tone" is required to prevent excessive vascular dilatation and undue reduction in blood pressure. Otherwise, this might result in severe gravitational pooling of blood in the lower limbs, thereby reducing blood flow to the brain and causing fainting spells. The sympathetic nervous system responds to environmental heat in another important way: The rise in body temperature is sensed by the hypothalamic center from which stimuli emanate via sympathetic nerves to the sweat glands, resulting in sweating. This serves to cool the body by the loss of heat resulting from evaporation of the sweat.

Control of the rate and strength of cardiac contractions is also under the control of the autonomic nervous system. Thus, a fall in blood pressure resulting from traumatic injury causing blood loss is sensed by pressure-

sensitive parts of the arteries called *baroreceptors*. Evidence of reduced arterial distension is sensed by these baroreceptors and conveyed by the parasympathetic (mainly the glossopharyngeal) nerves to the cardiovascular control center in the medulla. From these nuclei, sympathetic stimuli transmitted by the cardiac nerves cause acceleration of the heart rate, complemented by simultaneous reduction in the parasympathetic stimuli via the vagus nerves which slow the heart rate. For example, a stimulus to contraction of the blood vessels is required in order to maintain the blood pressure when we arise from bed in the morning, so as to prevent fainting from excessive pooling of blood in the lower body. This stimulus is conveyed by norepinephrine release within the walls of the blood vessels from the nerve endings of the sympathetic nerves that innervate each blood vessel.

Dysautonomia indicates malfunction of the autonomic nervous system. It can be a serious and even terminal disease. Some of the specific disorders that fall within the group are Postural Orthostatic Tachycardia Syndrome (POTS), Neurocardiogenic Syncope, Mitral Valve Prolapse Dysautonomia, Pure Autonomic Failure, and Multiple System Atrophy (Shy–Drager Syndrome).

To get more details, visit the website by Dr. David H.P. Streeten, from SUNY Health Science Center, Syracuse, NY 13210, who graciously presents a very didactic section on this subject.

2.7.5. The Synapse

The word "synapse" refers to the site of functional apposition of two neurons. We already stated that the nervous system is formed by neurons that control all the conscious and unconscious activities of our life. The neurons have a central body or *soma* and short projections called *dendrites*. Besides, there is a long projection called *axon*. With these two projections, neurons make connections, or synapse, with other neurons.

The nervous impulses are generated in the neural soma and propagate along the axon at high speed (3 to 50 m/s). We have dealt with this conductive property above, in the Electrophysiology Section. When the axon synapses, that is, connects to another neuron, it makes possible the transmission of the message originated in the cellular body. Such transmission is **unidirectional**. Specific substances (chemical transmitters) are released at the pre-synaptic side, which, after traversing the space

between the two neurons, reach the post-synaptic side where there are receptors sensitive to the transmitter. Thus, a synapse means a discontinuity from an histological point of view. Typical chemical transmitters are *acetylcholine*, *serotonin*, *noradrenaline* (or *norepinephrine*) and *dopamine*. Acetylcholine was the first known transmitter and is one of the most important.

The process of transmission of the action potential through a synapse can be summarized in the following steps: The neurotransmitter is manufactured by the neuron and stored in vesicles at the axon terminal. When the action potential reaches the axon terminal, it causes the vesicles to release the neurotransmitter molecules into the synaptic cleft. The neurotransmitter diffuses across the cleft and binds to receptors on the post-synaptic cell. The activated receptors cause changes in the activity of the post-synaptic neuron. The neurotransmitter molecules are released from the receptors and diffuse back into the synaptic cleft. The neurotransmitter is reabsorbed by the postsynaptic neuron. This process is known as *reuptake*.

There are several websites where the interested student can find detailed information about synapses and the process of transmission across them. For the time being, the subject is beyond the scope of this book.

2.7.6. Cerebrospinal Fluid (CSF)

It is a liquid compartment that is small but significant part of the ECF, however, because of its importance in anesthesiology it is better to treat it as a subject of the CNS. The French physiologist Magendie rediscovered it in 1825 and one of the holes through which it flows carries his name, Magendie's foramen. A young Italian physician, Domenico Cotugno, in Naples, made the first description in 1764 (Brazier, 1959).

The CSF is bounded by two membranes, the *piamater*, which covers all the brain and spinal cord structures, and the *arachnoid*, which is situated above the former covering the vascular network. Above the arachnoid there is the duramater. Thus, when the skull is penetrated by a surgical instrument (a *trephanum*) or when the spinal column is punctured, say, to practice *peridural anesthesia*, the duramater is the first membrane to encounter. Besides, the CSF occupies all spaces of the CNS, including the four cerebral ventricles.

Specialized capillaries, the choroidal plexuses, passively (as a ultrafiltrate) and actively (by secretion) extrude from blood through the capillary wall the CSF into the lateral cerebral ventricles, circulating via the interventricular foramens to the third ventricle and through the cerebral aqueduct into the fourth ventricle. From the latter, fluid passes via many apertures to the subarachnoid space to diffuse into all cerebral and spinal cord spaces. CSF is absorbed by the subarachnoid cilae, which project into the dural sinuses to return to the venous circulation. In the spinal cord the CSF returns to the general circulation via the spinal veins.

The volume is in the order 135 mL at a pressure of 10 mmHg. It flows at a very slow rate of 0.3 mL/min following the path described above. Cushing call it the "third circulation" (the lymphatic being the second and the blood circuit the first). Sodium, chloride, potassium and calcium plus glucose and some proteins are components of the fluid and it is isosmotic with plasma (i.e., 289 mOsm/L).

The main function of the CSF is purely mechanic: it acts as a buffer, some sort of pillow, to significantly decrease the traumatic effect of hits. The human adult brain weight is about 1,400 g, since it is floating in its own sea, the actual effective weight goes down to a bare 50 g.

Problem: What basic physical principle explains the brain weight reduction mentioned above? Search for the necessary numerical information to validate the numbers quoted above. Do not blindly accept them as an absolute truth.

It also plays a regulatory function that sows up in some pathologies of the CNS. For example, an increase in its hydraulic pressure leads to an increase in arterial blood pressure. The latter tends to stabilize at a value slightly higher than the pressure exerted against the medulla oblongata. This is known as Cushing's law or reflex, described in 1901–2. This would be a protective mechanism to preserve blood flow to the central nervous structures in pathologies characterized by higher than normal intracranial pressure, such as hydrocephalia or brain tumors. The phenomenon can be beautifully demonstrated in the experimental animal (Evans and Geddes, 1969).

2.7.7. Closing Remarks of Section 7

Knowledge about the CNS is still incomplete. The process of quantification referred to in previous sections requires more evolution with a better

definition of the variables. Ablation and stimulation techniques and the appearance of pathologies and the results of accidents as well have permitted finding causal relationships among centers and anatomical areas.
Certain pathologies demonstrated the existence of feedback loops, as for example in the circuit cerebral cortex-basal ganglia-thalamus-cerebral cortex, which, when malfunctioning, is involved in diseases such as Parkinson's, balism, chorea, and atetosis, all of them collectively named as *dyskinesias* (from Greek, *dys*, incorrect, and *kinesis*, movement). Another CNS feedback circuit is cerebral cortex-protuberance-cerebellum-thalamus-cerebral cortex, which would lead to *ataxia* or *intention tremor*. In the Middle Age, some of these diseases were indiscriminately called Saint Vitus' Dance. History in this respect is quite attractive and sometimes sad.
Studies on the synapse are still going on since it appears as an extremely important and relevant subject to the Nervous System. The reader is encouraged to proceed further in it for its many connotations in the concepts of memory, adaptation, behavior, and also as inspiration in neural modeling. For the latter subject and for the Nervous System in general, we suggest the website
http://www.emc.maricopa.edu/faculty/farabee/BIOBK/BioBookNERV.html,
where the Text ©1992, 1994, 1997, 1998, 2000, 2001, by M.J. Farabee (all rights reserved) is displayed. Its use for educational purposes is encouraged. Email: mj.farabee@emcmail.maricopa.edu

2.8. Muscular System

> *We look at it almost with worship, it is the driving engine to move us about and beyond, to fight and run; it gives us joy and pleasure, and yet, how far behind the mind it is!*

2.8.1. Introduction

The skeleton and muscles function together as the *musculoskeletal* system. This system (often treated as two separate ones, *muscular* and *skeletal*) plays an important homeostatic role allowing the individual to move, for example, to more favorable external conditions. There are, however,

other not negligible functions, as the production of heat during muscle contraction (sometimes, in a cold day, we rub our hands together or jog swiftly on the same spot), and the dynamic and static support to the whole body supplied by the bony structure (otherwise, we would collapse as some sort of jelly).

Most of the tissues in the body is muscle. When the hand is set on the body surface, most probably it will seat on a muscle. The beef we eat is muscle. The handsome, powerful and frequently admired or sought after physique displayed by an athlete is essentially a well-organized and developed complex bundle of skeletal muscles.

In this section, we describe just a few selected topics of the Muscular System, specifically those which are perhaps more attractive to the future biomedical engineer or may be better amenable to the engineering methodology. The skeleton is left out, more properly treated in a course of anatomy or biomechanics.

Many websites are available for the curious mind, among them, we recommend http://www.umds.ac.uk/physiology/mcal/spinmain.html. For the intrinsic mechanism of muscular contraction, the classic article by H.E. Huxley (1969) is suggested, where the author describes a revealing model for cross-bridge action at variable muscular filament spacing. The paper by Uwe Proske (1997) is an excellent and well-referenced update about the muscle spindle. Keep also in mind that the muscular system is also intimately related to the nervous system, so much that you may also speak of the *neuromusculoskeletal system*.

2.8.2. Functional Unit: The Neuromuscular Junction

The spinal cord gives off efferent motor fibers that innervate specific skeletal muscles. One fiber controls several muscle fibers following orders emanated, say, voluntarily from the motor cortex, or reflexly from a spinal segment. In big strong muscles, as the biceps, or the triceps, or muscles of the leg, the ratio is very large, that is, a single nerve fiber may take care of 100 or 200 muscle fibers, while in fine movement muscles (as in the fingers or lips or the vibrissae of a rodent's snout or cat whiskers), the ratio can be as small as 1 to 2 or 1 to 6. One nerve fiber and its set of innervated muscle fibers, be it large or small, is a *functional motor unit*. **Stimulation of that nerve fiber produces contraction of all its associated muscle fibers**. Removal of any of the component elements

(say, any of the fibers) destroys or damages the unit. For the understanding of the pathophysiology of muscle as well as for the development of adequate prostheses and motor assist devices, this concept is extremely important and must be clearly grasped.

A nerve is composed of many fibers and, thus, there is a multiplication effect as it enters the muscle because of the nerve fiber to muscle fiber ratio. However, even within the same muscle, not necessarily each fiber innervates exactly the same number of muscle fibers. There is some spread of values. Moreover, one nerve may innervate more than one muscle. For example, the *musculocutaneous nerve*, with its roots at cervical levels C5, C6 and C7, innervates three muscles: the *biceps brachii*, the *brachialis* and the *coracobrachialis*. One way of controlling the force of contraction is by recruiting more fibers in order, say, to lift a given weight.

The student is encouraged to search in the web for more information about this nerve and its three associated muscles. Find out what kind of movements they are responsible for. Besides, he/she should find examples of nerve fiber to muscle fiber ratios.

Nerve fibers connect to muscle fibers through a modified synapse, the so-called *neuromuscular or myoneural junction*, and, as in its nervous analogy, there is also a discontinuity: on one side, the nerve membrane can be seen under the electro microscope, on the other, the muscle membrane, and both are separated by extracellular fluid. Classic neuromuscular physiology describes three phenomena (one physiological, another biochemical and a third pathological) that clearly indicate the potential weakness of the junction, as if it were a "bottleneck" of the system:

1-If the nerve is stimulated repeatedly to obtain contraction of its muscle, soon the muscle will not respond appropriately. Contractions will become definitely weaker. However, the nerve will show a normal response to stimulation and the muscle will contract also normally when directly stimulated. This is a classical experiment in physiology. Obviously, there is some failure at the junction. It is now known that there is exhaustion of the acetylcholine (ACh) vesicles at the pre-junction site. This is called *fatigue*, and it is a physiological, that is, an expected normal phenomenon.

2-There are substances, as for example *curare* (an extract obtained from various species of *Strychnos*, which are plants found in Southamerica),

that have a reversible blocking action on the myoneural junction leading to paralysis of the skeletal muscles. The Indians poisoned their arrows to hunt animals or against the Spanish *conquistadores*. Death from curare is caused by asphyxia, because the skeletal muscles become relaxed and then paralyzed. However, the poison only works in the blood; poisoned animals have no harmful effects on humans if ingested (orally). Its vapors are not poisonous. During curare poisoning the heart continues to beat, even after breathing stops, which means that heart function is not stopped by curare. The horror of curare poisoning is that the victim is very much awake and aware of what is happening until the loss of consciousness. Consequently, the victim can feel the progressive paralysis but cannot do anything to call out or gesture. If artificial respiration is performed throughout the ordeal, the victim will recover and have no ill effects.

3-There is a disease, *myasthenia gravis*, precisely located at these junctions. The action potentials are not efficiently transmitted through the gap (also called *cleft*) and the net result is a weak muscular contraction (thus, the etymology of the term is well explained, from Greek, *myo*, muscle, *asthenia*, weakness, *gravis*, serious). The usual cause is an acquired immunological abnormality, but some cases result from genetic abnormalities. The prevalence of myasthenia gravis in the United States is estimated at 14/100,000 population. Patients with myasthenia gravis come to the physician complaining of muscle weakness. The course of disease is variable but usually progressive. After 15 to 20 years, weakness often becomes fixed and the most severely involved muscles are frequently atrophied. The normal neuromuscular junction releases acetylcholine (ACh) from the motor nerve terminal in discrete packages. ACh diffuses across the synaptic cleft and binds to receptors on the muscle end-plate membrane. Stimulation of the motor nerve releases ACh that depolarizes the muscle end-plate region and, thereafter, the muscle membrane causing muscle contraction. In myasthenia, the concentration of ACh receptors on the muscle end-plate membrane is reduced because antibodies are attached to the membrane competing with acetylholine. ACh is released normally, but its effect on the post-synaptic membrane is reduced. The post-junctional membrane is less sensitive to applied ACh, and the prob-

ability that any nerve impulse will cause a muscle action potential is reduced.

A little police story to check your Sherlock Holmes aptitudes: Two persons have dinner together and eat exactly the same food poisoned with curare. One of them dies soon thereafter while the other one survives with no after-effect of any kind. There was suspicion of murder. What disease had the victim? Who was probably the murderer and what knowledge did he/she have?
To think over: Curare is used in anesthesiology, but very well controlled. What for?

The motor unit (neuron and its associated muscles fibers) constitutes and essential component of the muscular system. In fact, it can be considered as its building block. The myoneural junction, in turn, appears as a unique sensitive site where complex electro-biochemical phenomena take place. Similar to its relative, the nerve synapse (connecting nerve to nerve fibers), it is also unidirectional, that is, conduction of the action potential only occurs in this case from nerve to muscle.

2.8.3. Muscle Spindle and Golgi Organ: Posture

Muscles, small or large, are composed of bundle of fibers. Usually, they are strongly attached to bones by *tendons*, which are strong ligaments to keep muscles in place and where stretch sensing cells conforming the *organ of Golgi* are located. The bibliography is full of excellent histological and anatomical details. A particular differentiated portion of the muscle is the *muscle spindle*, which is found within the belly of the muscle and runs in parallel with the main muscle fibers. The spindle is a biological transducer, for it senses muscle length and changes in length, in fact, it is sensitive to stretch. It has sensory nerve terminals whose discharge rate increases as the sensory endings are stretched. This nerve terminal is known as the *annulospiral ending*, so named because it is composed of a set of rings in a spiral configuration. These terminals are wrapped around specialized muscle fibers that belong to the muscle spindle (*intrafusal fibres*) and are quite separate from the fibers that make up the bulk of the muscle (*extrafusal fibres*). The muscle spindle itself contracts very little as compared to the highly contractile muscle fibers.
Let us explain in a very simplified manner the *stretch reflex*, the *inverse stretch reflex* and the biasing action of the *gamma efferent fibers* (Figure 2.73). When the muscle is stretched as a whole (as for example during an

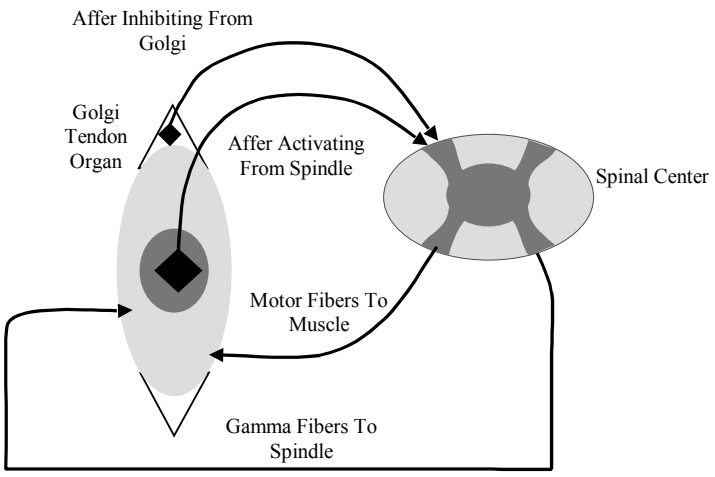


Figure 2.73. THE STRETCH AND THE INVERSE STRETCH REFLEXES. Simplified diagram of these two essential reflexes that play a role in posture and protection of the structures. The gamma fibers to the muscle spindle adjust its sensitivity to stretch. The triangles represent the tendons, where the Golgi sensing organ is. The central portion of the muscle is the spindle, another stretch sensing place, with its intrafusal fibers. See text.

excessive movement of a limb), the spindle also stretches generating action potentials that travel up via afferent fibers to the appropriate spinal center where they activate the efferent motor signal to the muscle to elicit its contraction and, thus, counteracting the initial stretching stimulation. Hence, it protects the muscle. However, if the stretch is too strong, it is this time also sensed by the Golgi organ, a stretch sensitive transducer set at a higher level than the muscle spindle. Its output follows afferent pathways to the same segmental spinal level where **it inhibits** now the motor response to the muscle. It is also a protective reflex: the stretch was too strong and it is better to yield rather to contract trying to counteract it; hence, the name of *inverse stretch reflex*. The constant dynamic and complex combination of these kind of responses all over the musculature help in keeping the posture, say, to maintain balance as gravity tends to pull the body down.

However, Figure 2.73 shows another neural pathway to the muscle spindle, the so-called *gamma efferent system*. Even though the intrafusal spindle fibers are not very contractile, they mildly do respond to stimula-

tion, and the gamma fibers are precisely motor fibers to the muscle spindle. Activation of these fibers produce contraction of the spindle and, therefore, as the structure becomes slightly shorter, it is more sensitive to the overall stretch of the whole muscle. This is a way to adjust the sensitivity to stretch or, expressed in other words, to introduce bias to the system. Since the gamma efferent neural pathways recognize their origin in the higher centers, many times psychological stress increases their outflow leading to painful muscle spasms because their spindles have been sensitized. Neck muscles are frequently the targets of these spasms. This is to be distinguished from the local and also painful muscular contracture (cramps) typically taken place in the legs. The latter may be caused by a variety of maladjustments, such as low glucose, fluid loss, electrolyte imbalance, overexertion, or fatigue. The exact physiologic mechanisms underlying cramps are not fully understood.

2.8.4. Diseases of the Neuromuscular System

The motor unit is a complex and delicate structure that, when in normal operation, performs exquisitely and with precision during a lifetime. Pablo Casals (1876–1973), Arthur Rubinstein (1886–1982) and Yehudi Menuhin (1916–1999), all three famous cellist, pianist, and violinist, re-

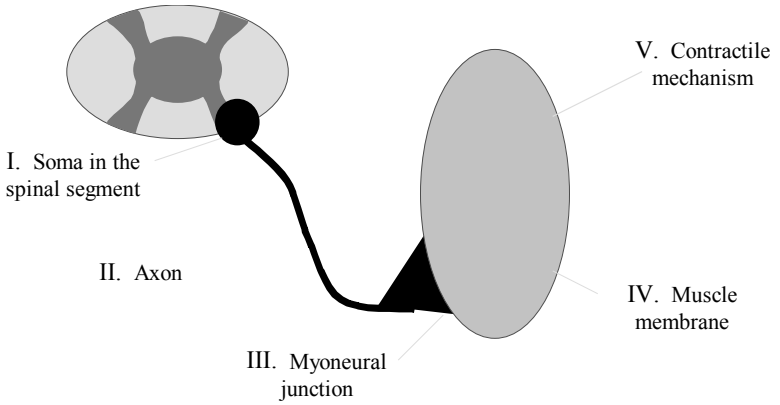


Figure 2.74. SITES OF DISEASE ALONG THE MOTOR UNIT. There are five vulnerable sites along the efferent motor unit that can be affected by serious diseases: the nerve fiber itself (soma and axon), the myoneural junction, the muscle membrane and the contractile mechanism of the muscle fibers. See text.

spectively, kept giving concerts until very old ages always capturing the audiences with their virtuosity. The finesse and dexterity of their finger movements was more than ample proof of the perfect functionality of the motor neurons involved.

However, irreversible diseases that may lead to serious crippling of the individual can hit the motor unit. Figure 2.74 marks the five vulnerable points. First, the motor fiber itself, either at the level of the cell body or soma within the spinal cord segments, or at its axon, or at both places. *Poliomyelitis* is the traditional and most dreaded example. Since there is peripheral nerve injury, the motor unit becomes non-operative and soon progressive muscular atrophy shows up.

First the injectable Jonas Salk's and later on the oral Albert Sabin's vaccines, developed and tested between 1952 and 1962, provided the effective weapon to put polio down in the United States and almost completely worldwide. In 1964, only 121 cases were reported nationally. Currently, there are typically fewer than ten new cases per year. A true-life saver, indeed.

The third vulnerable point is the neuromuscular junction already referred to above. *Myasthenia gravis* is its disease, fortunately with a very low incidence.

Myotonia congenita is a genetic, neuromuscular disorder characterized by the slow relaxation of the muscles. It attacks the muscle membrane manifested by a failure to relax normally. Symptoms may include muscle stiffness and hypertrophy (enlargement). The disorder is caused by a genetic mutation involving the chloride channel of the muscles. The muscle stiffness, which particularly occurs in the leg muscles, may be enhanced by cold and inactivity, and is often relieved by exercise.

The last vulnerable point is the highly complex contractile machinery. *Muscular dystrophy* is a typical example. It refers to a group of genetic diseases too characterized by progressive weakness and degeneration of the skeletal muscles. Although some forms first become apparent in infancy or childhood, others may not appear until middle age or later.

Dermatomyositis affects also the contractile mechanism. It is one of a group of acquired muscle diseases called *inflammatory myopathies* and is characterized by a rash accompanying, or more often, preceding muscle weakness. The most common symptom is the latter, usually affecting those muscles that are closest to the trunk of the body (proximal mus-

cles). Eventually, patients have difficulty rising from a sitting position, climbing stairs, lifting objects, or reaching overhead. Trouble with swallowing (*dysphagia*) may occur. Occasionally, the muscles ache and are tender to touch. Patients may also feel fatigue and discomfort and have weight loss.

2.8.5. Closing Remarks of Section 8

In this highly partialized introduction to the muscular system, we underlined the *motor unit*, its *myoneural junction*, the basic control mechanisms involving the *muscle splindle* and *Golgi organ* actions and the diseases associated with five vulnerable points of the unit. It would be better to speak of the *neuromuscular system* or, perhaps, of the *neuromusculoskeletal system*, because it is not easy to separate them fully out.

The biomedical engineer faces a tremendous modeling task to improve the overall understanding of these inter-related systems, with non-linear mathematics and control systems as essential tools.

Artificial limbs or prostheses of different kind appear as challenging aids to the amputee. Such devices precisely intend to replace one part of the musculoskeletal system (a hand, an arm, a leg). One of them is the so-called *myoelectric limb*, in which the control signal is derived from surface electrodes picking up nerve or muscle action potentials still under voluntary control of the patient. The signal is processed by electronic circuitry to reach a decision as to whether a function is to be activated, usually a predetermined movement of the artificial limb.

However, for the signal to be of clinical value in controlling a prosthesis, an estimation process is required that must not take much more than one tenth of a second, otherwise, an unacceptable delay will occur between the amputees' contraction and the action of the prosthetic. The task facing the amputee in using myoelectric prosthesis is difficult. It involves generating a myoelectric signal level corresponding to a specific target. The difference between the signal actually generated and what was intended is termed "operator's error". Thus, the actual behaviour of the myoelectric prosthesis results from the intent of the amputee modified by two sources of error, the system error and the operator's error. Neither of them can be eliminated. The subject is quite a challenge for the bioengineer, for the physician and, above all, for the patient, who also generates the need of psychological support studies.

Another related subject is Functional Electrical Stimulation (FES). It is well beyond the scope of this text, but we should mention a special issue devoted to it where prostheses and the neuromusculoskeletal system are constantly dealt with (Popovic, 2003).

2.9. The Cell

As bricks are building blocks for a house, so are cells for tissues and organs.

2.9.1. Introduction

The cell is as fundamental to biology as the atom is to chemistry. All organisms are made of cells. Behind everything we discussed up to here, there are cells, and ultimately, they supply the properties that characterize the systems of our initial general Figure 2.1. Cells are systems in themselves, with complex dynamic properties, well kept and contained in the manifold inner space. The student has probably already taken a biology course; however, a quick refreshment will help him/her out in the general understanding of this source material for the bioengineer. We will also probe into some basic concepts of molecular biology, which contains undoubtedly essential answers to the many questions about life.
As a good reference to widen and improve concepts, we recommend the textbook by Campbell, Reece and Mitchell (1999), and that by Pevzner (2000) to dig into something newer and more specific. The CD by ROCHE, *Genetics*, of its Educational Program is another suggested reference (see www.rochegenetics.com). As always, browse the *IEEE Proceedings of the Engineering in Medicine and Biology Society*, say of 2002 or 2003, respectively, to find contributions on cellular engineering, cell-cell interactions in blood, stem cell engineering, molecular engineering, and molecular biomechanics, to mention just a few. INTERNET is another permanent and unavoidable source to consult because all books, without exception, are already outdated by the time they come out.

2.9.2. A Panoramic View

Robert Hooke discovered and coarsely described cells in 1665, however, a full detailed description was largely void until the last decades. Most subcellular structures are too small to be resolved by the light microscope. That was the task reserved for the electron microscope in the 1950's and thereafter. Nowadays, imaging techniques have advanced even further with the introduction of the scanning microscope complemented with signal processing methods.

There are two types of structurally different cells: *prokaryotes* and *eukaryotes* (from Greek, *pro*, before, *karyon*, kernel or nucleus, and *eu*, true or correct). Only bacteria are prokaryotic cells, the rest (plants, animals) are eukaryotic.

The prokaryotic cell has no nucleus. Its genetic material is concentrated in a region called *nucleoid* without membrane. The eukaryotic cell, instead, is surrounded by a membrane. The region between the nuclear membrane and the cell membrane (or *plasma membrane*) is called *cytoplasm*, where the *organelles* are located. The latter are largely absent in prokaryotic cells.

Here, we are mainly concerned with eukaryotic cells. Their size is a general feature that strongly relates to function. Its plasma membrane acts as a selective barrier that allows sufficient passage of oxygen, nutrients, and wastes to service the entire cell. We have already referred to this membrane in the electrophysiology section and are familiar with the important ionic exchange that takes place across it. In addition to the external membrane, a eukaryotic cell has extensive and elaborately arranged internal membranes, which partition the cell into compartments. They participate in the cell's metabolism.

2.9.3. The Nucleus

The nucleus (about 5μ in diameter), the most conspicuous *organelle*, contains the majority of the genes that control the eukaryotic cell. The nuclear envelope is a double membrane separated by a space in the order of $30\times10^{-3}\mu$. The envelope is perforated by pores of about $100\times10^{-3}\mu$. At the lip of each pore, the two membranes fuse. Let us remind that *organelles* are subcellular or intracellular structures. By and large, they are too small to be resolved by the light microscope.

Within the nucleus, the DNA (*deoxyriboneucleic acid*) is organized along with proteins into material called *chromatin*. As a cell prepares to divide (reproduce), chromatin coils up becoming thick enough to be discerned as separate structures called *chromosomes* (*chromos* means "color" in Greek, and the name originates in the staining procedures to visualize these structures when observed under the microscope). The *nucleolus* is the most prominent structure within the nucleus.

The nucleus controls protein synthesis in the cytoplasm by sending molecular messengers in the form of RNA (*ribonucleic acid*). This messenger RNA, or mRNA, is synthesized in the nucleus following instructions supplied by the DNA. Once in the cytoplasm, mRNA attaches to *ribosomes*, where the genetic message is translated into the primary structure of a specific protein. Ribosomes are particles whose initial components are synthesized within the nucleus, too. Such components traverse the nuclear envelope pores into the cytoplasm where they combine to form these small organelles.

DNA is the molecule which contains genetic information and makes up our genes. It consists of two complementary nucleotide chains containing the bases *adenine* (A), *thymine* (T), *guanine* (G) and *cytosine* (C), held in a double stranded helix by bonds between base-pairs. A with T and G with C.

The chromosomes are structures made up of DNA and proteins. They contain the genetic information in a linear array. The order of nucleotide bases along a DNA strand is known as the *sequence*. The genetic information is encoded in the precise order of the base-pairs. DNA sequencing is the laboratory process designed to precisely determine the sequence of bases in the DNA. During cell division, the entire DNA of the cell is copied.

Human cells have 23 pairs of chromosomes, one of each pair inherited from each parent. The basic unit of heredity is the *gene*; they are ordered sequences of DNA base-pairs, located in specific positions on chromosomes. Genes contain the information for producing proteins or functional RNA molecules, which make up the structure of cells, and control their functions.

The complete genetic material of an organism conforms the *genome*. The *haploid* human genome contains 3 billion base-pairs of DNA organized

into 23 chromosomes. *Diploid* is the state of having two copies of each chromosome. Most human cells, except the gametes, are diploid with 46 chromosomes (23 pairs), including the sex chromosomes. Conversely, *haploid* is the state of having one copy of each chromosome. The reproductive cells are haploid, with 23 chromosomes in human egg and sperm cells.

In short, genes are sequences of base-pairs that encode information for proteins. They can range in size from less than 100 base-pairs to several million base-pairs. A diploid genome of $6x10^9$ bases, if stretched, would measure about 1.8 meters long. Chromosomes are relatively large structures made up of DNA and proteins that contain genes, that is, the genetic information. The nucleus acts as a master station.

2.9.4. Closing Remarks of Section 9

Cell anatomy and cell physiology are inter-related disciplines of basic biology standing on their own feet. This section is a mere refreshment of some concepts or just a bare preliminary set. We should briefly mention the remaining parts.

The different membranes of an eukaryotic cell form the *endomembrane system*. It includes the *nuclear envelope*, the *endoplasmic reticulum*, the *Golgi apparatus* (not related at all to the Golgi tendon organ already described before), *lysosomes*, and different types of *vacuoles*. Although the plasma membrane is the cellular boundary, it is part of the system because it strongly interacts with the endoplasmic reticulum and other internal membranes. Functions of this system are many, complex and important. The student is invited to dig deeper, even though it is not fully necessary for the time being.

Another cellular system is the *cytoskeleton*. Its most obvious unction is to provide mechanical support to the cell and maintain its shape.

Mitochondrion is the singular form of *mitochondria*. Their major role is synthesis of substances from sugars, fats, and other fuels with the help of oxygen to supply energy to the cellular metabolic processes. These are mobile and flexible organelles, although in some cells they tend to stay in a fixed position (as for example in muscle fibers). Mitochondria are also self-reproducing, they have their own circular DNA.

Cells of multi-cellular organisms also receive signals from other cells, including signals for cell division and differentiation. The majority of

cells in our bodies must constantly receive signals that keep them alive and functioning. The key concept is that the many signaling systems of biology have very similar or related steps. The same signaling system can lead to very different responses in different cells or different organisms. Studies of the mechanisms of cell signaling are leading to new understanding of many diseases, and to new strategies for therapy.

Chapter 3

Signals: What They Are

"Observe, think, and observe again", used to say Dr. J.J. Izquierdo, an old well-reputed professor of physiology at the Universidad Nacional Autónoma de México (UNAM), in Mexico City, back in the 1960's ... "and physiological signals are one way to observe", we may add.

3.1. Introduction

The biological source of information has been already reviewed in the preceding chapter. Now it is the time to consider the signals it generates. Researchers always struggled to somehow "see" the physiological events they were dealing with, so that some kind of visualization means (such as records on paper, records on an oscilloscope, records on a monitor, or records as images) became mandatory, for the first processing of that graphic information is made by eye, using the observer's knowledge and experience.

This chapter deals with the nature of the biological/physiological signals always or mostly presented as experimental functions of time (even though something was already anticipated in the preceding chapter, especially when describing the cardiovascular system). The explicit mathematical relationship is rarely known, if ever; at most it is an empirical approximation. They contain the information to be interpreted later on and, thus, their true shape, timing and location are essential characteristics to preserve. Hence, the first objective is to give an overview of them. A second objective is to glance into their fields, showing how bioengineering unexpectedly projects out when, by sheer necessity, new techniques must be developed in order to recover information otherwise hid-

den within the signal. A typical example of the latter is the fetal electrocardiogram or the signal obtained from an image.
By and large, physiological systems generate five types of signals, i.e., electrical, mechanical, magnetic, thermic and chemical. This criterion to classify them is based on the kind of energy they handle; hence, it is a physical criterion. For that matter, we will discuss the bioelectric and the biohydraulic events; thereafter, signals produced by biomechanical systems and implanted biomaterials, cellular signals and images treated as signals. As required in the course of the text, reference will be made to Chapter 2 (source of these signals).

3.2. Bioelectric Events and their Signals

It carries information, it triggers contractions, it produces or inhibits secretions, it elicits sensation; it is, indeed, a factotum.
At the end of Section 2.2.1. of Chapter 2, in small case, we briefly mentioned the discovery of animal electricity by Luigi Galvani and the famous controversy with Alessandro Volta (with his invention of the electric pile) that led to two major disciplines of enormous scientific, technological, social and economic projections: the establishment of Electrophysiology and of Electrical Engineering. That was the beginning of the saga to know what the injury potential and the elicited and reversible change in polarity (the action potential) really were. It took about 150 years to unravel most, but not all, of the secrets, from 1796 until 1948, when the microelectrode was introduced.
The bioelectric events are electrical signals manifested as differences of potential between two points located in some places of the living organism, either inside or on its surface. The injury potential of an excitable tissue is an incomplete manifestation of the membrane potential. In engineering terms, it is a direct current (or dc) event, because it remains stable as long as the cell is alive. As already mentioned above, the typical resting value is about 90 mV (negative inside the cell and positive outside). The action potential, instead, is a spike; the duration is shorter or longer according to the tissue it comes from. In engineering terms, it is a pulse that sometimes is modeled by a rectangular wave with maximum amplitude in the order of 100 mV (see 2.2.2). Excitable tissues (nerve,

skeletal muscle, cardiac muscle, smooth muscle) are the traditional and main generators of these electrical activities; however, there are other tissues, too, able to produce small differences of potential (such as the eye, the skin or an oocyte). Let us call the latter non-traditional sources of bioelectricity, even though in some instances a contribution of excitable cells may exist.

3.2.1. Non-Traditional Bioeletrical Signals

3.2.1.1. From the eye

– *Electro-oculogram* (*EOG*)

The eye has a standing electrical potential across it, as some kind of weak battery, with the front of the globe (cornea) positive and the back negative. This resting or "standing potential" is generated largely by the transepithelial potential across the retinal pigment. It varies from one to several millivolts (up to 10 in some cases), depending upon the state of retinal illumination, because light leads to a polarization change of the basal pigment membrane. Thus, it is not generated by excitable tissue but, rather, is attributed to the higher metabolic rate of the retina. Interestingly enough, the polarity of this potential difference in the eyes of invertebrates is opposite to that of vertebrates. In fact, this field may be detected with the eye in total darkness and/or with the eyes closed. It was first observed by Emil du Bois–Raymond (1818–1896), in 1848, a renowned German physiologist (yes, with a French name). The standing eye potential is not really constant, but slowly varying and is the basis for the *electro-oculogram* (EOG); its chief application is in the measurement of eye movement. A particular application of the EOG is in the measurement of *nystagmus*, which denotes small movements of the eye. The resulting signal is called an *electronystagmogram*. It depends both on the *visual system* and the *vestibular system* and provides useful clinical information concerning each.

Study subject. Search for the *vestibular system*. What is its function? It is a beautiful and complex structure both from the physiological and engineering point of view.

Skin electrodes placed nasal and temporal to the eye can detect the standing potential. Another pair of electrodes can also pick up the signal, one on the eyebrows and the other below, on the lower eyelid, that is, with

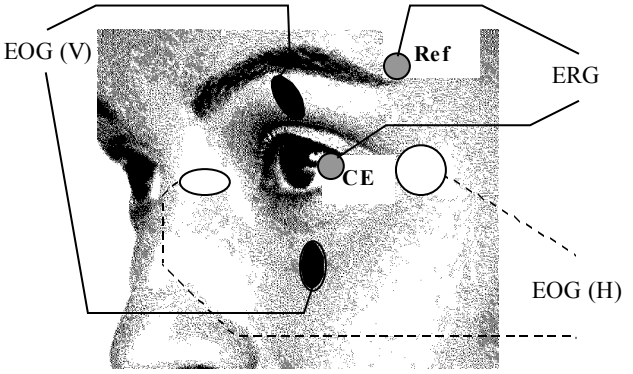


Figure 3.1. ELECTRO-OCULOGRAM. This figure partially represents a face (left eyebrow, left eye). The EOG is a record of the standing potential of the eye, manifested when the eye moves horizontally or vertically or in a compounded way. The horizontal and vertical pairs of electrodes record, respectively, those components of the EOG. The ERG, instead, is recorded with a corneal electrode CE against a reference electrode Ref. This signal is mounted on the standing potential or on the EOG.

the electrode axis perpendicular to the horizontal axis determined by the two lateral electrodes. The system reminds the cardiac dipole (in this case is the eye dipole). As the eye moves upward or downward, or laterally (*saccadic movement*), each pair of electrodes will detect the corresponding projection of the dipole. Saccadic movements or response (the latter elicited by adequate stimulation) are involuntary, abrupt, rapid, small jerks of both eyes simultaneously in changing the point of fixation on a visualized object, such as the series of jumps the eyes make in scanning a line of print when reading. The electron beam of an oscilloscope or a TV set precisely moves in a similar way driven by a specifically generated sawtooth waveform. Such eye movements produce the so called *Rapid Eye Movement* (REM) signal that have significance in certain stages of sleep and of the level of anesthesia. It becomes clear that there is a horizontal and a vertical component. Figure 3.1 depicts the arrangement.

Let us consider the horizontal lead. The rotation of the eye to the right results in a difference of potential, with the electrode in the direction of movement (i.e., the right canthus) becoming positive relative to the second electrode. The difference in potential should be proportional to the sine of the angle. The opposite effect results from a rotation to the left. The calibration of the signal may be achieved by having the patient look

consecutively at two different fixation points located a known angle apart and recording the concomitant EOGs. Typical achievable accuracy is ±2°, and maximum rotation is ±70°; however, linearity becomes progressively worse for angles beyond 30°. The student can find in the web records of this kind that clearly illustrate the procedure and its outcome. The recording of the EOG is a routinely applied diagnostic method in investigating the human oculomotor system (Geddes & Baker, 1989). For example, cranial trauma may cause injury to the 3rd pair (*common oculomotor nerve*) leading to total or partial paralysis of the eyes.

– *Electroretinogram* (*ERG*)

It is obtained by retinal illumination, the latter causing a triphasic variation in the standing potential (three components or waves called *a*, *b* and *c*, respectively) with an overall duration in the order of 100 ms. However, such duration depends on the stimulus length. Geddes and Baker (1989), for example, illustrate with a record of 2.5 s obtained in response to a long stimulus (2 s). The ERG alone reflects the summation of electrical responses generated by neurons and non-neuronal cells in the retina and pigment epithelium in response to light. The ERG, then, is superimposed on the standing potential of the eye (SP). A corneal electrode and neutral electrodes placed on the skin around the eye are required to obtain the record (Figure 3.1 and Figure 3.2).

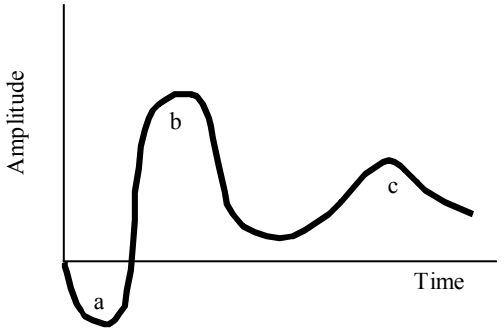


Figure 3.2. ELECTRORETINOGRAM (ERG). It is triggered by a light stimulus, usually showing three components, *a*, *b* and *c*, of variable amplitude and duration. All the response is superimposed on the EOG, which is a stable potential. Hence, it has a dc component not shown in the figure. The total duration from the *a*-wave to the *c*-wave may be in the order of 100 ms if the stimulus is just a flash. See text.

The eye resembles a camera. There are lenses in the front focusing light over the *retina*, located in the back. Within the retina is an area known as the *macula*, which is the portion responsible for detailed and color vision. The remainder of the retina is responsible for night vision and side vision. The retina sits on a pigmented layer of cells known as the *retinal pigment epithelium* (RPE). The retinal pigment epithelium absorbs stray light, provides the retina with nourishment and helps clear waste products from the eye. Beneath the RPE is the *choroid*. The choroid is a thin tissue with a high rate of blood flow. The optic nerve and brain tract connect the retina in the eye with the rest of the brain. Diseases of the retina, the retinal pigment epithelium or the optic nerve can result in poor vision and can also cause an abnormal electric pattern in these tissues. There are numerous retinal disorders (such as *cone dystrophy* and *retinitis pigmentosa*), or choroidal and optic nerve diseases that can be diagnosed by electric testing. The electro-oculogram is often used together with the electroretinogram.

Holmque, working with frogs, discovered the ERG in 1865. Huma, in 1877, and Dewar, in the 1940's, made substantial advances, already with recordings from humans. For this particular subject, EOG and ERG, we suggest to visit
www.city.ac.uk/optometry/Lab_Electrodiagnostics.pdf.
Besides, check the appropriate references given in the list at the end of the book.

– *Visual Evoked Potentials* (*VEP*)

An effective stimulus delivered to a sense organ results in a response detectable on the brain cortex. Auditory, somatosensory and visual responses to appropriate stimuli are relatively easy to pick up by placing electrodes over the corresponding cortical areas. Such electrical signals are called *evoked potentials*; with them, the integrity of the neural pathway from the sense organ to the cerebral cortex can be evaluated.

Scalp electrodes located on the occiput detect the response to a flash of light or a patterned image (say, a checkerboard). Visual evoked potentials can also be recorded from other scalp regions; however, the waveforms differ. The retinal receptors send information via the optic nerves (which converge at the *chiasma*, at the base of the brain); fibers from the nasal areas decussate precisely in that place while those from the temporal sides do not. Beyond the chiasma, fibers form the optic tract, synapse

at the *geniculate body*, and continue to the occipital cortex, where there is a topographic representation of the retina (Picton, 1974). This evoked response is also a good test of macula function.

Hence, the eye appears as a complex generator of electrical signals that can be used in the diagnosis of many diseases and in the evaluation of several systems and their respective functions. Detailed information is available in the WEB, for example, at

http://www.emedicine.com/neuro/topic69.htm

or directly from the International Society for Clinical Electrophysiology of Vision (ISCEV) on specific applications and on standard recording techniques.

3.2.1.2. From the skin: the electrodermogram (EDG)

The term EDG should be considered a generic name for a family of signals recordable from the skin. Perhaps, it should be classified as "traditional" because its ancestry goes rather back in time. A typical example is the *electrodermal response* (EDR), that is, a voltage measurable between one electrode in an area richly supplied by sweat glands and another in a region devoid of them. Cholinergic stimulation via fibers from the sympathetic nervous system constitutes the major influence on the production of sweat by these *eccrine glands*. It shows after an emotional stimulus and reflects a change in the activity of the autonomic nervous system. Besides, this response is also associated to a change in resistance (another form of manifesting the EDG), the so-called *galvanic skin response* or GSR (Geddes & Baker, 1989). These phenomena are still not fully understood; in spite of this shortcoming, EDR and GSR are nevertheless widely used, especially in the so-called "lie detectors".

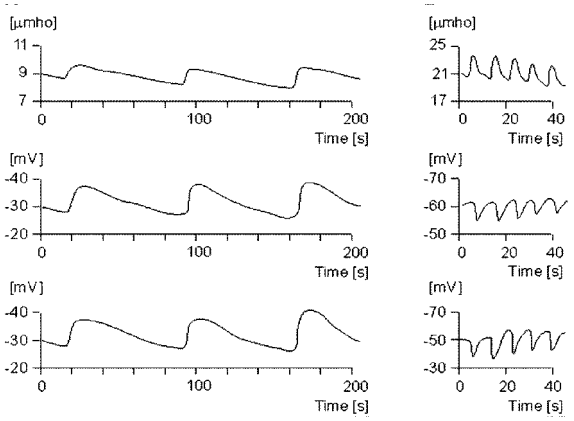
The student is encouraged to review anatomical details of the skin to better understand these electrodermal modifications. He or she will see, for example, that the most superficial layer is called the *epidermis*. Somewhat deeper, blood vessels are found in the *dermis* whereas the *eccrine* sweat gland secretory cells appear at the boundary between the dermis and the *panniculus adiposus*, also referred to as *hypodermis*. The excretory duct of the eccrine sweat glands consists of a simple tube made up of a single or double layer of epithelial cells; this ascends to and opens on the surface of the skin. It is undulating in the dermis but then follows a spiral and inverted conical path through the epidermis to terminate in a pore on the skin surface (Ebling, Eady and Leigh, 1992). The epidermis ordinarily has a high electrical resistance due to the thick layer of dead cells with thickened keratin membranes. This aspect is not surprising, since the function of skin is to provide a barrier

and protection against abrasion and mechanical assaults. However, sweat ducts from underlying cells when filled with a relatively good conductor (sweat can be considered the equivalent of a 0.3% NaCl salt solution and, hence, a weak electrolyte) provide many low-resistance parallel pathways. As a consequence, the effective skin conductance can vary greatly, depending on present and past eccrine activity. The afore-mentioned behavior is particularly significant in the palmar and plantar regions. Furthermore, when sweat gland activity is abolished, there is an absence of EDG signals (Fowles, 1986).

Measurement of the output of the sweat glands, which EDR and GSR are thought to do, provides a simple gauge of the level and extent of sympathetic activity. This is the basic concept underlying EDG and its application to psychophysiology.

As mentioned above, there are two major measures of the electrodermal response. The first chronologically to appear involved the measurement of resistance or conductance between two electrodes placed in the palmar region; Charles Féré, in France, originally suggested it in 1888. It is possible also to detect voltages between these electrodes, as already said before; these potential waveforms appear to be similar to the passive resistance changes, though its interpretation is less well understood. J. Tarchanoff, in Russia, pioneered this measurement in 1889. Sometimes, it is called the Tarchanoff phenomenon. The first type of measurement is referred to as *exosomatic*, because an external current must be injected. The second type, which is less commonly used, is called *endosomatic*, since the source of voltage is internal. Researchers also distinguish whether the measurement is of the tonic background level L, or the time-varying phasic response R type. Thus, there are a number of specific measures, each described by a three letter-abbreviation such as, EDA or Electrodermal Activity, EDL or Electodermal Level, EDR or Electrodermal Response, SCL or Skin Conductance Level, SCR or Skin Conductance Response, SRL or Skin Resistance Level, SRR or Skin Resistance Response, SPL or Skin Potential Level and SPR or Skin Potential Response. The GSR is an older terminology slowly disappearing. Electrodermography, or EDG, can be used as a generic name encompassing all of the previous ones.

Figure 3.3 shows signals characteristic of SCR and SPR waveforms. Those identified as having *slow recovery*, shown in Figure 3.3a, have a duration of around 40 s, with phasic amplitudes of around 2 μS (microsiemens = micromho) for conductance and 10–20 mV for potential. Data



(a) (b)

Figure 3.3. ELECTRODERMOGRAM. (a) Upper trace: Slow-recovery SCR; middle and lower traces are monophasic SPR's. (b) Upper trace: Rapid-recovery SCR; middle and lower traces are more rapid monophasic SPRs. See text for definitions of abreviations. Downloaded from INTERNET, after Chapter 27, The Electrodermal Response, in Principles and Applications of Bioelectric and Biomagnetic Fields, by J. Malmivuo and R. Plonsey, Oxford University Press, 1995, courtesy of these authors, and originally from Fowles, 1974.

collected by Venables and Christie (1980) give a mean SCL of 0.3 μS and SCR of 0.52 μS in a study of a particular population (N = 500–600). *Rapid-recovery* SCRs and SPRs are shown in Figure 3.3b. This information is freely available in the WEB.

Recommendations for electrodermal measurements were drawn up by a committee selected by the editor of *Psychophysiology* and published by that journal (Fowles, Christie, Edelberg *et al*., 1981). The paper by MacPherson, MacNeil, and Marble (1976) on measurement devices may also be useful. Several electrical models have been proposed to explain and validate electrodermal and dermoconductive phenomena; however, their predictive and quantitative values have not been fully established and, at best, are only relative. In the meantime, electrodermal activity (EDA) appears to be useful as an empirical tool for registering the level of sympathetic activity in psycho-physiological experiments. One measure of the extent of interest in EDR is the references to papers that list EDR as a keyword. In the *Science Citation Index* for 1991, for example, one finds approximately 25 such references (i.e., publications). The importance attached to such measurements includes the statement that palmar sweat is one of the most salient symptoms of an anxiety state and, for some, the single most noticeable bodily reaction.

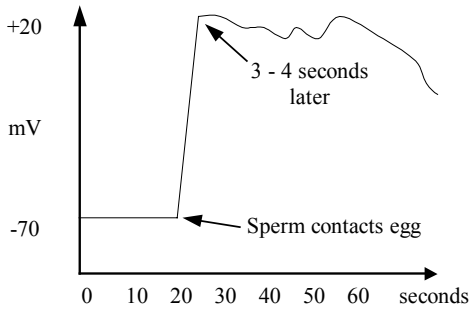


Figure 3.4. ELECTRICAL BLOCK TO POLYSPERMY IN SEA URCHIN EGGS. Observe the reversal of polarity, from –70 mV to +20 mV after the sperm contacts the egg. Redrawn after a picture downloaded from the WEB.

3.2.1.3. From the oocyte

The *oocyte* is the female reproductive cell, also called egg or ovum. It has been found, for example, that in ovarian oocytes of amphibians (say, *Rana pipiens*), the cytoplasm is 50 to 80 mV negative, relative to isotonic Ringer's solution. Microelectrodes must be used for its recording. When the spermatozoid penetrates the oocyte, this potential switches reversing its polarity to block the entrance of another sperm (Figure 3.4).

In vitro maturation of mammalian oocytes is an area of great interest due to its potential application in the treatment of infertility. The morphological and physiological changes that occur during oocyte development are still poorly understood. Emery, Miller and Carrell (2001) evaluated the membrane potential and the sodium/potassium permeability ratio of oocytes acutely isolated, and cumulus-oocyte complexes in *metaphase II* and *preantral follicle stages*. Thus, it is a subject of paramount importance in developmental biology.

The student is urged to review the basic concepts of developmental cell biology in order to refresh the different terms in current use.

The functional unit of the mammalian ovary is the developing follicle. The follicle is comprised of somatic *granulosa cells* of two categories, *thecal cell*s and *cumulus cells*, and a single gamete cell or oocyte. If the follicle continues to develop and does not undergo *atresia*, it will yield an oocyte competent to undergo fertilization and eventually form a new

organism. Several changes in the morphology and physiology of the oocyte and surrounding cumulus cells occur at the different stages of oocyte development. These include dynamic changes in gap junctions, cytokine release, morphology, and membrane physiology. Here we are mainly focused on the latter.

The function of the granulosa cells, and specifically cumulus cells, as they pertain to oocyte maturation within the follicle is an area of interest. The presence of both *gap junctional communication* and *autocrine/paracrine effect* has been established. These avenues of communication may play a role in the maintenance of membrane properties, providing a means for initiating progression or imposing regression of the oocyte's *meiotic stage*. It is also possible for the opposite effect to be present; membrane function of the oocyte to have an effect on gap junction regulation and cytokine release. The membrane physiology of the follicular cells includes the electrical properties of the cell, such as the membrane potential, cell coupling, ion channel activity and ion permeability. The membrane physiology of the oocyte and surrounding cumulus cells has been shown to change during germinal vesicle breakdown, a late developmental stage. In order to address this area of follicular maturation, Emery, Miller and Carrell (2001) measured the membrane potential (E_m) and permeability ratio of sodium (Na^+) and potassium (K^+) (P_{Na}/P_K) of oocytes denuded of surrounding cells, and in complex together. Metaphase II (MII) and preantral follicle oocyte-cumulus complexes were used for evaluation.

Intracellular electrical recordings revealed that cumulus-enclosed oocytes have a membrane potential significantly more negative at the preantral follicle stage than at metaphase II stage (–38.4 versus –19.7 mV). The membrane potential of the cumulus-free oocytes was not different between the preantral and metaphase II stages. The membrane potential of the cumulus cells forming preantral stage follicles was shown to be significantly different from that of the oocyte within the follicle (–28.6 versus –38.4 mV). The sodium/potassium permeability measured in cumulus-enclosed oocytes at the preantral stage equaled a mean value of 0.33. The ratio was significantly lower when measured in oocytes denuded of cumulus cells or cumulus-enclosed metaphase II oocytes. These data show a change in the membrane potential and Na^+/K^+ permeability ratio

during ooycte development from the preantral stage oocyte to the metaphase II stage.
Electrical measurements were recorded as the potential difference between a 3 M KCl Ag/AgCl microelectrode (tip impedance of 20 MΩ) inserted into the oocyte and an external 3 M KCl glass electrode located downstream of the oocyte. Electrodes were coupled to a dual channel high input impedance amplifier. Recordings were displayed on a digital storage oscilloscope and stored to the hard drive of a PC.
The paper of Emery, Miller and Carrell (2001) has been the main source of most of this material. The interested student can find more details about the subject in the papers by Okamoto, Takahashi and Yamashita (1977), McCullough and Levitan (1987), and Mattioli, Barboni, Bacci *et al.* (1990).

3.2.2. Traditional Bioelectrical Signals

In electrophysiological studies, concerned with the traditional excitable tissues mentioned above, action potentials are commonly recorded by means of microelectrodes (in research laboratories) or small electrodes (in research or clinical laboratories). The duration of these signals vary enormously from one tissue to the other; say, from 0.1 ms in nerve and skeletal muscle up to 300 or 400 ms in cardiac muscle. In smooth muscle, the duration can reach 500 ms with rather unstable resting potential. Smooth muscle is responsible for the contractility of hollow organs, such as blood vessels, the gastrointestinal tract, the bladder, or the uterus. Its structure differs greatly from that of skeletal muscle, although it can develop isometric force per cross-sectional area that is equal to that of skeletal muscle. However, the speed of smooth muscle contraction is only a small fraction of that of skeletal muscle.
The traditional electrical signals recorded with relatively large electrodes are the electrocardiogram (ECG), the electromyogram (EMG) and the electroencephalogram (EEG).
The ECG was described and discussed with enough detail in Chapter 2, Section 2, Part 2 and, thus, there is no need to elaborate further in the subject. Its origin is located in the heart cells. It has been fully established during the course of the XXth Century and is widely and routinely used all over the world as important, easy and inexpensive diagnostic tool.

Let us now go into the electromyogram (EMG) and the electroencephalogram (EEG), as the other two bioelectric events that have won wide clinical acceptance.

3.2.2.1. The EMG

When a muscle contracts, small electric potentials are produced. If normally innervated, the muscle shows no electrical activity at rest. Surface or needle electrodes can sense such potentials, which are usually proportional to the level of the muscle activity. The signal detected by the electrodes is amplified and recorded and is known as the electromyogram. The EMG signal is virtually meaningless "as is;" however, when properly analyzed, considerable information can be retrieved from it (Figure

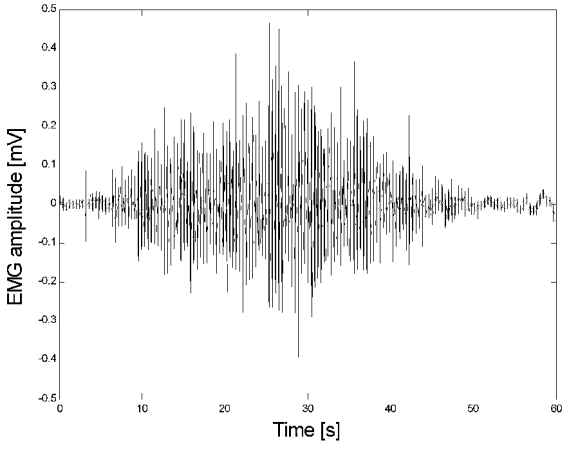


Figure 3.5. ELECTROMYOGRAM. Typical EMG recorded from a 40 yrs old female. The subject stood up to perform lateral abduction-adduction of one arm. Initially, both arms rested hanging down. Hence, the record displays the sequence of lifting and lowering the arm from 0° to 90° and back to 0°. The amplitude of the signal increases with abduction and decreases with adduction. For the same angle, such amplitudes were smaller when the arm was returning to its resting position than when it was being lifted. An angular scale was drawn behind the subject in 10° steps, from 0 (hanging arm) to 90° (the latter corresponding to a horizontal extended arm). A pair of sterile stainless steel acupuncture needles was used as electrodes (0.2 mm in diameter), inserted in the belly of the deltoid medium with a separation of 2 cm. Abduction corresponds to the first left half and adduction to the second portion. Sampling frequency was 3 kHz. The vertical axis is calibrated in rms mV while the horizontal axis corresponds to time. Record obtained at the Department of Bioengineering, UNT, 1997.

3.5).

The EMG is a rather ubiquitous tool used, for example, as a measure of muscular effort in traumatology, as an evaluation parameter in rehabilitation engineering, as the control input to myoelectric prostheses and as another diagnostic piece of information in nerve damage or injury caused by a compressed disk in the neck or back, nerve compression from carpal tunnel syndrome, myopathies many times faced in clinical medicine, neuromuscular diseases, inflammation or degeneration of peripheral nerves caused by conditions such as diabetes, pernicious anemia, and heavy metal toxicity. It can also be used to estimate the muscle action potential propagation velocity (Geddes and Baker, 1989; Spinelli, Felice, Mayosky *et al*., 2001).

The presence, size, and shape of the waveform produced on the oscilloscope (often combined with sound output via a loudspeaker) provide information about the ability of the muscle to respond to nervous stimulation. Each muscle fiber that contracts produces an action potential and the EMG is the compounded signal of many individual action potentials. This electrical activity of skeletal muscle can be described mathematically as a random process, which is amplitude modulated. When muscular effort is low, the amplitude of EMG is low; when muscular effort is high, the amplitude of EMG is high. Thus, better estimates of EMG amplitude improve the ability to determine the activation level of muscles. Active research in applied signal processing and stochastic estimation is being used to improve methods for estimating the amplitude of EMG signals (Clancy, 1999; Clancy & Hogan, 1994; 1995; 1997; Clancy and Farry, 2000; Clancy, Morin and Merletti 2002).

Since EMG is essentially a by-product of muscle contraction, it is logical to try to relate the electrical activity of muscle to its mechanical activity. Joint torque is frequently selected as the mechanical activity, as it is easy and reliable to measure. By and large, as the number of active motor units is increased and/or the average firing rate of active motor units is increased, both EMG amplitude and total muscle tension increase. However, the relationship is dynamic and may also need to be treated as nonlinear.

When a muscle fiber loses its nerve supply, it exhibits a characteristic irritability manifested as spontaneous discharges at rest. Single muscle

discharge, called fibrillations (not related at all with cardiac fibrillation) have a short duration (.5 to 1.5 msec), low amplitude (50–300 microvolts) and a rather regular rhythm. After denervation, it takes some time for fibrillation to appear; this appearance time is species-dependent (about 3 days in mouse, 4 in the rat, 6 in the rabbit, 8 in the monkey and up to 18 days in humans). Thus, the higher the rank in the zoological scale, the longer the time. If reinnervation does not occur, the muscle fibers atrophy and the fibrillation potentials disappear. If reinnervation takes place, the fibers cease to atrophy and the fibrillation potential gradually disappear, being replaced slowly by normal muscle action potentials which show up when a voluntary effort is made or when the muscle is contracted reflexively (Geddes & Baker, 1989).

Muscular fibrillation is abnormal but is not characteristic of any single disease. They may be seen whenever a muscle cell membrane becomes hyperirritable. If they are widespread in all four extremities and consistent with the clinical history, *amyotrophic lateral sclerosis* (or Lou Gherig's disease) should be suspected. However, there is a great deal of subjectivity in the interpretation of an EMG. *Polyphasicity* (that is, more than 5 baseline crossings on an EMG) and increased serration may be seen in reinnervated muscles or primary myopathy; in most of them (such as *myotonic dystrophy*), however, amplitude is reduced and the action potential prolonged.

The student is advised to review the background material in electrophysiology, the concept of motor unit and the basics of skeletal muscle physiology. Something can be found in the preceding chapter, in Geddes and Baker (1989), in any physiology text, or in the WEB.

3.2.2.2. The EEG

Geddes and Baker (1989) produced a clear account of this subject. Some of the material that follows has been taken from these authors. See also Nuñez (1981) and Sharbrough, Chatrian, Lesser *et al*. (1991) As usual, the WEB is another place where more information can be readily found.

The electrical activity of the brain is recorded with three types of electrodes: scalp, cortical and depth. With scalp electrodes, the recording is called an electroencephalogram (EEG). If the electrodes instead are placed on the surface of the brain, the recording is an electrocortigram (ECoG). If electrodes are advanced into the brain, the term *depth re-*

cording (DREEG) may be used. With the latter, and surprisingly, little damage if any is caused. Any of these signals recognizes its origin in the activity of numerous neurons with fluctuating membrane and action potentials. All three techniques are examples of extracellular recording.

R. Caton, in 1875 and 1887, was apparently the first one to report spontaneous electrical activity from the brain of a variety of animals (Geddes and Baker, 1989); however, he only provided good verbal description of the phenomenon. It was Berger, in 1929, who obtained graphic records using Einthoven's string galvanometer and scalp electrodes. His papers went largely unnoticed until Adrian and Matthews, in 1934, in Great Britain, and Jasper and Carmichael, in 1935, in the USA, reviewed and confirmed his earlier findings. A complete account of the development of electroencephalography was published under the editorship of Rémond (1971).

As figures 3.6 and 3.7 illustrate, the electroencephalographic signal looks like noise, although a normal EEG shows a dominant frequency of about 10 Hz and amplitude in the range of 20 to 200 μV. This activity, which is called the *alpha rhythm*, ranges in frequency from about 8 to 13 Hz and is most prominent in the occipital and parietal areas. The alpha rhythm increases in frequency with age from birth and attains its adult appearance by 15–20 years of age. It is most salient with the eyes closed and in

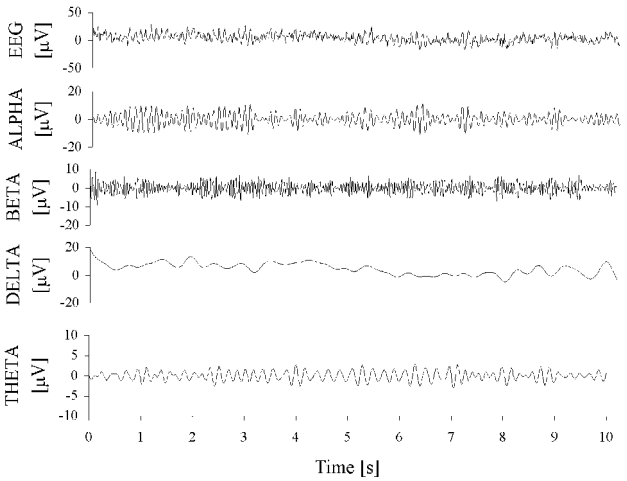


Figure 3.6. ELECTROENCEPHALOGRAM. EEG´s showing a regular normal signal and typical alpha, beta, theta, and delta rhythms. Records obtained at the Department of Bioengineering, UNT, 2003.

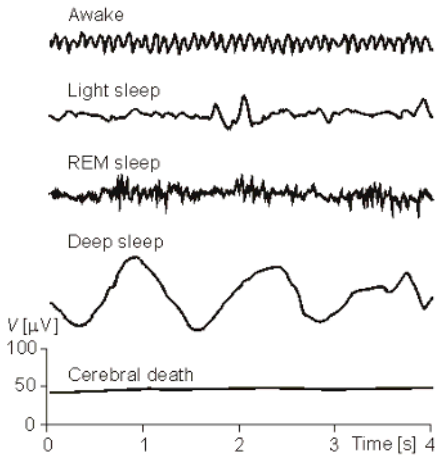


Figure 3.7. EEGs DURING DIFFERENT CEREBRAL CONDITIONS. The subject is awake (upper channel), lightly asleep (second channel), asleep under conditions of rapid eye movement or REM (third channel), and in deep sleep (fourth channel). A flat EEG means cerebral death and represents the criterion followed to decide donation of an organ. Downloaded from the WEB, freely available.

the absence of concentration. Opening the eyes or performing a cerebral activity, such as mental arithmetic, decreases or abolishes the alpha rhythm. There are other waveforms such as the *beta*, *delta*, *theta* and *spikes* (Figure 3.6). Their interpretations belong to the specialist and lie beyond the introductory scope of this text. Sleep has also important reflection on the EEG (Figure 3.7), so much that there are laboratories that specialize in it, particularly to study its many disturbances.

From a physical point of view, the EEG can be modeled as a finite sum of harmonic oscillations at discrete rates triggered by a central pacemaker. Hence, brain waves can be interpreted as consisting of a fundamental oscillation superimposed to higher harmonics. In terms of this concept, brain waves are composed of a series of simple sinusoidal waves ranging in frequency between 0.25Hz and 64Hz (7 octaves; doubling a given frequency defines an octave, say, from 250 Hz to 500 Hz, and it is a term widely used in music), whereby the full composition essentially depends on the state of consciousness, such as wakefulness or sleep stages. Thus, the EEG can help diagnose seizure disorders like epilepsy, head injuries, brain tumors, encephalitis, some kinds of infections, metabolic disturbances, and sleep disorders.

The number of nerve cells in the brain has been estimated to be on the order of 10^{11}. Cortical neurons are strongly interconnected. Here, the surface of a single neuron may be covered with 1,000–100,000 synapses. The electric behavior of the neuron corresponds to the description of ex-

citable cells introduced earlier in the book. The current density associated with neuronal activation produces an electric field, which can be measured, as mentioned before, on the surface of the head or directly on the brain tissue. For most excitable tissue the basis for the current density is the propagating action potential, for the EEG, instead, it appears to arise from the action of chemical transmitters on postsynaptic cortical neurons. The action causes localized depolarization or hyperpolarization with the net result in either case of a spatially continuous volume source distribution, because neural tissue is generally composed of a very large number of small, densely packed cells. Theoretical models have been developed using electric field theory.

There is an international standard system to record the spontaneous EEG. In it, 21 electrodes are located on the surface of the scalp. The positions are determined using as reference points the *nasion* (delve at the top of the nose), and the *inion* (bony lump at the base of the skull on the midline at the back of the head). From these points, the skull perimeters are determined and they are divided into given intervals. Besides the international system, many other lead systems exist for recording electric potentials on the scalp. Bipolar or unipolar electrodes can be used in the EEG measurement. In the first method the potential difference between a pair of electrodes is measured. In the latter method the potential of each electrode is compared either to a neutral electrode or to the average of all electrodes. Guidelines for EEG-recording are periodically revised and published by the American Electroencephalographic Society (see, for example, *Journal of Clinical Neurophysiology*, 1994, January issue).

3.2.3. Biomagnetic Signals

Magnetobiology and *Biomagnetism*, the former dealing with the effects of external magnetic fields on biological systems while the latter studies only magnetic fields generated by these systems, has taken a clear objective and well-defined path in the last 40 or 50 years, after being mixed unfortunately for a very long time with pseudoscience and charlatanism (Valentinuzzi, 2002). There are several research groups in the USA, Finland, Japan and Brazil seriously working in this fascinating area applying top technologies.

Biomagnetic sources can be found in electric currents (as action potentials generated by excitable tissues), in *diamagnetic* and *paramagnetic*

substances in the body, and also in the presence of *ferromagnetic* substances found or injected as tracers in the organism. We should mention also the well-documented existence of magnetic bacteria, found sometimes in sewage waters. These signals are akin to electrical signals so that, collectively, we might speak of bioelectromagnetic signals.
The following smaller case paragraphs very succinctly describe the three types of substances found in nature and in living organisms according to their magnetic properties.

Diamagnetic substances are expelled from a magnetic field, i.e., they are displaced from the points of greater intensity to those of lesser intensity. If the body they form is elongated in shape, then it is positioned (while suspended) with its longest axis oriented transversely to the direction of the field. The field inside these substances is weakened. Antoine Cesar Becquerel (1788–1878) —founder of a dynasty of distinguished scientists— worked and corresponded with Faraday on diamagnetism. He had noticed examples of it (around 1827) before Faraday (who did it circa 1847) but had failed to generalize from them. Michael Faraday (1791–1867), instead, had a clearer understanding of the phenomenon. In fact, he coined the words “dia” and “paramagnetic”. Alexandre Edmond Becquerel (1820–1891) was the second son of Antoine Cesar. From 1845 to 1855, Edmond devoted most of his attention to the investigation of diamagnetism.
Paramagnetic substances are sucked in by a magnetic field, i.e., they are displaced from the points of lesser to the points of greater intensity, and a body of such material is lined up with its greater axis in the direction of the field, reinforced in its interior (field lines are more concentrated). The study of *paramagnetism* is mainly due to Pierre Curie (1859–1906). In fact, he studied the three groups of substances: ferromagnetic, such as iron, that always magnetize to a high degree; low magnetic, or paramagnetic substances, such as oxygen, palladium, platinum, manganese, and several salts, which magnetize in the same direction as iron but much more weakly; and also diamagnetic substances, which include the largest number of elements and compounds (many are biomaterials). Curie concluded that diamagnetism must be a specific property of atoms. It exists also in ferromagnetic or paramagnetic substances but is little apparent there because of its weakness. Ferromagnetism and paramagnetism, on the other hand, are properties of aggregates of atoms and are closely related. The ferromagnetism of a given substance decreases when the temperature rises and gives way to a weak paramagnetism at a temperature characteristic of the substance and known as its “Curie point”. *Paramagnetism is inversely proportional to the absolute temperature*. This is Curie’s law. A little later, Paul Langevin, who had been Curie’s student at l’ École de Physique et Chimie, proposed a satisfactory theory by postulating a thermal excitation of the atoms in the phenomena of magnetization. These are concepts that still constitute the basis of modern theories of magnetism.
Ferromagnetic substances intensify the field extraordinarily. The reinforcement can be thousands of times greater than in the case of paramagnetic substances. The bodies made of

of these substances deform the lines of force to such an extent that the former tend to get nearer to the poles of the magnet generating the field.

When a paramagnetic body is placed in a magnetic field, it also acquires diamagnetism, so that there is a superposition of both properties. To elucidate accurately the value of paramagnetism, it is necessary to make corrections relative to diamagnetism. Paramagnetism always conceals diamagnetism by virtue of its greater intensity. Ferromagnetic bodies are less common in nature than the paramagnetic bodies. Ferromagnetism can be considered a special form of paramagnetism, which manifests when matter acquires a cristalline structure and depends on the intensity of the magnetic field.

The three variants of magnetism (dia-, para- and ferro-) may be characterized by the concept of *magnetic permeability*, which becomes larger when going from diamagnetism to paramagnetism to ferromagnetism. *Magnetic permeability* is related to the concept of *magnetic susceptibility*. There is a well-known mathematical formulation that the student probably already saw in physics courses. This magnetic characterization of substances or materials in general becomes relevant to their respective behaviors when an external magnetic field is applied to them. Physiological or injected magnetic tracers can be detected with susceptometric measurements, as for example iron in liver (Carneiro, Baffa *et al*., 2002). Such signals may help in the study of stomach motility (Carneiro, Baffa & Oliveira, 1999) or in the generation of images of the gastrointestinal tract (Moreira, Murta and Baffa, 2000).

Akin to the electrocardiogram, the electromyogram and the electroencephalogram, we can easily envision the existence of a *magnetocardiogram* (MCG), a *magnetomyogram* (MMG) and a *magnetoencephalogram* (MEG), since the three former, respectively, are time changing electric phenomena. In other words, ion currents in excitable tissues sustain very weak magnetic fields, and this phenomenon stands above the shoulders of Hans Oersted, who back in 1820, demonstrated that an electric charge in motion through a conductor (i.e., a current) sets up a magnetic field. This is basic physics knowledge.

Baule and McFee, in 1963, recorded the fluctuating magnetic field produced by the heart so taking the first *magnetocardiogram* (MCG). They found it had a maximum of about 5×10^{-7}gauss at the peak of the QRS complex. The earth's steady magnetic field strength is in the order of 0.5 gauss; by comparison, the heart maximum magnetic field, at several centimeters from the torso, is only about one millionth of the earth's field. Magnetic disturbances or noise in an urban environment can easily be greater than 10^{-4} gauss. Thus, sophisticated technology is mandatory to extract the weak MCG from this much larger background interference. Since its inception, important advances have been made in biomagnetic instrumentation. The most important is the development of shielded rooms and ultrasensitive magnetometers based on the *superconducting quantum interference device* (SQUID). Description and discussion of the necessary technologies are by far beyond the objectives of this textbook and the reader, if interested, should search the specialized literature. The literature is extensive and several reviews are available, as for example, Williamson, Hoke, Stroink *et al*. (1989). Despite the similar appearance of MCG and ECG waveforms, MCG typically shows greater spatial variation and can potentially provide improved localization of sources (Cohen and Chandler, 1969). In addition, MCG is sensitive primarily to tangential sources while the ECG is more sensitive to radial sources. Thus, additional information may be expected (Verzola, Baffa, Wakai *et al*., 1995). Figure 3.8 illustrates a MCG picked up with a coil perpendicular to the chest of a subject.

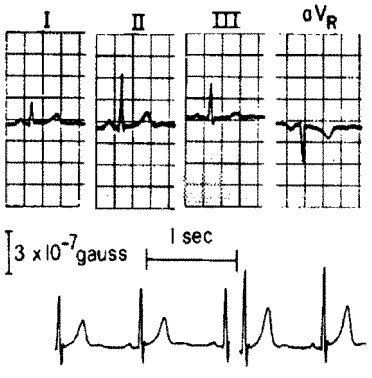


Figure 3.8. MAGNETOCARDIOGRAM. This MCG was taken with the coil axis normal to the chest in two different positions relative to the heart; the ordinate is proportional to the induced magnetic field. The upper channel shows the simultaneously recorded ECG. Courtesy of Dr. David Cohen, MIT, Boston.

The *magnetomyogram* (MMG) is defined as the recording of one component of the magnetic field vector as a function of time due to currents generated by skeletal muscle. Slowly changing or even dc components are recordable by this technique; besides, since no electrodes are needed, contact potentials and interface problems are totally non-existent. Figure 3.9 displays an example. Two MMG's are shown recorded from the elbow and the palm of the hand (Cohen & Givler, 1972).

The *magnetoencephalogram* (MEG) represents the third signal associated to physiological electrical temporal changes, this time from the cerebral neurons. There is abundant bibliography reporting it. Cohen (1972), one of the pioneers in the field, reported the records shown in Figure 3.10, where the traditional effect of closing the eyes is clearly seen. There are possible situations in which the measurement of the brain's magnetic field can reveal information that is either unavailable or very difficult to obtain with the EEG, i.e., dc components and existence of closed loop currents that do not produce differences of potential on the scalp. In both cases, there are magnetic fields.

3.2.4. Bioimpedancimetric Signals

Impedance, in simple words, is a quantitative measure of the hindrance offered by a given system when an applied force-like quantity tries to cause or maintain the passage of a fluid-like quantity through the system. By and large, impedance is a complex relation of the first quantity (as

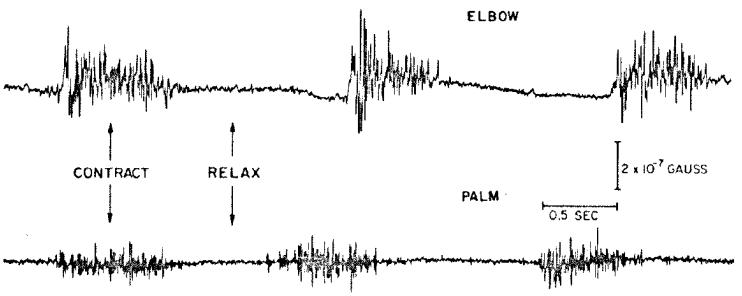


Figure 3.9. MAGNETOMYOGRAM. Records obtained from the elbow and palm of the hand regions showing clearly voluntary contraction and relaxation of the muscles involved in the movements. Observe the smallness of its amplitude. By permission, after Cohen and Givler, 1972.

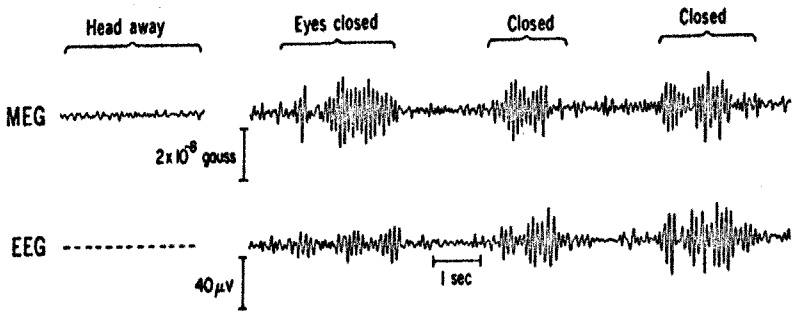


Figure 3.10. MAGNETOENCEPHALOGRAM. The upper channel shows the MEG and the lower one the EEG. Both display the alpha rhythm from the brain of a normal subject. The magnetometer was located at the left occipital region, as were the bipolar set of EEG leads. With the eyes closed, as expected, both signals are greatly enhanced. With permission, modified after Cohen (1972).

mechanical force, hydraulic pressure, electric voltage, magnetomotive force or temperature gradient) to the second one (as velocity, volume flux, electric current, magnetic flux or heat). Possible units in each case are (dynes×s/cm), (dynes×s/cm^5), (volts×s/coulombs = ohms), (amp/volts×s), (°C×s/cal), with the proviso that the second, third and fourth cases correspond, respectively, to Poiseuille's, Ohm's and Hopkinson's laws (Morucci, Valentinuzzi, Rigaud *et al*., 1996).

The impedance concept finds applications in physiology and medicine, as for example, as described above in Chapter 2 in the cardiovascular system (impediment to the blood flow in any vessel, especially in the aorta), or in the respiratory system (impediment to the air flow through the airways), or in otolaryngology, where the acoustic impedance is one of the parameters frequently used to evaluate ear function). Probably, other examples could be found, say, in the gastrointestinal tract or in the reproductive system. However, in this section, we deal exclusively with electrical impedance, as defined by the generalized Ohm's law. Its inverse, admittance, can also be used.

3.2.4.1. Mechanisms to change impedance

A physiological variable induces modifications in the electrical impedance offered by a biological system between any two points. Moreover,

an impedance meter could be connected to such points yielding, at its output, a signal proportional to the physiological variable, but in terms of impedance, either modulus or phase or both. In other words, the impedance meter plays the role of a transducer in the sense that it transforms one kind of change into another kind (Figure 3.11). This signal is obviously of electrical nature, for either a voltage or a current usually manifests it.

To understand the mechanisms by which a physiological variable can produce modifications in the biological impedance opposing the passage of an electric current, two very simple and common electrical engineering models will be considered: the *cylindrical resistor* and the *capacitor of parallel plates*.

In the first case, the resistance between the two ends of a cylinder of length L and cross-sectional area A is given by

$$R = \rho \frac{L}{A} \tag{3.1}$$

where ρ, in (ohms×cm), represents the resistivity of the material the cylinder is made of. Its inverse, 1/ρ, is the conductivity σ. For example, the resistivity of human blood and plasma is, respectively, about 135 and 55 ohms×cm. Values for other tissues have been reported in the literature (Geddes & Baker, 1967). The constancy of this parameter and its numerical determination for biological materials is still a valid subject of

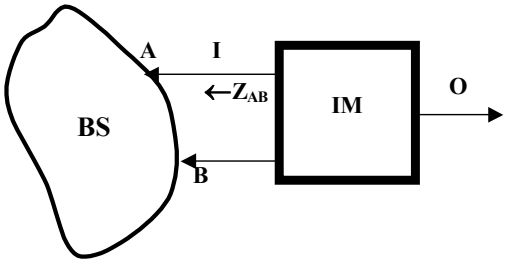


Figure 3.11. THE IMPEDANCIMETRIC SIGNAL. An impedance meter IM looks into the impedance Z_{AB} offered by the biological system BS. To detect the impedance, a voltage E_{AB} between points A and B must be applied forcing a current I through BS. Some physiological variable in the latter changes the impedance and the meter produces an output O. A similar diagram can be drawn in terms of admittance. The student is encouraged to do it.

research, especially *in vivo*. Equation (3.1) clearly shows that any geometrical change will also cause a change in resistance.

In the second case, we consider a capacitor of parallel plates of area A, separated by a distance d, with an insulating material in between. If the material is vacuum, the capacitance value is C_o; if, instead, some other insulator is placed, C will be the value of the capacitance. The relationship $K = C/C_o$ defines the so-called *dielectric constant* of the insulating material. Besides, the capacitance C_o is given by a simple and well-known expression,

$$C_o = \varepsilon_o \frac{A}{d} \tag{3.2}$$

where $\varepsilon_o = 8.85\times10^{-12}$ coulombs^2/N×m^2 stands for the permittivity of vacuum (Tippens, 1973). From the above-said definition of dielectric constant and by the use of equation (3.2), the following is obtained,

$$C = KC_o = K\varepsilon_o \frac{A}{d} = \varepsilon \frac{A}{d} \tag{3.3}$$

where ε represents the permittivity of the insulating medium. It is easy to verify that $K = \varepsilon/\varepsilon_o$ allowing the dielectric constant of the insulating medium to also be called *relative permittivity*. There is still a paucity of full knowledge of this characteristic property in biological materials. By the same token, as mentioned above for the resistive case, if a physiological variable somehow induces changes in the permittivity or in the geometry of the system under study, modifications will also be produced in the reactive component of the biological impedance presented by the system. Foster and Schwan (1989) offered a list of permittivity values for some biological tissues. For example, the relative permittivity of blood they reported is 4×10^3, measured at 100 kHz, while that of skeletal muscle is 1×10^7 at 10 Hz and 8×10^4 at 10 kHz. In other words, it falls as the testing frequency increases. Such behavior is typical of any biological tissue. Observe that relative permittivity (as any relative parameter), because of its very definition, is just a dimensionless number.

3.2.4.2. Concept of biological impedance

If a voltage E_{AB} is applied across any biological tissue, a current I will tend to traverse the tissue, finding electrical impedance that is mathematically described by the complex relation between E_{AB} and I (Figure

3.11). This is similar to the technological impedance definition. The biological impedance Z_B usually shows a constant component Z_o and a time-dependent variational part ΔZ which may vary with different factors of the living tissue, such as its geometry, temperature, biochemistry, or others. Thus,

$$Z_B = Z_0 + \Delta Z \tag{3.4}$$

which can also be broken down into the resistive and reactive components. Usually, the variational part delta is what really matters because it contains the information about the changes induced by the physiological variable. Hence, most of the time the constant component Z_o is discarded.

3.2.4.3. Some examples of physiological impedancimetric signals

– *Respiration*

The impedance technique as a method of detecting respiration is based upon the close correlation found experimentally between changes in the respired volumes and changes in transthoracic impedance; i.e., during inspiration, the impedance increases, while during expiration, it decreases. The precise origin of these impedance modifications has not been well explained yet, but the idea has proved to be useful and easy to implement (Morucci, Valentinuzzi, Rigaud *et al*., 1996).

Figure 3.12 shows records obtained from an 8-year-old girl using the same pair of electrodes placed bilaterally at the xiphoid level and on the mid-axillary line. The first channel is the surface electrocardiogram and the second is respiration (inspiration upward). The figure shows also the time marks at 1 s each. There is an increase in cardiac frequency during inspiration and a decrease during expiration. This physiological phenomenon is known as the respiratory heart rate response (RHRR), which is more marked in children than in the adult. The respiratory impedancimetric signal is quite handy for the newborn, for children in general and also for weak people who may find difficulties breathing into a spirometer. It is also useful in animals, from the small one, as the mouse, to big animals like the cow or horse. Calibration in absolute terms, that is, in ohms per cubic centimeter, remains a problem with a highly restricted answer (Geddes and Baker, 1989; Morucci, Valentinuzzi, Rigaud *et al*., 1996).

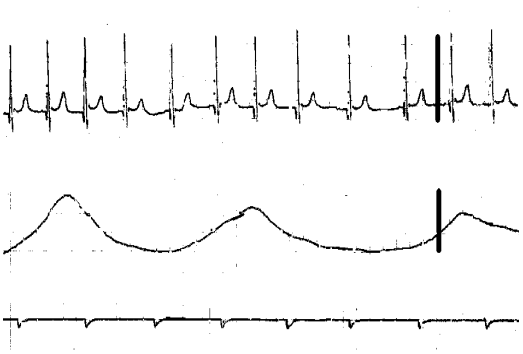


Figure 3.12. RESPIRATION RECORDED WITH IMPEDANCE. The upper channel is the surface ECG and the second channel is the transthoracic impedance. Channel 3 indicates the time marks 1 s apart. The heavy dark vertical bars display calibrations: 1mV for the ECG and 5 ohms for impedance. Records obtained by the author at the Department of Physiology, Baylor College of Medicine, Houston, TX, 1968.

– *Impedance cardiography*

The respiratory signal recorded with the impedance concept contains also a faster and much smaller component, which is contributed by the heart contractions. When adequately recorded and processed, it is possible to obtain stroke volume by application of a simple theory (Geddes and Baker, 1989; Morucci, Valentinuzzi, Rigaud *et al*., 1996).

– *Other bioimpedancimetric signals*

The impedance technique in biology is a subspecialty in itself by far beyond the scope of this textbook. Suffice it to say that it can be used to evaluate the circulation of the extremities, intraventricular volumes, brain circulation, total body water, muscular contractions and a variety of physiological and biological events. In fact, it is only limited by human ingenuity (Geddes and Baker, 1989; Morucci, Valentinuzzi, Rigaud *et al*., 1996; Riu, Rossell, Bragós *et al*., 1999). The latter authors are active in this subject at the Universitat Politècnica de Catalunya, Barcelona, Spain.

The interested student is invited to search in the literature for the multiple applications of the impedance technique describing the kind of signals they produce. There are several websites with ample information. Use the word “bioimpedance” to find them.

As an exercise, try to see why a skeletal muscle, the uterus, the stomach, the vagina, a microbial culture, may cause a change in their respective impedances when either a bipolar or tetrapolar lead system of electrodes are adequately placed. The first three should detect movement, the fourth should give information regarding ovulation in women and estral conditions in animals like mares or cows, and the latter should be related to bacterial growth.

3.2.5. Biohydraulic Events and their Signals

> *The electrical activity gives the* "*go*", *but pressure sustains the flow*, *which*, *in the end*, *keeps the tissues alive*.

This term —*biohydraulics*— is not common, however, it is descriptive. It refers to the pressure and flow developed by fluids in the very many body cavities. In particular, we associate it mainly with the so-called *hemodynamics* in medical practice, i.e., with cardiovascular compartments and their moving blood contents.

The variables playing a role in the cardiovascular system (CVS) have been introduced and discussed in Chapter 2. One of them, *blood pressure* in general, represents a matter of concern and study, but *arterial blood pressure* is particularly important for the wealth of information it contains and because of the current easiness to measure it. Almost everybody is familiar with the arm cuff and the small instrument used daily by practicioners, nurses, and even laypeople at hospitals and homes.

Stephen Hales (1677–1761) —minister at Teddington in Middlessex— was the first to actually measure arterial blood pressure in the horse publishing his results in a famous opera entitled *Statical Essays*, in 1733 (Willius and Keys, 1941).

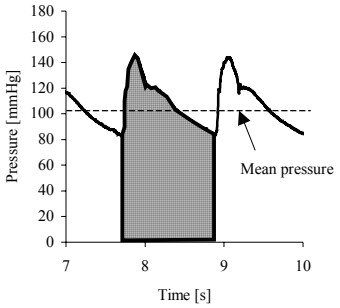


Figure 3.13. ARTERIAL BLOOD PRESSURE. It represents and ideal record showing the maximum or systolic pressure, the minimum or diastolic pressure and the well-known mean pressure defined mathematically as the area under one cycle divided by its duration. Record obtained in the Department of Bioengineering, UNT, 1996.

His verbatim description goes like this: "In December I caused a mare to be tied down alive on her back; she was 14 hands high, and about 14 years of age, had a fistula on her withers, was neither very lean nor yet lusty. Having laid open the left crural artery about 3 inches from her belly, I inserted into it a brass pipe whose bore was 1/6 of an inch in diameter; and to that, by means of another brass pipe which was fitly adapted to it, I fixed a glass tube, of nearly the same diameter, which was 9 feet in length. Then untying the ligature on the artery, the blood rose in the tube 8 feet 3 inches perpendicular above the level of the left ventricle of the heart, but it not attain its full height at once."

He had performed the first measurement of blood pressure by direct cannulation and had invented the open tube vertical manometer. The unit of measurement was inches of blood, nearly equal to water for blood is slightly denser than the former. The student should find out what the withers and the crural artery are. Besides, he/she should think whether the animal was anesthetized or not. When was anesthesia introduced?

The development, which ensued in the 250 years that followed after the scientific event described above, constitutes a fascinating piece of history nicely told by Geddes (1970) in a delightful book. Nowadays, we are able to record blood pressure continuously as a function of time (Figure 3.13). This signal is routinely obtained in research laboratories and during many types of surgical procedures. One of the technical problems

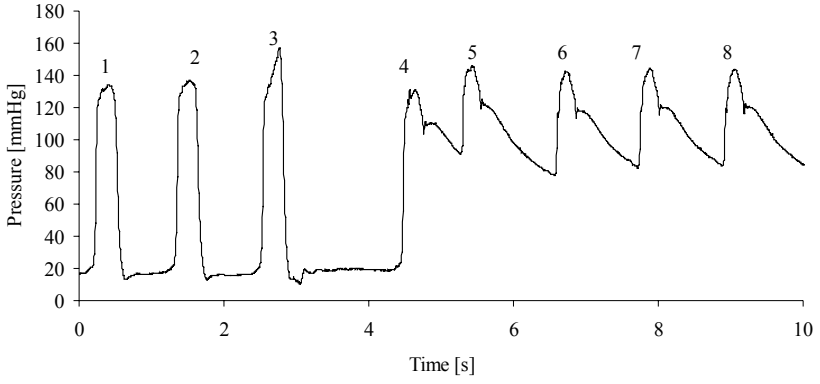


Figure 3.14. BLOOD PRESSURE AS A TIME COURSE EVENT. Experimental canine blood pressure record obtained form the left ventricle (beats 1,2,3) and the aorta (beats 4–8) as a microtip catheter was withdrawn from the ventricle. The dicrotic notch is clearly seen, meaning an acceptable faithful reproduction. If the notch disappears, it may mean incorrect instrumental detection, that is, not enough harmonics are included by the recording system (see text). Records obtained at the Department of Bioengineering, 1996.

encountered is the faithful reproduction of the event, which is related to the harmonic content of the signal. It has been determined that in this particular case of arterial pressure, the spectrum must go from zero frequency (or dc) up to about the 10th harmonic (Geddes, 1970; Geddes, 1984). Figure 3.14 displays an actual arterial blood pressure record from an anesthetized dog.

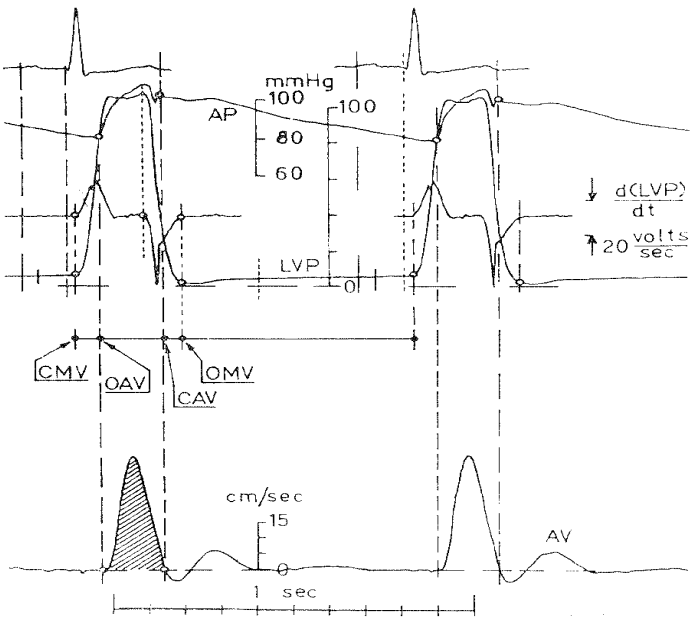


Figure 3.15. SET OF SIGNALS FROM THE CVS (two electrical and three of biohydraulic origin). These are five channels obtained from an anesthetized dog (with fentanyl, droperidol and sodium pentobarbital) using a high-speed recorder. Channel 1 is the ECG, lead II, clearly showing two consecutive beats that were the trigger of the other events. The second channel AP is arterial blood pressure picked up at the root of the aorta with a microtip transducer. It clearly shows the dicrotic notch, which marks the closure of the aortic valve. The third channel is the time derivative (electronically obtained) of the fourth channel, the intraventricular pressure. That derivative is to evaluate myocardial contractility. Finally, the last channel is blood velocity in the upper portion of the aorta detected with a miniature electromagnetic transducer. The lower horizontal bar shows time marks 1s apart. Records obtained by the author at the Department of Physiology, Baylor College of Medicine, Houston, TX, 1976.

Study subject: Determine the conditions for the faithful reproduction of arterial blood pressure in the dog, the rabbit, the mouse and the hummingbird. Hint: Take into account the normal expected cardiac frequency. Search in the literature for different arterial blood pressure records, from different species and also from different arteries within the same subject.

Blood flow is the other biohydraulic event of great significance but not easy to measure or record. Figure 3.15 depicts an excellent set of five signals from the heart. The left ventricle ejects blood in cyclic spurts (see Chapter 2) when the aortic valve is opened (from OAV to CAV in the figure). The area under the velocity curve (dashed) measures stroke volume (velocity in cm/s × ejection time in s) when multiplied by the aortic cross-sectional area in square cm. In this particular case, a postmortem examination gave an aortic diameter of 1.5 cm meaning a cross-sectional area of 1.77 cm^2. This led to a stroke volume of 5.8 cm^3/beat and a cardiac output of 377 cm^3/min at 65 beats/min. The animal (15 kg of body weight) had *filaria* (heart worms in the common language, rather frequent in the state of Texas, where the experiment was carried out) so accounting for the low value. A dog of this weight should have given in the order of 1,200 cm^3/min for cardiac output. CMV and OMV in the figure stand for closure and opening, respectively, of the mitral valve. The student is invited to review the section on the CVS in Chapter 2 to better understand Figure 3.15.

Find out what heartworms are (filaria). As mentioned above, it is a rather common disease in dogs in the State of Texas. How is it transmitted?

Heart sounds are classified here as a biohydraulic kind of signal, admitting that the criterion is rather arbitrary. It might be called a *paraphenomenon* because, as the prefix *para-* indicates, it suggests something auxiliary or derivative from the main word, and precisely, heart sounds are the audible manifestation of important biohydraulic events supplying significant pathophysiological information of clinical value.

The origins of these sounds can be traced back to several causes:

1. Closure and opening of the cardiac valves, the former more than the latter, which produce sudden *acceleration* and *deceleration* of blood leading to *vibrations* that propagate to the anatomical vicinity of the source.

2. Muscular contraction through the *actine-myosin* sliding mechanism and the *friction* associated with it. Sometimes this is termed muscular noise. Skeletal muscle also generates it.
3. *Turbulences* generated when the movement of blood passes from the laminar to the non-laminar regime, that is, when blood velocity is greater than certain limit value usually related to *Reynolds number*.
4. *Friction* in the blood bulk and against the arterial and intracardiac walls.
5. *Cavitations* produced by collapsing bubbles due to vortices or turbulences. This is a well-known phenomenon that occurs in any system of rapidly moving fluid. It is found everywhere from hydraulic turbines to boat propellers and human joint articulations. Pockets of air or gas or even vacuum form and then disappear with a potentially loud sound.
6. Narrowings of different kind that produce the *Venturi effect* (an acceleration).

The student is invited to dig a little deeper into the concepts of actine and myosine, two extremely important muscular proteins; also to review laminar and turbulent flow and the meaning of Reynold's number. Any physiology textbook will serve for the former and physics textbook for the latter. What cavitations are is more elusive and a visit to the WEB is recommended. The phenomenon may take place during deep-sea diving and is associated with the so-called *caisson disease* or decompression sickness (when the diver returns to sea level too quickly). Another physics subject to review is the *Venturi effect*.

In fact, heart sounds and noises consist of a complex set of acoustical signals mostly and commonly detected either with the naked ear or, much better, by means of the traditional *stethoscope*. Clinicians are specifically trained to interpret and to differentiate them from the normal expected pattern. Graphic registration of these acoustical signals, either on paper or on a monitor screen, constitutes the *phonocardiogram* (PCG).
Four transient heart sounds, associated with the above-mentioned basic physical phenomena, occur during the normal cardiac cycle. The two sounds which mark the beginning and end of systole, and which are heard in all normal subjects or mammalian animals, are the first and second heart sounds. In many subjects, the third heart sound will be heard *shortly after the second*. In most, the fourth heart sound will be *heard immediately before* the first heart sound. However, all listeners do not necessarily hear the third and the fourth sounds always.

Because the transient heart sounds correlate to the electrical and mechanical events of the cycle, they provide information on their timing and on the presence or absence of some of the events. Usually the mechanical events on the left and right sides of the heart occur almost simultaneously. Often, the interval is so short that it could only be detected using phonocardiography; however, there are circumstances under which the interval is long enough for it to be detected on auscultation. Even so, it requires good training. If the interval between the events on the left and right sides is sufficient (>30 ms), the sound can be *split* into two elements.

The intensity of the transient sounds varies in different subjects. For example, they are best heard in fit thin persons, but may be very quiet in fat, barrel-chested individuals. They are louder in an excited subject or during tachycardia. Sounds may be muffled by fluid in the thoracic or

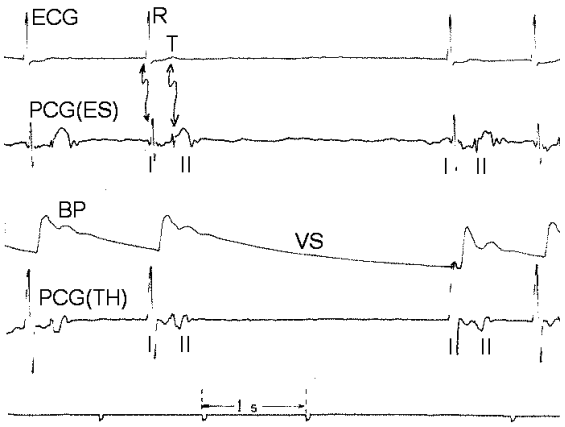


Figure 3.16. PHONOCARDIOGRAMS. These records were obtained from an experimental dog. Channel 1 displays the ECG (II) while channel 2 is the PCG detected with an esophageal probe carrying a small microphone on its tip. The third channel is the carotid blood pressure obtained by means of an external transducer connected to the tubing in turn inserted into the artery. The last channel is also the PCG but this time picked up by a thoracic microphone. The lower time marks, as indicated, give the 1 s calibration. One of the vagus nerves was stimulated triggering a well depicted slowing of the heart VS with a typical diastolic run off. The first sound I (or S1) correlates with the R-wave while the second sound II (or S2) correlates with the T-wave and the dicrotic notch. Records obtained by the author at the Department Bioengineering, UNT, 1978.

pericardial cavities. Solid tissue, such as tumors or abscesses, which displace the heart or are present between the heart and the chest wall, may result in the sounds being reduced or increased in intensity, depending on the position of the heart valves in relation to the body surface.
Auscultation is that part of the physical examination involving the act of listening with a stethoscope to sounds made by the heart, lungs, and blood. It may be complemented by the phonocardiogram.

It was William Harvey (1578–1657) who, over 400 years ago, referred indirectly to the valve noise in any hydraulic system when he wrote ... "as by the clacks of a water-bellows used to rayse water". Obviously, he had noticed the phenomenon and was making a descriptive analogy (McGrady, Hoff and Geddes, 1966). But a systematic study of the thoracic sounds and noises is due to René Théophile Hyacinthe Laennec (1781–1826), the French physician who, in 1816, introduced the stethoscope and the method of auscultation producing, thereafter, a famous book entitled *De l'Auscultation Médiate*, in 1819 (Kliegfield, 1981). James Hope provided a definite demonstration of the valvular origin of cardiac sounds, between 1828 and 1841, after a masterful series of experiments in horses. He hindered the valve's action with a catheter and the sound disappeared (McGrady, Hoff and Geddes, 1966).

Figure 3.16 displays two phonocardiographic channels showing the approximate correlation in time between the first sound S1 with the electrocardiographic R-wave. S1 is mainly produced by the closure of the two atrioventricular valves (left and right heart). There is also a minor contribution from the opening of the arterial valves, which, shortly after, start both ventricular ejections. The second sound S2 recognizes its origin mainly in the closure of the two arterial valves (aortic and pulmonic valves). It is associated with the electrocardiographic T-wave and, as expected, with the dicrotic notch, too. Opening of the two atrioventricular valves slightly after probably add some components to the sound. Thus, the time interval between the first and the second heart sounds is a rough measure of the ejection period. The long diastolic run off due to the vagal slowing of the heart shows that the signals of the PCG channels are not artifacts. These two sounds easily heard in everyone have been phonetically described as *lub-dub*. Figure 3.17 offers another view of the PCG, as intravascular sounds detected with microtransducers implanted at the tip of a catheter within the aorta.

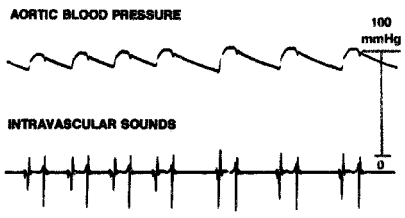


Figure 3.17. FIRST AND SECOND HEART SOUNDS. They were recorded with microtransducers installed at the tip of a catheter inserted via the femoral artery into the ascending aorta. The first sound is the double spike coincident with the diastolic pressure and the second sound corresponds to the other double spike coincident with the dicrotic notch. The distance between each pair is a measure of the ejection duration. These microtransducers are currently being manufactured by Millar Instruments, in Houston, TX. Figure after a brochure of this factory, 1972, by permission, with the author's acknowledgment for the courtesy.

Heart sounds should be distinguished from *murmurs*. The latter, sometimes called *pronoises*, are heard during a normally silent period of the cardiac cycle and mostly relate to pathological conditions. A murmur is like a prolonged "whoosh". It could be due to a narrowing of the aortic or pulmonary valve or to a leak through the mitral or tricuspid valves (both due to acquired or congenital causes), during systole of the ventricles. Figure 3.18 well exemplifies a systolic murmur due to aortic stenosis. But heart murmurs may also be heard when there is signicant narrowing of the mitral or tricuspid valve during diastole (that is during relaxation, when the blood is flowing from the left or right atrium into the re-

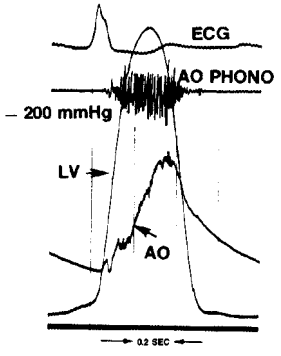


Figure 3.18. SYSTOLIC MURMUR. ECG, PCG, left intraventricular pressure LV and aortic pressure AO recorded with microtip transducers. Narrowing of the aortic valve (stenosis) produces a prolonged murmur, which extends all over systole. The deficient valvular closure affects the rising aortic pressure phase, as clearly manifested in the record. Figure after a brochure of Millar Instrument, Houston, TX, 1972, by permission, with the author's acknowledgment for the courtesy.

spective ventricle) due to acquired (say, rheumatic fever) or congenital causes. Similarly, diastolic murmurs may be heard when the aortic or pulmonary valve leaflets do not adequately oppose each other.

The first heart sound (S1) as recorded by a high-resolution phonocardiography consist of 4 sequential components: (1) small low frequency vibrations, usually inaudible, that coincide with the beginning of left ventricular contraction and felt to be muscular in origin; (2) a large high-frequency vibration, easily audible related to mitral valve closure (M1); (3) followed closely by a second high frequency component related to tricuspid valve closure T1; (4) small frequency vibrations that coincide with the acceleration of blood into the great vessel. The two major components audible at the bedside are the louder M1 best heard at the apex followed by T1 heard best at the left lower sternal border. They are separated by only 20–30ms.

The second heart sound S2 has two components: A2 and P2, coincident with the incisuras of the aorta and pulmonary artery pressure traces, respectively, and terminating the right and left ventricular ejection periods. Right ventricular ejection begins prior to left ventricular ejection, has a longer duration, and terminates after left ventricular ejection, resulting in P2 normally occurring after the A2. Right and left systole are nearly equal in duration, and the pulmonary artery incisura is delayed relative to the aortic incisura, primarily due to a larger interval separating the pulmonary artery incisura from the right ventricular pressure, compared with the same left-sided event. This interval has been called the "hangout interval". Its duration is felt to be a reflection of the impedance of the vascular bed into which the blood is being received. Normally, it is less than 15ms in the systemic circulation and only slightly prolongs the left ventricular ejection time. In the low resistance, high- capacitance pulmonary bed, however, this interval is normally much greater than on the left, varying between 43 and 86ms, and therefore contributes significantly to the duration of right ventricular ejection.

The third heart sound S3 is due to ventricular turbulences generated during the period of rapid ventricular filling, difficult to be heard and not easily recorded; however, in children and adolescents, chances for its detection are better, probably because of the contractions vigor. Obviously, it is an early diastolic event with low frequency content. It occurs

about 0.14–0.16 seconds after the second sound. The S3 reflects decreased ventricular compliance or increased ventricular diastolic volume. The abnormal S3 is heard in individuals with coronary artery disease, cardiomyopathies or incompetent valves.
The fourth heart S4 sound is attributed to atrial contraction. It is a low frequency sound heard just before the first heart sound, occurring during the late diastolic phase. It is seldom heard in normal hearts. Under certain pathological conditions, however, it may show up, as for example when the ventricles have a decreased compliance or are receiving an increased diastolic volume, also in myocarditis or atrioventricular block. It is usually accentuated with inspiration and may be due to pulmonary valve obstruction, pulmonary stenosis, pulmonary hypertension, or right ventricular myocardial infarction.

Phonocardiography constitutes a subspecialty within the broader field of cardiology. The student is invited to visit the WEB in order to search for more information, be it pathophysiological, clinical or technological.

3.3. Signals Produced by Biomechanical Systems

> *Diseases*, *natural catastrophes*, *wars*, *human aggressions*, *and* ***mainly traffic accidents***, *are producing a growing number of physically maimed young people. Biomechanics*, *at the base of rehabilitation engineering*, *can supply devices to help them out along with considerable background knowledge*.

Any moving organ or tissue has a biomechanics associated with it: the skeletal voluntary muscles, the involuntary rhythmically contracting myocardium, all smooth muscles (involuntary too) covering blood vessels, urinary canals, gastrointestinal system, reproductive organs at large, and the articulated skeleton in its manifold complex arrangement. In all of them, somehow, force, length and angular changes are manifested as basic events while slightly more complex ones as tension, acceleration and torques also take place. Production of work is a salient characteristic of mechanical systems. We will briefly review in this section the most

important signals that can be recorded either in the laboratory environment or within the clinical setting.

The electrical signals of skeletal, cardiac and smooth muscles trigger their respective contractions and, thus, they develop force F (measured, say, in dynes, newtons or kilograms-force, the latter unit to be distinguished from kilogram-mass) usually accompanied by a change in muscular length L. The time derivatives of F and L are often used as estimators of contractility; they quantify velocity of contraction (in dynes/s or in cm/s, respectively). In experimental physiology, the myogram and the cardiomyogram are routinely recorded.

Figure 3.19 shows records obtained from a frog gastronecmius muscle increasing the stimulus frequency. It starts with a twitch on the far left fusing to a sustained pull at 50 pulses per second on the far right. Although the stimulus intensity was not altered, greater tension was produced with repetitive stimuli. This is a characteristic property of skeletal muscle and is referred to as the twitch/tetanus ratio. In this case, the tetanic stimulation developed 1.7 times that developed when single stimuli were given; thus, the efficiency of the tetanic contraction is better than that of a twitch. It is a good example of biomechanics because it shows a combination of a physiological property with the actual production of force and work.

Figure 3.20 shows a set of experimental records obtained from a terrapin (the Argentine *Geochelonia chilensis*, mostly found in the NW region of

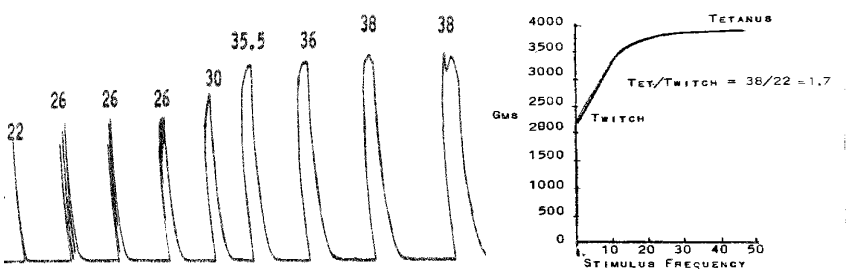


Figure 3.19. SKELETAL MUSCLE CONTRACTION. Lef: Sequence from twitch to full tetanic contraction. Numbers on top indicate the force developed (from 22 to 38 grams). Right: The graph relates tetanic/twitch ratio to stimulus frequency. Record obtained in the Department of Physiology, Baylor College of Medicine, in 1966, when Dr. Hebbel E. Hoff was its Chairman.

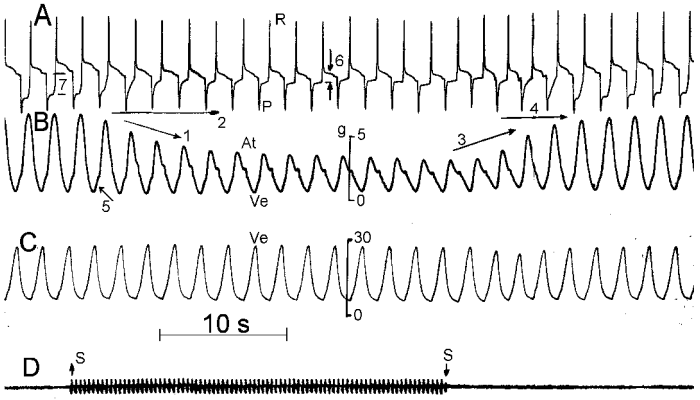


Figure 3.20. CARDIOGRAMS FROM A TERRAPIN (*Geochelonia chilensis*). Channels from top to bottom are: atrio-ventricular direct electrogram, atrial myogram, ventricular myogram and finally below the interval of right vagal simulation. The two calibration bars, 5 and 30 grams-force, quantify force of contraction, respectively, for the atrium and for the ventricle. Records obtained by the author at the Department of Bioengineering, UNT, 1972.

this country). After demedulation (pithing), the plastron was carefully cut open with a circular saw and the pericardium gently dissected through to expose the heart. An atrio-ventricular electrographic lead was recorded from one hook-electrode inserted in the right atrial appendage and another similar electrode inserted into the ventricular apex. This is the record shown by the first channel A. In it, the two main electrical cardiac components are seen: a monophasic-inverted atrial P-wave —with a depression marked as 7— and the ventricular R-wave with an elevation marked as 6. The depression and the elevation indicate damage to both, the atrium and the ventricle, probably due to the small wound produced by the hooks. These two hook-electrodes were also attached to two myographs, respectively, via a non-extensible string to simultaneously record the mechanical atrial and ventricular cardiograms (channels B and C). There are two vertical bars that show both calibrations in grams-force. Obviously, the ventricular pull was much stronger than the atrial one.

Cardiac frequency can easily be calculated as about 57 beats/s, a typical value for this kind of animal during the springtime season in a subtropi-

cal region as the province of Tucuman is in Argentina. The right vagus nerve was mildly stimulated (SS interval in channel D) to elicit the classical and clearly distinguished vagal cardiac effects: The negative bathmotropic effec, or decrease of the duration of the atrial monophasic wave (follow arrow 2) and the negative inotropic effect or decrease of the force of atrial contraction (follow arrow 1). Please notice that the atrial myogram has two waves: one, At, upward, is the true atrial contraction that follows the electrical P-wave; the other is interference from the ventricle, the downwards Ve-wave because since both chambers are anatomically coupled, the ventricle pulls the atrium down when it contracts. The Ve-wave clearly follows the electrical R-component. On the right hand side of the figure, arrows 3 and 4 signal the recovery of both effects after vagal stimulation was off. Thus, these records represent a good example of physiological cardiac biomechanics (see Chapter 2, Section 2).

Surface tension, defined as F/L, where the force F is applied perpendicularly to a unit length over the surface of a tissue, can be studied over the epicardial surface with a multichannel arrangement of small strain gauges (see also Chapter 2, Section 2). Strain is another important geometrical parameter defined as relative change in length, say, during contraction (see Chapter 2, Laplace's Law). Usually, such experimental results of complex design supply background material for the development of models. For example, M.P. Nash and P.J. Hunter, from the University of Auckland, New Zealand (check the internet site for more details) studied heart mechanics through mathematical modeling based on *in vitro* biaxial tests and observations of cardiac microstructure to describe the stress-strain properties of myocardial tissue. They presented stress and strain distributions for the passive heart.

The group of Tatsushi Tokuyasu, Shin'ichiro Oota, Ken'ichi Asami, Tadashi Kitamura, Gen'ichi Sakaguchi, Tadaaki Koyama, and Masashi Komeda (Kyushu Institute of Technology, Dept. of Mechanical Systems Eng., 680–4 Kawazu, Iizuka, Fukuoka 820–8502, Japan, http://www.imcs.mse.kyutech.ac.jp) developed a most interesting and fascinating training system for cardiac muscle palpation. Touching the cardiac muscle is necessary to assess the mechanical conditions of muscle before cardiac surgery. Cardiac palpation is the only way to make surgical plans for left ventricular plastic surgery. The training system for cardiac palpa-

tion these authors have developed consists of a Magnetic Resonance Imaging (MRI)-based virtual left ventricular image and a one-dimensional manipulator as a haptic device. Mechanical properties of the cardiac muscles of a dog and a pig are embedded in the virtual heart. Experiments showed that the developed training system enables users to feel the reactional force to the virtual heart surface from the manipulator in real time. Obviously, epicardial and wall stress constitute biomechanical input signals to the detecting system.

Haptics (from Greek, *haptesthai*, to touch, refers to something relating to or based on the sense of touch) is the study of how to couple the human sense of touch with a computer-generated world. One problem with current virtual reality systems is the lack of stimulus for the sense of touch. For example, if a user tries to grab a virtual cup there is not a non-visual way to let the user know that the cup is in contact with the user's virtual hand. Also, there is not a mechanism to keep the virtual hand from passing through the cup. Haptic research attempts to solve these problems and can be subdivided into two sub-fields, force (kinesthetic) feedback and tactile feedback. The student is encouraged to search for more information on this attractive and relatively new field.

Another type of useful signal is the combination of, say, force with its time derivative or length or pressure also with their respective time derivatives. These are sometimes called *phase diagrams* (Eucker, Lisauskas, Singh *et al*., 1999). They may supply complementary information to evaluating a given biomechanical system. However, their interpretation is not straightforward and, so far, they have limited use. Besides, the term "phase diagrams" is also used in physical chemistry and in chaos theory tending to be confusing or misleading if not precisely identified.

The pressure exerted in a joint, as for example the knee or a vertebral disk, is well documented (Goel & Weinstein, 1990). Sometimes, the values may reach the breaking point of the materials involved. Wilke, Neef, Caimi *et al*. (1999), in Germany, for example, conducted intradiscal pressure measurements with one volunteer performing various activities normally found in daily life, sports, and spinal therapy. The goal was to measure that pressure because loading of the spine still is not well understood. The most important *in vivo* data are from pioneering intradiscal pressure measurements recorded by Nachemson during the 1960's. Since that time, there have been few data to corroborate or dispute those find-

ings. Under sterile surgical conditions, these authors (Wilke, Neef, Caimi *et al*., 1999) implanted a pressure transducer in the nucleus pulposus of a normal L4-L5 disc of a male volunteer. Pressure was recorded with a telemetry system during a period of approximately 24 hours for various lying positions; sitting positions in a chair, in an arm-chair, and on a pezziball (ergonomic sitting ball); during sneezing, laughing, walking, jogging, stair climbing, load lifting, and others. The following values were measured: lying prone, 0.1 MPa; lying laterally, 0.12 MPa; relaxed standing, 0.5 MPa; standing flexed forward, 1.1 MPa; sitting unsupported, 0.46 MPa; lifting a 20 kg weight with round flexed back, 2.3 MPa; with flexed knees, 1.7 MPa. Recall that 1 Pa = 1Newton/m^2; 1000 mmHg = 133,300 Pa = 0.133 MPa, where M stands for mega). They found a good correlation with most of Nachemson's data during many exercises, cautiously concluding that the intradiscal pressure during sitting may be less than that in erect standing, that muscle activity increases pressure, that constantly changing position is important to promote flow of fluid (nutrition) to the disc, and that many of the physiotherapy methods studied are valid, but a number of them should be reevaluated.

Human locomotion and human gait biomechanics is a subject well covered in the literature and of utmost relevance (Medved, 2001). In this field, the importance of signals and measurements is well recognized. Engineering solutions, systems and procedures facilitate a more objective evaluation and a better understanding of the locomotor function, which is yet to be fully understood. It is possible to acquire new and better insight into the mechanism of action and function of the neuro-musculo-skeletal system, both in normal and pathological conditions, which from the engineering point of view, is one of the most complex automatic control systems. Special attention must be paid to kinematic variables, searching for answers as for example the question of which variables have to be measured, and why these can be determined via biomechanical modelling of the human body, thereby enabling quantitative characterization of locomotion by treating the body as a complex multisegmental mechanical system. Limbs, trunk, neck, and head are divided in segments with angular movements developing specific torques.

Data collected through the U.S. Department of Labor's Annual Survey of Occupational Injuries and Illnesses demonstrate the high morbidity of

work-related injuries and disorders. Estimates involving days away from work, approximately 34% of the total, result simply from overexertion or repetitive motion. The true cost of work-related overexertion injuries and disorders in the United States is not known. Conservative estimates of annual expenditures, based on workers compensation payments (indemnity and medical services) and other direct costs, range between $13 to 20 billion. The total cost to society is believed to be substantially higher due to various indirect costs (e.g., lost productivity, costs of hiring and training replacement workers, overtime, administrative costs, and miscellaneous transfer payments) that are not included in the conservative estimates. The total annual societal cost has been estimated to be as high as $100 billion.

Epidemiological and laboratory-based research methods have both been used to evaluate the significance of various risk factors associated with work-related musculoskeletal disorders. These studies are designed to look for significant associations between exposure to ergonomic risk factors such as force, repetition, and/or posture. The latter closely relates to excessive tension in muscles, ligaments, tendons and even bone structures. There is strong evidence also of a causal relationship between low-back pain and whole-body vibration, and between segmental vibration and hand-arm vibration. However, there is an inherent difficulty in the determination of the most adequate variables. Biomechanical models and laboratory studies do not replace epidemiological studies, but these approaches provide important complementary information in the quest toward understanding the complex process of how exposures to ergonomic risk factors result in physiological responses that may ultimately lead to work-related injuries and illnesses.

3.4. Signals Produced by Biomaterials

Whenever an external technological piece is introduced in the biosystem, in close contact with it, the problem of compatibility/rejection shows up. The new material does not have to react neither chemically nor biologically nor the tissue cells must attack the first one in any way.

The biomaterials, which currently are utilized in the majority of tissue engineering approaches, function to provide mechanical support and/or

deliver cells to a desired anatomic site. However, they are not designed to specifically interact with desired cell populations and guide the resultant structure and function of engineered or regenerated tissues. Biomaterials are also designed to provide several different types of information to cells in their environment and, thus, mimic many functions of the native extracellular matrices found in tissues. This aspect involves the design and synthesis of new polymers, the development of new processing approaches for existing polymers, and extensive physical and biological characterization of the resulting tissue engineering scaffolds.

A variety of types of information may be conveyed to cells from the biomaterial. The specific mechanism of cell adhesion to a material (e.g., specific ligand-cell receptor bonds) may be designed into a material to promote a specific pattern of gene expression in the adherent cells. Mechanical signals may be conveyed to the adherent cells via the biomaterial, and the specific receptors used to convey the mechanical signal may be varied to tune the cellular response to the mechanical signal. In addition, specific combinations or sequences of growth factors may be locally released from the biomaterials to affect local cell populations, and drive various processes such as migration or proliferation. Both biodegradable polymers and ceramic systems are also being developed to control local tissue formation with biomaterials.

There is a tremendous medical need for new tissues and organs for people suffering from a variety of diseases and accidents. New ways are being studied to grow, or engineer (thus, this has come to be called *tissue engineering*) new tissues and organs using biomaterials that serve to guide new tissue formation following placement in the body. There are a variety of approaches to achieve this goal such as transplanting cells, delivering proteins which make cells already in the body form new tissues, using gene therapy to grow tissues, and using the biomaterial by itself to drive regeneration. A common theme in all of these approaches is combining basic studies on the mechanisms by which cells interact with materials and synthesizing new polymers that mimic natural materials in the body. These include bone, cartilage, smooth muscle, skeletal muscle, liver and other soft tissues such as intestinal tissue, salivary glands, and pancreas.

The biomaterials field is certainly dynamic and diverse. Medical device manufacturers benefit from the availability of biocompatible materials — particularly biocompatible coatings and implant materials— but may in fact find their devices competing with entirely new treatment modalities. A piece of equipment that mimics organ function may eventually give way to a complete synthetic organ, and a hip implant could be rendered unnecessary thanks to a newfound ability to regenerate the patient's natural bone. Such developments are still in the future, but medical device manufacturers should start considering how they can compete with or complement these emerging technologies.

For example, researchers at the Massachusetts Institute of Technology and New York University reported in the June 6, 2000 issue of the *Proceedings of the National Academy of Sciences* (PNAS) that they have made a biomaterial that supports living nerve cells. This peptide-based scaffold, on which neurons grow fibers to communicate with each other and establish functional synapses, may be the long-sought ideal medium for growing replacement nerve cells for victims of spinal cord injuries and other forms of nerve damage.

New biological materials based on the tiny protein linkages called peptides will become increasingly important in developing approaches for a wide range of innovative medical technologies, including controlled drug release, new scaffolds for cell-based therapies, tissue engineering and biomineralization. These peptide-based biomaterials can be designed at the molecular level. Although parts of animal cells such as collagen can be extracted as a basis for growing cells, such animal-derived materials may carry and pass on viruses to the attached growing cells. In contrast, if the peptide-based material is not extracted from animal cells, the latter problem does not exist; however, unlike other synthetic materials, these peptides are still completely biological. The peptides are composed of amino acids, which are the building blocks of all proteins. The peptides do not evoke an immune response or inflammation in living animals, and they can be used for a variety of applications. They can be individually tailored to grow virtually every type of cell in the body.

While a successful scaffold has to be cell-friendly, it has to be particularly hospitable to grow nerve cells. The brain seems to keep firm control on the reproduction of its cells. Neurons grown on the wrong substrate

die. After an initial period of development at the very beginning of life, very few new neurons are produced by the adult central nervous system. Researchers have found that peptides can self-assemble into non-protein-like structures such as fibers, tubules, sheets and thin layers. They can be made responsive to changes in acidity, mechanical forces, temperature, pressure, electrical and magnetic fields and light. They are stable at temperatures up to 350 degrees Celsius and can be produced up to a ton at a time at affordable cost. They can be programmed to biodegrade.

3.5. Cellular Signals

Mechanical signals regulate the development of a variety of tissues (e.g., blood vessels, bone, cartilage) in our bodies, and may play a role in various diseases as well. We hypothesize that mechanical signals will be critical to the development of engineered tissues as well, and will be required to achieve full function. At the most basic level, studies of how externally applied mechanical signals are sensed by cells and are transferred intracellularly (mechanotransduction) are being performed with muscle and bone cells. The role of focal contacts, both as pathways for mechanical signal propagation, and as targets of the mechanical signal, are being studied. There is also the possibility that the cytoskeletal assembly is directly regulated by external mechanical signals. In other words, the cytoskeleton would be able to translate mechanical signals into changes in biochemical signaling pathways. At the tissue level, the response of three-dimensional engineered tissues to external mechanical loading is being delineated in a variety of *in vitro* model systems. Altogether, these studies should define mechanisms by which external mechanical signals regulate gene expression and lead to strategies to exploit these mechanisms in the context tissue engineering and regeneration. Obviously, this subject strongly intermixes with tissue engineering.

3.6. Image as a Signal

by Emilce Moler and Max E. Valentinuzzi

> *What we hear, feel, taste or smell certainly may cause a lasting impression on us, but what we see really catches our inner self.*

3.6.1. What is an Image?

The retina acts as the instant film that continuously plays the changing surroundings where we act and move about. Many times, the scenery is somehow, permanently or temporarily, fixed on a physical vehicle (piece of paper or screen) taking the form of a picture, painting, radiography or the like. In this way, examination can be carried out at leisure and supposedly in detail. As opposed to a static view (say, a photograph), a dynamic event (as shown in a movie) is even better. Obviously, seeing the object and/or seeing the phenomenon going on appear as most attractive to the scientist at large and to the biological researcher and clinician in particular. This is the main reason the effort invested in the development of imaging techniques has been enormous, constant and still growing, and the accomplishments have been outstanding and even astonishing, from the already centenary X-rays to the newer systems introduced in the last decades, be it computed assisted tomographies (CAT´s) of different kind, nuclear magnetic resonance (NMR), or ultrasounds.

In simple words, an image is a representation, likeness, or imitation of an object or thing, a vivid or graphic description, something introduced to act in place of something else. An image contains descriptive information about the object it symbolizes. For example, a photograph displays information in a manner that allows the human eye and brain to visualize the subject.

3.6.2. An Image Can Be Converted to a Sequential Time Event

Most images originate in the form of optical light energy; this is the image form that we deal with every day. Optical light energy images are the type we perceive *at once* with our eyes, having the *simultaneity effect*, as some kind of on-off phenomenon: the landscape is either perceived as a whole or not, in two dimensions (the third dimension is not real, but per-

ceived as such, that is, it belongs to the integration process carried out at higher centers). However, an optical image may be converted to an electrical sequential signal with a video camera or similar device, so becoming a time event. This conversion changes the representation of the image from an optical light form to a continuously varying electrical signal that can be displayed on an oscilloscope screen. Of course, the signal seen on the scope is not perceived any more as an image, but it does contain the necessary information about it. Furthermore, the analogue image signal can be digitized and turned into discrete data form. Image processing operations can be applied to an image either in the optical, analogue or digital form.

Besides the true images that can be seen and perceived by eye (as photographs, drawings, or paintings), there are also some non-visible physical properties (for instance, temperature, pressure, population density maps, mountainous geographic regions, sea depths, or the like) that may be converted in some special type of image. They are, instead, distributions of measurable physical properties. In other words, an image (as a landscape) is created from a given characteristic distribution.

3.6.3. Mathematical Model of an Image

In order to represent and manipulate images on the computer, we must define appropriate mathematical models where we try to find schemes to allow the discrete representation of these models, with the purpose of obtaining an encoding of the image on the computer.

Since the image is presented as a two-dimensional signal from which an abstract model is to be derived, the conceptual framework of signal theory appears acceptable and useful. It becomes necessary to make an extension of the sampling theorem of Shannon–Whittaker because this theorem was derived assuming that the signal is one-dimenensional. It establishes that if the signal is limited to a frequency band ranging from 0 to Ω cycles per second, it is completely determined by samples taken at uniform intervals at most $1/(2\Omega)$ seconds apart. Thus, we must sample the signal at least twice every full cycle. The sampling rate limit $1/(2\Omega)$ is known as the Nyquist limit, in honor of H. Nyquist, who pointed out to the importance of such value in telegraphy. In other words, *the theorem relates the high frequencies in the signal with the sampling rate*. Intuitively speaking, the higher the frequencies present in the signal, the

higher the sampling rate must be if we want to have a faithful reconstruction. It is easy to extend the theorem to two-dimensional signals for its application to images.

The term *image* refers to a two-dimensional light intensity function, denoted by $f(x,y)$ where the value of f at spatial coordinates (x,y) gives the intensity or brightness of the image at that point. As light is a form of energy, $f(x,y)$ must be finite and a non-zero number, that is,

$$0 < f(x,y) < \infty \tag{3.5}$$

The images people perceive in every day visual activities normally consist of light reflected from objects. The basic nature of $f(x,y)$ may be characterized by two components: *illumination* and *reflectance*. They are combined as a product to form $f(x,y)$, that is,

$$f(x,y) = i(x,y) \times r(x,y) \tag{3.6}$$

where $i(x,y)$ is the illumination component. It represents the amount of source light incident on the scene. The nature of this component is determined by the light source. The second function $r(x,y)$ is the reflectance component. It represents the amount of light reflected by the objects in the scene.

It is common to call intensity of a monochrome image f at coordinates $f(x,y)$ the gray level l of the image at that point, which lies within the interval,

$$L_{\min} \leq l \leq L_{\max} \tag{3.7}$$

Such interval $[L_{\min}, L_{\max}]$ defines the *gray scale*. Common practice is to shift this interval numerically to the interval $[0, L]$, where level l, equated to level 0, is considered black, and level l becomes equal to level L, is considered white in the scale. All intermediate values are shades of gray varying continuously from black to white.

The use of color in image processing is a powerful descriptor that often simplifies object identification from a scene. Besides, the human eye can discern thousands of colours, shades and intensities, as compared to only no more than two-dozen shades of gray. The process of the human brain color perception is a psycho-physiological phenomenon not yet fully understood. Basically, the colors perceived in an object are determined by the nature of the light reflected from the object. If the light is achromatic, i.e., void of color, its only attribute is its intensity. Thus, the term gray

level refers to a scalar measure of intensity that ranges from black to white.

Three basic quantities are used to describe the quality of a chromatic light source: *radiance*, *luminance and brightness*. *Radiance* is the total amount of energy that flows from the light source; it is usually measured in watts (W). *Luminance* gives a measure of the amount of the energy an observer perceives from a light source; it is measured in lumens (lm). *Brightness* is a subjective descriptor that it is very difficult to measure; it represents the achromatic notion of intensity and is one of the major factors in describing color sensation.

Owing to the structure of the human eye, all colors are seen as variable combinations of the three so-called primary colors: red (R), green (G) and blue (B). These primary colors can be added to produced the secondary colours of light: magenta (read + blue), cyan (green + blue) and yellow (red + green).

The characteristics generally used to distinguish one color from another are *brightness*, *hue and saturation*. As already said, *brightness* indicates the chromatic notion of intensity, while *hue* is an attribute associated with the dominant wavelength in a mixture of light waves, that is, it represents the dominant color perceived by an observer. When we call an object red or blue, we actually refer to its hue. *Saturation* refers to relativity purity or to the amount of white light mixed with the hue. The pure spectrum colors are fully saturated. Hue and saturation taken together are called chromaticity; hence, a colour may be characterized by its chromaticity and brightness.

3.6.4. Image as a Tool in Medicine and Biology

The analysis of different types of images is an invaluable tool in biology research and medicine in these days. These technologies have greatly increased knowledge of normal and diseased anatomy for medical research and are a critical component in diagnosis and treatment planning. In the late 1960s, the medical diagnostic imaging field began to apply digital image processing techniques to X-ray images. Digital biological image analysis techniques were first applied in the 1970s and, with the advent of microprocessor–based personal computers, it increased in popularity in the 1980's.

Biological and medical uses for digital image processing techniques have expanded through the 1980's and 1990's to include Computed Tomography (TC), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET) and ultrasound imaging applications. These images provide an effective means for non-invasively mapping the anatomy of the subject.

We can apply various techniques for improving the visibility of features that are not evident or clear in the original image, such as contrast balancing and edge sharpening. In some cases, digital image processing techniques provide totally automated systems for specimen analysis, automatic counting and classification of cell structures and other objects meeting prescribed characteristics (as bone, tissue and cell analysis). Analysis, classification, and matching of DNA material through these techniques is a common practice

Medical radiological imaging looks at the internal components of the human body. X-ray imaging and computed tomography techniques make intensive use of digital image processing too. For instance, Digital Subtraction Angiography (to enhance blood vessel) or Computed Tomography (to create images using multiple projections).

3.6.5. Digital Image Processing

No matter what the origin of an image is, it must ultimately exist in a form the digital processor understands: the digital form. So, we must create a representation universe, where we try to find schemes to allow the discrete representation of these models, with the purpose of obtaining an encoding of the image on the computer. We will now specialize those concepts to the case of images. We have three abstraction levels, corresponding, respectively, to continuous models, discrete representations, and symbolic encoding of images.

Within the digital domain, an image is represented by discrete points of numerically defined brightness. By manipulating these brightness values, the digital computer carry out immensely complex programming process that allows us to adapt and modify digital image processing operations quickly.

Digital Image processing has become a significant form of image processing because of continuing improvements in sophisticated semiconductor technologies. This has led to computer hardware performance in-

creases and declining costs. Digital image processing algorithm research also continues at fast pace, spurred on by increasing commercial demands. When coupled with an expanding applications base, digital image processing techniques are becoming key tools in diverse new industries. The interested student may find detailed background material in the classic books by Castleman (1979), in Gonzalez and Woods (1992) and in Glasbey and Horgan (1995). Also, he/she may check papers such as Chalama and Kim (1997) and/or Chittaro (2001).

Emilce Moler was born in La Plata, Argentina. She received the undergraduate degree in Mathematics and the Master of Sciences from the Faculty of Sciences of the Universidad de Mar del Plata, Argentina, in 1983 and 1995, respectively. She joined the Department of Mathematics of the same university in 1987, where she is currently Assistant Professor of Computing and Image Processing. In 1988, she became also member of the Signal Processing Laboratory of the Department of Electronics. Recently, Emilce obtained the PhD degree at the Universidad Nacional de Tucumán with a dissertation supervised by the author of this book. Her research interests include image segmentation and mathematical morphology with applications in biomedicine.

3.7. Concluding Remarks of Chapter 3

Physiological signals have been reviewed to introduce the student to this rich set full of invaluable information. Bioelectric events were arbitrarily divided in non-traditional and traditional signals, the former originating in the eye, the skin and the oocyte, while the latter have their respective stems in the heart, brain and muscle. Signals related to bioelectric events are the biomagnetic ones (such as the magnetocardiogram, magnetoencephalogram and magnetomyogram) and the bioimpedancimetric signals derived from changes in a bioimpedance. Thereafter, we moved on to biohydraulics, with its two main representatives: blood pressure and blood flow. Heart sounds, perhaps finding opposition from some people for their inclusion in this section, were considered as a paraphenomenon of the biohydraulic event. Section 3.4 dealt with biomechanical signals. In them we included force, length, torques, acceleration, surface tension, mostly generated by muscles in general. Biomaterials, cells and tissues are signal sources that do not seem to comply with traditional and more commonly handled concepts, but they do exist and are responsible for

quite a revolution in the biomedical sciences. Their signals are mostly of chemical nature. Their characterization, however, may not be an easy task. The chapter is closed with a brief description of the signals produced by an image, perhaps the most powerful form to transmit information. The thermic signal was not described. Its use is mostly limited in the detection of breast cancer using infrared sensors. Unfortunately, it is also used in war to locate enemies in dark or foggy environments.

Chapter 4

Signal Pick Up

by Carmelo J. Felice, Rossana E. Madrid and Max E. Valentinuzzi

Physiologists and biomedical engineers are very much like the fisherman, who throws the line as far and as deep as possible to catch the known and also the unexpected piece.

4.1. Introduction

Once the signal is identified, the task is to pick it up in order to study it at leisure, to characterize it and sometimes to better determine its true source. It may not be an easy enterprise. History of science has taught us several sweat-and-blood examples, as the nerve action potential and arterial pressure, to mention only two. Obviously, a gadget is required, something equivalent to the fishpole-fishhook-bait assembly. This is exactly what *electrodes*, *sensors*, *transducers* and *biosensors* are for.

The latter three words are often used interchangeably, however, they have slightly different meanings. Strictly speaking, a sensor just detects the signal under the original type of energy (electrical, mechanical, thermic, magnetic, chemical, luminic) while the transducer only transforms the small amount of energy (whatever the type) contained in a signal into another type of energy, the second usually being electrical. Thus, it literally "translates"; but it needs a sensor. However, the sensor proper is often so well immersed in the transducer as to make it essentially impossible to dissect them out. Electrodes are specific electrical sensors; in a way, they might be considered as electric-electric transducers because their input signal is in the form of electrical energy and the output is also of the same kind. A biosensor, as its prefix indicates, includes in it some kind of biological material, as for example an enzyme, which reacts with an external substrate to catalyze the production of a given substance that,

in turn, generates a transducible signal. The whole may turn into a semiconductor chip. The final output will be always electrical. Many textbooks on bioinstrumentation have dealt with this relatively new and still developing subject.

Herein, the objective is to introduce the student to another section of the biomedical engineering world recalling, from the very starting line, that no recording system can be better than its picking up gage, for sensors-transducers, in general, are the bottleneck of the system. Tremendous advances have been made in the XXth century and they are still subject of research and development with new ideas being incorporated almost every day to scientific and technical knowledge. Most of this material belongs to the area of bioinstrumentation, one of the traditional and early specialties of bioengineering/biomedical engineering.

4.2. Electrodes: The Electric-Electric Transducer

This section describes and discusses the entrance gate for bioelectric signals to the electronic and computer environment. Electrodes represent such gate; they are the place where ionic currents (typical of living organisms) are transformed into electronic currents (usually circulating in technological equipments).

As mentioned in the preceding chapter and also above, by and large, physiological systems generate five types of signals, i.e., *electrical*, *mechanical*, *magnetic*, *thermic and chemical*; electrical and chemical ones have received a great deal of attention. We have referred also to *biohydraulic* signals; however, they do not encompass a different kind because they belong to the mechanical group.

The human being is in many respects an *electric animal*; ferromagnetic materials in it are only present in tracer amounts, as for example, iron in the liver. If biological magnetic fields were of higher intensity, perhaps simpler technologies would have been developed earlier. However, this is not the case, and biomagnetic fields sustained by bioelectric currents are so feeble that only within the last four decades or so high level technologies were developed to sense and transduce them adequately, as briefly described in the preceding chapter.

Very simplistically speaking, to be alive, a living organism must keep dc and ac electric potentials with currents traversing fluids and membranes, must maintain an enormous variety of biochemical reactions and carefully watch over the concentration of essential substances. Collectively, all such actions are encompassed by the term *metabolism*. Medical science uses the variables involved in these processes and the biomedical engineer constantly digs into them searching for better and more reliable probes, that is, *sensors-transducers*.

Electrodes are the front-end gadget to sense and measure electric potentials or currents. They touch tissues, usually directly but sometimes indirectly (as the case is when electrolytic bridges are interposed). For example, the ECG requires surface electrodes directly touching the patient's skin; if membrane potential is measured, instead, a very small electrode (a microelectrode) needs to puncture the cell membrane to get in contact with the intracellular fluid. In any case, always at least two electrodes are mandatory, one of them acting as the return pathway. Sometimes, three electrodes are used, the third to electrically balance out the system. The output signal is calibrated either in volts or amperes, depending on whether voltage or current is measured. As stated already above, the electric potential or current sensed by the electrodes in fact also transduce ionic currents into electronic currents that excite the electronic input circuitry.

Electrodes can generally be classified into three groups:

1. noninvasive, or those that are applied to skin surfaces, such as conventional ECG limb electrodes or EEG scalp electrodes;
2. semi-invasive, such as nasopharyngeal and tympanic electrodes or EMG acupuncture needle-type electrodes; and
3. invasive, which include depth, grids, and subdermal needles.

The classic book by Geddes and Baker (1989) carefully describes, with examples, a wide variety of commonly used electrodes. Many are even commercially available. Sizes cover an ample range, from centimeters down to minute and sometimes rather elaborate pieces, as the case is with double–barrel microelectrodes or small multielectrodic arrays for research use in electrophysiology.

Student task: Search in the literature, or perhaps with a specialized professional, for different types of electrodes. Try to examine them identifying their parts.

In the case of biochemical reactions, even at biomolecular level, detection is indirect. Technology usually does not deal with transducers that produce chemical signals as output because it is more difficult to handle and interpret them. Chemical signals, instead, frequently via enzyme concentrations, are translated into electrical counterparts. The preceding chapter, devoted to signals, offers a variety of examples where electrodes are an indispensable requisite. We can include in them the measurement of pH, oxygen and carbon dioxide levels in blood, and also many other substances such as specific ions.

4.2.1. The Interface: Two Lion's World

Absolutely all cases wherein electrodes pick up an electric signal fall in either of two categories: Either the sought signal is *within* the very close neighborhood of the electrodes themselves, as when measuring pH or bacterial concentration (Felice, Madrid, Olivera *et al*., 1999; Felice & Valentinuzzi, 1999), or the desired signal *crosses* the electrodic surface, as when recording ECG or EEG or EMG.

Both cases require good understanding of the electrode behavior when placed in touch with a biological tissue or with a cell suspension. A well-known way to study it makes use of metallic electrodes immersed in a simple *electrolytic medium*, as for example, sodium chloride solution (NaCl), which is called saline solution, for short. Over the last hundred years, say from Warburg, in 1899, to Geddes, in 1997, and many other authors in-between, abundant electrochemical research has produced models to at least partially elucidate the real system. Remember that models are neither right nor wrong; they are just approximations that may describe better or worse the system under study depending on the specific conditions of the latter.

The contact region between an electrode and the surrounding medium is ill-defined; since it is located in the midst of two entirely different worlds, namely the electrolytic environment and the electronic side, or the solution and the metallic face, it was named *electrode-electrolyte interface*, the latter word meaning literally "between faces". Colloquially speaking, two faces looking at each other and exchanging kisses.

How can an electrode transduce, or translate, electricity into electricity? When two stainless steel small pieces are immersed in 0.9 % NaCl and an external battery is connected to them, a measurable electric current

will circulate from the positive pole through the solution and back to the battery via the negative side; a very simple circuit, indeed. However, and to our surprise, the moving electric charges or *carriers* are not the same depending where in the circuit we look at. In all metallic portions such carriers are electrons, each with an equal and well-known amount of charge; instead, in the bulk of the electrolyte the carriers are Na^+ and Cl^- ions, also with a definite and known electric charge. In both places, the current is defined as the number of charges (in coulombs) per unit time (in seconds) and, thus, we could speak in terms of charge flow (in amperes). Since the circuit forms a closed loop, the total current sustained by the battery must be the same at any place, be it the solution, the electrodes, the wires or within the battery, each place always considered as a whole (to distinguish from current per unit cross-sectional area or current density which may differ at different points). Somewhere in the circuit carrier's change from one type (electrons) to the other (ions) or vice-versa and that critical place is precisely the *electrode-electrolyte interface*. There is an exchange phenomenon that consists of ions either receiving or delivering electrons. The ionic species can be as simple as the above-mentioned Na^+ and Cl^- or as complex as those derived from aminoacids or other biochemical compounds.

An interface is formed spontaneously the very moment an electrode is immersed in an electrolytic solution. The metal can be considered as a cloud of free electrons around positive ions fixed or attached to a crystalline network. Instead, free positive and negative ions and polar molecules that behave as orientable dipoles usually form the solution. Once the electrode is submerged in the solution, a redistribution of charges takes place on both sides of the limiting faces building up a metal-solution difference of potential. Two basic phenomena account for it,

appearance of superficial either free or induced charges, *be it in excess or defect*, *with respect to the bulk of each face*, or

formation of a layer of oriented dipoles towards the electrode surface.

The complex system of charges and oriented dipoles has been named the double electric layer or simply the double layer. Another name is *Helmholtz double layer*, for it was this German scientist who first studied it in 1879. In the early times of this knowledge, it was thought that only two layers of charge formed such system (one positive and the other nega-

tive), thus its name. However, now it is well documented that the charge distribution is far more complex and only exceptionally is formed by just two layers. No matter how complex the distribution is, the interface region keeps an overall electrical balance; in other words, the principle of neutrality is maintained.

The electric-electric transduction concept becomes clear now with the *transference of charges* occurring in the double layer region (the term is still used in spite of its complex distribution). The charge transference process is one of the several electrochemical reactions that take place at the interface. Say that we look at the solution side. An ion must first arrive in the interface region. Such arrival can be due to any of three different transport processes:

- *migration* or *drift*, when the driving cause is an electric field;
- *diffusion*, when concentration gradients are present, irrespective of the particle having or not an electric charge; and
- *convection*, which happens when movement is due to temperature gradients or to mechanical stirring of the medium. Combinations of them are also possible.

Once at the interface, ions either deliver electrons to the electrode or receive electrons from it. The exchange is collectively called the *oxidation-reduction process* (oxidation refers to loss of electrons and reduction to gain of electrons). In the case of sodium, potassium, or chloride, reduction or gain of an electron is a simple reaction. With proteins the situation appears as more complex, sometimes involving short-lived intermediate substances before reaching the final product.

Diffusion usually is directed toward the interface because the ionic concentration at the double layer is lower than in the solution bulk. This is due to the very transference, for once it takes place, the ion disappears as such, it is replaced by another species, and its concentration decreases.

Hermann von Helmholtz (1821–1894) can perhaps be considered as the first biomedical engineer; thus, a word about him should be said. It is educational. He attended the Potsdam Gymnasium showing interest mainly in physics and he would have liked to study that subject at the university. The family financial position, however, meant that higher education was conditioned to a scholarship. Such financial support was only available for particular topics and Hermann's father persuaded him to study medicine, which was supported by the government. In 1837, Helmholtz (he was 16) was awarded a grant to enter the Royal Friedrich–Wilhelm Institute of Medicine and Surgery, in Berlin. But he had to sign a document

promising to work for ten years as physician in the Prussian army after graduating. Besides, he went into mathematics on his own and also read philosophy, particularly Kant (not easy task for a young guy, indeed). His research career began in 1841, with his dissertation. The direction taken by physiology though, based on the so-called “vital forces, was against his view, maintaining that physiology had to be set on the principles of physics and chemistry. He graduated from the Medical Institute in 1843 and was assigned to a military regiment at Potsdam, but spent all his spare time doing research. He concentrated on showing that muscle force was derived from chemical and physical principles. In 1847, he published a very important paper, *Über die Erhaltung der Kraft* (On the conservation of energy). Many of his ideas were based on concepts previously advanced by Clapeyron and Sadi Carnot (students: find out who they were). He demonstrated that in various situations where energy appears to be lost, it is in fact converted into heat. That paper was an important contribution and it was quickly seen as such. In fact, it played a large role in Helmholtz's career for the following year he was released from his obligation to serve as an army doctor so that he could accept the chair of physiology at Königsberg. His career progressed rapidly there, publishing important contributions to physiological optics and acoustics. He received great acclaim for his invention of the ophthalmoscope in 1851 and rapidly gained a strong international reputation. In 1852, he published his theory of color vision. In 1855, he was appointed to the chair of anatomy and physiology in Bonn. Hydrodynamics was also subject of his concern with contributions that had a strong impact, although somewhat delayed. Helmholtz agreed in 1858 to setting up a new Physiology Institute in Heidelberg and this event was mixed with some personal problems (his father’s death in 1858 followed by his wife’s in 1859). He was left to bring up two young children and within eighteen months he married Anna von Mohl, daughter of a professor at Heidelberg. Some of his most important work was carried out while he held this post in Heidelberg. He studied mathematical physics and acoustics producing a major study in 1862 that looked at musical theory and the perception of sound. In 1843, Ohm had stated the fundamental principle of physiological acoustics, concerned with the way in which one hears combination tones; Helmholtz improved it formulating a resonance theory of hearing. From around 1866, Helmholtz began to move away from physiology and towards physics. When the chair of physics in Berlin became vacant in 1870, he indicated interest in the position, which he took in 1871. A major topic that occupied Helmholtz was electrodynamics, searching for a compatibility with the principle of conservation of energy. Helmholtz devoted his life to seeking the great unifying principles underlying nature. His career began with one such principle, that of energy, and concluded with another, that of least action. He longed to understand the ultimate, subjective sources of knowledge. That longing found expression in his determination to understand the role of the sense organs, as mediators of experience, in the synthesis of knowledge. Helmholtz owed the depth characteristic of his contributions largely to mathematical and experimental expertise. He was a great scholar in the tradition of Leibniz, embracing all the sciences, as well as philosophy and the fine arts.

4.2.2. Interface as an Electrical Circuit

An array of parallel RC circuits that for simplicity are usually lumped into one RC network is frequently used to model the electrode-electrolyte interface (Figure 4.1). Its main components are the *electrolytic medium resistance* R_m, to account for ionic conduction in the solution bulk, the *double layer capacitance* C_{dl} formed by ionic accumulation and/or by polarized particles which give rise to the half-cell potential (the other half is located at the companion electrode completing the circuit), the *charge transference resistance* R_{ct} to model the hindrance faced by the electrons when moving to and from the electrodes, and the *diffusion impedance* Z_w, also called the *Warburg impedance* after the investigator who proposed it back in 1899; the latter considers how difficult is for the charges to diffuse towards the interface. Often, the electric elements of similar kind are lumped into a single one (say, the Warburg resistance is combined with the transference resistance) but conceptually this is not quite correct because different phenomena are being mixed even though it is electrically

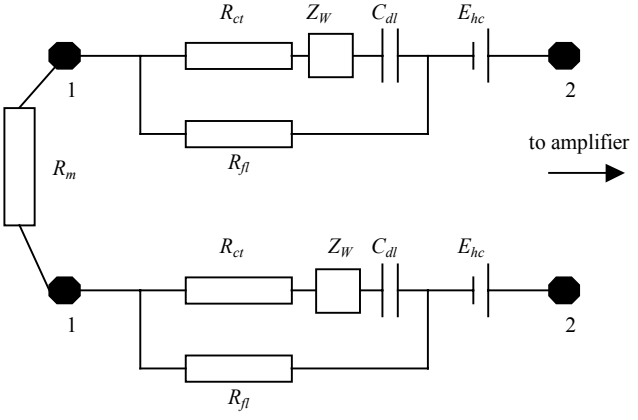


Figure 4.1. ELECTRODE-ELECTROLYTE INTERFACE EQUIVALENT CIRCUIT. R_{ct} or *charge transfer resistance*; C_{dl} or *double layer capacitance*. These components together are called *polarization elements*. E_{hc} or *half-cell potential* and R_{fl} or *resistive faradic leak*, the latter to provide for the dc and low frequency interface behavior. The return electrode (lower part of figure) is similar so that the two points called 1 are immersed in the electrolytic solution with a resistance R_m. Points 2 are connected to the input amplifier. In general, both electrodes will have different values of the equivalent circuit elements.

simpler. The Warburg impedance has a complex nature, for a resistance in series with a capacitance forms it. These components vary with the frequency of the signal applied to the interface. Besides, it must be underlined that the half-cell potential is characteristic of each metal-electrolyte combination and is not able to sustain any current. The total voltage between a pair of electrode terminals is the algebraic sum of the two half-cell potentials; usually, they are numerically different, and they may even display opposite signs depending on the metals the electrodes are made of. If the metal is the same for both electrodes and since the half-cell potentials appear in series, at least theoretically, they should tend to cancel out. Changes in acidity (a desired signal in pH-meters), bacterial growth (also a desired change in impedance microbiology), movements of the electrodes (undesired in ECG or in EMG records) modify the half-cell potentials, which are detected and amplified by the recording system. To measure a single half-cell potential is impossible; thus, an arbitrary standard electrode has been chosen (the *hydrogen electrode*) and electrode potentials are measured and tested against it in electrochemistry laboratories. There are other types of reference electrodes. The bibliography describing constructive and technical details is abundant (Geddes and Baker, 1989).

Emil Daniel Warburg (1846–1931) was one of the first to investigate the components of the electrode-electrolyte interface. He was a German physicist connected to several other physicists of great accomplishments and prestige, like Kohlrausch, Einstein and Planck (students: find out who they were). Warburg postulated that the capacitance of an electrode-electrolyte interface varies inversely with the square root of frequency, i.e., $C = Kf^{-\alpha}$, where K is a constant depending on the metal species, electrolyte concentration and temperature. Experimentally, it has been found that many times the exponent α ranges between 0.22 and 0.79 (Geddes, 1972).

Student task: Search for information about the hydrogen electrode. Describe it. Search for other types, such as the calomel and the silver-silver chloride electrodes.

The circuit shown in Figure 4.1 holds only for ideally smooth interfaces and does not take into account electrode rugosities. Thus, such model can be used in a few experimental situations, as the case may be with a mercury electrode. Some authors (Geddes & Baker, 1989; Webster, 1992) often make use of purely empirical models for the interface, calling it Z_i, the *interface impedance*, or Z_p, the *polarization impedance*, usually composed of a resistance and a reactance in series plus a battery, also in se-

ries, to account for the *half-cell potential*. The simple model proposed by Geddes and Baker (1968) is acceptable and enough when just interface stability or impedance are analyzed, say, during ECG or EEG recordings, but may prove insufficient in more stringent requirements. Somehow and hidden in it, this impedance Z_i includes implicitly the electrochemical parameters of Figure 4.1 and also the distortion caused by surface roughness (Felice, 1995).

A still unresolved problem of many interface models, such as those by Warburg (1899), Fricke (1932), Randles (1947), Sluyters–Rehbach & Sluyters, (1970) or Liu (1985), is the lack of ability to account for a direct current pathway. An exception is Geddes' above-mentioned model with a faradic resistance R_{fl} in parallel with Z_i. However, since the model is empirical, it does not supply an explanation of the phenomenon. It must be noted that the properties of Z_i **are not** constant. Such behavior happens because the polarization elements vary with time, frequency and current density. As an example, Figure 4.2 displays in log-log scale the interface reactance decline as the frequency of the applied signal increases.

The width of an interface is physically small. If the electrode is ideally smooth or perfectly polished, the interface size falls within molecular ranges, that is, in the order of 10^{-9} to 10^{-10} meters. Instead, when real electrodes are dealt with, the surface metal always shows a certain degree of roughness that can lie anywhere from 10^{-8} to 10^{-4} meters. These rugosity degrees can be obtained with different final polishing suspensions (see www.buehler.com). Rugosity produces a rather curious effect: the rougher the electrode surface, the smaller the action on the measurements

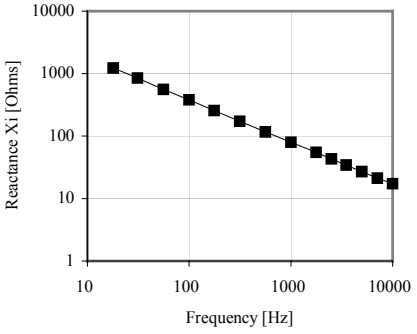


Figure 4.2. INTERFACE REACTANCE X_i VS FREQUENCY. Stainless steel parallel plane electrodes placed in a 1ml syringe filled with BHI (Brain Heart Infusion) at 25ºC. Experimental data obtained at the Department of Bioengineering, UNT.

made through the interface. To make electrodes rougher is common practice when bioelectric signals such as EEG or ECG are recorded. An important benefit is a significant decrease in the unavoidable impedance interposed between the source (heart, brain or other) and the electronic amplifier. As example, Figure 4.3 shows a platinum wire electrode roughened by using low frequency square-wave potential perturbations (Pajkossy, 1991).

This kind of electrodes has interface impedance 10 to 100 times lower than those that are untreated. Only recently the structures of this figure were approximately modelled by fractals (fractional elements). The concept, introduced by Mandelbrot (1975), has self-similarity as principal characteristic, that is, no matter how much an imaginary microscope power is increased to observe a given pattern, the pattern keeps being similar to itself. Fractal structures seem to be part of our inner biological nature, be it blood flow cerebral distribution (Nagao *et al*., 2001) or the helicoidal DNA. A true challenge was for physicists and electrochemists the development of the equations to predict, at least qualitatively, the behavior of a rough electrode-electrolyte interface (Liu, 1985; Nyikos & Pajkossy, 1985). The subject goes beyond the objectives of the book and suffice it just this short mention for the curious student.

Student Task: Fractal legumes. Buy a cauliflower and analyze its parts as you cut it into smaller and smaller pieces. Looking at it with a lens might help. Search in INTERNET for more information.

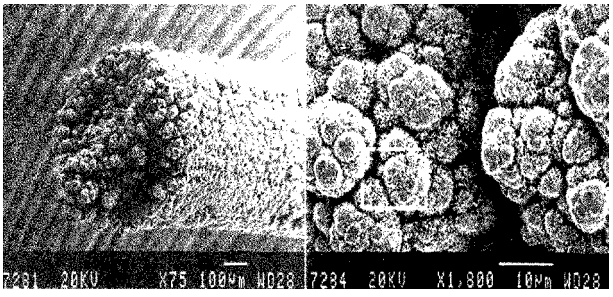


Figure 4.3. ROUGHENED PLATINUM WIRE TIP. Images were obtained using a scanning electron microscope. Left view: scale 100 µm. Right view scale 10 µm. Reproduced with permission from Electrochemistry at fractal surfaces, *J. Electroanal. Chem.*, 300 (1991), pp:1–11.

The interface does not embrace only the surface metal and its rugosities; as mentioned previously, it includes also the associated electric charges. Hence, an ideally smooth electrode (with an interface width anywhere between 10^{-9} and 10^{-10} meters), but submerged in diluted saline solution, increases its interface width up to 10^{-6} meters. In high concentration solutions, instead, the double-layer size is much smaller than the rugosities of the electrode metal. The above-mentioned interface width is defined as the perpendicular distance to the electrode measured from the deepest metallic valley to the last molecular level **associated with the interface**.

4.2.3. Useful and Annoying Interfaces

An electrode-electrolyte interface (EEI) may in some cases supply useful information while in others may act as a disturbance.

The first situation, from an electrochemical point of view and being somewhat repetitious, includes the measurement of pH, CO_2 and/or O_2 concentration, say, in blood or plasma. These physiological parameters are essential in clinical practice and during surgical interventions (Peura, 1992). Currently there are in the literature a great variety of gages, called biosensors, where an electrochemical transducer is combined with enzymes, nucleic acids, cellular receptors, antibodies or intact cells so producing myriads of gadgets to measure glucose, cholesterol, urea, lactate, creatinine or other substances (Zhang *et al.*, 2000). In impedance microbiology the interface is a natural ally too, providing quantitative and qualitative information about microorganisms growth (Felice & Valentinuzzi, 1999). In the dairy industry, evaluation of either the reactive or the resistive components or both of cow milk samples using a bipolar technique can be used to assess possible contamination (Felice, Madrid, Olivera *et al.*, 1999).

In the second situation, instead, the interface impedance is not wanted for it interferes with the desired signal, as exemplified in the recording of the ECG, EEG, EMG, or other similar signals; the biopotential appears in series with the EEI and, thus, the latter tends to modify the former both in amplitude and in shape.

The EEI must also be either avoided or decreased in biomass measurements. Electrodes are introduced in fermentators where high concentrations of microorganisms, fungii, yeasts or other types of living cells are present (Davey, Davey & Kell, 1993). The best way to avoid it is by

means of the tetrapolar impedance technique; in it, one electrode pair injects current while the other pair (which does not take current) detects a voltage difference (Morucci, Valentinuzzi, Rigaud *et al.*, 1996). However, even though this procedure is enough in most biomedical applications (Webster, 1992), it may not be fully satisfactory in some biomass applications and must be complemented with other techniques (Davey & Kell, 1998).

4.2.4. Interface Behavior with Low and High Current Density

When a small sinusoidal potential of a given frequency is applied to the interface, the circulating current is also of the same frequency with amplitude proportional to the applied potential. Thus, the current is predictable if the interface impedance is known; the system is said to have a linear behavior. However, if the potential increases, beyond certain value the resulting crossing current is not any more sinusoidal and harmonics generated at the interface show up. Thus, the behavior becomes non linear. Besides, the EEI decreases as the applied potential increases. The proportional relationship between potential and current is lost. Figure 4.4 describes such situation, where the impedance drop is dramatic beyond 0.7–0.8 mA. The origin of this behavior is still subject of research and it is believed that, at least partially, is due to the charge transference phenomenon (McAdams, Lackermeier, McLaughlin *et al.*, 1995). The fractal geometry may also play an added role (Ruiz and Felice, 2003). The nonlinearity may occasionally be useful, as the case is when monitoring materials corrosion, or may turn into a highly upsetting effect when microorganisms are measured in a fermentator (Yardley, Kell *et al*., 2000).

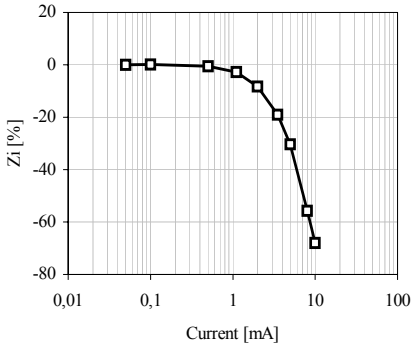


Figure 4.4. INTERFACE IMPEDANCE MODULUS VERSUS APPLIED CURRENT. Stainless steel electrodes (area = 0.3 cm^2) immersed in *Brain Heart Infusion* at 37°C. The impedance decrease is shown as percentage of the initial value at low current. Experimental data obtained at the Department of Bioengineering, UNT.

Leslie Geddes, Herman Schwan and their collaborators carefully studied the EEI as function of current density. Their many contributions, classics in the literature, are frequently used for bioinstrumentation design (Schwan, 1968; Ragheb and Geddes, 1991; Mayer, Geddes, Bourland *et al*., 1992; Geddes, 1997; Geddes and Roeder, 2001).

4.2.5. Bioelectric Signals Picked Up with Electrodes: Four Different Situations

For the sake of illustration, in this section we describe four interface bioengineering applications: pH measurement (when the generated dc potential difference at the interface is used to quantitate hydrogen ion concentration), interface impedance measurement during bacterial growth (to quantitate bacterial concentration), continuous membrane potential measurements (when the pH dc potential becomes a disturbance), and finally, measurement of classical variable bioevents (when the interface impedance appears also as a disturbing phenomenon). In brief: in two situations the interface is welcome and in the other two it is declared as nondesirable.

4.2.5.1. Measurement of pH (dc signal **at** the interface)

Let us first review as introduction the concept of pH. The following paragraph can be skipped if the student is already familiar with these basic concepts. Søren Peter Lauritz Sørensen (1868–1939), Danish biochemist, introduced it as a convenient way to quantitate the acidity of a solution. A numerical scale can be established by taking the negative logarithm of hydrogen ion concentration. The letters pH stand for "pondus hydrogenii" (literally hydrogen weight, in Latin), as acidity is caused by a predominance of hydrogen ions [H+]. Originally, Sørensen wrote both letters with capitals, but W. M. Clark (inventor of the oxygen electrode that carries his name), in 1920, suggested the current notation "pH" for typographical convenience; as an extension, "p-functions" have also been adopted for other concentrations and concentration-related numbers. For example, a pH of 4.5 refers to a molar H ion concentration of 3.2x104 and a pCa = 5.0 means a concentration of calcium ions of 10–5 M. Thus, in its original description, the acid potential of aqueous solutions is expressed in terms of the pH scale, where the symbol "p" means "take the negative logarithm of whatever follows in the formula", say, for pH, pOH, or p[anything], i.e., **pH = –log [H+] = log {1/[H+]}**. Note that the hydrogen ion concentration must be ascertained before the pH can be calculated. Tremendous swings in *hydrogen ion or hydronium ion concentration* occur in water when acids or bases are mixed in it (see below for the conceptual chemical difference between the two terms). These changes can be as big as 1×10^{14} meaning that concentrations can change by multiples as big as one hundred trillion. The pH

scale is a logarithmic scale. Every multiple of ten in H+ concentration equals one unit on the logarithm scale. Physically, the pH is intended to tell what the acid "potential or weight" is for a solution. In a sense, the system is inverted; so, a low pH value indicates a great acid potential while a high pH indicates a low acid potential (this is upside down and counter-intuitive.)

A more modern definition makes use of the molar concentration of *hydronium ions* **[H_3O^+]** in solution, that is, **pH = – log([H_3O^+])**. Pure water autoionizes to produce equal concentrations of hydronium and hydroxide ions **[OH^-]**. This is described by, **2 H_2O = H_3O^+ + OH^-** whose equilibrium obeys the law of mass action in the form Kw = [$H3O^+$] [OH^-] = 1.0×10^{-14} (at 25ºC). This modern form of the equation for water autoionization recognizes that protons do not exist in solution but instead are bound to an electron lone pair in water: $H^+ + H_2O = H_3O^+$. The obsolete term hydrogen ion and its concentration [H^+] have been replaced by "hydronium ion" and [H_3O^+] but we continue to use pH (and not pH_3O). Since the hydronium ion concentration and the hydroxide ion concentration are equal in pure water, it follows that [H_3O^+] = [OH^-] = 1.0×10^{-7}. Then, the pH of pure water is pH = – log(1.0×10^{-7}) = 7.00. In dilute acid solution the hydronium ion concentration is higher; e.g., in micromolar hydrochloric acid, 10^{-6} M HCl, the hydronium ion concentration is [H_3O^+] = 10^{-6} mol/L so that the pH = 6.00. That is, one step lower (higher) on the pH scale represents 10 times higher (or lower) hydronium ion concentration. In a similar way to pH, the concentration of hydroxide ion is also expressed on a logarithmic negative power scale: pOH = – log([OH^-]). Further, since water autoionization equilibrium relates [H_3O^+] to [OH^-], the pH and pOH are related: [H_3O^+] [OH^-] = 1.0×10^{-14}, so that –log([H_3O^+]) – log([OH^-]) = –14.00 or pH + pOH = 14.00. The hydrogen ion concentration in pure water around room temperature is about 1.0×10^{-7} M. A pH of 7 is considered "neutral", because the concentration of hydrogen ions is exactly equal to the concentration of hydroxide (OH^-) ions produced by dissociation of the water. Increasing the concentration of hydrogen ions above 1.0×10^{-7} M produces a solution with a pH of less than 7, and the solution is considered "acidic". Decreasing the concentration below 1.0×10^{-7} M produces a solution with a pH above 7, and the solution is considered "alkaline" or "basic".

In 1906, Max Cremer found that a difference in hydrogen ionic concentration caused a difference of potential across a glass membrane. The circuit contains two Ag/AgCl reference electrodes and a glass membrane sensitive to hydrogen ions (Figure 4.5). There are two interfaces at stake: one between the glass and the electrolyte within the bulb (left electrode in Figure 4.5) sustaining a constant potential difference and another between the glass and the sample with a potential related to its acidity. The interface potentials of the two reference electrodes are constant because of the high concentration of HCl and KCl. Thus, any change in the sample solution has negligible effect on them. Moreover, the electrode metal within that electrolyte must have a very low drift, so giving good stabil-

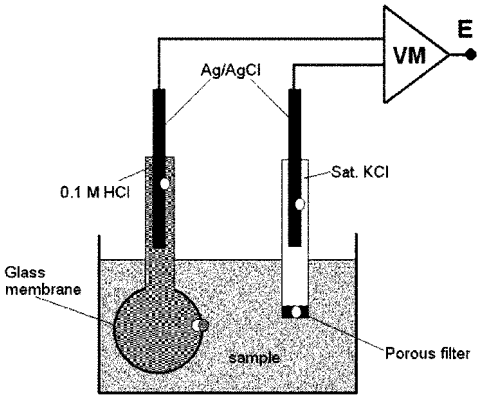


Figure 4.5. BASIC SYSTEM TO MEASURE pH OF A SOLUTION. White points at the EEI's mark dc interface potentials; their sum E is given at the voltmeter VM output. The double circle area at the glass membrane is the only interfacial potential that changes with hydrogen ion concentration. The rest remains constant.

ity to the system. The electrode in HCl is separated from the sample by a glass membrane sensitive to H^+. The other electrode, really the reference in this circuit, is separated from the sample by an orifice covered with a porous filter that lets the electric current flow but it prevents mixing of the electrolytes.

The liquid-liquid interface at the filter sustains a minute potential due to two reasons, on one hand, the sample ionic concentration is much lower than K^+ or Cl^- concentration within the electrode; thus, it does not generate an apreciable difusión potential as ions traverse the porous filter towards the KCl solution. On the other hand, K^+ and Cl^- difusión to the sample through the filter does not generate significant electrical potential because they involve charges of equal magnitude and mobility.

In summary, the voltmeter records the algebraic sum of all interface potentials but the only one that varies with the hydrogen ionic concentration originates at the glass-sample interface.

4.2.5.2. Bacterial concentration (ac signals **at** the interface)

In this application the interface impedance is continuously measured with electrodes immersed in innoculated broth cultures. The growth of microorganisms leads to proton and/or ionic production causing an increment of the interface capacity C_i. Such parameter better follows bacterial development than monitoring the broth ohmic resistance (Noble, Dziuba, Harrison *et al.*,1999; Felice, Madrid, Olivera *et al.*, 1999; Felice & Valentinuzzi, 1999). Figure 4.6 shows this capacitance for different microorganism concentrations in raw cow milk, where contamination is

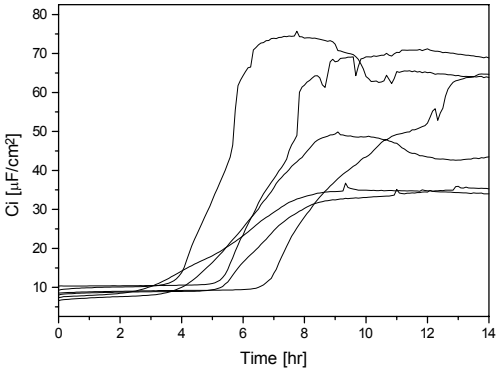


Figure 4.6. INTERFACE CAPACITANCE C_i GROWTH CURVES. Two stainless steel electrodes immersed in culture broth. Assorted set from different cells. From Felice, Madrid, Olivera *et al.* (1999), experimental data obtained at the Department of Bioengineering, UNT.

well known and composed of several bateria. The upward inflexión point is used to estimate the initial bacterial concentration; a documented law states that the lower the initial concentration, the higher the initial contamination.

4.2.5.3. Membrane potential (dc signal **through** the interface)

The resting membrane potential of an excitable tissue is a dc signal measured by a very small electrode (microelectrode) piercing the membrane to get into electrical contact with the intracellular fluid against a big return electrode placed in the extracellular fluid. The inside is about 80 to 90 mV negative to the outside. Since the interface half-cell potentials interfere and tend to heavily distort the desired electrophysiological signal, they must be balanced out with a previous short-circuit procedure, otherwise, there is no way to separate them out (Geddes & Baker, 1989). The bibliography is abundant and the student is encouraged to search for more information regarding the subject.

4.2.5.4. ECG, EMG and EEG (ac signals **through** the interface)

This application has been already been dealt with in a preceding chapter. We only emphasize here that these physiologycal signals are of the ac type that traverse the electrode-electrolyte interfaces; thus, changes in electrode impedance introduce distortion, especially when the latter is

too high. Typical simple solution is the use of conductive gels (Geddes & Baker,1989).

4.3. Sensors/Transducers

It was said above that the sensor senses the signal and the transducer translates it into another type of energy. By and large, the sensor is also part of the transducer and mostly so much embedded in it that physical separation becomes extremely difficult if not impossible. Geddes and Baker (1989), already in the first edition of their classical book (1968), make a clear and quite sensible distinction that helps out in the understanding of this important subject: the concepts of *transducible property* and *principle of transduction*.

A transducible property is represented by a singular characteristic of a substance or event to which a principle of transduction can be applied. A principle of transduction is any one of the many methods employed to convert the transducible property to an electrical signal, the latter thereon treatable by electronic means.

For example, pressure is well defined as force per unit surface area (say, measured in dynes/cm^2). It appears, thus, as a mechanical transducible property of the cardiovascular system. If pressure is exerted to a diaphragm, the latter will be slightly displaced and may be made to change the value of a parallel-plates capacitor so becoming a principle of transduction. Such capacitor is an easy element in an adequate electronic circuitry to produce a sensitive electric signal proportional to pressure changes. However, other principles of transduction could be applied to the same transducible property as well, the only limiting factor being the ingenuity of the designer. It becomes clear that the *singularity* of the transducible property and the *selectivity* of the principle of transduction are intimately related, so reflecting in the quality of the transformation.

If the transducible property is not clearly and quantitatively defined, no transducible property can be applied. That is the case with diffuse, not well-established psychophysiological variables such as anxiety, pain, or fear. Some areas (cardiovascular physiology, respiratory physiology, muscular biomechanics) lend themselves better to quantitative analysis

because their variables are mathematically defined based frequently on physical concepts.

However, not always a principle of transduction is at hand. In that case, transduction is not possible and technology should find new ways. History of science exemplifies abundantly such search (Geddes, 1970).

Student task: Define quantitatively the different variables of the cardiovascular and respiratory system, the muscular system and the endocrine system. Refer to Chapter 2 if necessary. Expand the task to other physiological systems. Try to rank them according to their possible level of quantification. Recall the example given above of psychophysiology, a discipline where variables are not well set yet.

From now on, we will speak in terms of transducers, that is, a device translating the energy of a signal into another type of energy. Let us illustrate with a few examples:

- Microphones: air pressure changes due to sound become an electrical signal, thus, they are mechano-electric transducers.
- Loudspeakers: electric variations are transformed into sound via air pressure changes. They are electro-mechanic transducers and can be considered as the opposite of the formers.
- Photoelectric devices: a luminic signal is transformed into an electric signal, as clearly the very name implies.

Student task: Think of other transducers. Even though the most frequent output signal is electric, it may not necessarily be so. Remember, as one possible example, the old Marey's tambour system to detect blood pressure. The recording apparatus was the smoked kymograph.

Ideally, the information carried by the signal must be preserved. In other words, the event has to be faithfully reproduced. Such requirement leads to three basic and essential characteristics:

Amplitude linearity. *Any change in the amplitude of the input signal must be reflected in a proportional change of the output signal*.

Good frequency response. *The transducer should be able to follow slow and fast changes*, *sometimes even from zero frequency*. *How fast the change can be is determined by the highest frequency the device is able to pass*.

No phase distortion. *This characteristic refers to the relative delays the different frequency components of the signal have at the input*. *Those*

delays should be preserved at the output; *otherwise there would be phase distortion*.

A fourth important characteristic from the point of view of quantitative measurements is *calibrability*, i.e., and obvious as it may sound, the electric output signal should be expressed in terms of the original signal units. Continuing with the blood pressure example, the electric signal output in volts or millivolts should be mathematically related to the biohydraulic input signal measured, say, in mmHg.

The book by Geddes and Baker (1989) presents a thorough description of the different and most important principles of transduction, namely, resistive, inductive, capacitive, photoelectric, piezoelectric, thermoelectric and chemical. Each shows a large variety of transducer types often associated with specific electronic circuits. The student is invited to visit the WEB in order to better update his/her knowledge.

Student chore: Find a resistive, capacitive, inductive, photoelectric and thermoelectric transducer by searching either in the literature or in the WEB. Comment on each.

Another student chore: Two transducers that deserve some deeper thought are the microtip blood pressure transducer and the microtip velocity probe, the latter to assess blood flow. Search the literature and the WEB.

4.3.1. Characteristics of the Wheatstone Bridge

Many resistive transducers apply this classical configuration because, when out of balance, it produces a good signal.

Sir Charles Wheatstone is most famous for it but never claimed the invention, however, he did more than anyone else to invent uses for it when he 'found' the description of the device in 1843. The first description of the bridge was by Samuel Hunter Christie (1784–1865), in 1833.

Typical examples in physiology and biomedical engineering connected according to this arrangement are strain gauges of different types, as those used in the old Staham units or the microtip blood pressure devices. Good understanding of the bridge characteristics is essential for optimum design and construction of transducer.

In this circuit (Figure 4.7) there are three loops, from which three loop equations are easily written down, namely,

$$E = (R_1 + R_3)i_1 - R_1 i_2 - R_3 i_3 \tag{4.1}$$

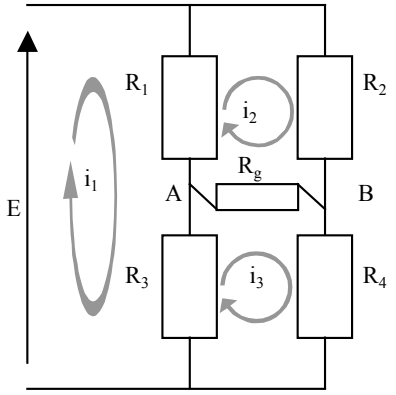


Figure 4.7. WHEATSTONE BRIDGE CONFIGURATION. *E* is an external applied voltage, either dc or ac. When in balance, the difference of potential between points *A* and *B* is zero and no current flows through the transverse branch R_g.

$$0 = -R_1 i_1 + \left(R_1 + R_2 + R_g\right) i_2 - R_g i_3 \tag{4.2}$$

$$0 = -R_3 i_1 - R_g i_2 + \left(R_3 + R_4 + R_g\right) i_3 \tag{4.3}$$

Solving the latter for the currents, say, by Cramer's rule, and considering that in balance the current $i_g = i_2 - i_3$ through the transverse resistance R_g must be zero, leads to the balance condition $R_1R_4 = R_2R_3$, or to the well-known statement that the diagonal resistor products are equal. The student is invited to actually work out the details of this derivation.

Student task. Obtain the balance condition by removing the transverse resistor R_g and considering the circuit a voltage divider E_A/E and E_B/E. The difference of potential between *A* and *B* should be zero when the circuit is balanced out.

The sensitivity of the bridge is an important characteristic to take into account and will orient the criterion to choose the four resistors, for actually there are infinite combinations to meet equality of the diagonal products, as found above. Let us go back to Figure 4.7 removing the resistor R_g and introducing a slight imbalance, such that resistor R_4 is increased by a fraction pR_4. In that case, there will be a minor difference of potential $\Delta E = E_A - E_B$. Both potentials can be calculated by the voltage divider rule and the difference ΔE will be expressed in terms of E and the entire resistor network including the fraction p. If the ratios $R_3/R_1 = R_4/R_2$ obtained in balance are equated to r, an equation relating ΔE to E will be obtained, that is,

$$\Delta E / E = pr / (1 + r)^2 \tag{4.4}$$

meaning that to reach a hight sensitivity, for a given E, the deviation delta from zero must be large when a small variation p is produced. In other words,

$$\Delta E/p = E\left[r/(1+r)^2\right] \tag{4.5}$$

the latter is expressed in volts per fraction of resistance change. Since we are interested in maximum sensitivity, the equation above is solved for $\Delta E/E$, derived with respect to p and equated to zero, leading to $r = 1$, i.e., maximum sensitivity is obtained when all four resistors are equal (or nearly equal). The student should work out the mathematical details of this important derivation.

4.4. Biosensors

For this subject, a visit to www.ornl.gov/info/ornlreview/rev29_3/ (Biosensors and Other Medical and Environmental Probes, by K. Bruce Jacobson) is recommended. Biological sensors, essential for the constant communication between live entities in general and the surrounding environment, have been real models to imitate in order to accomplish functions that man wants to control. Sensors created from tissues, cells, proteins and enzimes and integrated into technological frames are collectively called *biosensors*; the prefix indicates that some kind of biological component is involved in it (Tran–Minh, 1993). The final output signal is by and large electrical (Pallàs–Areny & Webster, 1991). Biosensors not only help in the understanding of biological proceses but also have many biomedical and industrial applications. Acidity, oxygen, carbon dioxide, glucose, urea sensors are commonly found in hospital practice. Other metabolites as needed targets are included almost daily (Zhang, Wright &Yang, 2000).

The first contributions can be traced back to the late 1960's, when Updike and Hicks (1967) introduced the first enzymatic electrode for glucose detection. Preservation of the biological sample integrity along with maintenance of enzyme activity are major challenges that not always find an easy solution; suffice it to say that in many cases the enzyme can be used only once. Obviously, the latter difficulty increases

costs. Occasionally, more than one enzyme is required, and the problem is multiplied. Currently, the biological material is placed on a semiconductor chip.

Thus, biosensors are analytical tools combining a biochemical recognition component with a physical transducer. The biological sensing element can be an enzyme, antibody, DNA sequence, or even microorganism. The biochemical component serves to selectively catalyze a reaction or facilitate a binding event. The selectivity of the biochemical recognition event allows for the operation of biosensors in a complex sample matrix, i.e., a body fluid. The transducer converts the biochemical event into a measurable signal, thus providing the means for detecting it. Measurable events range from spectral changes, which are due to production or consumption of an enzymatic reaction's product/substrate, to mass change upon biochemical complexation.

Enzymes are nature's catalysts. Like all catalysts, they increase the rate at which a reaction reaches equilibrium by providing a low-activation energy reaction pathway. They usually operate in approximately neutral pH at mild temperatures, generate no by-products, and are highly selective. Enzyme-catalyzed reactions can be selective for one substrate or a group of substrates. Several thousand enzymes have been isolated, and several hundred are available commercially. They are classified by the reactions they catalyze. With amperometric enzyme electrodes, for example, oxidoreductase enzymes are most frequently used. Oxidoreductases catalyze the oxidation (removal of electrons) or reduction (addition of electrons) of the enzyme substrate. Since oxidoreductases are most closely associated with electrochemical processes, their turnover is easiest to observe by electrochemical detection. Enzymes, like all proteins, are made of amino acid chains folded into specific three-dimensional structures. They range in size from 10,000 to several million daltons. Besides aminoacids, many enzymes also contain prosthetic groups, as nicotinamide adenine dinucleotide (NADH), flavin (FAD), heme, Mg^{+2}, and Ca^{+2} that enhance enzyme activity. With oxidoreductases, the prosthetic groups serve as temporary traps of electrons or electron vacancies.

The prosthetic groups can be near the surface or deep within the enzyme's protein structure. In the latter case, the trapped charge is not easily transferred to an electrode. According to Marcus theory, which is

used to explain electrochemical reaction rate, electron transfer decays exponentially with distance. Therefore, electron transfer from an active site near the center of a 100 kD protein to an electrode is highly improbable. From a biological perspective, concealment of the active site is often necessary for selective targeting or redox reactions toward a specific synthetic or degradative route. For electrochemical biosensor applications, however, the difficulty of electron transfer into or out of the active site poses a real problem. Mediators capable of accessing the active site are frequently used to assist in the transduction of the enzyme activity into a measurable amperometric response.

The biomedical sector makes up of an important portion of the biosensors market. NASA, for example, studies all types to be connected to future astronauts (Soller *et al*., 2002). Glucose, lactate, urea and creatinine sensors are essential for the physiological follow up of space crews. Other sensors make use of antibodies to determine neurotransmitters or hormones (Hines, 1996).

Biosensors can also make use of a neural interface technology to detect nerve and muscle activity. Currently, there are some to measure muscle electrical activity, brain electrical activity, and eye movement. Electrodes sit on the skin over the muscle or nerve being sampled. Eye movement, for example, is determined from biosensors placed strategically on the forehead and under the eyes (for all this, check the WEB, Biosensors, by Cindy Tonnesen and Gary Withrow, 2003).

Biosensors have the potential to affect field application areas including medicine, physical therapy, music, and the video game industry. Although biosensors are not limited to any group of people, they are particulary useful for the handicapped. Even completely paralyzed individuals have electrical activity in their bodies that can be detected. One biosensor application developed for the handicapped is an electronic instrument that produces music from bioelectric signals. Signal inputs such as eye movements, muscle tensions, and muscle relaxations are converted to MIDI (Musical Instrument Data Interface) and output to a synthesizer. Before being mapped to MIDI, the signals are analyzed for specific intensity and spectral characteristics for the particular individual. For dysfunctional or weak muscles the signals can be amplified according the level of tension and relaxation. These signal inputs are then interpreted to

control volume, pitch, tempo, and other aspects of musical composition (Knapp & Lusted, 1990).

Medical applications are presently seen in the diagnosis and correction of eye disorders. Strabismus is a condition in which an individual's eyes are not aligned properly, and thus do not move in conjunction with one another. This can be corrected by surgery but the current use of prisms to determine the degree of correction necessary is not very accurate. Biosensors tracking the eye movements can determine with high accuracy the number of degrees in both the X and Y planes that the eyes need to be adjusted.

Just as biosensors can be used to determine amounts of eye correction, they can also be used to train the eye, as they can be an input device to video game exercises to realign eye tracking. This same method of muscle training through a video game could be used for rehabilitation of potentially any muscle group, as biosensors can be individually customized to detect levels of muscle activity for most muscle groups. In the same way that patients undergoing rehabilitation could use biosensors as an input device for their video exercises, the video game industry could use biosensors as yet another powerful input device for entertainment. Also contributing to physical therapy, biosensors can help to create custom exercise programs for injured patients and athletes, can be used by athletes to check muscle condition, and can be connected to a multitude of external monitoring devices.

Biosensors potentially have a number of uses in the emerging field of Virtual Reality, particularly when the environment is interactive, that is, the entire body is immersed in it. This environment could react to hand or arm gestures, eye movements, or any muscle or nerve as input. These forms of input are attractive as they are somewhat more natural and intuitive to the user; they are being researched with input devices such as gloves and body suits. Biosensors strategically placed on the body could provide an alternative way to provide this interface. They also could be utilized by handicapped people who may not be able to use a glove or body suit. Possible use of prosthetic limbs where just the bioelectric activity to the nerve endings of a missing limb could be used to control an artificial limb. In cases of paralysis, the nerves, prior to loss of transport ability, or brainwaves might be electrically monitored for instructions to

control/move a mechanical device attached to the paralyzed limb. When brainwaves can be reliably monitored, we can study relationships between EEG (brain activity) and specific cognitive activities such as sleep behaviors and sleep states. Simple brain wave detection has been successful in early research stages, but breaking through the use of subvocal commands would be perhaps the most powerful input controller we have yet seen. The interested reader may find more detailed information in Guilbault & Lubrano (1972), Heller (1988), Rechnits (1988), Ho & Rechnitz (1992), Scheller, Wollenberger, Pfeiffer *et al*. (1996).

4.5. Comments and Conclusions

In terms of chapters, this is roughly one half of the book, hence, it is a good place to glance back at the trodden path to, thereafter, quickly take a look at the half of the way ahead of us. The INTRODUCTION (Chapter 1) set the objectives and offers some definitions to start with. Chapter 2, the longest so far, presented the SOURCE of the information we are dealing with; this is why physiology dominated the scene in it. Sources give off information in the form of SIGNALS (Chapter 3) and they are essential for the physiologist and the biomedical engineer. This Chapter 4 gave a view of how those SIGNALS CAN BE PICKED UP, a basic task of the bioengineer because he/she constantly is in search of new ways, techniques and even tricks to fish out of the darkness that piece of needed information. The chapter's motto should be by now well understood.

Electrodes were introduced as electric-electric transducers, in itself a nice clear concept, where the two faces of the interface play a significant complex role, with ions on one side and electrons on the other, but both kinds shaping up the electric current through the circuit. Charge migration or drift, diffusion and convection appear as electrochemical processes behind the charge transference phenomenon, which, in the end, amounts to the long studied oxydation-reduction process.

Electric equivalent circuits have been proposed to more or less simplistically analyze the interface. We emphasize that all these models, without exception, are both good and bad because they are just approximate mod-

models that may serve our purposes better or worse. In a sense, this can be viewed as if looking through a spyglass.

We learned also that interfaces might be helpful for some applications and quite annoying for others. Obviously, it is also a question of where we stand and what we want from them. Current through them can certainly distort their behavior a lot and, as a general rule, we had better keep it as small as possible

Rather quickly, we scanned some signals picked up by electrodes. In the previous chapters something was anticipated, but here the concepts of signals generated **at** the interface and that of signals **crossing** it was underlined. Measurements of acidity and of bacterial concentration are excellent examples of the first type; for the former, the concept of pH was also refreshed for those who perhaps forgot it.

Sensors and transducers are the subject of section 4.3 underlining that, since the sensor represents the sensitive portion of a transducer, both tend to merge, and many times it may not be easy or it may even be impossible to separate them out. Basic concepts and characteristics are presented while a few tasks are suggested to the student. No more details are needed for the time being because specific courses on the subject are waiting further down in the student's road. The Wheatstone bridge deserves a little bit of attention; it is a widely used configuration with characteristics that indicate design criteria. Finally we get to the biosensors. A convenient overview tries to show how much future and possibilities there are making once more biomedical engineering a fascinating discipline. References are ample to satisfy curiosity and to go deeper in the subjects.

Rossana E. Madrid was born in 1964. She received the Electronics Engineering degree (1991) and the Ph.D. in Bioengineering (1998) from the University of Tucumán (UNT), the latter degree under the supervision of Dr. Max Valentinuzzi, with a dissertation about the design and development of a system to measure impedance and turbidity in biological suspensions. As a postdoctoral fellow, she spent a short time in the *Institut für Technishe Chemie* of the University of Hannover, Germany, with the group of biosensors led by Dr. Thomas Scheper. Currently, she is an instructor at the Bioengineering Department of UNT. Her main research interests include the development of equipment for food industry and bioprocesses as well as the study of non-linear impedance measurements in biological suspensions. She has published several papers in these subjects.

Carmelo J. Felice got the Electronics Engineering undergraduate degree and the doctorate in Bioengineering from the Universidad Nacional de Tucumán, Tucumán, Argentina. Currently, he is Professor and Head of the Department of Bioengineering of the same university and Career Investigator of the National Research Council of Argentina (CONICET). His main interest deals with impedance microbiology and spectral analysis, subjects in which lie most of his scientific production. Besides, he holds three patents in the same field. Carmelo is full member of the Sociedad Argentina de Bioingeniería (SABI).

Chapter 5

Biological Amplifier

Like a spying glass, it magnifies tiny electrical bioevents.

by Myriam C. Herrera and Max E. Valentinuzzi

5.1. Introduction

Biological electrical signals show characteristics requiring a special design of the electronic device that receives them. In general, the biosignal is a voltage generated by excitable tissues (like nerves or muscles) or by other live structures not necessarily within the latter category (as the retina, bone, ovocytes, or some renal epithelial cells). The preceding chapters describe them with considerable details.

Typically, the voltage range spreads from the tenths of microvolts to may be a few hundreds of millivolts. For example, the electroencephalogram can be in the order of 20 μV, the electrocardiogram lies between 1 and 3 mV, and the resting muscle membrane potential is about 100 mV. To make these amplitudes compatible with the different recording and/or monitoring systems, amplification is needed, as the first and simplest step of signal processing. Moreover, such biosignals are always more or less immersed in or mixed with interference and noise of various characteristics and sources. Thus, amplification must be selective and/or must somehow reject or decrease the level of undesired disturbance.

This amplifier, via electrodes, is in direct contact with the biological system (see Chapter 4). It may produce damage due to unwanted circulation of current or to the application of relatively high potentials. As a consequence, protection against any kind of mishap is another characteristic to take care of.

The electronic vacuum tubes —diode and triode— were introduced by John Ambrose Fleming, the former, and by Lee De Forest, the latter, early at the beginning of the XXth century, in 1904–6. However, it took considerable time to develop and recognize the potentialities of the electronic amplifier with a triode (the audion, as it was first called) and, later on, with pentodes in radio communications. In physiology, Joseph Erlanger and Herbert Gasser were the first to specifically make use of an electronic amplifier and an oscilloscope, designed and built by themselves, in 1928. With them, they were able to record nerve compound action potentials and measure the velocity of propagation of this basic bioevent. They were awarded the Nobel Prize in 1944 (Physiology or Medicine) for their fundamental contributions and, besides having been electrophysiologists, they should be honored also as biomedical engineers. The biological amplifier is a differential amplifier that can be traced back to Matthew, in 1934 in the UK, and to Otto Schmitt, in 1938, in the USA, specifically designed to reject ground-referred interference. It originated in the life sciences gaining status later on in industry. Its predecessor, the push-pull amplifier, came from radio engineering as a means for increasing the output power (Geddes, 1996). The transistor made its appearance much later, in 1947, as the invention of William Shockley, John Bardeen and Walter Brattain, also Nobel Prize winners (Physics, 1956). It took considerable time (c.a. 1960, or even later) for it to enter into physiology and medicine, as for example in electrocardiography. The integrated circuit (IC), technological base for the nowadays operational amplifier boom and indisputable *prima donna* in everything touched by electronics, entered into scene in the early 1960's. The first microprocessors, glaring outgrowth of the former obviously centered on integrated electronics, made their appearance around 1975. No doubt, the XXth century has rocketed off technology, although perhaps the human values have stayed well behind, unfortunately (Valentinuzzi, 2002).

5.2. Basic Requirements

They can be ideally listed as follows,

- Only the desired signal is to be amplified;
- The signal should not be distorted;
- Any disturbance should be rejected;
- The device should not affect in any way the biological system;
- It should be protected against external discharges (defibrillators, electrosurgery);

and these requirements are to be taken as the **general design criteria** of the device.

5.3. Instrumentation Amplifiers (IA)

When an electronics engineer has to amplify a given signal, probably his/her first idea is to use a low cost IC operational amplifier (called op-amp in the daily electronic jargon). It is linear because the output appears as an enlarged replicated version of the input, with a constant ratio — the gain — between output and input over all the frequency band of interest. Due to practical limitations of these IC's — high but not infinite gain, flat frequency response within specific ranges — (Dostál, 1993), the simple and convenient op-amp needs a few external components (resistors and capacitors) to better define its behavior. In this way, it is possible to implement functions that the op-amp could not perform by itself. Figure 5.1 displays what is probably the simplest arrangement, where Z_1 and Z_2 are two known impedances.

If the *Z*'s are replaced by ohmic resistors R_1 and R_2, respectively, applying the two basic rules of the ideal operational amplifiers (i.e., the two input terminals are at the same potential while no current flows into those terminals), Ohm's law leads to the output/input relationship, or gain,

$$V_o / E_i = -(R_2 / R_1) \tag{5.1a}$$

amply referred to and used in the current literature, technical notes and manuals.

This configuration is handy because simultaneously it allows amplification and sign inversion; thus, it is an *inverter*. If now the input potential

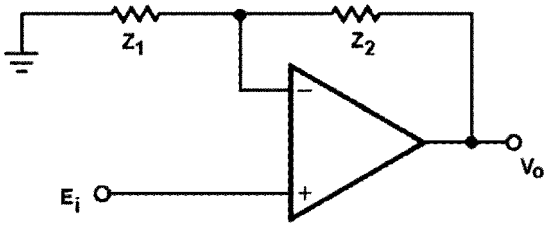


Figure 5.1. BASIC OPERATIONAL AMPLIFIER: SINGLE INVERTER. It is easily seen that V_o/R_2 (or current through R_2) is equal to $-E_i/R_1$ (or current through R_1) since no current goes into the negative terminal of the amplifier. Besides, the two input terminals (negative and positive) are essentially at zero potential. The two impedances are being taken as ohmic resistors. If, instead, impedance Z_1 (or resistor R_1) is grounded and the positive terminal connected to an input voltaje E_i, there is no inversión and the gain is given by $1 + R_2/R_1$. The student should work out the demonstration.

E_i is connected to the positive terminal and R_1 is grounded, in other words, if the two input terminals are swapped, the resulting equation becomes,

$$V_o/E_i = \left(1 + R_2/R_1\right) \tag{5.1b}$$

clearly showing that the circuit amplifies without inversion and, for that reason, is usually called *non-inverting amplifier*. The student is advised to work out the derivation of equation (5.1b) by considering the basic operational amplifier rules. By modifying the external elements in these arrangements (as for example with complex impedances, usually RC-networks), the range of constant gain can be delimited as a function of the frequency in a selective way.

Are these simple circuits good enough for the detection of biopotentials, fulfilling the basic requirements listed above? The answer is "no". Its main limitation resides in the common terminal employed for both, the input and the output signals. Besides, such common terminal is also shared with the power supply of the op-amp. In practice, very often neither of the input terminals coincide with the common point and, thus, they cannot be connected to the latter. Most resistance sensor bridges and some kind of transducer configurations illustrate such situation. As a consequence, **an amplifier with two floating terminals is required**, **amplifying the voltage difference between the two**. This is precisely a differential amplifier. An instrumentation amplifier, or amplifier for biopotentials, is a precision differential voltage gain device optimized to operate in a noisy environment.

5.3.1. Differential Amplifiers Based on a Single Op-Amp

Figure 5.2 shows a typical and simple configuration to implement a balanced differential amplifier based on a single op amp. Assuming the operational amplifier is ideal ($V_1 = V_2$), the output voltage V_o can be calculated by the superposition theorem (Skilling, 1965). The output for E_1, when E_2 is grounded, is given by,

$$V_{o1} = E_1\left[R_2/\left(R_1 + R_2\right)\right] \times \left[\left(R_3 + R_4\right)/R_3\right] \tag{5.2}$$

where two factors are included: $\left[\left(R_3 + R_4\right)/R_3\right]$, the gain of the *non-inverting amplifier*, and the voltage dividing effect of the $\left(R_1, R_2\right)$ network.

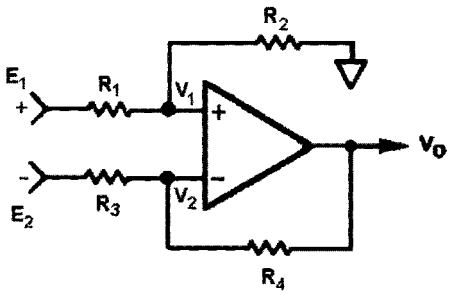


Figure 5.2. BALANCED DIFFERENTIAL AMPLIFIER. When E_2 is grounded, $V_o = V_{o1}$, and the voltage divider rule leads to $E_1/V_1 = (R_1+R_2)/R_2$ and $V_{o1}/V_2 = (R_3 + R_4)/R_3$; from the two latter, equation (5.2) is obtained after solving them for V_1 and V_2, respectively, and equating them. When the other input terminal is grounded, instead, the arrangement coincides with that of Figure 5.1, and $V_o = V_{o2}$ is given by equation (5.1a), which now should be written as equation (5.3) is. See text for more details and final result. The student should work out all the algebraic manipulations.

Conversely, when E_1 is grounded, we are left with the first case and its output V_{o2} is described by eq. (5.1a), or

$$V_{o2} = -E_2\left(R_4 / R_3\right) \tag{5.3}$$

Superposition of the two latter quickly leads to $V_o = V_{o1} + V_{o2}$, or,

$$V_o = E_1\left[R_2/\left(R_1 + R_2\right)\right]\times\left[\left(R_3 + R_4\right)/R_3\right] - E_2 \times \left(R_4/R_3\right) \tag{5.4}$$

If $R_2 = R_4, R_1 = R_3$, and the input difference is defined as $E_d = E_1 - E_2$, the output voltage can be written as

$$V_o = \left(E_1 - E_2\right)\times\left(R_4/R_3\right) = E_d\left(R_4/R_3\right) \tag{5.5}$$

after proper replacements in eq. (5.4), clearly indicating a proportional relationship between output and input, where the ratio R_4 / R_3 obviously stands for the circuit differential gain.

Let us now obtain a general expression of the differential gain, without imposing resistor symmetry. We merely have to make $E_2 = -E_1 = -E_i/2$ in equation (5.4) because, precisely, a differential input is characterized by one half of the signal as positive on one of the terminals and the other half as negative on the other terminal, and both

portions centered against the reference. Moreover, no common input is assumed. Under this condition, equation (5.4) becomes,

$$V_o = (E_i/2)\times\{[R_2/(R_1+R_2)]\times[(R_3+R_4)/R_3]+(R_4/R_3)\} \quad (5.6)$$

from which it is easily seen that the differential gain V_o/E_i is quantitatively described by,

$$G_d = (1/2)\times\{[R_2/(R_1+R_2)]\times[(R_3+R_4)/R_3]+(R_4/R_3)\} \quad (5.7)$$

Both eqs. (5.5) and (5.7) coincide when the symmetry condition is imposed on the latter. Besides, they are straightforward and nice (yes, when beauty is found in a mathematical derivation, it means we are understanding and beginning to like the subject).

5.3.1.1. Amplifier loading effect

However, there are some restrictions to consider. One of them relates to the amplifier input impedance presented to the biological system. It may become critical when electrodes are directly applied, as in electrocardiography, requiring the concept of impedance matching, or adaptation, in order to guarantee efficient signal transference. Keeping always in mind the ideal op-amp (Figure 5.2), the input impedances for E_1 and for E_2, respectively, are (R_1+R_2) and R_3 (remember the potential V_2 is a virtual zero); in fact, since V_1 is nearly equal to V_2, the former could be reduced to only R_1. Thus, when looking from the biological system, the amplifier's input impedance is given by (R_1+R_3), which combined with the output impedance of the biological source, acts as voltage divider so decreasing the actual input signal (Figure 5.3). In other words, the latter loads the original voltage.

Thus, it is important to make the amplifier input impedance significantly

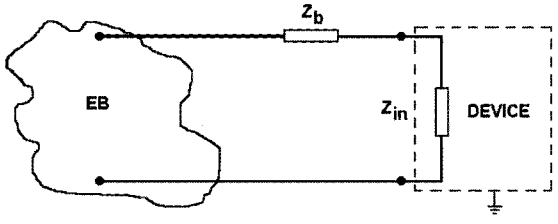


Figure 5.3. LOADING EFFECT OF THE INPUT IMPEDANCE OF THE AMPLIFIER.

higher than the output impedance of the biological element, that is, $Z_{in} >> Z_b$; the higher, the better. Since gains much larger than 10 are desirable, due to the high input impedance requirement, R_2 and R_4 should take high values. However, values greater than 10 MΩ are not recommendable because susceptibility to noise also goes up. Their tolerance is usually large (about 20 %) and their quality tends to be poorer. On the other hand, if one lowers tolerance to improve circuit performance, the whole configuration becomes more expensive. Hence, a compromise should be reached.

Suggested exercise: Design a symmetric amplifier according to the diagram of Figure 5.2 connected to a biological system with a 500 Ω output impedance. The required gain is 50 and the actual signal between E_1 and E_2 should not be lower than 0.95 the original source signal E_b. Resistors larger than 200 KΩ are not recommended. Redesign the resistors if the actual signal were to be 0.9 and 0.99 of E_b, respectively. Discuss the practical limitations, if any, in all three cases.

5.3.1.2. External resistors mismatch

Another limitation of this arrangement rests on the intrinsic practical mismatch of all four external resistors (ratios should be kept, aiming at the best possible circuit symmetry, which means the so called **resistor's pairing or matching**); otherwise, an imbalanced output is produced (Pallàs–Areny & Webster, 1999). Let us define the common mode input voltage E_c as

$$E_c = \left(E_1 + E_2\right)/2 \tag{5.8}$$

which is reflected at the output multiplied by the **common mode gain** G_c, while the differential input voltage E_d appears multiplied by the **differential gain** G_d, so that the total output, by superposition, is in fact,

$$V_o = G_c E_c + G_d E_d \tag{5.9}$$

However, by definition of differential amplifier, the common mode gain should be zero or at least tend to zero, getting again a relationship as shown in equation (5.5). In practice, G_c has a finite value different from zero and can be experimentally obtained as the quotient between the output voltage V_{ocm} and the input voltage, when the same value $E_1 = E_2 = E_i$

with respect to the reference is applied to both terminals. Thus, by making use of equation (5.4), we get

$$V_{ocm} = E_i \times \{[R_2/(R_1 + R_2)] \times [(R_3 + R_4)/R_3] - (R_4/R_3)\} \tag{5.10}$$

which, after dividing through by E_i, leads into

$$G_c = [R_2/(R_1 + R_2)] \times [(R_3 + R_4)/R_3] - (R_4/R_3) \tag{5.11}$$

as a full expression of the common mode gain.

Suggested exercise: Say that the four resistors in Figure 5.2 are equal ($R_1 = R_3 = R_4 = R$), but one of them, R_2, is 0.1 % off its nominal value, that is, it is either 0.99xR or 1.01xR, depending on whether it deviates in excess or defect. Replacing these values in eq. (5.11) yields a common gain G_c = –0.00503 and 0.00498, respectively, obviously small, finite and different from zero. Repeat the exercise when the deviation is 0.01, 1 and 5%. Discuss.

5.3.1.3. Common Mode Rejection (CMR)

Since in real life we have to deal with imperfect technological elements, this amplifier always handles the two gains introduced above, one large and the other small. Hence, it appears convenient to define a combined parameter in order to quantify how high the differential gain and how low the common mode gain are. Such parameter, extremely important in the biological amplifier, is the **Common Mode Rejection Ratio** (**CMRR**), or

$$CMRR = G_d/G_c \tag{5.12a}$$

simply defined as the ratio of the two gains. Another definition also in use is as the logarithm of the ratio, or

$$CMRR[\text{dB}] = 20 \times \log(G_d/G_c) \tag{5.12b}$$

obviously measured in decibels. Conceptually, it does not add anything; it is just an alternative. Both equations (5.12) must be emphasized because of their paramount importance in this kind of amplifier.

In the numerical example given above, since the differential gain is 1 ($G_d = R_4/R_3$, with both resistors equal to R), and rounding up an absolute common gain of 0.005, we end up with a CMR = 46 dB. In most of their applications, instrumentation amplifiers require CMR's > 80 dB, while in cases where a resistive transducer bridge is involved, 120 dB or even more may be needed. The example shows that an increase of the differential gain to 10, with the same single resistance offset of only 0.1

%, raises the *CMR* to 66 dB without reaching yet the necessary levels. The student is invited to try different numbers in order to get a feeling of the restrictions and possible practical difficulties.

Pairing of the external elements is a desirable condition and definitely should be sought. Maximum pairing is reached when G_c in equation (5.11) is forced to be zero, or,

$$\left[R_2/(R_1+R_2)\right]\times\left[(R_3+R_4)/R_3\right]-(R_4/R_3)=0 \tag{5.13}$$

easily leading to,

$$R_4/R_3=R_2/R_1 \tag{5.14}$$

as the practical condition to be met, better explaining the reason to search for symmetry. Nonthe less, usually it is not easy to obtain, thereby limiting the actual common mode rejection. Besides, as already mentioned, its input impedance is also limited. For these reasons, the single op-amp configuration is rarely used, if ever.

5.3.2. Differential Amplifiers Based on Two Op-Amps

The two-amplifier approach shown in Figure 5.4 is basically formed by

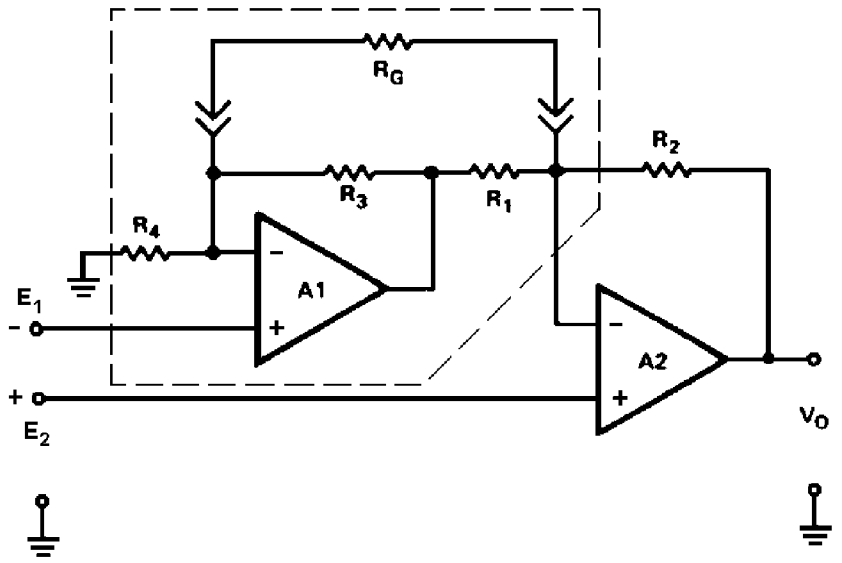


Figure 5.4. TWO OP-AMP BIOLOGICAL AMPLIFIER.

the cascade connection of two single units also linked by an overall feedback resistor R_G. The latter supplies highly linear adjustable gain without introducing common mode disturbances. Let us set up Kirchhoff's node equations at both extremes of this resistor (which coincide with the negative input points, respectively, of each amplifier), say point I to A_1 and point II to A_2; that is,

$$\left(E_2 - E_1\right)/R_G + \left(V_{o1} - E_1\right)/R_3 = E_1/R_4 \tag{5.15}$$

$$\left(V_o - E_2\right)/R_2 = \left(E_2 - V_{o1}\right)/R_1 + \left(E_2 - E\right)_1/R_G \tag{5.16}$$

The first equation (5.15) corresponds to point I and the second one (5.16) to point II. Take proper notice of the signs, as for example the first term in the first equation is positive because the difference of potential $\left(E_2 - E_1\right)$ sustains a current towards the node. Besides, recall always that potentials at the amplifiers inputs are equal so that E_1 and E_2 are used as respectively equivalent to those of the negative terminals.

Let us briefly indicate the way to obtain the differential gain G_d of this circuit defined as $V_o/\left(E_2 - E_1\right)$: First solve the first equation for V_{o1} and, thereafter, solve the second equation also for V_{o1}; then, equate them and solve for V_o to get

$$\begin{aligned} V_o &= E_2 \times \left(R_2/R_1 + R_2/R_G + R_2R_3/R_1R_G + 1\right) \\ &- E_1 \times \left(R_2/R_1 + R_2/R_G + R_2R_3/R_1R_G + R_2R_3/R_1R_4\right) \end{aligned} \tag{5.17}$$

Comparison of equation (5.17) right-hand two terms quickly shows that both factors between parentheses would be exactly the same if the subterm R_2R_3/R_1R_4 were forced to be 1, meaning that $R_2/R_1 = R_4/R_3 = k$, that is, when the resistors' pairing or matching condition is met. In that case, the equation can be rewritten as

$$V_o = \left(E_2 - E_1\right) \times \left(R_2/R_1 + R_2/R_G + R_2R_3/R_1R_G + 1\right) \tag{5.18}$$

and the differential gain G_d appears as,

$$G_d = \left(R_2/R_1 + R_2/R_G + R_2R_3/R_1R_G + 1\right) \tag{5.19}$$

Recalling the relationship *k* imposed above, the third term of this equation reduces to R_4/R_G and (5.19) becomes

$1 + R_2/R_1 + R_2/R_G + R_4/R_G = 1 + k + \left(R_2 + R_4\right)/R_G$. Finally, all four resistors

can be made equal, i.e., $R_1 = R_2 = R_3 = R_4 = R$, and

$$G_d = 2 \times \left(1 + R/R_G\right) \tag{5.20}$$

showing that the differential gain of the amplifier based on two operational amplifiers depends **only** on the feedback resistor R_G. The normal operational amplifier circuit errors, resistor matching, and common mode swing characteristics limit the performance of this circuit. In this case, the common mode rejection errors of the two amplifiers tend to cancel each other out. Resistors must also be paired to optimize CMRR. Such condition is obtained, as before, by equating to zero the common mode gain. The relationship is as that given by equations (5.12). Modifying the gain with R_G does not affect pairing and, as a consequence, does not affect the CMRR either. Input impedance is high. All these added features make the circuit better than the previous one. A possible disadvantage is that the circuit is not symmetric, in other words, E_1 is amplified by A_1 and A_2 while E_2 is only amplified by A_2. Thus, both common mode gains are different making impossible to reach zero level. However, the circuit can be implemented with inexpensive discrete components.

Based on equation (5.20), calculate the components needed to get a differential gain of 100. Remember that resistors higher than 200 or 250 kΩ are not recommendable. Repeat everything but for gains equal to 50 and to 200, respectively. Discuss.

5.3.3. Differential Amplifiers Based on Three Op-Amps

Figure 5.5 shows the most popular and already classic configuration of an instrumentation amplifier. The objective is to obtain the overall gain (which could also be called *transfer function*). Applying the principle of superposition, say for $E_1 = 0$, and the simple rules of the two basic arrangements, that is, equation (5.1a) — the *inverter* — and equation (5.1b) — the *non-inverter* — at the output of each amplifier, we get,

$$V_{a2} = E_2\left[\left(R_1 + R_G\right)/R_G\right] \tag{5.21}$$

$$V_{b2} = -E_2\left(R_1^{'}/R_G\right) \tag{5.22}$$

where the former corresponds to the non-inverter and the latter to the inverter.

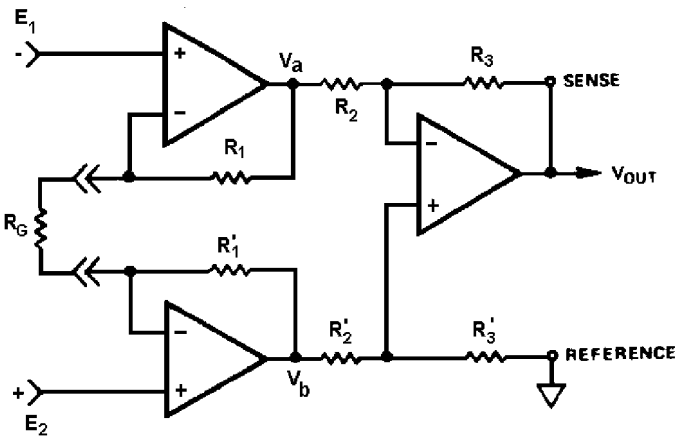


Figure 5.5. DIFFERENTIAL AMPLIFIER BASED ON THREE OPERATIONAL AMPLIFIERS. Also commonly called *instrumentation amplifier*.

Similarly, with terminal E_2 grounded, the partial voltages V_{a1} and V_{b1} are given by,

$$V_{a1} = -E_1\left(R_1^{'}/R_G\right) \tag{5.23}$$

$$V_{b1} = E_1\left[\left(R_1^{'} + R_G\right)/R_G\right] \tag{5.24}$$

This time the first equation corresponds to the inverter and the second equation (5.24) to the non-inverter. From all four equations (5.21, 5.22, 5.23 and 5.24), the output voltages at points *a* and *b* are given as algebraic additions, or,

$$V_a = E_2\left[\left(R_1 + R_G\right)/R_G\right] - E_1\left(R_1^{'}/R_G\right) \tag{5.25}$$

$$V_b = E_1\left[\left(R_1^{'} + R_G\right)/R_G\right] - E_2\left(R_1^{'}/R_G\right) \tag{5.26}$$

The final output voltage, V_o, after the third amplifier and applying the superposition theorem, is obtained by grounding first point V_b and, thereafter, point V_a. In mathematical terms,

$$V_o = -\left(R_3/R_2\right)\times V_a + \left[R_3^{'}/\left(R_2^{'} + R_3^{'}\right)\right]\times\left[\left(R_3 + R_2\right)/R_2\right]\times V_b \tag{5.27}$$

We have already all the needed relationships. In order to manifest the differential action of the circuit, a better symmetry is brought about by making $R_3 = R_3^{'}$, $R_2 = R_2^{'}$ and $R_1 = R_1^{'}$, changing the latter equation into a simpler one, or,

$$V_o = (V_b - V_a) \times (R_3 / R_2) \tag{5.28}$$

which, after substituting for V_a and V_b with equations (5.25) and (5.26) above, becomes

$$V_o = (E_1 - E_2) \times [(2R_1 / R_G) + 1] \times (R_3 / R_2) \tag{5.29}$$

so establishing a clean-cut linear relationship between the output voltage and a differential input in a device made up with just three IC's and four resistors (without counting the power supply). The differential gain is only a matter of moving the differential input as divisor of V_o to the left.

This arrangement is divided in two stages: The first one (composed of two op-amps) provides high input impedance and some gain while the second stage (only one op-amp) improves the common mode rejection. Often, field effect transistors (FET) or bipolar input operational amplifiers are used for the first stage reaching values of Z_{in} in the order of gigaohms (1 GΩ = 10^{12} Ω). Input-FET op-amps, however, generally have poorer CMRR than bipolar amplifiers. The second stage (or last op-amp) is decisive in the final determination of the CMRR. The curious student may find more details in the current literature or in the WEB or may wait until reaching the bioinstrumentation courses.

5.3.3.1. Common Mode Performance

The rationale is similar to that described above, that is, CMRR and gain, too, only depend on resistor's pairing (R_3, R_3', R_2 and R_2'). Besides, it can be demonstrated that CMRR is not influenced by R_1 and R_1' pairing. Let us see,

The common mode output voltage V_{ocm} is obtained applying at the input a common mode signal $V_{incm} = E_1 = E_2$, that is, by hooking together both terminals and connecting them to a source against the reference. Thus, equation (5.28) has to be used, giving,

$$V_{ocm} = (V_{acm} - V_{bcm}) \times (R_3 / R_2) \tag{5.30}$$

where V_{acm} and $-V_{bcm}$ should be replaced, respectively, by equations (5.25) and (5.26) yielding,

$$V_{acm} = V_{incm} \times [(R_1 + R_G) / R_G] - V_{incm} \times (R_1 / R_G) \tag{5.31}$$

$$V_{bcm} = V_{incm} \times [(R_1' + R_G) / R_G] - V_{incm} \times (R_1' / R_G) \tag{5.32}$$

Considering that $R_1 = R_1'$, algebraic manipulation immediately shows a null factor which makes $V_{ocm} = 0$, so leading to the conclusion that the first stage resistors do not affect the CMRR. Hence, its gain, determined in part by R_G, can be modified without increasing the common mode signal.

Resistors R_1', R_1 and R_G (Figure 5.5) can be complex impedances whose effect in the frequency domain may be useful, as for example to reduce gain at the higher side of the spectrum. Possibilities are, in this respect, only limited by the designer's ingenuity. This first stage should have a relatively low gain, to avoid saturation from common mode inputs, whereas higher gain is reserved for the second stage, where saturation is not a problem. An instrumentation amplifier like this may reach voltage gains of 1,000 or higher with CMRR better than 80 dB, input impedance of a few GΩ and bias currents in the order of nA. In the last two decades or so, new and more efficient integrated circuit technologies have allowed significant reductions in production costs. There are devices in the market that include three op-amps. These are the basics of the subject that should permit the student to get around and even to solve some problems with.

Suggested exercise: Assuming the circuit shown in Figure 5.5, find the resistor values if the first stage has a gain equal to 5 and the second stage a gain equal to 10. Thereafter, consider a common mode input signal of 10 volts and calculate the undesired output signal with a CMRR of 66 dB and of 84 dB. Play a little with different numerical figures. Discuss.

5.4. Instrumentation Amplifier Specifications

An engineer makes measurements to know what kind of numbers he/she is dealing with, as for example to have an idea of the signals to handle or the environment the equipment is supposed to operate in (temperature, humidity, a vehicle, a highly noisy place) or any other special restriction (as space or weight). Based on these numbers and also on objectives clearly established, he/she sets a list of requirements, called *specifications*, the device is supposed to meet. Hence, the biological amplifier (perhaps to be placed in a surgery theater) and the instrumentation amplifier to implement it, always are accompanied by the *specification sheet*.

In this section we will introduce and discuss the different data and their relative significance.

5.4.1. Basic Requirements

The manufacturer usually provides the data sheet or *spec sheet* (as sometimes is referred to in the daily jargon). It is obviously the end product core information; it concisely and numerically describes the unit to be installed in its working place. Even though the basic specifications cannot be improved, at least, some parameter compensation can be made. For example, the spec sheet may list, say,

Typical $V_S = \pm 15V$ Power supply

$R_L = 2K\Omega$ Load resistance

$T_A = +25°C$ Ambient temperature

Deviations from these conditions might degrade (or in some cases, improve, as for example with less stringent temperature requirements) the device's performance. The word “typical” often found in the spec sheet means that the datum was obtained in the factory during bench tests by averaging, in fact implying a probable different value around the stated one for the particular chip we bought. Other specifications have more precise definitions, or are based on several electronic determinations or knowledge and, thus, are beyond discussion.

5.4.2. Gain Range

The data sheet provides the *gain equation* to be applied for a given IC. It usually takes the form,

$$G = 1 + 2 \times 10^5 / R_G \tag{5.33}$$

where R_G [Ω] stands for the external resistor to be connected to a pair of stated pins. From it, for example, $R_G = 200{,}000/(G-1)$, yielding values of $G = 1$ with $R_G = \infty$ (open circuit) to $G = 1001$ for $R_G = 200\ \Omega$. In practice, the external resistor is an R-network with at least a switch in order to choose a particular gain from a set of pre-established values. Resistors quality, wiring, contact resistances and stray capacitances will influence

the accuracy of the design and should be also considered. A single potentiometer is not a good solution.

Suggested exercise: Calculate a network to obtain, for a given IC (for example, an AD520; look it up in a manual), gain values of 10, 20, 50, 100, 200 and 500, including the switch connecting the resistor to the circuit terminals.

High gains (say, > 500) are in general not recommended because, when there is marked input noise or drift, they tend to worsen the situation. Eventually, if the information is to be digitized for further processing, gain can be compensated for in a subsequent stage (the so called *gain trimming*). Another piece of information refers to changes in gain due to changes in temperature. It is measured in parts per million per degree centigrade [ppm/°C]. Intelligent systems may correct for this error by means of a feedback loop.

Suggested experimental exercise: Test an IA (say, that mentioned before) by heating it up with a hair-drier and drawing the gain versus temperature curve.

5.4.3. Non-Linearity and Distortion

Non-linearity is the deviation from the expected straight line in a diagram plotting the output parameter versus the input parameter. When there is non-linearity, there is distortion of the reproduced waveform. Figure 5.6 depicts a transfer function (or gain) of a hypothetical device.

The horizontal axis represents the input voltage (say, in volts) and the vertical axis the output voltage (also in volts) but divided by the gain. This is an artifice to produce the ideal straight line at 45°, that is, bisecting the orthogonal system. When the horizontal axis is scanned from left to right measuring at each position the difference between the actual point on the curve and the theoretical point on the straight line, two maxima are found, Δ^-_{max} and Δ^+_{max}. The first one underestimates ("negative" deviation) while the second overestimates ("positive" deviation). The larger of the two in absolute terms divided by the corresponding input voltage (read below on the graph) is a measure of non-linearity and is expressed as "equal or smaller than the calculated percent". Another way is by dividing that maximum deviation by the full-scale output range. Non-linearity can be corrected with a straight line fitting procedure in which the slope and offset voltage are modified until $|\Delta^+_{max}| = |\Delta^-_{max}|$.

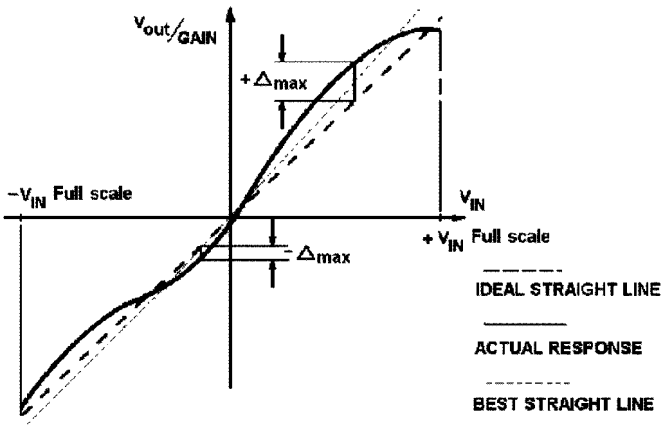


Figure 5.6. TRANSFER FUNCTION ILLUSTRATING NON-LINEARITY.

Figure 5.6 shows such line (dotted). Such criterion is the base of the sometimes-called *Best Straight Line Method*.

5.4.4. Input Characteristics

They can be described in terms of voltage, current and impedance and, as such, should be separately treated although, in fact, they are inter-related.
Input Voltage Range refers to the maximum applicable value. For example, the type AD524 accepts up to ±10 Volts. Many times, this parameter is given both for differential and common mode inputs.
Input Current is zero in the ideal op-amp, but this is real life and, no matter how much we may dislike it, it is small and finite; at least two types are to be accounted for:
1) *Input Bias Current* is sustained by the power supply, which keeps the active elements alive in any circuit. Devices with FET at the input have very low bias current but they are strongly dependent on temperature. Thus, it is important to determine their sensitivity to it. Typical values, say for an AD620A range between 0.5 to 2.0 nA. Often, only the maximum value is supplied.
2) *Input Offset Current* is defined as the difference between the two bias currents (one per terminal); hence, it is a residual current, some kind of left-over of an imperfect circuit. In spite of the differential input, the bias

current requires a return pathway to the reference terminal. Otherwise, amplifier saturation and drift could reach unacceptable levels.
In both types of current (bias and offset), their sensitivity to temperature changes is given in pA or nA per degree centigrade.
Input Impedance (Differential and Common Mode), appears as critical if one recalls the loading effect it has on the desired biological signal to be recorded. Once more, it is a fact of the biomedical engineer's life making us gently smile of the preached ideal infinity. For the IA depicted in Figure 5.5, the first stage two op-amps input impedances, in turn dependent on the factory technology, define the overall input impedance. It may reach from a few MΩ to a few GΩ. The IC AD524, as an example, has a specified value of 10^9 Ω, in the differential and in the common mode as well.

5.4.5. Common Mode Rejection Ratio

The higher this parameter, the better the behavior of the device. Usually, it is specified for full range input voltage at a given source imbalance. Since it depends on the gain, the latter is also given (Bronzino, 1995). We suggest to experimentally test an IA so that the student can have a hands-on feeling. Review the definitions and equations given above and remember the two input terminals must be hooked together to inject a common mode signal.

5.4.6. Voltage Offset and Drift

Above, we introduce the offset current. We should not be surprise if there is also an offset voltage. If a direct current (dc) amplifier output — with its input terminals short-circuited and grounded — is connected to a paper recorder running at very slow speed during several hours, a line moving slowly up and down will be obtained; that is, the recording pen did not stay at the same level, it **drifted** aimlessly. Moreover, as soon as the output is connected to the recorder, the output voltage (which should be zero for a zero input) shows a well measurable **offset**.
Offsets are due to intrinsic imbalance of the input circuits (FET's or other) and are also dependent on the temperature. Drift is a consequence of instabilities and is also dependent on temperature.
The total offset at the output has two components; the input-offset voltage multiplied by the amplifier's gain and the output-offset voltage inde-

pendent on the gain. For higher gains, there is a dominance of the former, while for lower gains the latter tends to have more weight. The technologies behind the IC's implementing the amplifiers have decidedly a strong influence on the actual values. Drift is quantitatively expressed in terms of rate, as volts per day or per month and also as volts per day per degree centigrade.

Offset voltage can be compensated for in some types with an internal adjustment via an *ad-hoc* external potentiometer. However, it may not be a full solution because the bias condition is usually modified and/or there is thermal interaction with the device. The commonest way is by means of an external voltage applied to the reference terminal. The data sheet itself usually hints this kind of arrangement. Finally, the offset voltage is many times referred to the voltage supply, so attempting to quantify fluctuations effects of the latter on the former.

5.4.7. Frequency Response

The dynamic range of the amplifier depends on the frequency band it can handle. Biological signals have low amplitude and extend from 0 Hz (dc) up to a few hundred Hz in most of the cases, or even up to 10 kHz when the electromyogram is recorded or in high fidelity electrocardiography. In comparison with amplifiers in other fields, the biological amplifier requires a reduced bandwidth and is always shifted to the low side.

The data sheet often provides the gain versus frequency plot or it gives the high cut-off frequency, that is, the frequency at which gain drops 3 dB. For a few ac biological amplifiers, as in electrocardiography, the low cut-off is also given (in the order of fraction of Hz). Typical values for IA's are well above the requirements of any physiological signal.

There are two time related parameters, which conceptually belong to the category of dynamic characteristics: *Settling Time and Recovery Time*.

The first one is defined as the total time required for the output to respond to a fast full-scale input step. It is some kind of lag time or "inertia-like" phenomenon found in many other systems. Typical values are in the order of 10 to 75 μs (rather fast, indeed) and it does not necessarily show gain dependency.

Transient discharges (as from cardiac pacemakers, defibrillators, electrosugery equipments) may get into the biological amplifier, imposing an overload, and driving it into saturation. How long does it take to recover,

i.e., to return to the normal operating range? This is the *recovery time*. By and large, it depends on the transient characteristics (amplitude and duration) and also on the overall amplifier frequency response features. It can reach several seconds. There are protecting circuits, which automatically disconnect the amplifier once an overload is detected.

Suggested exercise: Find examples where the concept may (or may not apply). Hints: A force suddenly applied to a body; the lag time of a muscle to respond to a stimulus. Design possible experiments. It does not matter whether they are or not realizable. Discuss critically. Einstein used to imagine experiments and his results were highly productive.

5.5. Noise and Interference

> *The weed always has to be separated out from the good herb; besides, never mix two good herbs, choose only one; otherwise, you may get a tummy ache. Old Mother Rose's advice.*

The input signal to a biological amplifier has several components: The desired biopotential to be measured plus unwanted disturbances, such as noise or interference of different characteristics — usually originated in external sources different from the biological system — interference from other normal and always present biopotentials, and the ever present potentials at the electrode-electrolyte interface (see the preceding Chapter). The latter could reach amplitudes significantly higher than the desired signal and may become particularly annoying with motion (the so called movement artifact).

Generally speaking, any unwanted signal mixed with or superimposed on the desired one is called noise or interference or disturbance. We will not make any distinction among these three words although, under certain circumstances, noise could be differentiated from interference or disturbance (the first would be a continuous random signal of very wide frequency content, while the second could be manifested as single or multiple pulses of undetermined shape and duration). Its origins recognize a variety of sources:

External sources are located outside of the biological system and of the amplifier. They usually come from electrical and electronic appliances, but the atmospheric discharge is also possible.
Internal sources are located within the biological system and, most of the time, produce normal potentials which do not belong to the desired biological signal.
Intrinsic sources essentially refer to the amplifier itself. It is noise of the integrated circuits and its associated circuitry.
An adequate amplifier design must tend to fully eliminate or at least minimize all the above-mentioned disturbances.

5.5.1. Coupling of External Interference

First, we will recall three fundamental physics postulates, which permit the modeling of noisy systems:
Electric fields are confined within capacitors;
Magnetic fields are confined within inductors;
Circuit dimensions are always much smaller than the noise wavelengths.
If for some reason they do not hold, or the analyst does not want to make use of them, the situation changes drastically transforming the system enormously more complex and essentially impossible to resolve. Models are the only way to approach complex systems and, as such, models **must** turn to simplifying assumptions and hypotheses, sometimes even oversimplifications. Remember what was said about models in the preceding chapter.
According to the way a disturbance gets into the amplifier, we can consider different types of coupling: *conductive*, *capacitive*, *inductive* and by *electromagnetic radiation.*

5.5.1.1. Noise counductively coupled

It is a very common situation. A cable laid on the floor of a room may pick up noise. The wiring from the power supply to its amplifiers is an easy route. A common resistor or impedance taking current from two different circuits can introduce interference from one to the other and vice-versa. Figure 5.7 depicts such situation: The current due to the second circuit modulates voltage *VG1*, developed by the first current. Conversely, the first one also modulates voltage *VG2*, developed by the second circuit. The net result is mutual introduction of noise from one to the

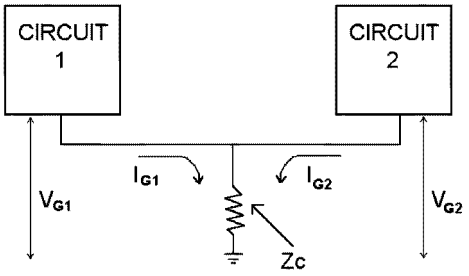


Figure 5.7. CONDUCTIVELY COUPLED NOISE.

other. Since the pathway is a line or a wire or resistor network, either overtly or covertly (the latter may no be easy to uncover), this kind of coupling noise is called *conductive.*

5.5.1.2. Noise capacitively coupled

As the word indicates, the culprit here is a capacitor but … it is not a well-behaved "touchable" capacitor built by man, it is a stray capacitance (called also distributed or parasitic) formed by wires, connections, the animal or the patient and the noise source. Typical values, to illustrate, of interest to the biomedical engineer in the medical environment are:

Human subject standing up on an insulator, between the subject and the noise source	700 pF
Adapter of ac input power to ±15V dc output, between primary and secondary	100 pF
Shielded two-wire cable, between both wires, per foot	40 pF
Between one wire and shield, per foot	65 pF
Coaxial cable (RG58), between central conductor and shield, per foot,	33 pF
Optic isolator, photodetector LED, between input and output	2.0 pF
Resistor, 1/2 w, tip to tip	1.5 pF

Figure 5.8 gives a simple model visualizing a capacitive coupling between two conductors when there is a generator connected to one of the wires and a resistive load to the other. The reference is ground. Voltage

V_1 and conductor 1 act as interference source while conductor 2 is its receptor. There is a stray capacitance, C_{12}, coupling the first to the second. C_{1G} and C_{2G} are capacitances from each wire to ground. The second conductor has a load R. The noise voltage V_N can be calculated by considering the complex voltage divider (see Figure 5.8) formed between V_1 and V_N via impedances Z_{1G} and Z_{2G}, respectively, offered by the stray capacitance C_{12} and the parallel combination of C_{2G} and the load R; hence, obtain first the ratio $Z_{2G}/(Z_{12}+Z_{2G})$, and work it out algebraically to find,

$$V_N/V_1 = j\omega[C_{12}/(C_{12}+C_{2G})]/[j\omega+1/R(C_{12}+C_{2G})] \tag{5.34}$$

where, interestingly enough, capacitance C_{1G} does not play a part. Besides, if the load resistance has a much lower value than the combined impedance presented by $C_{12} + C_{2G}$, a rather common situation in practice, the equation of above reduces to

$$V_N/V_1 = j\omega C_{12}R \tag{5.35}$$

In conclusion: The noise voltage due to capacitive coupling is a function of the amplitude and frequency of the noise source and also of the coupling capacitance and the resistance to ground of the affected circuit. Since the first two parameters usually cannot be controlled, the only way of reducing noise y by acting on the two latter.

The student is urged to work out the last derivation. First divide through by jω, thereafter,

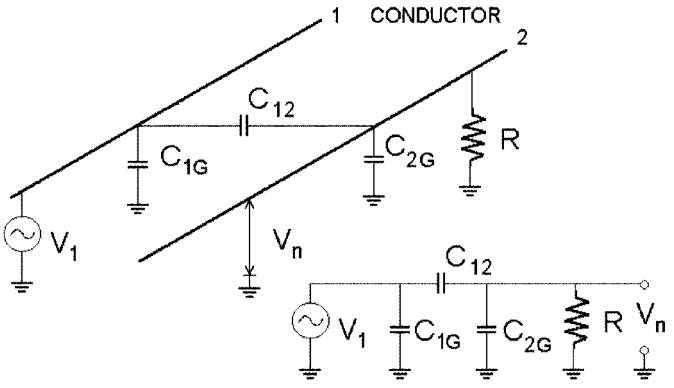


Figure 5.8. INDUCTIVE COUPLING. The circuit to the right is an equivalent representation of the two wires to the left.

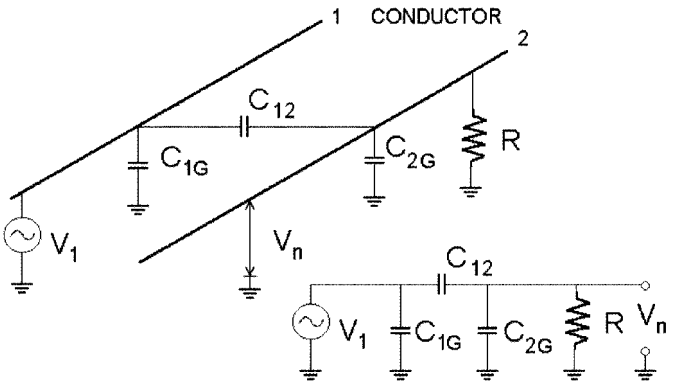


Figure 5.9. CAPACITIVE COUPLING, A PATHWAY TO NOISE. Impedance Z_{12} is represented by the stray capacitance C_{12} between the two wires while impedance Z_{2G} appears as the parallel combination of the load R and the stray capacitance to ground C_{2G}.

multiply through by *R*, and in the new expression apply the above-mentioned simplification. We use the words resistor and resistance, capacitor and capacitance, inductor and inductance. Pay attention how they are placed in the sentence recalling that those ending in *-or* refer to the physical element containing the electrical property whose English term ends in *-ance*. Very rarely, if ever, we can hold in our hands, say, a resistor showing pure resistance, for it is always contaminated by the other two properties!

5.5.1.3. Noise coupled inductively (or magnetically)

Current flowing into a circuit generates a magnetic field that, in turn, may induce voltages in a neighbor circuit via a mutual inductance. Let us consider Figure 5.9, where such situation is depicted. Magnetic flux is $\Phi = L_1 I_1$, partially transferred to the second circuit by the mutual inductance M_{12} so inducing, by Faraday's law, the noise voltage $V_N = \mathrm{j}\omega M_{12} I_1$. An equivalent circuit (a model) is also shown.

The mutual inductance depends on the geometry of the system and on the magnetic properties of the medium (most probably air in the physiology laboratory or the hospital environment). As in the previous capacitive case, the source amplitude and frequency directly influence the amount of noise induced. The only way to reduce it is by reducing the mutual coupling. The task may face great practical difficulties though, especially

when there are power distribution lines, transformers and ac machinery in the proximity of the working place.

5.5.1.4. Noise coupled by electromagnetic radiation

Physiology laboratories, intensive care units, coronary units, catheterization labs an other medical departments placed in the neighborhood of radio (either AM or FM), TV or radar stations, are prone to nicely blur their records with this very annoying interference. With due high respect, we should unload the original responsibility to Maxwell, Herz, Tesla, Marconi and many other wireless communication pioneers and, at the same time, recognize the interference as part of the price we pay for living the luxury of an almost simultaneous world. However, and do no take it all on these fine men, electrocautery and diathermy equipments (which use RF in the MHz range) may also be powerful and very close sources of disturbing signals. Some piece of the medical equipment acts as a *de facto* antenna (perhaps the cables hooked to the patient) while, somehow, the amplifier acts as an unwilling detector. Techniques to reduce this disturbance are complex, not easy to implement, require a good physical construction of the operating theater along with a good arrangement of the instruments within it, plus adequate filtering.

5.5.2. Network Interference

> *Take a deep breath, buddy, 'cause this is gonna be a thorny morsel to swallow*

The domestic urban power line is without doubt the main source of interference detected by our biosignal systems. Instrumentation amplifiers are specifically designed to reject, as much as possible, this 60 or 50 Hz (frequency depends on the country) disturbance coupled either capacitively or inductively or both to either the patient or the experimental animal.
Let us consider an ECG recording via surface electrodes. Figure 5.10 schematizes a usual arrangement, where always the right leg is grounded. Coupling possibilities in the hospital or laboratory environment are manifold. Just merely think of the countless ac-lines crossing walls, ceilings and floors, getting into power supplies, equipments of all sorts, including those located in the very working place and other as pedestrian as vac-

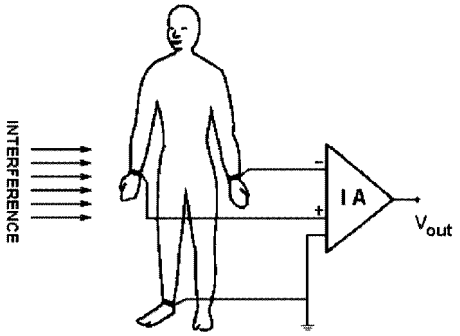


Figure 5.10. INTERFERENCE PICKED UP BY AN ELECTRO-CARDIOGRAPHIC SYSTEM.

uum cleaners or floor-polishers, and easily you will realize that our poor biological amplifier is literally swimming in a noise-sea. In fact, you may even believe the successful measurement of what you want, say, the ECG, is an impossible mission. Well, for the Biomedical Engineer no mission is impossible.

A little gossiping to relax you: Do you remember or have you ever seen a replay of *Mission Impossible*, that breath-holding TV series of the sixties? For that team no mission was impossible. Lalo Schiffrin, a great and talented Argentine-born musician and composer, wrote — in a not common rhythm (5/4 time signature, i.d., five quart beats in a measure) — its background beautiful music. He was high-school classmate of this book's author (MEV), and once, as teenagers, he played in MY piano (still in my hands and in excellent conditions); that is the best musical accomplishment found in my cv!

5.5.2.1. Capacitively Coupled Interference

Figure 5.11 intends to represent Figure 5.10 in a slightly more real world: The patient is replaced by a volume conductor holding the biological generator (for example, the heart in this case), V_{BIOL}, which develops surface potentials A, B, and C, detected by the electrode leads. Between these points and the generator there are resistances, R_A, R_B and R_C, materialized by the body fluids. Their values can be in the order of 500 Ω. Besides, Z_{e1}, Z_{e2} and Z_{eG}, represent the respective contact electrode impedances, that is, the impedance between the patient's skin and the electrode metal; expected values lie somewhere between 1 KΩ and 20 KΩ. The amplifier receives the signal (any, the good and the bad one) with its two input impedances, the differential Z_d and the common mode one Z_{cm}. It would be a good idea to go back in the chapter to review these concepts.

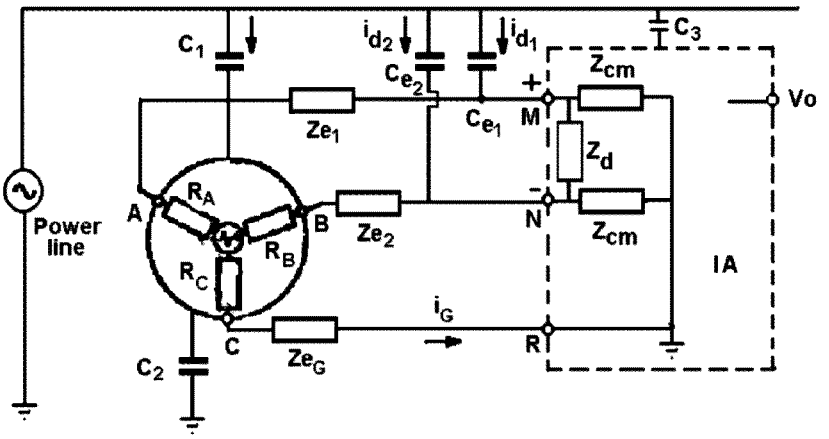


Figure 5.11. MODEL TO ANALYZE CAPACITIVE PATHWAYS. The circle represents a volume conductor (the patient) with the heart in its center (the biological generator). Surface electrodes *A*, *B* and *C* pick up a fraction of that signal. Currents i_{c1} (through capacitor C_1), i_{d1} and i_{d2} return as i_G to ground; they do contribute to interference, but whatever currents may traverse C_2 and C_3 do not. Besides, they are usually rather small. See text for more details.

The possible capacitive coupling pathways are defined as follows:
C_{e2} and C_{e1}, between the power line and the lead wires;
C_1, between the power line and the patient;
C_2, between the patient and the reference;
C_3, between the power line and the equipment case.
Coupling magnitude depends on how close the active and ground conductors are with respect to the measuring system and patient. Capacitive stray elements, respectively, offer pathways for currents i_{d1}, i_{d2} and i_{c1}, returning to ground as i_G via Z_{eG}; the latter is always lower than the impedance offered by C_2 while the differential input impedance to the amplifier is very high. The other two stray capacitive pathways, C_2 and C_3, are not interference contributors. Hence, the total current through the return connection is,

$$i_G = i_{d1} + i_{d2} + i_{c1} \tag{5.36}$$

Moreover, since the patient's impedance is much lower than the electrode impedance, the net interference voltage applied to the amplifier appears expressed as,

$$V_M - V_N = i_{d1} \times Z_{e1} - i_{d2} \times Z_{e2} \tag{5.37}$$

Assuming $i_{d1} \approx i_{d2} \approx i_d$, equation (5.36) takes the simple forms,

$$V_M - V_N = i_d \left(Z_{e1} - Z_{e2} \right) \tag{5.38}$$

leading to the unfortunate but not surprising conclusion that **there is a noise differential voltage at the amplifier's input due to capacitive coupling**. At the output, it is reflected as,

$$V_o = G_d \times \left(V_M - V_N \right) = G_d i_d \times \left(Z_{e1} - Z_{e2} \right) \tag{5.39}$$

Well … after a regret sigh, the question springing up right away is, and how big this nasty interference can actually be? Let us make some estimates of a hypothetical but possible situation. Biomedical engineers, as any other engineer, always have to play with numbers to get an idea of where they really stand.

Let us consider a unipolar cable, 9 m long, with which a current $i_d \approx 6$ nA was measured in a given medical office. Even though theoretically the electrode impedance should be equal for the pair, experience has taught us that there is always a difference. Perfect matching does not exist. Assuming a good installation of the electrodes (good electrolytic paste, good pressure, clean surface), the difference can range from 1 KΩ (excellent contact) to 20 KΩ (poor contact), producing 6 μV < ($V_M - V_N$) < 120 μV, as the probable range of the undesirable signal. Some authors cite differences even above 50 KΩ. We think it is way to high, … or the technician installing the electrodes did not know what he/she was doing! The typical ECG input signal competing with the intruder is in the order of 1 to 2 mV, that is, interference is about 1 % to 10 % of the signal to measure. The first level is barely acceptable, the second is definitely not because at the output, if the gain is 500, we would be getting from 0.3 mV to 60 mV … of **noise**! We seem to be almost "kinda stuck". What do we do then? Experience from many people, careful analysis of labs, operating rooms, coronary units and similar working shops in many places can be summarized in a few advices, and beware, because there is for the time being no other way out,

- **worsen the coupling conditions by decreasing the input cable length**, **increasing the distance between the input cable and the power line**, **shielding the input cable**, **and grounding the shield**;

- **decrease the electrode impedance by improving the electrodes**, **cleaning them well**, **using adequate paste and**, **very important**, **installing them properly**.

Still we have the common mode signals entering the system by the very same capacitive coupling. Our amplifier is well prepared for them with a high rejection ratio, typically 80 dB (see above). The common mode potential between any point over the patient and ground is,

$$V_{cm} = i_G \times Z_{eG} \tag{5.40}$$

The latter equation is based on the fact that biological impedances are always much lower than contact impedances. That common mode interference, taking into account the voltage divisors shown in Figure 5.11, produces a difference between points M and N that can be written as,

$$V_M - V_N = V_{cm}\{[Z_{cm}/(Z_{cm} + Z_{e1})] - [Z_{cm}/(Z_{cm} + Z_{e2})]\} \tag{5.41}$$

but, since both electrode impedances are much smaller than Z_{cm}, the latter equations reduces to,

$$V_M - V_N = V_{cm} \times [(Z_{e2} - Z_{e1})/Z_{cm}] \tag{5.42}$$

Verify the latter equation, but be careful how you do it. Hint: Fully develop the fraction within the brackets in (5.40).

In brief, equations (5.37) and (5.38) describe the differential interference while (5.39) and (5.41) deal with common mode noise; both types are capacitively coupled.

To illustrate now the latter, let us assume i_G = 0.5μA and Z_{eG} = 10 kΩ, leading to V_{cm} = 5 mV by applying equation (5.39). If the electrode impedance imbalance is ($Z_{e2} - Z_{e1}$) = 5 kΩ and Z_{cm} = 1 MΩ, the amp would be presented with $V_M - V_N$ = 5 mV [5 kΩ/1 MΩ] = 25 μV, after equation (5.40). This is a border value. If you play a little with possible numbers, easily the unwanted signal may reach 100 μV or even more, which is unacceptable or at least objectionable. To reduce the level of interference still sneaking in, in spite of the common mode rejection, we have to,

- **increase the CMR of the amplifier;**

- **use low impedance electrodes trying to decrease the imbalance.**

For the former, try to get a CMR better than 100 dB with amplifier's input impedance between 10^9 and 10^{10} Ω at 50–60 Hz. For the latter, try to keep the difference below 10 KΩ or, even better, below 5 KΩ. Yes, we know, it is not an easy task, but you belong to the younger generation full of motivation and pep!

5.5.2.2. Inductively Coupled Noise

We were so entangled with the previous analysis and, up to a point, happy with our results, that we forgot the magnetic field generated by the flowing electric currents. If the patient, cables and equipment are arranged in such a way that a closed loop is formed near the magnetic lines (Figure 5.12), a voltage will be induce in the loop, as described by,

$$V_n = 2\pi f B A k \cos\theta \quad [\text{V}] \tag{5.43}$$

where

f is frequency, usually 50 or 60 Hz;

B stands for the root mean square value (RMS) of the flux density, in Gauss;

A represents the area of the loop, in cm^2;

θ is the angle formed by the magnetic field and the plane of the loop;

$k = 10^{-8}$, constant to be consistent with the CGS system.

Suggestion: Go back to your physics notes and review Faraday's induction law, the associ-

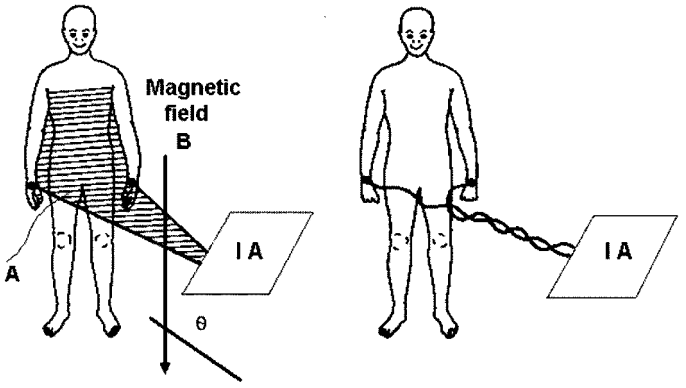


Figure 5.12. MAGNETICALLY COUPLED NOISE. IA stands for instrumentation amplifier.

ated units and how the cosine gets into the equation.

The equation teaches that a reduction in any of the factors brings about a decrease in the induced voltage. Hence, getting away from the magnetic generator (the power line), twisting the power cables to cause an alternation in the magnetic field direction, putting the signal cables very close together to reduce the loop area, searching for an optimal orientation to minimize the effect of the angle, are all *on the spot* measures to improve the signal to noise ratio.

5.5.3. Electrode Contact Potentials

Every time a metal is placed into contact with an electrolytic solution a redistribution of electric charges takes place. This forms the so-called Helmholtz double layer. We refer to Chapter 4 for a more detailed discussion of the subject. This phenomenon happens even when the electrodes are placed on the skin, either with paste or without it, for the skin always has a fine film of salty water (our perspiration). Movement of the electrodes, pressure applied to them (even lightly), modifies such distribution and along with it the contact potential which is cleanly detected by the amplifier producing large shifts in the recorded signal.

5.5.4. Biopotential Interferences

Other normal electrical signals from the same patient or animal can act as a disturbance blurring or even hiding the desired bioevent. Sometimes, it is easy to determine where the interference comes from; sometimes, it may not. Typical examples are,

- Muscle activity (electromyogram or EMG) in the electrocardiogram (ECG);
- EMG in brain electrical activity (electroencephalogram or EEG);
- Maternal ECG in fetal ECG;
- Respiratory movements in ECG
- Respiratory movements in thoracic impedance records (impedance cardiogram or ICG).

Suggestión: Find other examples. Remember the different physiological systems and their many signals. Take a look at a physiology textbook or go back above.

The first two situations, extremely common, may be confused with interference from other external sources (such as RF interference) because of

the high frequency of the EMG. It is advisable to ask the patient to relax and keep quiet. In an experimental animal it is more difficult and, occasionally, a mild muscle relaxant is injected. The situation becomes more difficult in patients showing intrinsic tremors (as with some cases of Parkinson disease). Thus, adequate filtering is required favored by the fortunate higher frequency of these disturbances with respect to the ECG and EEG.

The third case has produced innumerable papers for the fetal cardiac signal supplies an essential piece of information to the obstetrician. One disadvantage is the large amplitude of the mother's ECG as compared to the fetus' ECG. Conversely, one advantage is the lower frequency of the mother's ECG in relation to the fetus' (roughly, 1:2). Hence, good filtering (usually digital) combined with averaging techniques can successfully unbury the fetal signal from its mom's.

In the fourth example, respiratory movements cause a shift in the electrode position (contact potential change) and modulation of the thoracic basal impedance in ICG records. All this takes place at respiration rate (about 10–15 breaths/min), which is rather slow and, thus, easy to remove by filtering it out.

5.5.5. Ground Loops

The situation depicted in Figure 5.13 gives rise to a ground loop manifested by the appearance of a difference of potential ($V_A - V_B$), that is, between the ground connection, say, of an instrumentation amplifier and the ground connection of another medical machine, both monitoring the same patient. Obviously, both ground potentials are not the same, and the resulting current produces a common mode voltage drop at the amplifier's input to be reflected at its output, as previously shown above. If you are lucky, V_A may be equal to V_B, and your problem automatically vanishes. However, frequently the situation is not so favorable and the only way to remove this kind of interference is by connecting both pieces of equipment to the same ground point (the sometimes so-called *star-connection*). Besides, one single reference electrode must be used on the patient (Webster, 1992; Tomi Engdahl, 1997–2000, *Ground loop problems and how to get rid of them*, in
www.epanorama.net/documents/groundloop/index.html;

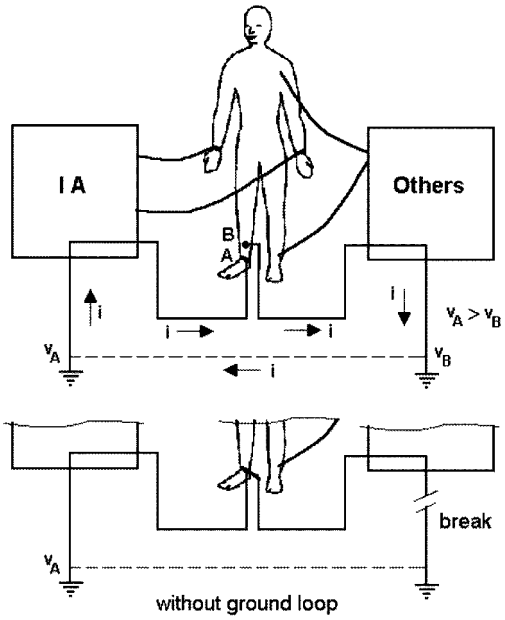


Figure 5.13. GROUND LOOPS

another interesting piece is to be found in http://www.deerfieldlab.com/ana.html).

5.5.6. Other Types of Noise

Still, there are other forms of interference, although their incidence is by and large quantitatively less important or significant. However, they deserve brief description at least (Togawa, Tamura, Öberg *et al*., 1997).

Static electricity due to frictional movements (against blankets, sheets, pillows), either of the patient, medical personnel or other people around, which may be named *triboelectricity* (*tribo*, friction), is able to sustain currents through the patient and any electrode on him/her. If the electrode impedance is high, the voltage drop through them can become an input signal to the IA, as explained before.

Kinetic energy imparted to electrons by temperature produces an internal type of ohmic noise (Pallàs–Areny & Webster, 1999) usually referred to as *thermic* or *Johnson noise* (perhaps, it should be better named Schottky–Johnson noise; see historical note below) mostly manifested at the

amplifier input stage. Its root mean square (rms) voltage value is expressed as,

$$e_n = (4kTBR)^{1/2} \tag{5.44}$$

where $k = 1.38\times10^{-23}$ joules/kcal is Boltzmann constant, T in Kelvin degrees represents the absolute temperature of the resistance involved, B in Hz stands for the bandwidth of the system, and R is the resistance value in ohms. Its spectral distribution is uniform and very wide for which reason sometimes is called *white noise*; besides, it does not depend on the manufacture technology.

Another internal noise is the *shot* (or *Schottky*) *noise*, associated to the carriers flow through a semiconductor junction. Its rms value is given by,

$$i_n = (2qIB)^{1/2} \tag{5.45}$$

where q stands for the electron charge, I is the current through in amps, and B represents the bandwidth in Hz. Similar to the previous one, it is a typical white noise, too. The term *shot* is derived from the fact that the noise sounds like a handful of B–B shot thrown against a metal surface.

Contact or *flicker noise* is caused by conductivity fluctuations due to imperfect contact between two different materials (interface effects). It is proportional to the current through, characterized by low frequency components and shows up in any device where such discontinuities are present, as relays, metal/insulators or gates of FET transistors. Flicker noise is proportional to $1/f$, where f is the frequency.

Finally, the *popcorn-like noise* originates in manufacturing defects, mostly metallic impurities in semiconductor junctions. Since current flow at the quantum level is not smooth and predictable, an intermittent burst phenomenon sometimes occurs. This noise, called *popcorn noise*, consists of pulses of many milliseconds' duration (check in the WEB SITE, Joseph J. Carr and John M. Brown, Introduction to Biomedical Equipment Technology, Prentice Hall PTR).

The student is invited to check and work out the units associated with the formulas above, (5.43) and (5.44). Full details of this rather difficult subject are usually developed in specific electronic courses.

Historical note: In the early years of electronic circuitry, engineers and physicists were trying to solve problems involved in making better vacuum valves. Although many of the problems were related to design and manufacturing techniques, the fundamental problem of noise

was gaining recognition. J.B. Johnson and Harry Nyquist, who worked at Bell laboratories in the United States, would provide some of the answers in the 1920's (Johnson, J.B. "Electronic noise, the first two decades," *IEEE Spectrum*, February, 1971). However, Walter Schottky (1886–1976), in Germany, had made before important contributions in a classic paper on noise in valve amplifiers, which was published in 1918. He had reached the conclusion that there would be two sources of noise of a fundamental nature in an amplifier. The first would occur in the input circuit and would result from the random motion of charge caused by the thermal motion of the molecules in the conductors, or what is now known as thermal noise. Since the noise is originated in the input circuit and would appear amplified in the output circuit, he deduced that it would be proportional to Boltzmann constant multiplied by the absolute temperature. In the mid–1920's, Johnson experimentally identified thermal noise and Nyquist analyzed the discovery mathematically, producing a formula of 4kT watts per unit of bandwidth, confirming Schottky's deduction. Schottky's second fundamental source of noise would be caused by the randomness of the emission from the cathode and the randomness of the velocity of the emitted electrons, which is now know as shot noise. This noise was first experimentally identified and measured in Schottky's laboratory. Later studies showed it was linked to factors such as the material and design of the cathode and extended to semiconductor junctions.

5.5.7. Characterization of an Amplifier's Noise

There are two common ways to quantifying noise of a given amplifier: by the *equivalent rms voltage value at the output when the input is short circuited* and by the *equivalent rms current value also at the output*, *when the input is open*.

The first one is expressed in $nV/Hz^{½}$ for a given frequency. Due to the 1/*f*-type of noise (flicker), its value increases at low frequencies. The point where the low frequency noise equals the thermic noise is called *frequency cross point*; it defines the frequency below which noise becomes frequency-dependent. If the technology of the IA belongs to the bipolar transistor kind, total noise originates essentially as thermic noise of the base resistance and as shot noise of the collector current. When the technology includes FET's, noise is dominated by the Johnson type at the channel resistance and is usually higher than in the bipolar case; its cross point frequency is also larger.

The second form is a measure of noise due to currents and is expressed in $pA/Hz^{½}$. In bipolar transistors, the source is shot effect of the base current and also to flicker. Amplifiers with a FET input stage show lower noise currents.

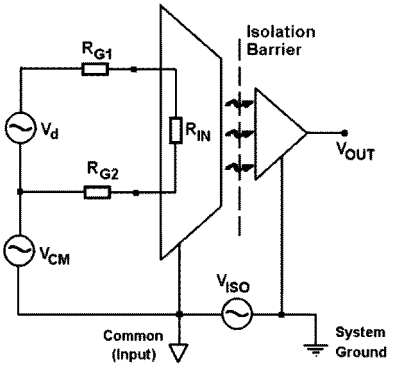


Figure 5.14. ISOLATION AMPLIFIER

A third way to quantifying noise is the total equivalent voltage, V_{nt}, which combines all noise types,

$$V_{nt} = \left[4kTBR_s + V_n^2 + \left(I_n R_s\right)^2\right]^{1/2} \tag{5.46}$$

The student will easily recognize three components: thermic noise, voltage noise and current noise. Total noise measurement and/or evaluation is a subject of great interest and amply described in the virtual literature. The student should search on his/her own.

5.5.8. Isolation Amplifier

This kind of amplifier offers ohmic isolation between its input terminals and its output. Parasitic currents and leaks must be very low and the dielectric breakdown voltage V_{ISO} (Figure 5.14) has to be as high as possible. In other words, it must show high input resistance, say, about 10 TΩ, and low input capacitance, say, <10 pF (Sheingold, 1980; Pallàs–Areny & Webster, 1999).

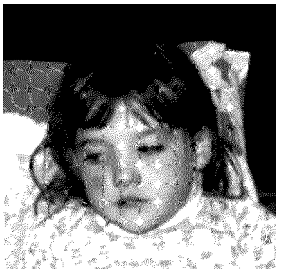
The biological isolation amplifier has as main objective protection of the patient by removing all possible electric shock risks that could result from the interaction between the patient, the amplifier and other equipment, specifically defibrillators and electrosurgery machines. Besides, such isolation often prevents interference from the main.

Isolation acts as a barrier between the patient and any technological piece connected to him/her. It can be achieved using either a transformer, a capacitor or an optical system. Ideally, in an isolated system (Figure 5.14) there is no current flowing through the barrier. This is described by

a parameter called the isolation voltage V_{ISO}, between the input reference point and the output reference. Commercial pieces specify a continuous voltage, say 2000 V, actually meaning that the device was tested at least at twice such rated value, or 4000 V.
Since the isolation amp is a safety device, the professional biomedical engineer and any industry manufacturing it must comply in this respect with standards and regulations periodically updated by the American Association for Medical Instrumentation (AAMI).

5.6. Conclusions

We have introduced the student to the basics of the biological amplifier describing its main characteristics and its possible implementation either with one, two or three operational amplifiers. Thereafter, the usual specifications found in the data sheets were explained to proceed thereon with the delicate and always present problem of noise and interference. Practice and study are mandatory to fully master the subject and, for that matter no doubt, the student will be required to take at least a specific bioinstrumentation course. This chapter will serve, we think, as an adequate launching platform and we sincerely hope that, after "enjoying" it, you did not end up as sick as the the poor girl below.



Myriam C. Herrera was born in Tucumán, Argentina, in 1959. She received the Electronics Engineering degree and PhD in Bioengineering from the University of Tucumán (UNT)[1], in 1984 and 1996, respectively. Currently, she is Associate Professor at the Department of Bioengineering (DBI), Director of the Biomedical Engineering Undergraduate Program of the *Facultad de Ciencias Exactas y Tecnología*, UNT, and Principal Professional Research Assistant at INSIBIO[2] supported by CONICET[3]. From 1981 on, she was active at different academic levels. Besides, she is member of the Argentine Society of Bioengineering and President of the Bioengineering Committee of the Argentine Cardiology Federation. In

1999, she was awarded a special mention of the prize “New Engineers: Medical Engineering” by the Scientific and Technological Promotion Agency, Argentina. She has published several research papers, presented communications in conferences and has spent short training periods in the USA, Bolivia, Colombia and Perú. Her interest lies mainly in bioimpedance, new medical technologies in the areas of cardiac and vascular mechanics and education, too. [1]UNT: Universidad Nacional de Tucumán; [2]INSIBIO: *Instituto Superior de Investigaciones Biológicas*; [3]CONICET: National Research Council of Argentina.

Chapter 6

The Interpreter: Reading the Signals

"*We look at physiological signals*, *watch them carefully*, *and many times they even show a nice appearance*; *the main problem*, *however*, *lies on how we read and understand the message carried by them*."

by Max E. Valentinuzzi and Ron S. Leder

6.1. Introduction

The first step of the interpreter in understanding the physiological significance of a signal is observation. The five human physiological senses (vision, audition, touch, smell and taste), probably in this order of importance, are still the human information input channels. Training, education and experience provide the tools with which to observe, analyze and interpret biological signals. It took time and effort to learn the physiological significance of the P, QRS and T components of the normal electrocardiogram or the significance of the dicrotic notch in the arterial blood pressure tracing. Modern signal processing and signal analysis are powerful computational tools to aid in the task of signal interpretation; they can single out small dents, incisures or notches that, otherwise, would pass unnoticed but which are important because they may have significant clinical meaning. Nonetheless, interpretation should not be confused with signal processing and analysis. The latter are techniques that subject the signal to a number of numerical operations (processing) in order to break it down into components of different type and kind (analysis), often following unique shapes and distributions (patterns) so helping in the identification of pathologies or conditions. **No new information is produced**, **for information is usually well hidden in the signal and the task is to uncover it.** Interpretation, instead, requires the human intellect, to correlate those components and/or patterns or both with a particu-

lar physiopathological situation. We will introduce the frequency and the time domain concepts referring to specific literature for detailed descriptions. The concepts of discretization and algorithm will be also introduced. The student will take in due time courses devoted to this matter. Herein, we just want to offer an overview so that the newcomer knows where he/she stands and what to expect.

6.2. Pattern Reading

Karl Ludwig, in Leipzig, initiated the modern era of continuous records in physiology when he introduced the kymograph in 1847 (Geddes, 1970). It was a landmark, indeed, that changed the course of the physiological sciences. Mechanical and electrical events, especially those associated with the cardiovascular system, started a long road to dissecting out each of their respective components. Physiologists and physicians alike had to learn by sheer hands-on experience what is known as *pattern reading* (that is, to extract information from the shape of the tracing by visual inspection), perhaps the first step of the interpreter. Let us recall some specific examples while we recommend the student to review Chapters 2 and 3, especially sections dealing with signals.

To read the ECG, Wilhelm Einthoven, physiologist and physician (1903), with his string galvanometer as recording apparatus, and Thomas

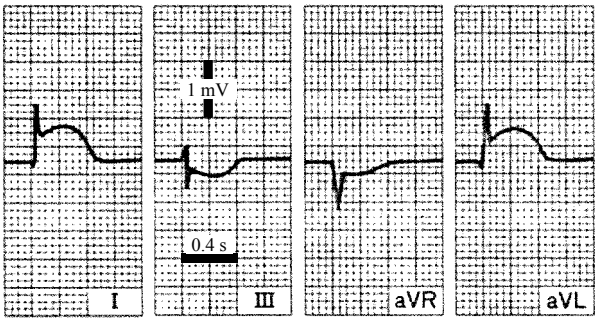


Figure 6.1. MYOCARDIAL INFARCTION. It is marked by the appearance of Q waves, elevation or depression of the S-T segment and inversion of T waves. These records depict an anterior wall infarction after a few hours it took place. There is a clear and significant S-T shift. After an ECG course given by Dr. Carlos Vallbona, Department of Physiology, Baylor College of Medicine, Houston, Texas, 1968.

Lewis (1911), with his classical treatise (*The Mechanism of the Heart Beat*), are to be considered founders and great teachers that patiently established criteria and rules still valid in clinical electrocardiology (Figure 2.31). Contributions of many others that followed improved knowledge and understanding of different aspects of this signal; say, an upward or downward deviation of the ST segment, depending on its extension, is associated with myocardial ischemia or infarct (Figure 6.1) while a long QRS and slurring of the trace are very likely related to bundle branch block and to desynchronization in the pumping action of both ventricles (Figure 6.2). A long PR interval speaks of atrio-ventricular partial block anticipating a probable second-degree block (Goldman, 1970).

Direct cannulation is required to record arterial blood pressure as a function of time (see Figures 2.14, 3.13 and 3.14). The pressure may range from 20 mm Hg to 320 mm Hg. The shape of the dicrotic notch, associated with the aortic valve closure, can show abnormalities (as insufficiency or stenosis) while large differential (or pulse) pressure is an indicator of arterial hardening or arteriosclerosis. A highly compliant aorta

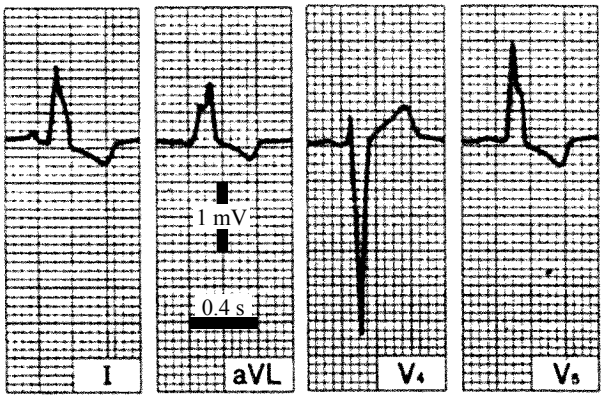


Figure 6.2. BUNDLE BRANCH BLOCK. It occurs when there is an impediment to the conduction of the cardiac electrical impulse along one of the branches of the bundle of His. Prolonged QRS and slurring are two characteristics well depicted in the records shown above. Right bundle branch, while still having the same features as left bundle branch block (shown in these records), must be diagnosed by other added details present in different leads, especially precordial. After an ECG course given by Dr. Carlos Vallbona, Department of Physiology, Baylor College of Medicine, Houston, Texas, 1968.

(i.e., less stiff) will have a smaller pulse pressure for a given cardiac stroke volume. There are non-invasive techniques to record pulse pressure with the goal of better patient screening based on pattern changes. Thus, pulse pressure seems to be a potent and clinically useful predictor of the risk of cardiovascular disease in a variety of populations. Increasing pulse pressure is strongly associated with future coronary heart disease risk in middle-aged normotensives as well as hypertensives, in African Americans as well as in whites (see Klabunde's website).

Other examples of biological signals include the electroencephalogram (EEG), the electromyogram (EMG), and heart sounds, all collected over the years essentially in analog form, i.e., signals defined over all time with infinite precision; they are continuous-time, continuous-amplitude recordings. Nowadays, there are digital databanks that store and update this physiological information and experimenters and practitioners are being trained with these modern tools while also using traditional philosophy (www.physionet.org).

6.3. Discretization of a Signal

Digital computers, as the name indicates, do not deal with continuous signals. The analog signal is converted to discrete numerical pieces, both in time and in amplitude. Hence, the recorded experimental or clinical trace is discretized. Often, the words digitalization or digitized are used as synonyms.

6.3.1. Horizontal Discretization: Sampling Time Interval

Referring to Figure 6.3, suppose a continuous blood pressure record is to be digitized. The first step consists of selecting a *sampling interval* Δt to pick up amplitudes of the signal which correspond to instants 0, 1, 2, …, (n–1), n; since Δt is kept constant, the time difference ($T_n - T_{n-1}$) = Δt, always. Obviously, $1/\Delta t$ defines the *sampling rate* f_s. Vertical bars in Figure 6.3 make up the *sequence* or collection of numbers that now represent the record. We could connect sequentially with straight segments the bar tops in an attempt to reconstruct the original trace; however, even though the first 5 points and the last 2 connected by 4 and 1 straight lines, respectively, essentially coincide with the record, the 5th, 6th and 7th bars miss the minimum A and the maximum B of the dicrotic

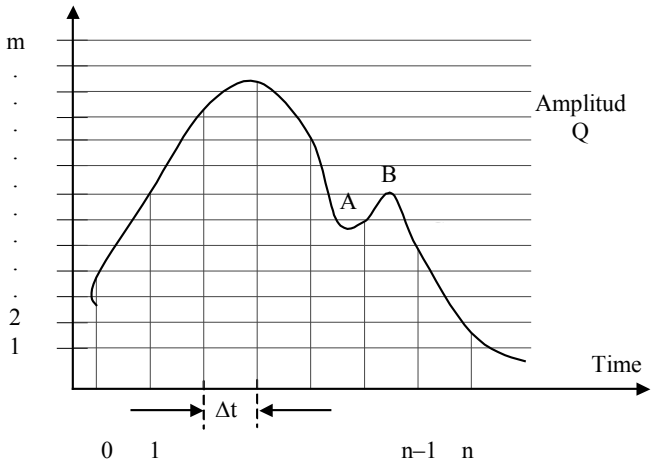


Figure 6.3. CONCEPT OF SAMPLING. It means converting an analog signal to a discrete-time sequence of numbers. The trace above mimics a blood pressure record where A and B correspond, respectively, to the minimum and maximun of the dicrotic region.

notch. It appears clear that, roughly, just by eyeballing the figure, the sampling rate should have been at least twice as fast (meaning a shorter sampling interval) to capture all of the features of the signal.

In fact, the Nyquist or Sampling Theorem tells us that under ideal conditions we can successfully sample and reconstruct or play back frequency components up to one-half the sampling frequency. Thus, a sinusoidal signal (or component of a more complex signal, such as the dicrotic notch shown in Figure 6.3) can be correctly reconstructed from values sampled at discrete, uniform intervals as long as the highest signal frequency is less than half the sampling frequency. Any component of a sampled signal with a frequency above this limit, often referred to as the folding frequency, is subject to a distorting phenomenon named *aliasing* (*alias*, in Latin, refers to a fake individual or character; it does not really exist). In other words, frequency components less than the sampling frequency that do not exist are created by undersampling and are folded back into the true data. In practice sampling is usually done at least three times the maximum frequency that will appear in the data and a special low pass or anti-aliasing filter is used to guarantee aliasing will not occur.

Exercise: Obtain the limiting sampling rate. Find the maximum sampling period for a sinusoid of 100 Hz. Choose any musical CD and find its sampling frequency. Wich is the human audible frequency range? Does the CD digital music meet the sampling theorem requirement?

Aliasing can be explained easily and didactically in terms of a visual sampling system: movies. We all have watched a western where the wheel of a rolling wagon, desperately trying to get away from a bunch of mean outlaws, appears to be going backwards. This moving image is aliasing. The movie's frame rate undersamples the rotational frequency of the wheel, and our eyes are deceived by the misinformation. This also happens when we try to record and play back frequencies higher than one-half the sampling rate.

Consider a digital audio system with a sample rate of 50 kHz, recording a steadily rising sine wave tone. At lower frequency, the tone is sampled with many points per cycle. As the tone rises in frequency, the cycles get shorter and fewer and fewer points are available to describe each shorter cycle. At a frequency of 25 kHz, only two sample points are available per cycle; this is Nyquist sampling limit. In music, with its many frequencies and harmonics, aliased components mix with the real frequencies to yield a particular unpleasant distortion. If the pitch of the tone continues to rise, the number of samples per cycle is not adequate to describe the waveform, and the inadequate description is equivalent to one describing a lower frequency tone, and this is the alias or artificially created tone. In fact, the tone folds back into the data about the 25 kHz point. A 26 kHz tone sampled at 50 kHz becomes indistinguishable from a 24 kHz tone, a 30 kHz tone becomes a 20 kHz tone, which are obnoxious forms of distortion. And there is no way to undo the damage. That is why steps must be taken to avoid aliasing from the beginning (see http://www.earlevel.com/Digital%20Audio/Aliasing.html; for a demonstration of sinusoidal aliasing, see http://www.dsptutor.freeuk.com/aliasing/AD102.html).

6.3.2. Vertical Discretization: Amplitude Quantization

The signal amplitude of the trace in Figure 6.3 is continuous. The amplitude values of the sampled data are whatever value the trace has at the particular sampling instant. The full-scale amplitude range of the signal in volts must be divided in a finite number of equal amplitude steps, Δq, to complete the discretization process. Thus, ***quantization*** *is the conver-*

sion of discrete-time, *continuous-amplitude signal to discrete-time*, *discrete-amplitude signal.* The calibrated trace in pressure units is expressed in counts (so many Δq steps), the number of which is determined by the number bits (or binary digits) used to encode the full-scale amplitude. The more bits (hardware) the smaller the Δq step and the more accurate the amplitude representation. This process leads to a vertical resolution that, as with the horizontal time axis, may miss small variations of the signal. The relationship used to establish the vertical step Δq is,

$$\Delta q = Q_{fs} \Big/ \left(2^b - 1\right) \tag{6.1}$$

where "b" stands for the number of bits and Q_{fs} is the full scale signal amplitude at the sampling time, which for design calculation purpose is taken as the maximum expected value. For example, if only 3 bits are to be used and the expected maximun systolic pressure is fixed at, say, 240 mmHg, Δq would be a voltage approximately equal to a pressure of 30 mmHg, obviously giving a rather coarse resolution. Three bits can code eight unique binary numbers (000, 001, 010, 011, 100, 101, 110, 111), or seven intervals, corresponding to eight pressures of (0, 30, 60, 90, 120, 150, 180, 210) mmHg levels. Observe that the last level of 240 mmHg is missed by the 3-bit code. Should we want that last level, we would have to calculate the vertical step using 320 mmHg as maximum, but in that case the resolution would worsen to 40 mmHg. Typically, analog to digital converters use a minimum of 8 bits for a resolution of one part in 255 leading to an accuracy of better than one percent. If the full scale voltage were 5 volts the system could resolve approximately 20 millivolts.

In short, the continuous signal recording space is covered with a grid formed by rectangles ($\Delta q \times \Delta t$); only corners of them closest to the continuous trace are recorded by the discretization process and, eventually, the trace may cross one of such corners. Obviously, those corners correspond to the intersections of the pre-established vertical and horizontal lines, which are the boundaries of Δt and Δq.

Exercise: Take any of the records displayed in Chapters 2 and 3 and design an adequate discretization grid. Repeat with several examples of your own and discuss.

6.4. What Do We Do With the Signals?

Why did we bother to digitize the signal? We need a digital representation of the signal to be able to use computer software to help us interpret the signal with a view toward better understanding of the physiological process(es) by which the signal was created.

Well, here is the signal, discretized, stored in a computer memory, and ready as raw material to be processed in order to get into the interpretation stage. Physiological signals are naturally presented as graphic non-explicit functions of time, that is, we never know the mathematical expression relating the event (action potential, blood pressure, substance concentration, or other) with the independent variable *t*. The discretization process introduced above is the beginning of a *time domain analysis*. Any other procedure applied to that collection of sequentially ordered numbers following a given mathematical rule (an *algorithm*), usually based on a mathematical model, either general or specific, enters into the actual analysis of that time course signal in order to search for significant information. Models, then, become an integral part of the analytical stage. Sometimes, they may be based on physical or physiological principles to set up the equations to start with or the latter can only be supported by empirical facts. However, in all cases, the experimentally recorded signal is the guiding light that keeps us tethered to the real live world. The physiological signal, however, is a complex event, many times periodic or quasi-periodic; at times, it shows itself as a more or less isolated burst or pulse (so increasing its complexity), as the case is with nerve or muscle action potentials or with some neuroendocrine hormones. As a consequence, frequency domain analysis, such as Fourier's, teaches us that complex physiological signals may be decomposed into a series of simple sinusoids; such techniques are borrowed from the engineering sciences and permit breaking up of the ECG, EEG, EMG and the like. Mathematical models become also an integral part of this *frequency domain analysis*. In the end, both types of analysis should produce similar or equal results. This subject belongs to a more advanced course called Physiological Control Systems, where topics like linear and nonlinear modeling are dealt with.

In signal analysis the ingenuity, knowledge and experience of the interpreter are important. Classical methods, as convolution, Laplace trans-

forms, and Fourier methods, are used to analyze and represent biomedical signals. Linear systems are represented by transfer functions providing the basis for system identification in the time and frequency domains. Important analytical techniques include the Fast Fourier Transform (FFT) and z-transform. Statistical procedures are also used, such as autocorrelation or cross-correlation. One of the main characteristics of real-life signals is their non-stationary, multicomponent nature and, many times, conventional methods fail to fully analyze these signals; investigators have used a joint time-frequency analysis method like the Wigner distributon to better reveal information contained in recordable biological signals.

Hence, one first and simple conclusion states that, to help in reading the signal, **processing of biomedical signals refers to the application of signal processing methods to biomedical signals**. It even sounds naïve and repetitious. A well-known sequence of processing steps for signal analysis is: (1) artifact rejection; (2) feature extraction; and (3) pattern recognition.

A second simple conclusion says that **any and all possible processing algorithms may be used**. In other words, do not self-inhibit if an apparent crazy idea for analysis suddenly pops up.

Obviously, good and sensible biomedical signal processing requires an understanding of the needs (e.g., biomedical and clinical requirements) and adequate selection and application of suitable methods to meet these needs. Within the rationale for biomedical signal processing, we should mention signal acquisition and signal processing to extract *a priori* desired information aiming at interpreting the nature of a physiological process based either on (a) observation of a signal (explorative nature), or (b) observation of how the process alters the characteristics of a signal (say, by monitoring a change of a predefined characteristic).

Goals for biomedical signal processing can be summarized as follows:

1. Quantification and compensation for the effects of measuring devices and noise on signal;
2. Identification and separation of desired and unwanted components of a signal;
3. Uncovering the nature of phenomena responsible for generating the signal on the basis of the analysis of the signal characteristics;

4. Eventually leading to modelling, but often more pragmatic than pure modelling.
5. Test the understanding one gains via signal analysis by trying to predict the effect on the signal of an experimental manipulation

Student task: Find the definition of *algorithm*. Dissect into its respective simple operations the algorithm of multiplication and the algorithm of división. Etymologically, the word derives from arabic.

Another area of interest into which signal processing can bring light is **biomedical signal classification**, based on signal characteristics, signal source, and applications. In this respect, signals may be deterministic (accurately described mathematically and rather predictable), periodic or almost periodic, transient, stationary (statistical properties do not change over time), stochastic (defined by their statistical distribution), ergodic (statistical properties may be computed along time distributions). Some people claim, for example, that all real biosignals may be considered stochastic, a helpful concept indeed when trying to model, for example, the EEG.

Student study subject: Dig a little deeper into the concepts of deterministic, stochastic and ergodic signals. Search for their essential characteristics. They have complex definitions, even if one tries simple versions. For example, "a phenomenon is stochastic (random) in nature if it obeys the laws of probability", or else, "a stochastic process is a statistical process involving a number of random variables depending on a variable parameter (which is usually time)". For ergodic you will run into definitions such as "an attribute of stochastic systems; generally, a system that tends in probability to a limiting form that is independent of the initial conditions". However, when going into mathematical definitions, things become rather complex and should be better left for the specialist unless the student decides to put considerable effort and time in the subject. Related words are haphazardness, randomness and noise, all of them always associated with "lacking any predictable order or plan".

6.5. Final Remarks and Conclusions

The biomedical signal must be read, must be understood and must be interpreted in its relationship with the pathophysiological event. This is the main message of the chapter. Noise, interference and/or artifacts have to be identified and removed; otherwise, we run the risk of attaching biological meaning to a non-existent event. Techniques and procedures to

better read the signal, starting with simple visual inspection up to using the most sophisticated algorithm based perhaps on complex mathematical models and assumptions, belong to the area of *signal processing*. You are going to take courses devoted to signal processing and this chapter should offer you a platform with which to begin.

Feature extraction and pattern detection are still the beginning and main background of the interpreter. Thereafter, armed with his/her personal computer, the signal must be discretized, as explained above. Usually, this *acquisition process* is performed by analog-to-digital converters (A/D converters) to finally enter into the processing proper of the signal, as outlined in section 6.4 above. But remember, in the end interpretation is made only by the human being.

Among other points to mention about biological signals is the characteristic spontaneous and huge amount of variability in amplitude and frequency. The so-called physiological periodic signals are not periodic; at best they are quasi-periodic, for the ECG, blood pressure, and respiration, for example, change frequency and also amplitude wih time. Other physiological events are more variable and, in some cases, rather difficult to predict. Very likely, such variability contains information of potential clinical value. For that purpose, special numerical techniques have been developed and are constantly under research, most of them of the nonlinear type (Cerutti, 2002). This area offers attractive possibilities for the young investigator.

Ron S. Leder is a native of New York City. He obtained the PhD degree in biomedical engineering from Wisconsin University, Madison, WI, under the direction of Dr. John Webster. His has experience in data acquisition and signal processing where he contributed with several papers. Besides, he has been an active participant in the IEEE/EMBS. Currently, he is with the Department of Biomedical Engineering at the *Universidad Autónoma Metropolitana-Iztapalapa* in Mexico, DF.

Chapter 7

Feedback: The Need of Mathematical Models

Theoretical studies in biology and physiology may sound weird, especially to the young student, as they frequently involve appalling oversimplifications. However, mathematical physics also makes use of amazing oversimplifications. The works of Maxwell and Einstein once were considered purely speculative. Yet it was Maxwell's work that made our present day electrical industry possible. It was Einstein's speculations about four-dimensional space-time that made the atomic age possible. The thousands of ungenerous patents that followed them, mostly buried in oblivion and many giving healthy profits to their inventors and more comfortable life to the users, represent the practical arm. Biology as yet has not had its Maxwell or its Einstein ... but they may come ... perhaps sooner than expected ... perhaps they are in front of these modest lines. Paraphrased from Nicholas Rashevsky (*1964*).

7.1. Introduction

The line of thought of this undergraduate introductory textbook started with definitions and general concepts of bioengineering/biomedical engineering (Chapter 1), it proceeded to describe and explain the sources of physiological signals (Chapter 2), entering thereafter into the signals themselves (Chapter 3). Chapter 4 dealt with the way a signal is picked up while Chapter 5 went into considerable details of the first step in signal processing, that is, amplification, acting as some kind of magnifying lens so that the interpreter (Chapter 6) is better equipped to read the signal making use of a number of algorithms mostly processed by computer aid. If one takes a look at the recording channel, we easily realize that it has given us the blueprint for the book. Now feedback is needed, as the stage to test our reading and interpretation, where mathematical models play a central role trying to improve the quantification process. Many

times the model must be disposed of or modified because its results did not fit acceptably the real situation, but this is a calculated risk and should not bring discouragement. Thus, this Chapter 7 signals the time to look back, review, revise, study, search, and experiment again, initiating an endless feedback loop to improve knowledge and understanding. The chapter speaks of the need of mathematical models but it does not contain mathematics, as the preceding one spoke of signal processing but did not enter into the subject matter proper because that is the task of future courses. The student now is getting an overall view of bioengineering, as anticipated in the introduction of the book.

Ever since the early 17th century, there has been a growing explosion of the physical sciences. Biology could not stay as a mere observer or as an outsider to such fascinating phenomenon and it started to try enlightening its own problems by adopting physical and mathematical perspectives, so definitely entering into the quantification stage. Descartes' mechanical views were a beginning followed by several other scientists, even though earlier and more isolated attempts can be found in the history of science. Electricity and magnetism also influenced experimental and theoretical biology for long periods. During the 20th century, the work of Lotka represents another set of physical arguments and analogies to produce biologically significant understanding of several biological events. Differential equations and linear algebra, field and bifurcation theories, computation and game theories, differential geometry and group theory, are background knowledge applicable to modern theoretical biology (Lumsen, Brandts & Trainor, 1997), which should be considered a relatively new branch slowly but firmly setting its feet in the scientific scenario.

If there is a good theory, predictions can be made, and that means better knowledge of the subject in question. Thus, in experimental research the mere collection of facts is not sufficient. The Human Genome Project dramatically exemplifies the concept; in it, the amount of data has reached huge quantities while quantitative models and clear interpretation still stay quite behind and definitely not keeping pace, at least for the time being. Perhaps we should credit Nicholas Rashevsky, a Russian born physicist, as the founder and propeller of modern Mathematical Biology (also called Biomathematics), when in the 1930–1940 decades cre-

ated a strong working group, later a department (which was called "committee"), at the University of Chicago. Besides, he founded the B*ulletin of Mathematical Biology*, still published, where outstanding contributions saw the light. The amount of theoretical material collected in it over more than 65 years is, no doubt, a consultation and inspirational source.

One of the purposes of a theory is to determine certain quantities that cannot be measured directly, thus, it may provide an indirect measurement. Another aim of a theory is to explain a given set of facts answering question such as "how does it happen" and "why does it happen", one step ahead of the simpler "what happens", usually explained by the experimental act. A third and more provocative outcome of a theory is the suggestion of further experimental studies, perhaps even predicting the existence of new events or effects. However, no theory in biology or physiology or medicine has been developed to such an extent as the theory of celestial mechanics, but there is no reason to doubt, however, that predictions of this type will eventually be made. Advances in the last 20 years have shown possible avenues and anticipated significant discoveries. Nontheless, even though optimism well dresses up a good attitude, always remember that no theory can describe a set of phenomena in its entire complexity and no theory should be condemned because it explains certain phenomena only approximately. The requirement that a theory should be realistic in all its aspects is itself unrealistic. Such a requirement is against the very nature of theoretical thinking. When a theory has stood the acid test of numerous successful comparisons with experiments or observations, we can rely on it within a wide range, even though this range is always limited. Models born in the biomathematics cradle, old or new, whatever simple or complex, linear, non-linear or statistical, offer the background theoretical material for the signal processing employed by the interpreter (Rashevsky, 1938).

7.2. Linear versus Non-Linear Models

A system is termed linear when its output signal exactly reproduces the input signal, except for a magnifying or demagnifying constant factor. Such factor takes many times the form of a relationship called the trans-

fer function which characterizes the system. If the output signal is distorted, the system becomes non-linear. The system is, in any case, deterministic, because given an input there is always an output.
A system is classified as statistical when, after an input signal is applied, the output signal shows up with a given probability; thus, it may not appear. If it shows up, it may be either linear or non-linear, but it is not deterministic as above.
A system is called oscillatory when, given an input, the output is periodic with time; it may be either a sustained or a fading oscillation. Obviously, the system is non-linear, but it may be treated as linear; it is also deterministic. However, some systems may not even need an input to produce an oscillation; thus, they are true biological oscillators (as, for example, the heart). Structurally and functionally, periodicity is a frequent characteristic in biology and physiology; striated muscle, the Ranvier nodes sequence in mielinated nerves, the repetition of bone laminae, the periodic DNA organization, are clear examples of the former spatial type, while the latter shows itself with respect to time, as in the heart and respiration, in the cerebral cortex, sexual cycles, circadian rhythms, migrational seasons in birds and fish, and others.
The definitions given above are oversimplified in their wordings but briefly summarize concepts frequently handled in modeling and signal processing. The concepts of linearity and non-linearity are at times a bit elusive and may lead to long discussions, perhaps more philosophical in themselves than real. Contributions amount to thousands of papers if somewhat arbitrarily we take 1930 as the starting modern date. Textbooks are also many (perhaps 40 or 50), and some of them must be referred to because they contain fundamental and basic material.

Something about the available literature of models in biology: An early treatise was published in 1938 (Rashevsky, with updated editions in 1948 and 1960). It truly is an almost (or partially) successful attempt of building-up a systematic mathematical biology, similar in its structure and aims to mathematical physics. The book looks for physical interpretation of the biological phenomenon, sometimes with a grain of naivety when looked at from a 50-years after perspective, but straight and highly stimulating, even in controversial subjects. A simplified, more accessible and practical version of biomedical mathematics appeared a few years later (Rashevsky, 1964); it is recommended for the beginner, for it has good examples in the cardiovascular, endocrine and nervous systems, including pharmacological applications. A posthumous, more advanced and rather speculative piece saw the light the

year of Rashevsky's death (Rashevsky, 1972). As one of his disciples said, "he made a step towards unity of science by the construction of the conceptual structure which he called 'world set', of which physics, biology and sociology are subsets". Defares and Sneddon (1961) produced a nice didactic book very much in the line of thought established by Rashevsky but of a more elementary level. Its applications section is also recommended for the beginner. Thereafter, we find a series of more engineering oriented books that can be considered as the biological control branch of engineering; they contain several excellent illustrating examples of different levels of complexity. Mostly, they are based on linear models with the Laplace Transform being the principal mathematical tool (Riggs, 1963, 1970; Grodins, 1963; Milhorn, 1966; Milsum, 1966). Curiously, Rashevsky's contributions are not recognized in this literature, perhaps because he and his school were considered too theoretical. Bailey (1967) is less engineering-like, somewhat more mathematically biased and with a hint of clinical engineering; it is one of the very few referring to Rashevsky's. Blesser (1969) produced a good integrative attempt where very basic definitions are introduced. Schwan (1969), Clynes and Milsum (1970), and Brown, Jacobs and Stark, (1971) inaugurated the edited multiauthored work covering selected lists of subjects and, thus, as any text of this kind, lacking a comprehensive nature although with outstanding chapters while Jones (1973), Gold (1977) and Finkelstein and Carson (1979), already well in the 70's, went back to a more integrative bioengineering look. Edelstein–Keshet (1988) aimed at presenting instances of interaction between biology and mathematics following a line of thought that encompasses more biology at large than the way we have focused it herein; its contents must be considered as of advanced level. There are books dealing with models in specific areas, as compartmental analysis (Rescigno & Segre, 1966; Welch, Potchen & Welch, 1972; Jacquez, 1972), which represents an important concept and rather successful analytical technique. Transport phenomena in the cardiovascular system, —say, within blood vessels or across capillaries or among body compartments—, were well handled by Middleman (1972) and, thus, it becomes a good complement and expansion to deepen the pure descriptive physiological event. Mechanical systems of different types (Ghista, 1979) and the organs of equilibrium, as superb control devices (Valentinuzzi, 1980; Highstein, Cohen & Büttner–Ennever, 1996), are appealing and attractive for the advanced student. Other books have unique or more focalized approaches, as Lieberstein (1973) or Schneck (1990), and are in some respects of a higher level. This brief review does not pretend to be exhaustive since other texts perhaps could have been included.

Characterization of a linear behavior is usually carried out in any of four equivalent ways: impulse response, step response, frequency response, and transfer function. All require an input or forcing function to elicit an output or response (Glantz, 1979). As time advanced, the nonlinear approach gained more and more predominance, especially because computer power and versatility increased significantly and at tremendous pace making tractable problems that before could only be set as a challenge to solve in the future; well, we seem to already be witnessing the

future. Khoo (2000) is a relatively recent example that didactically proceeds from the simple traditional linear approach to the more complex and sophisticated non-linear case.

Is linearity bad or something that we should be ashamed of? Not at all, and do not shy away from it; quoting and paraphrasing from Grodins (1963): "Linear differential equations are usually backing up the linear systems. These equations constitute a minute fraction of the totality of differential equations and they are the only ones for which a complete analytical theory exists and for which general analytical solutions can be obtained. It is fortunate that many physical systems are sufficiently linear to permit their satisfactory description by linear differential equations." Besides, most of the times the parameters involved are physically concentrated in space; the so-called lumped-constant systems. In other words, time is the only variable. Space coordinates are not considered and, if they are, partial differential equations must be used. This is the case of the distributed-parameter system. However, non-linearity is the rule and not the exception in biology. As first approximation, linear models work surprisingly well in many instances, but one can find other instances in which the non-linear features are critical for the proper understanding and description of the system. The principle of superposition is not applicable in non-linear systems and that appears as clear disadvantage (Khoo, 2000). The impulse response is no longer usable as local solutions cannot be extrapolated to global scale. Sometimes, non-linearity can be treated by means of the piece-wise analysis, that is, any curve is always decomposable in sufficiently small straight lines; in that way, addition of several linear equations can account for the phenomenon.

Let us now suggest a series of illustrating tasks or exercises that the student should try to work out checking, thereafter, with the indicated bibliography. Some other simple examples have previously been developed in Chapter 2, as in the case of the kidney equations and the calculation of total and partial blood flow.

Student task 1: Read, study and discuss linear and non-linear models of lung mechanics as described by Khoo (2000) in its Chapter 9, pp 229–230. It briefly shows how in some instances the linear model fails or is not accurate enough.

Student task 2: The glucose concentration-time curve in blood during continuous intrave-

nous injection of glucose represents a clear case of clinical interest. Assuming that the concentration rate increases with the rate of infusion and decreases by regular elimination, a simple linear differential equation is obtained which leads to the typical exponential growth. See Defares & Sneddon, 1961, Chapter X, pp 527–9.

Student task 3: The form of the arterial pulse can also be modeled with a linear differential equation considering first the systolic phase and, thereafter, the diastolic run-off. See Defares & Sneddon, 1961, Chapter X, pp 529–533.

Student task 4: The three exercises above serve as introduction to the uptake of radioactive potassium by human erythrocytes, growth of an isolated population and growth of a non-isolated one. See also Defares & Sneddon, 1961, Chapter X, pp 533–540.

Student task 5: Blair's Theory of nervous excitation. By now, the student is sharp enough to go into this a little more complicated example. The source is the same, that is, Defares & Sneddon, 1961, Chapter X, pp 593–597. The theory has been largely superseded by new knowledge, however, it represents an excellent step in the teaching-learning process. The motivated student, already full of enthusiasm, would not waste his/her time if the whole Chapter X in Defares & Sneddon's book is taken up; if not, he/she may proceed to the next task.

Student task 6: The reflex regulation of pupil area by the iris is a typical example of a biological control system. For this subject see now Chapter 2, pp 24–49, in Milsum, 1966.

Student task 7: About the semicircular canals. Higher animals possess sensitive transducers of translational and rotational skull motion, including Coriolis effects. They have profound actions on the postural control of the body. See Chapter 8, pp 186–191, in Milsum, 1966. It is a beautiful example of second-order linear lumped model. The student should also find out about Coriolis acceleration and the meaning of second order systems. Experiment on yourself the Coriolis effect: Stand up, start rotating around your longitudinal axis (feet to head), as when dancing. After a few turns and having reached a little speed, tilt your head to one side. What do you feel? Be careful, because the dizziness may even cause your fall, especially if you are not fit and well trained. After recovery, explain. Who was Coriolis? For a more advanced description, see Valentinuzzi (1980).

7.3. Characteristics of a System Model

Linearity (or non-linearity) could have been listed as another characteristic of a system, perhaps the most important of them all. However, we preferred to introduce it as a classification criterion. Once the system is placed in the proper group, we have to determine the other different characteristics in order to better predict its performance. Let us define and briefly discuss them:

– *Stability*

This concept is more or less intrinsic to us. Anyone knows that a normal person stands and walks with stability while a one year old toddler is unstable in its standing and walking trials and, very likely, will eventually fall down; thus, it is unstable. However, as engineers, we need to do better than this. A system may show oscillations that fade out either as underdamped, overdamped or critically damped responses after stimulation with an impulse. All three represent stable cases. A fourth situation may give off an undamped or sustained oscillation termed conditionally or marginally stable. Still another situation may occur when the impulse stimulus elicits a growing oscillation, that is, the system output never returns to its original operating point prior to stimulation. The latter is clearly unstable. Hence, after the preceding examples, we can say that a stable dynamic system is one that will respond to a bounded input with a bounded response (Khoo, 2000). All these cases are based on a well-known mathematical model and are unequivocally described in those terms.

– *Identifiability*

Once again, the common language understanding of the word offers a useful hint: there is need to know the identity of what we are studying, as the police searches for data to clearly establish who the suspect is; and again, something more quantitative is required by the bioengineer.

First type of identification: Say that the equations modeling a given system are known based on physical or physiological principles and that, after a specific input function, the output response is predicted by means of those equations. This type of analysis is called the forward or prediction problem (Khoo, 2000). Predictions tell whether the model provides an accurate description of the process under study, especially if the results are compared against actual experimental records.

Second type of identification: Greater challenge is posed by the inverse problem, which arises when, having a model and the measured output, the input is not observable and must be deduced. Often it requires the application of deconvolution (opposite of convolution).

Third type of identification: It is the most difficult and elusive and usually referred to directly as the identifiability of the system. This is the case of systems about which knowledge is very limited or when only

some knowledge is available. There is no established strategy for these situations, not easy to handle, and experience plays a significant role; specialists speak often of black box or gray box approaches, according to the actual level of knowledge of the system one has. Identification is subject matter of advanced courses.

– *Optimality*

Before fully accepting a model after preliminary tests, its response to a given input must be compared to the response of the actual physiological system to the same input. Thus, a criterion is to be chosen in order to evaluate the goodness of fit between the two time series (measured and calculated or real and theoretical). In other words, by minimizing the deviations between both series we are actually optimizing parameters.

Study subject: Find the precise definitions of damped, undamped, overdamped, underdamped, and critically damped oscillations. Recall the mechanical system formed by a spring, a mass and a friction.

Study subject: Find at least one example in the literature in which the model parameters have a one-to-one correspondence with the underlying physiological entities. Black box and gray box approaches may not show a similar correspondence. Find out what convolution and deconvolution are, as mathematical operations.

Student task 8: Neuronal dynamics and the Hodgkin–Huxley Model. See Chapter 9, pp 257–260, in Khoo, 2000.

7.4. Partial Final Remarks

The chapter is just an outline of what the modeling stage is or may be. Lots of questions must and should be flooding the young mind, feeling perhaps some frustration and even getting angry with the author while asking “why not more, why not getting deep into the subject”. The answer is simple, it takes time, you are just at the beginning and this is the primer to start. Besides, there are excellent texts fully devoted to mathematical models in biology and their computer implementation.

Chapter 8

Rounding Up and Looking Ahead

As already mentioned, “observe, think, and observe again”, used to say Dr. J.J. Izquierdo, an old well-reputed professor of physiology in Mexico back in the 1960’s ... and physiological signals are one way to observe; reading, interpreting and modeling those signals represent thinking about them, while modeling means also looking back to observe again, we may add. Finally, it is always good to round up and try to look ahead.

In this book we have underlined some historical aspects, including a few anecdotal pieces, because that kind of information often offers insights of the many components that may shape up the development of a given subject. We hope that objective was reasonably fulfilled and was enjoyable for the reader. Trying to look ahead is also a good exercise, although riskier for predictions may fail. Many subjects were consciously left out; otherwise, the book would have become too extensive and almost unreadable. Besides, we wanted to keep versatility and flexibility along with an agile rhythm.

There are areas that, without doubt, represent emerging and challenging fields, as adaptive control, micro and nano-technologies, which also relate to biomaterials, tissue engineering, cell engineering, biomedical implants and bones. Nano-technology is a rapidly growing field that will affect us all at some stage in our lives whether we know it or not. Intensive research is currently under way looking at ways in which our quality of life can be improved through the engineering of bone, tissue and organs. Some applications are currently available, as for example cochlear implants to restore hearing. More in the future are opto-retinal implants for the blind, artificial skin, tissue reconstruction, tissues and organs artificially grown on nano-patterned scaffolds. All this finds application in wound healing, regeneration of tissues and replacement of organs that have failed through disease or old age.

Tissue engineering at the nano-scale level is leading to the development of viable substitutes, which can restore, maintain or improve the function of human tissues. Regenerating tissue can be achieved in several ways, e.g., by using biomaterials to convey signals to surrounding tissues to recruit cells that promote inherent regeneration or by using cells and a biomaterial scaffold to act as a framework for developing tissues. As an example, selected cells can be harvested from a patient and can be modified at the cellular level to prepare it for later transportation. Small biopsies from uninfected sites are used to isolate tissue specific cells which can then be encouraged to multiply. The cells can then be re-transplanted directly or are combined with an appropriate matrix for transplanting. A good source to consult is the special issue on this subject published by *IEEE Engineering in Medicine and Biology Magazine* (vol 22, N°5, September/October, 2003).

The more we know about how cells work, the more possible it becomes to engineer cell to do things that once sounded like science fiction. This is *cell engineering*; in it, chemistry, nano-technology and materials science are all-important tools used to study and control how cells behave. One powerful way of finding out how cells work is by re-designing the extra-cellular matrix, the scaffold that surrounds a cell acting as a glue that sticks cells together. Cells react on the micro and nano-scale to the shape and chemistry of their surrounding environment. Nano-scale grooves, no wider then the cells themselves, can act as templates causing cells to line up. Once cells can be induced to organize themselves, even more possibilities open up, ranging from wound repair to the future vision of growing whole organs. Artery replacement is also possible using structures made form natural polymer to act as a scaffold around a patients natural artery, which given the appropriate stimulus, can be encouraged to regenerate. These scaffolds being biodegradable melt away after the regenerated artery is in place.

Implants have been traditionally used, say, in artery aneurisms, in dentistry, in vision to replace for example the lens, and in other medical applications. Countless patients have been benefited in the last three decades at least. The area is in constant expansion. A problem with medical implants is their acceptance by surrounding tissues. There are already specially developed coatings using nano-scale techniques and nano-

textured surfaces to create a cell friendly environment which encourages tissue to bond to the implant. Consequently, the implant will last longer and feel more comfortable.

Nano-technology can also take the potential for the rehabilitation of the infirm and elderly to a totally new domain. The domain of intelligent learning prosthetic devices. Some of the devices presently being researched include retinal implants. Many people suffer of degeneration of the retina with old age. One solution is to use a photo-sensor array which will detect incoming light and connect it to a signal processor. A signal can then be transmitted to an implanted receiver at the retinal interface which is connected via micro-contact to the retinal nerve.

In present day cochlear implants, the connection between the sound amplifier and the inner ear, consists of less than 22 electrodes which do not even make direct contact with the ganglions in the ear. Now, an implanted transducer can connect to one of the delicate bones in the inner ear, vibrating the bone to create sound

Bone supports our body as the hard main material of the skeleton. Nanotechnolgy is being used in teeth and bone replacements copying the way nature itself lays down minerals. This process is called biomimicry; it is already the basis of new tough and light materials for bullet proof vests and other defense applications. The use of nano-patterned polymers could eliminate the long recovery times, scarring and infection associated with bone grafts. Researchers hope to use this technique to grow adult stem cells that will turn into bone. Once the process of growing tissue on patterned scaffolding is perfected, nano-structured devices can be attached to further improve bone growth rates and reduce healing time. The devices have electrodes to provide an electric current which has been shown to stimulate bone growth as well as tiny channels along which controlled doses of the kind of proteins which have been shown to enhance cell growth can be pumped (see *What is nanotechnology*, CD-ROM prepared by the Institute of Nanotechnology, USA, or check its INTERNET site).

Bioinformatics is definitely on its way, including disciplines as biological structure informatics, computational biology, microarrays, genomic ontologies, genomics, neuroinformatics, pharmacogenomics and proteomics. Within bioinformatics we find also clinical informatics, with

clinical systems in high intensity care, disease management, e-health and clinical communication, evaluation of health information systems, health data warehousing, health information systems, integrated health and financial systems, patient records, public health informatics. Perhaps in this listing there is some overlapping, but time will certainly settle things down as experience and knowledge are gained.

Education and Training rises as an everlasting need, changing and evolving according to new technologies: Computer-assisted medical education, consumer health information, e-learning or distance learning, library information systems (especially under the pressure of the virtual journals), medical informatics teaching, patient education and self-care, nursing informatics, and professional education.

Human information processing and organizational behavior lead to subjects such as cognitive models, data visualization, natural language understanding and text generation, human factors and user interface, human-computer interaction, models of social and organizational behavior, and natural language processing. Perhaps, some of these subjects fall somewhat outside the bioengineering range, but it happens that its reach is so wide and long that easily one finds invading other fields.

Imaging and signal analysis are permanent challenging needs. The image stands as the best means for conveying complex information because the eyes constitute the best input channel to the brain. Thus, image processing and transmission, image recognition, registration and segmentation methods, imaging and signal standards, knowledge representation and ontologies for imaging, model-based imaging, signal processing and transmission, virtual reality and active vision methods appear as well-needed attractors for the creative minds.

Innovative technologies in health care include computer-communication infrastructures, mobile computing and communication, portable patient records, security and data protection, telemedicine, wireless applications and handheld devices. The subjects are obviously very pragmatic and medically oriented.

Finally, we should mention knowledge management which considers areas such as automated learning and discovery, clinical guidelines and protocols, controlled terminology, vocabularies and ontologies, intelli-

gent data analysis, decision support systems, neural network techniques, and pattern recognition and classification.

The list is long and very likely there are missing subjects or others that should not have been included. Ethics is the largest and most important; thus, the last paragraph is devoted to it (Valentinuzzi, 2004). What about the scientist's ethics when involved in research and development of new weapons? Shall we follow the stand suggested by Tom Lehrer in the lyrics of his popular 1960's song?

The rocket goes up, *who knows where it comes down*? *That's not my department*, *says Werner von Braun*.

The XXth Century started with about 1,000 million inhabitants, the New XXIst Century is beginning with about 6,000 million people, and we know that before 2050 the number will climb up to may be 12,000 millions. Besides, in less than 30 years there will be severe limitations in the availability of fresh potable water. Science in general, technology, physiology and bioengineering in particular: what could they do or offer to solve or alleviate the pressures emerging from such demands? What can we, bioengineers, do to actually start a new beginning for human health? No doubt, hard and demanding is the challenge. Certainly, more and better weapons and more powerful armies are not the proper and sensible way. Perhaps, we had better remember that before anything else, we are simply men and women, that being is much better than having and that independently of how much richness, or power, or knowledge, or worldly glories we might collect and store, the really important and significant fact is and will be how much we love and how much we have loved. And the scientific endeavor calls for a lot of love.

My basic education —strongly founded in humanism— and graduate engineering formative years took place in Argentina. After working in telecommunications for 5 years, I went to the USA, where I spent 10 years and had the opportunity of studying and going into research. One of my daughters and her children live there. But I believed my possible modest contributions were more useful in my native country, a land so many times socially and economically stricken. Sometimes, when traveling through the Calchaqui Valleys route (Province of Tucumán, Argentina), at about 9,000 feet of altitude, I stop a few minutes to look at the children in that small *Infiernillo Elementary School*, isolated, surrounded

by winds and majestic silent huge mountains, far away from urban, selfish and arrogant centers, and ask to myself, what world and future do they face? Meanwhile, I admire the generous courage of their almost unsupported teachers. Certainly, if any of those children were to require, say, a cardiac pacemaker (a biomedical engineering product we are proud of), even the cheapest model would not be affordable for his/her family. How many children like them are there all over the world? Is that not our department? Think it over …

And this is all I have for the time being. I wish you enjoyed it and feel encouraged to proceed successfully in your career displaying a cheerful smiling face like these fellows below.

References

AAMI (1993) American Association for Medical Instrumentation; Standards and Recommended Practices: Biomedical Equipment. Vol 2, 4th ed, Arlington, Virginia, USA.

Albrecht U (2002) Invited review: Regulation of mammalian circadian clock genes. Journal Applied Physiology, 92:1348–1355.

Arendt J (1995) Melatonin and the Mammalian Pineal Gland. Chapman and Hall, London.

Armentano RL, Cabrera Fischer EI, Levenson J, Barra JG, Breitbart GJ, Pichel RH, Simon A (1990) Effects of muscarinic and β-adrenergic blockade on the aortic elastic response to epinephrine to induced acute hypertension on conscious dogs. American Journal of Hypertension, 3:476–481.

Armentano RL, Levenson J, Barra JG, Cabrera Fischer EI, Breitbart GJ, Pichel RH, Simon A (1991) Assessment of elastin and collagen contribution to aortic elasticity in conscious dogs. American J. Physiol 260 (Heart Circ Physiol 29): H1870–1877.

Bailey NTJ (1967) The Mathematical Approach to Biology and Medicine, John Wiley, New York, 296 pp. It refers to some of Rashevsky's contributions.

Bartsch C, Bartsch H, Blask DE, Cardinali DP, Hrushesky WJM, Mecke D (eds) (2001) The Pineal Gland and Cancer. Springer Verlag, Berlin.

Baylis DH, Padfield PL (eds) (1985) The Posterior Pituitary. Marcel Dekker, Inc., New York, NY, USA.

Bishop BP (1997) Mentoring: A position of responsibility. The Physiologist, 10(2):114–5.

Blesser WB (1969) A Systems Approach to Biomedicine, MacGraw–Hill, New York, USA, 615 pp.

Brazier MAB (1959) The historical development of neurophysiology. Chapter 1 in Handbook of Physiology, Section 1, Neurophysiology, volume 1, American Physiological Society, Washington DC, USA.

Brigell M, Kaufman DI, Bobak P, Beydoun A (1994) The pattern visual evoked potential. A multicenter study using standardized techniques. Documenta Ophthalmologica (Dordrecht), 86(1): 65–79.

Bronner F, Sammon PJ, Stacey RE, Shah BG (1967) Role of thyrocalcitonin in the regulation of the blood calcium level. Biochemical Medicine, 1:261–279.

Bronzino JD (1995) The Biomedical Engineering Handbook. CRC Press, Inc. Salem, MA USA.

Brown JHU, Jacobs JE, Stark L (eds) (1971) Biomedical Engineering, Davis Co., Philadelphia, USA, 435 pp.

Burger HC (1968) Heart and Vector, Philips Technical Library, 142 pp.

Burr Brown Corp (1994) Burr–Brown Integrated Circuits Data Book. Linear Products. Tucson, Arizona, USA.

Burton, AC (1957) The importance of the shape and size of the heart. Am Heart J, 54(6):801–810.

Cabrera Fischer EI, Levenson J, Barra, JG, Armentano RL, Pichel RH, Simon A (1993) Preventive effect of chronic converting enzyme inhibition on aortic stiffening induced by renovascular hypertension in conscious dogs. Cardiovascular Research, 27:1039–1044.

Campbell NA, Reece JB, Mitchell LG (1999) Biology, Fifth Edition, Addison Wesley Longman, Inc, Menlo Park, CA, USA, 1175 pp + appendices.

Cardinali DP, Pévet P (1998) Basic aspects of melatonin action. Sleep Medical Review, 2:175–190.

Carneiro AA, Baffa O, Fernandes JP, Zago MA (2002) Theoretical evaluation of the susceptometric measurement of iron in human liver by four different susceptometers. Physiological Measurement, 23(4):683–693.

Carneiro AA, Baffa O, Oliveira RB (1999) Study of stomach motility using the relaxation of magnetic tracers. Phys Med Biol, 44:1691–1697.

Castleman R (1979) Digital Image Processing, Prentice–Hall, Englewood Cliffs, New Jersey.

Cerutti (2002) Guest editorial of a special issue devoted to Non-linear Processing of Biomedical Signals. Medical Engineering & Physics, vol 24, number 1.

Chalama V, Kim Y (1997) A methodology for evaluation of boundary detection algorithms on medical images. IEEE Trans Med Imaging, 16:642–652.

Chittaro L (2001) Information visualization and its application to medicine, Artificial Intelligence in Medicine, 22:81–88.

Christensen J (2001) The Motility of the Gastrointestinal Tract. The University of Iowa College of Medicine, Iowa City, Iowa. Available through an INTERNET website.

Cirillo VJ (1985) The suicide of Thomas Addison. Journal History Medicine and Allied Sciences, 40(2):214–5.

Clancy EA (1999) Electromyogram amplitude estimation with adaptive smoothing window length. IEEE Trans Biomed Eng, 46(6):717–729.

Clancy EA, Farry KA (2000) Adaptive whitening of the electromyogram to improve amplitude estimation. IEEE Trans Biomed Eng, 47(6):709–719.

Clancy EA, Hogan N (1994) Single site electromyograph amplitude estimation. IEEE Trans Biomed Eng, 41(2):159–167.

Clancy EA, Hogan N (1995) Multiple site electromyograph amplitude estimation. IEEE Trans Biomed Eng, 42(2):203–211.

Clancy EA, Hogan N (1997) Relating agonist-antagonist electromyograms to joint torque during isometric, quasi-isotonic, non-fatiguing contractions. IEEE Trans Biomed Eng, 44(10):1024–1028.

Clancy EA, Morin EL, Merletti R (2002) Sampling, noise reduction and amplitude estimation issues in surface electromyography. Journal Electromyography and Kinesiology, 12(1):1–16.

Clements JA (1957) Surface tension of lung extracts. Proc Soc Exp Biol Med, 95: 1170.

Clynes M, Milsum JH (eds) (1970) Biomedical Engineering Systems, McGraw–Hill, New York, 665 pp.

Cohen D (1968) Magnetoencephalography: Evidence of magnetic fields produced by alpha-rhythm currents. Science, 161:784–6.

Cohen D (1972) Magnetoencephalography: Detection of the brain's electrical activity with a superconducting magnetometer. Science, 175:664–6.

Cohen D (1999) Magnetoencephalography [review]. In Encyclopedia of Neuroscience, Adelman G (ed), Elsevier, vol.II, 1079–83

Cohen D, Chandler L (1969) Measurements and a simplified interpretation of magnetocardiograms from humans. Circulation, 39(3):395–402.

Cohen D, Cuffin BN (1983) Demonstration of useful differences between MEG and EEG. Electroenceph Clin Neurophysiol, 56: 38–51.

Cohen D, Cuffin BN (1991) EEG vs MEG localization accuracy: theory and experiment. Brain Topography, 4:95–103.

Cohen D, Cuffin BN, Yunokuchi K, Maniewski R, Purcell C, Cosgrove GR, Ives J, Kennedy JG, Schomer DL (1990) MEG versus EEG localization test using implanted sources in the human brain. Annals of Neurology, 28:811–17.

Cohen D, Givler E (1972) Magnetomyography: magnetic fields around the human body produced by skeletal muscles. Appl Phys Letters, 21(3):114–6.

Cohen D, Palti Y, Cuffin BN, Schmid SJ (1980) Magnetic fields produced by steady currents in the body. Proc Natl Acad Sci, 77:1447–51.

Cole KS (1982) Squid axon membrane: Impedance decrease to voltage clamp. Annual Review Neuroscience, 5:305–323.

Comroe JH (1971) Physiology of Respiration, Year Book Medical Publishers, Inc., USA, 245 pp.

Conn JW (1955) Primary aldosteronism, a new clinical syndrome. Presidential address to the Central Society Clinical Research, Journal Laboratory and Clinical Medicine, St. Louis, 45:3–17.

Coraboeuf E, Weidmann S (1949) Potentiels d'action du muscle cardiaque obtenues a l'aide de microelectrodes intracellulaires. Presence d'une inversion de potentiel. Comptes Rendues de la Societé de Biologie de Paris, 143:1360–1.

Cormack DH (1998) Clinically Integrated Histology. Lippincott–Raven, Philadelphia, USA.

Davenport HW (1977) Physiology of the Digestive Tract. Year Book Medical Publishers, Inc., Chicago, IL, USA, 4th ed, 220 pp.

Davey CL, Davey HM, Kell DB (1993) Introduction to the dielectric estimation of cellular biomass in real time, with special emphasis on measurements at high volume fractions. Analytica Chimica Acta, 279:155–161.

Davey, C., Kell, DB (1998) The influence of electrode polarisation on dielectric spectra, with special reference to capacitive biomass measurements. II. Reduction in the contribution of electrode polarisation to dielectric spectra using a two-frequency method. Bioelectrochem Bioenerg, 46:105–114.

de Vroomen M, Steendijk P, Lopes Cardozo RH, Brouwrs HHA, van Bel F, Baan J (2001) Enhanced systolic function of the right ventricle during respiratory distress syndrome in newborn lambs. Am J Physiol Heart Circ Physiol 280:H392–H4000.

Defares JG, Sneddon IN (1961) The Mathematics of Medicine and Biology, Year Book Medical Publishers, Inc, 663 pp.

Di Francesco D, Noble, DA (1985) A model of the cardiac electrical activity incorporating ionic pumps and concentration changes. Simulations of ionic currents and concentration changes. Phil Trans Royal Soc Lond, B307, 353–398.

Donnan FG (1925–6) The theory of membrane equilibria, Chemical Reviews, 1–2:73–90.

Dostál J (1993) Operational Amplifiers. Butterworth–Heinemann, Boston USA, 2nd ed.

Ebling FJG, Eady RAJ, Leigh IM (1992) Anatomy and organization of the human skin. In Textbook of Dermatology, 5th ed, RH Champion, JL Burton, FJG Ebling, eds, p. 3160, Blackwell, London.

Edelberg R (1968) Biopotentials from the skin surface: The hydration effect. Ann New York Acad Sci, 148: 252–62.

Edelstein–Keshet L (1988) Mathematical Models in Biology, Random House, New York, USA, 586 pp.

Einthoven W (1903) Ein neues Galvanometer. Ann Phys, 4(suppl.):1059–1071.

Emery B R, Miller R L and Carrell D T (2001) Hamster oocyte membrane potential and ion permeability vary with preantral cumulus cell attachment and developmental stage. BMC Developmental Biology 1:14. The electronic version of this article is the complete one and can be found online at http://www.biomedcentral.com/1471–213X/1/14. Published Oct 10, 2001, Licensee BioMed Central Ltd. Verbatim copying and redistribution of this article are permitted in any medium for any non-commercial purpose, provided this notice is preserved.

Erlanger J, Gasser HS (1937) Electrical Signs of Nervous Activity. University of Pennsylvania Press, Philadelphia, USA.

Eucker S, Lisauskas JB, Singh J, Kovács SJ (1999) Human cardiac physiology in the phase plane: The ejection fraction to limit cycle area relation. Proc. BMES/EMBS Conference, Oct 13–16, Atlanta, GA, p 229 (in CD).

Evans AF, Geddes LA (1969) A vasomotor response to increased cerebrospinal fluid in the spinal animal. Cardiovascular Research Center Bulletin (Houston, TX), 7(3):100–110.

Feldman MD, Erikson JM, Mao Y, Korcarz CE, Lang RM, Freeman GL (2000) Validation of a mouse conductance system to determine LV volume: comparison to echocardiography and crystals. Am J Physiol Heart Circ Physiol 279:H1698–H1707.

Felice C.J. (1995) Monitor digital de microorganismos: Aspectos teóricos y prácticos. Doctoral Dissertation, Departamento de Bioingeniería, Facultad de Ciencias Exactas y Tecnología, Universidad Nacional de Tucumán, 125 pp.

Felice CJ, Madrid RE, Olivera JM, Rotger VI, Valentinuzzi M.E. (1999) Impedance microbiology: quantification of bacterial content in milk by means of capacitance growth curves. J Microb Meth, 35(1): 37–42.

Felice CJ, Valentinuzzi ME (1999) Medium and interface components in impedance microbiology. IEEE Trans Biomed Eng, 46(12):1483–1487.

Féré C (1888) Note sur les modifications de la résistance électrique sous l'influence des excitations sensorielles et des émotions. C R Soc Biol (Paris) 5: 217–9.

Finkelstein L, Carson ER (1979) Mathematical Modeling of Dynamic Biological Systems, Research Studies Press, Letchworth, Herts, England, 329 pp.

Foster KR, Schwan HP (1989) Dielectric properties of tissues and biological materials: A critical review. Crit Rev Biomed Eng, 17(1):25–104.

Fowles DC (1974): Mechanisms of electrodermal activity. In Methods in Physiological Psychology. Bioelectric Recording Techniques, vol. 1, RF Thompson, MM Patterson, eds, pp. 231–71, Academic Press, New York.

Fowles DC (1986): The eccrine system and electrodermal activity. In Psychophysiology, MGH Coles, E Donchin, SW Porges, eds, pp. 51–96, Guilford Press, New York.

Fowles DC, Christie MJ, Edelberg R, Grings WW, Lykken DT, Venables PH (1981) Committee report: Publication recommendations for electrodermal measurements. Psychophysiol, 18: 232–9.

Fox SI (1999) Human Physiology. McGraw–Hill, Boston, 6th ed.

Fricke H. (1932) The theory of electrolytic polarization. Phil Mag, 14:310–318.

Furness A (1975) Implantable cardiac pacemakers, and the post-operative requirements for assessing pacemaker performance (pp 1–47). Chapter 1 in IEE Medical Electronics, Monographs 13–17, edited by D.W. Hill and B.W. Watson, Peter Peregrinus, Stevenage, Herts SG1 1HQ, England. Note: Old but far from being useless. The student will find it didactic, with essentially all the early papers referred to and a good picture of the problems faced in those first years. Many have already been solved. For more updated information, just browse the scientific journal named PACE.

Ganong WF (1981) Review of Medical Physiology, Lange Medical Publications, Los Altos, CA, USA, 10th edition, 628 pp.

Gardner D. (1992) A time integral of membrane currents. Physiol Rev, 72:S1–S4. Forty Years of Membrane Current in Nerve, supplement to Physiological Reviews by the American Physiological Society.

Geddes LA (1970) The Direct and Indirect Measurement of Blood Pressure, Year Book Medical Publishers, Inc, Chicago, 196 pp. The book has also an excellent historical review.

Geddes LA (1972) Electrodes and the Measurement of Bioelectric Events, John Wiley, New York.

Geddes LA (1984) Cardiovascular Devices and their Applications, John Wiley, New York, 392 pp.

Geddes LA (1997) Historical evolution of circuit models for the electrode-electrolyte interface. Annals Biomedical Engineering, 25:1–14.

Geddes LA, Baker LE (1967) The specific resistance of biological materials — a compendium of data for the biomedical engineer and physiologist. Med & Biol Eng, 5:271–293.

Geddes LA, Baker LE (1989, 1975, 1968) Principles of Applied Biomedical Instrumentation, New York, John Wiley & Sons, 961 pp.

Geddes LA, Hoff HE (1961) The capillary electrometer. The first graphic recorder of bioelectric signals. Archives Internationales d'Histoire des Sciences, 56/57:275–290.

Geddes LA, Hoff HE (1968) Who was the first biomedical engineer? Biomed Eng, 3:551–558.

Geddes LA, Hoff HE (1971) The discovery of bioelectricity and current electricity. IEEE Spectrum, 8(12):38–46.

Geddes LA, Roeder R. (2001) Measurement of the direct-current (Faradic) resistance of the electrode-electrolyte interface for commonly used electrode materials, Ann Biomed Eng, 29(2):181–6.

Ghista DN (1979) Applied Physiological Mechanics, Harwood Academic Publishers, Chur, Schweiz, 901 pp.

Glantz S (1979) Mathematics for Biomedical Applications, University of California Press, Berkeley, USA, 423 pp.

Glasbey C, Horgan G (1995) Image Analysis for the Biological Sciences, John Wiley & Sons, New York.

Goel VK, Weinstein JN (eds) (1990) Biomechanics of the Spine: Clinical and Surgical Perspective. CRC Press, Inc, Boca Raton, FL, USA, 295 pp.

Gold HJ (1977) Mathematical Modeling of Biological Systems – An Introductory Guidebook, John Wiley, New York, 357 pp.

Goldberger E (1942) A simple, indifferent, electrocardiographic electrode of zero potential and a technique of obtaining augmented, unipolar, extremity leads. American Heart J, 23:483–492.

Goldman MJ (1970) Principles of Clinical Electrocardiography, Lange Medical Publications, Los Altos, CA, USA, 400 pp.

Gonzalez R, Woods R (1992) Digital Image Processing, Adison–Wesley, New York.

Graeme JG (1973) Applications of Operational Amplifiers. Third-generation techniques. Mc Graw Hill Book Company, The BB Electronics Series.

Grodins FS (1962) Basic concepts in the determination of vascular volumes by indicator-dilution methods. Circ Res, 10(3, part two):429–446.

Grodins FS (1963) Control Theory and Biological Systems, Columbia University Press, New York, 205 pp.

Grüsser OJ (1997) From Postdam to Berlin with detours: the life of Hermann von Helmholtz. News in Physiological Sciences, 12:90–4

Guedj D. (1997) La Mesure du Monde. Robert Laffont, SA, Paris, France, 302 pp. It tells the story of "The Meter" during the French Revolution.

Guilbault GG, Lubrano GJ (1972) An enzyme electrode for the amperometric determination of glucose, Anal Chim Acta, 64:439–455.

Guillemin R (1976) Biochemical and physiological correlates of hypothalamic peptides. The new endocrinology of the neuron. Proceedings Association Research in Nervous and Mental Disease (ARNMD), Raven Press, New York, USA.

Guillemin R (1998) The History of Neuroscience in Autobiography. Vol 2, The Society for Neuroscience, pp 96–131.

Guyton AC, Hall JE (1996) Textbook of Medical Physiology. WB Saunders, Philadelphia, 9th ed.

Heller A (1988) Electrical connection of enzyme redox centers to electrodes, J Phys Chem, 96:3579–3587.

Henderson JM (1996) Gastrointestinal Pathophysiology. Lippincott–Raven, Philadelphia.

Highstein SM, Cohen B, Büttner–Ennever JA (1996) New Directions in Vestibular Research, in Annals of the New York Academy of Sciences, vol 781, June 19, 739 pp.

Hines JW. (1996) Medical and surgical applications of space biosensor technology, Acta Astronaut, Feb–Apr, 38(4–8):261–267.

Ho M, Rechnitz GA (1992) An introduction to biosensors. In Immunochemical Assays and Biosensor Technology for the 1990s, Nakamura RM, Kasahara Y, and Rechnitz GA (eds), Washington, DC, American Society for Microbiology, pp 275-290.

Hodgkin AL (1964) The Conduction of the Nervous Impulse, The Sherrington Lectures VII, Charles C. Thomas, Publishers, Spingfield, IL, USA, 108 pp.

Hodgkin AL, Huxley AF (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. Journal Physiology, 117:500–544.

Hoff HE (1936) Galvani and the pre-galvanian electrophysiologists. Annals of Science, 1(2):157–172.

Hoff HE, Scott HJ (1948), Physiology, New England J Medicine, 239:26, where an English translation of Fick's 1870 original communication entitled “On the measurement of the amount of blood in the ventricle” can be found.

Hoffman, FB, Cranefield (1960) Electrophysiology of the Heart, McGraw-Hill Book Company, New York, 323 pp.

Hughes I (1982) Congenital and acquired disorders of the adrenal cortex. Clin Endocrinology and Metabolism, 11: 89.

Huxley HE (1969) The mechanism of muscular contraction. Science, 164:1356–1366.

Jacobson ED (1982) Physiology of the mesenteric circulation. The Physiologist, 25(5):439–443.

Jacquez JA (1972) Compartmental Analysis in Biolgy and Medicine, Elsevier Publishing Company, New York, 237 pp.

Jones RW (1973) Principles of Biological Regulation, Academic Press, New York, USA, 359 pp.

Kaplan S. (1982) Clinical Pediatrics and Adolescent Endocrinology. WB Saunders.

Katz B (1966) Nerve, Muscle and Synapse. McGraw–Hill Books, New York, 193 pp.

Khoo MCK (2000) Physiological Control Systems, IEEE Press, Series on Biomedical Engineering, New York, 318 pp.

Kligfield P (1981) Laennec and the discovery of mediate auscultation. The American Journal of Medicine, 70:275–277.

Knapp RB, Lusted HS (1990) Musical performance by the handicapped generated from bioelectric signals, 119th Meeting Acoustical Society of America, Pennsylvania State University, May 22.

Knobil E (1999) The Wisdom of the Body revisited. News in Physiological Sciences, 14:1–11.

Koestler A. (1964) The Act of Creation. Pan Books Limited, Cavaye Place, London, Great Britain, 491 pp.

Lam ZH, Hwang K-S, Lee J-Y, Chase JG, Wake GC (2002) Active insulin infusion using optimal and derivative-weighted control. Medical Engineering & Physics, 24:663–672.

Lewis T (1911) The Mechanism of the Heart Beat. Shaw & Sons, London. Dedicated to Willhem Einthoven.

Lieberstein HM (1973) Mathematical Physiology, American Elsevier Publishing Co, New York, 377 pp.

Liu S.H. (1985) Fractal model for the ac response of a rough interface. Physical Review Letters, 55(5):529–532.

Lubkin V, Beizai P, Sadum AS (2002) The eye as metronome of the body. Survey of Ophthalmology, 47(1):17–26.

Lumsen CJ, Brandts, WA, Trainor EH, editors (1997) Physical Theory in Biology, World Scientific Publishers, Singapore, 1997, xvi+486 pp.

MacDonald DD, McKubre MCH (1987) Corrosion of material. Section 4.3 in Chapter 4 of Impedance Spectroscopy. Emphasizing solid materials and systems, JR MacDonald ed, John Wiley & Sons, New York, USA.

MacPherson RD, MacNeil G, Marble AE (1976) Integrated circuit measurement of skin conductance. Behav Res Methods Instrum, 8: 361–4.

Malmivuo J, Plonsey, R (1995) Principles and Applications of Bioelectric and Biomagnetic Fields, Oxford University Press, 1995. Freely available at http://butler.cc.tut.fi/~malmivuo/bem/bembook/00/co.htm

Mandelbrot B (1975) Les objects fractals: forme, hasard et dimension. Flammarion, Paris.

Mattioli M, Barboni B, Bacci ML, Seren E (1990) Maturation of pig oocytes: observations on membrane potential. Biol Reprod, 43:318–322.

Mayer S, Geddes LA, Bourland JD, Ogborn L. (1992) Faradic resistance of the electrode/electrolyte interface. Med Biol Eng Comput, 30(5):538–42.

McAdams ET, Lackermeier A, McLaughlin JA, Macken D, Jossinet J (1995) The linear and non-linear electrical properties of the electrode-electrolyte interface. Biosensors & Bioelectronics, 10:67–74.

McArdle WD, Katch FI, Katch VL (1991) Exercise Physiology, Lea & Febiger, Philadelphia, 853 pp.

McCrady JD, Hoff, Geddes LA (1966) The contributions of the horse to knowledge of the heart and circulation. James Hope and the heart sounds, Connecticut Medicine, 30(2):126–131.

McCullough DH, Levitan H (1987) Rabbit oocyte maturation: changes of membrane resistance, capacitance, and the frequency of spontaneous transient depolarizations. Develop Biol, 120:162–169.

Medved V (2001) Measurement of Human Locomotion, CRC Press, Boca Raton, FL, ISBN 0-8493-7675-0.

Michael JA, Richardson D, Rovick A, Modell H, Bruce D, Horwitz B, Hudson M, Silverthorn D, Whitescarver S, Williams S (1999) Undergraduate students' misconceptions about respiratory physiology. Am J Physiol 277 (Adv Physiol Educ 22):S127–S135.

Middleman S (1972) Transport Phenomena in the Cardiovascular System, Wiley–Interscience, New York, 299 pp.

Milhorn HT (1966) The Application of Control Theory to Physiological Systems, Saunders, Philadelphia.

Milnor, W.R. (1982) Hemodynamics, Williams & Wilkins, Baltimore, 390 pp. (See Chapter 7, pp. 157–191).

Milsum JH (1966) Biological Control Systems, McGraw–Hill, New York, 466 pp.

Moreira M, Murta LO, Baffa O (2000) Imaging ferromagnetic tracers with an ac biosusceptometer. Rev Sci Instr, 71(6):2532–8.

Morucci JP, Valentinuzzi ME, Rigaud B, Felice CF, Chauveau N, Marsili PM (1996) Bioelectrical Impedance Techniques in Medicine. Critical Reviews in Biomedical Engineering, 24(4–6):223–681.

Nagao M, Murase K, Kikuchi T, Ikeda M, Nebu A, Fukuhara R, Sugawara Y, Miki H, Ikezoe J (2001) Fractal analysis of cerebral blood flow distribution in Alzheimer's disease. Journal Nuclear Medicine, 42(10):1446–1450.

Noble DA (1962) A modification of the Hodgkin–Huxley equations applicable to Purkinje fibre action and pacemaker potentials. J Physiol (London), 160:317–352.

Noble PA, Dziuba M, Harrison D J, Albritton WL (1999) Factors influencing capacitance-based monitoring of microbial growth, J Microb Meth 37:51–64.

Nunez PL (1981) Electric Fields of the Brain: The Neurophysics of EEG, Oxford University Press, New York, 484 pp.

Nyikos L, Pajkossy T (1985) Fractal dimension and fractional power frequency dependent impedance of blocking electrodes, Electrochim Acta, 30(11):1533–1540.

Okamoto H, Takahashi K, Yamashita N (1977) Ionic currents through the membrane of the mammalian oocyte and their comparison with those in the tunicate and sea urchin. J Physiol, 267:465–495.

Page L (1965) Introduction to Theoretical Physics, 3d ed, 6th printing, D van Nostrand Co, 701 pp.

Pajkossy T (1991) Electrochemistry at fractal surfaces. J Electroanal Chem, 300, pp 1–11.

Pallàs–Areny R, Webster JG (1991) Sensors and Signal Conditioning. John Wiley & Sons, Inc, USA, pp 398.

Pallàs–Areny R, Webster JG (1999) Analog Signal Processing. John Wiley & Sons Inc, 605 Third Avenue, New York, 1st ed.

Patel DJ, Greenfield JC Jr, Fry DL (1964) Hydraulic input impedance to aorta and pulmonary artery in dogs. J Applied Physiol, 18:134–140.

Patterson SW, Starling EH (1914) On the mechanical factors which determine the output of the ventricles. J Physiol, 48:357–379.

Perkins JF (1964) Respiration. In Handbook of Physiology, section 3, vol. I, American Physiological Society, Washington, DC, 62 pp.

Peura RA (1992) Chemical Biosensors, Chapter 10 in Medical Instrumentation: Application and Design, JG Webster ed, Houghton Mifflin Co, Boston, USA, pp 542–595.

Pevzner PA (2000) Computational Molecular Biology: An Algorithmic Approach, MIT Press, Massachussets, USA, 314 pp.

Picton TW (1974) Evoked cortical potentials, how, what and why? American J EEG Technol, 14:9–44.

Pitts RF (1966) Physiology of the Kidney and Body Fluids. Year Book Medical Publishers, Inc, Chicago, IL, 243 pp.

Plonsey R (1969) Bioelectric Phenomena. McGraw–Hill Book Co, New York, USA, 380 pp (see pages 94–99).

Politti JC, Felice CJ, Valentinuzzi ME (2003) Hysteresis cycle during arm abduction and adduction in static positions. Medical Eng & Physics (Elsevier, UK), 25(4):317–320.

Popovic DB (Guest Editor, 2003) Control Issues of Functional Electrical Stimulation: Current and Future Systems. Medical Engineering & Physics, 25:1–78. There are 9 articles in the issue.

Powell T, Valentinuzzi ME (1974) Calcium homeostasis: responses of a possible mathematical model. Med & Biol Eng, 12:287–294.

Proske U (1997) The mammalian muscle spindle. News in Physiological Sciences, 12:37–42.

Ragheb T, Geddes LA. (1991) The polarization impedance of common electrode metals operated at low current density, Ann Biomed Eng, 19(2):151–63.

Randles EB (1947) Rapid electrode reactions. Faraday Discuss Physics Chemistry, 1, pp 11–19.

Rash RW (1984) A microprocessor analog simulation of the operation of countercurrent multiplier systems. The Physiologist, 27(1):58–61.

Rashevsky N (1938) Mathematical Biophysics, The University of Chicago Press, Chicago, IL, USA, 669 pp, with editions in 1948 and 1960, the latter by Dover Publications, New York, in 2 volumes.

Rashevsky N (1960) Mathematical Biophysics. Physico-Mathematical Foundations of Biology. Volume One, Dover Publications, Inc, New York, 3rd edition, Part 2, pp 375–462.

Rashevsky N (1964) Some Medical Aspects of Mathematical Biology. Charles C. Thomas Publisher, Springfield, IL, USA, 314 pp.

Rashevsky N (1972) Organismic Sets, Mathematical Biology, Inc., Holland, Michigan, USA, 213 pp. Published posthumously by his wife Emily through William Clowes & Sons, London. Nicholas died in January 1972 of massive cardiac infarction.

Rechnits GA (1988) Biosensors, Chem Eng News, 5:24–36.

Rescigno A, Segre G (1966) Drug and Tracer Kinetics, Blaisdell Publishing Company, Waltham, MA, 209 pp.

Remond A (ed) (1971) Handbook of Electroencephalography and Clinical Neurophysiology, vol 1, Elsevier, Amsterdam.

Rhoades RA, Tanner GA (1995) Medical Physiology. Little, Brown and Company, Boston.

Riggs DS (1963) The Mathematical Approach to Physiological Problems. The Williams & Wilkins Co, Baltimore, 445 pp.

Riggs DS (1970) Control Theory and Physiological Feedback Mechanisms, Williams & Wilkins, Baltimore,

Riu PJ, Rosell J, Bragós R, Casas O (eds) (1999) Electrical Bioimpedance Methods: Applications to Medicine and Biotechnology. Proceedings X International Conference on Electrical Bioimpedance, published by NY Academy of Sciences, 600 pp, 58 papers. Full text of volume 837 is available on-line to Academy Members and to subscribing libraries.

Rosato E, Kyriacou CP (2001) Flies, clocks and evolution. Philosophical Transactions Royal Society London, Section B, Biological Sciences, 356:1769–1778.

Roston S (1959) Mathematical representation of some endocrinological systems. Bulletin Mathematical Biophysics, 21:271–282.

Ruch TC, Patton HD (eds) (1966) Physiology and Biophysics, WB Saunders Co, Philadelphia. See Chapter 2 and its Appendix, pp 26–72.

Ruiz G, Felice CJ (2003) Non linear fractal impedance model of an electrode-electrolyte interface, 25th Annual International Conference/IEEE Engineering in Medicine & Biology Society, Cancun, Mexico.

Sabah NH (1999) Origin of the resting potential. IEEE Eng in Medicine and Biol, Sept–Oct, pp100–5.

Sabah NH (2000a) Rectification in biological membranes. IEEE Eng in Medicine and Biol, Jan–Feb, pp 106–13.

Sabah NH (2000b) Reactance of biological membranes. IEEE Eng in Medicine and Biol, Jul–Aug, pp 89–95.

Sabah NH (2000c) Aspects of nerve conduction. IEEE Eng in Medicine and Biol, Nov–Dec, pp 111–18.

Sagawa, K, Maughan, L, Suga, H, Sunagawa, K (1988) Cardiac Contraction and the Pressure-Volume Relationship, Oxford University Press, 480 pp, 1988.

Scheller FW, Wollenberger U, Pfeiffer D (1996) Overview of biosensor technology. In Advances in Molecular and Cell Biology, Bittar E, Danielson B,

and Bülow L (eds), Biochemical Technology Series, vol 15B, Greenwich, CT, JAI Press, pp 353–363.

Schmidt–Nielsen B (1984) August and Marie Krogh and respiratory physiology. J Appl Physiol: Respiration, Environment, and Exercise Physiol, 57(2):293–303.

Schneck DJ (1990) Engineering Principles of Physiologic Function, New Yok University Press, New York, 539 pp.

Schwan HP (1968) Polarization impedance and measurements in biological materials. Trans NY Acad Sci, 148:191–209.

Schwan HP, editor (1969) Biological Engineering, McGraw–Hill, New York, 556 pp.

Schwedel JB (1968) ECG Interpretation. Warner–Chilcott, New Jersey, 41 pp.

Schweitzer VG, Thompson NW, Harness JK, Nishiyama RH (1979) Management of severe hypercalcemia caused by primary hyperparathyroidism. Archives of Surgery, 113:373–381.

Sharbrough F, Chatrian G-E, Lesser RP, Lüders H, Nuwer M, Picton TW (1991) American Electroencephalographic Society Guidelines for Standard Electrode Position Nomenclature. J Clin.Neurophysiol, 8: 200–2.

Sheingold DH (1980) Transducer Interfacing Handbook: A guide to analog signal conditioning. Analog Devices Inc, Norwood, Massachusetts 02062, USA, 1st ed, 230 pp.

Skilling HH (1965) Electrical Engineering Circuits. John Wiley & Sons, New York USA, 1st ed.

Sluyters–Rehbach M, Sluyters JH (1970). Sine wave methods in the study of electrode process, Chapter 4 in Electroanalytical Chemistry, vol 4, A Bard 9(ed), Marcel Dekker, Inc, New York.

Smith, HW (1959), From Fish to Philosopher, CIBA Editions, 304pp.

Soller BR, Cabrera M, Smith SM, Sutton JP (2002) Smart medical systems with application to nutrition and fitness in space, Nutrition, Oct 18, (10):930–936.

Spinelli E, Felice CJ, Mayosky M, Politti JC, Valentinuzzi ME (2001) Propagation velocity measurement: Autocorrelation technique applied to the electromyogram. Med & Biol Eng & Comput, 39(5):590–3, 2001.

Starling, EH (1918) The Linacre Lecture on the Law of the Heart. Longmans, Green and Co, London.

Tacker WA, Geddes LA (1980) Electrical Defibrillation. CRC Press, Inc, Boca Raton, Florida, USA, 192 pp. Note: For more updated information on defibrillation and defibrillators, just browse the scientific journal named PACE.

Tarchanoff J (1889) Décharges électriques dans la peau de l'homme sous l'influence de l'excitation des organes des sens et de différentes formes d'activité psychique. C R Soc.Biol (Paris) 41: 447–51.

Thurau K (ed) (1974, 1976) Kidney and Urinary Tract Physiology, first part, 12 chapters, 350 pp, second part 9 chapters, 340 pp, University Park Press, Baltimore, Maryland, USA.

Tippens PE (1973) Applied Physics, New York, McGraw–Hill.

Togawa T, Tamura T, Öberg PA (1997) Biomedical Transducers and Instruments. CRC Press, Boca Raton, FA, 1st ed, 366pp.

Tran-Minh C (1993) Biosensors, Chapman & Hall, United Kingdom, 217 pp.

Updike SJ, Hicks GP (1967) The enzyme electrode, Nature, 214:986–988.

Valentinuzzi M (1980) The Organs of Equilibrium and Orientation as a Control System, Harwood Academic Publishers, Chur, Schweiz, 194 pp.

Valentinuzzi M, Valentinuzzi ME, Posey JA (1972) J Assoc Advancement Med Instr, 6(5):335–343.

Valentinuzzi ME (1969) Electrophysiology and Mechanics of the Snake Heart Beat. Doctoral dissertation, Department of Physiology, Baylor College of Medicine, Houston, TX, USA. Library of Congress Catalog #QP341.V34. Copies from University Microfilms, Ann Arbor, Michigan 48106, USA.

Valentinuzzi ME (1971a) A mathematical model of the electrocardiographic QT-RR relationship. Med & Biol Eng, 9(4):255–261.

Valentinuzzi ME (1971b) A mathematical model of the hepatic portal system. Med & Biol Eng, 9:213–220.

Valentinuzzi ME (2002) Magnetobiology: A historical view. Report to The Bakken Museum and Library, Minneapolis, MN, 12 pp. It can be obtained from this library by contacting Ms Elizabeth Ihrig, Ihrig@thebakken.org.

Valentinuzzi ME (2004) A new beginning for human health, *IEEE/EMBS Magazine*, July Issue. This article is based on the Cancun Conference Celebration Lecture, September 2003.

Valentinuzzi ME, Geddes LA, Baker LE (1968) A simple mathematical derivation of the Stewart–Hamilton formula for the determination of cardiac output. Med & Biol Eng, 7:277–282.

Valentinuzzi ME, Geddes LA, Baker LE, Hoff HE (1968) The node equation and its applications in physiology. Med & Biol Eng, 6:387–397.

Valentinuzzi ME, Geddes LA, Hoff HE, Bourland J (1970) properties of the 30º haxaxial (Einthoven–Goldberger) system of vectorcardiography. Cardiovasc Reasearch Center Bulll (Houston), 9(2):64–73.

Valentinuzzi ME, Ghista DN, Nichols WW (1979) Vascular Elasticity: Its homeostatic role. Simplified parametric characterization and determination. Chapter 8 in Applied Physiological Mechanics, DN Ghista (ed), Harwood Academic Publishers, New York, pp 309–379.

Valentinuzzi ME, Hoff HE (1970) Catheterization in the snake: correlation of cardiac events. Cardiovasc Res Center Bull, 8(3):102–18.

Valentinuzzi ME, Hoff HE, Geddes LA (1969) Observations on the electrical activity of the snake heart. J Electrocardiology, 2(1):39–50.

Valentinuzzi ME, Spinelli JC (1989) Intracardiac measurements with the impedance technique. IEEE/EMB Mag, 8(1):27–34.

Venables PH, Christie MJ (1980) Electrodermal activity. In Techniques in Psychophysiology, I Martin, PH Venables (eds), pp. 2–67, John Wiley, New York.

Verzola RMM, Baffa O, Wakai RT, Sousa PL (1995) Magnetocardiography in normal subjects and in patients with right bundle branch block. Brazilian J Med & Biol Res, 28:1327–1331.

Warburg von E (1899) Ueber das verhalten sogenannter unpolarisirbarer elektroden gegen wechselstrom, Annalen der Physik und Chemie, 67, pp 493–499.

Wardener HD, Clarkson EM (1985) Concept of natriuretic hormone. Physiological Reviews, 65(3):658–759.

Webster JG (1992) Medical Instrumentation. Applications and Design. Houghton Mifflin Co, One Beacon Street, Boston, MA 021008, 2nd ed.

Welch T, Potchen J, Welch M (1972) Fundamental of the Tracer Method, WB Saunders Company, Philadelphia, USA, 187 pp.

Wetterberg L (1999) Melatonin and clinical application. Reproduction and Nutrition Development, 39 (1999), pp 367–382.

White PA, Brookes CIO, Ravn H, Hjortdal V, Chaturvedi RR, Redington AN (2001) Validation and utility of novel volume reduction technique for determination of parallel conductance. Am J Physiol Heart Circ Physiol 280:H475–H482.

Wilke HJ, Neef P, Caimi M, Hoogland T, Claes LE (1999) New in vivo measurements of pressures in the intervertebral disc in daily life. Spine, 24(8):755–762, April 15.

Williamson SJ, Hoke M, Stroink G, Kotani M (1989) Advances in Biomagnetism, Plenum Press, New York.

Willius FA, Keys TE (1941) Classics of Cardiology, two volumes, Dover Publications, Inc, New York. Note: For those interested in the history of cardiology, the collection found in these two volumes is recommended.

Wilson NF, Johnston FE, Macleod AG, Barker PS (1934) Electrocardiograms that represent the potential variations of a single electrode. Am Heart J, 9:447–458.

Wright FS, Briggs JP (1979) Feedback control of glomerular blood flow, pressure and filtration rate. Physiological Reviews, 59(4):958–1006.

Wright K (2002) Times of our lives. Scientific American, 287(3):41–7.

Yardley JE, Kell DB, Barrett J, Davey CL (2000) On-line, real-time measurements of cellular biomass using dielectric spectroscopy, Biotechnol Genet Eng Rev, 17:3–35.

Yin, FCP (ed) (1987) Ventricular-Vascular Coupling: Clinical, Physiological and Engineering Aspects, Springer Verlag, New York, 406 pp.

Zhang S, Wright G, Yang Y (2000) Materials and techniques for electrochemical biosensor design and construction. Biosensors & Bioelectronics, 15, pp 273–282.

Zierler KL (1962) Theoretical basis of indicator-dilution methods for measuring flow and volume. Circ Res, 10(3, part two):393–407.

Zierler KL (1963) Theory of use of indicators to measure blood flow and extracellular volume and calculation of transcapillary movement of tracers. Circ Res, 12:464–471.

Index

C

D

E

F

G

H

I

W

Y

List of Figures

Author's Biographical Note

Born in Buenos Aires in 1932, he obtained the Bachelor degree from the National College of Buenos Aires (1950), and graduated as Telecommunications Engineer at the University of Buenos Aires (1956). Later, he earned a Ph.D. in Physiology and Biophysics from Baylor College of Medicine (USA, 1969), where he also became Assistant Professor (1969–73). Professor of Bioengineering and Head of Laboratory (1972–2001) at the *Universidad Nacional de Tucumán* (UNT) and Career Investigator (1977–2000) of the *Consejo Nacional de Investigaciones Científicas y Técnicas* (CONICET), Argentina, where he currently continues under a contract. Cofounder and first director (1980–7) of the *Instituto Superior de Investigaciones Biológicas* (INSIBIO). Director of the Master of Bioengineering Program at the UNT (1996-2003). In 1973, he shared the Nightingale Prize of Bioengineering (IFMBE & BES), in 1980 the Houssay Prize of Biology (*Sociedad Argentina de Biología*) and, in 1985, the Catalina B. de Barón Accesit Prize of Cardiology (CORDIC, Buenos Aires). In 1984, he was awarded the Golden Route Prize in Science (*Sociedad de Distribuidores de Diarios*, Buenos Aires). The IEEE/EMBS gave him the 1996 Career Achievement Award. He is a Life Fellow Member of the IEEE, Member of the American Physiological Society, the Institution of Physics and Engineering in Biology and Medicine (England), the *Sociedad Científica Argentina* and the *Sociedad Argentina de Bioingeniería*. In 1989, he was inducted into the Argentine National Academy of Engineering, in 1990, into the Córdoba Academy of Medical Sciences, and in September 1997, into the International Academy for Medical and Biological Engineering. He authored or coauthored over 90 scientific and technical papers and more than 30 teaching and/or general articles, has collaborated in 8 books and was guest editor to 5 special issues of scientific journals. For different length periods, he lectured at several US and Latin American universities acting, also, in the editorial boards of Medical Progress through Technology (1979–1994) and Medical Engineering & Physics (1989; Associate Editor, 2001-3) and helping, now and then, as referee to other journals. Besides, he was Latin American Representative to IFMBE (1988–1991) and also to IEEE/EMBS (1988–1991). Between1991 and 1994, he was President to the Latin American Regional Council of Biomedical Engineering (CORAL). He has worked in the fields of bioimpedance, cardiovascular system, impedance microbiology, fibrillation-defibrillation, numerical deconvolution and biomedical engineering education. Max loves his students, likes children, animals and music (plays piano). He enjoys a simple outdoors life. Being is much better than having.

UNDERSTANDING THE HUMAN MACHINE:

A PRIMER TO BIOENGINEERING

By

MAX E. VALENTINUZZI, EE, PhD
Fellow Investigator of the National Research Council of Argentina
Honorary Professor of the Pontificial University of Perú

Department of Bioengineering
Universidad Nacional de Tucumán (UNT)
Tucumán, Argentina
Also associated with
Instituto Superior de Investigaciones Biológicas
(INSIBIO/CONICET[1]-UNT)

Tucumán City, April 26, 2004

[1] CONICET: Consejo Nacional de Investigaciones Científicas y Técnicas of Argentina