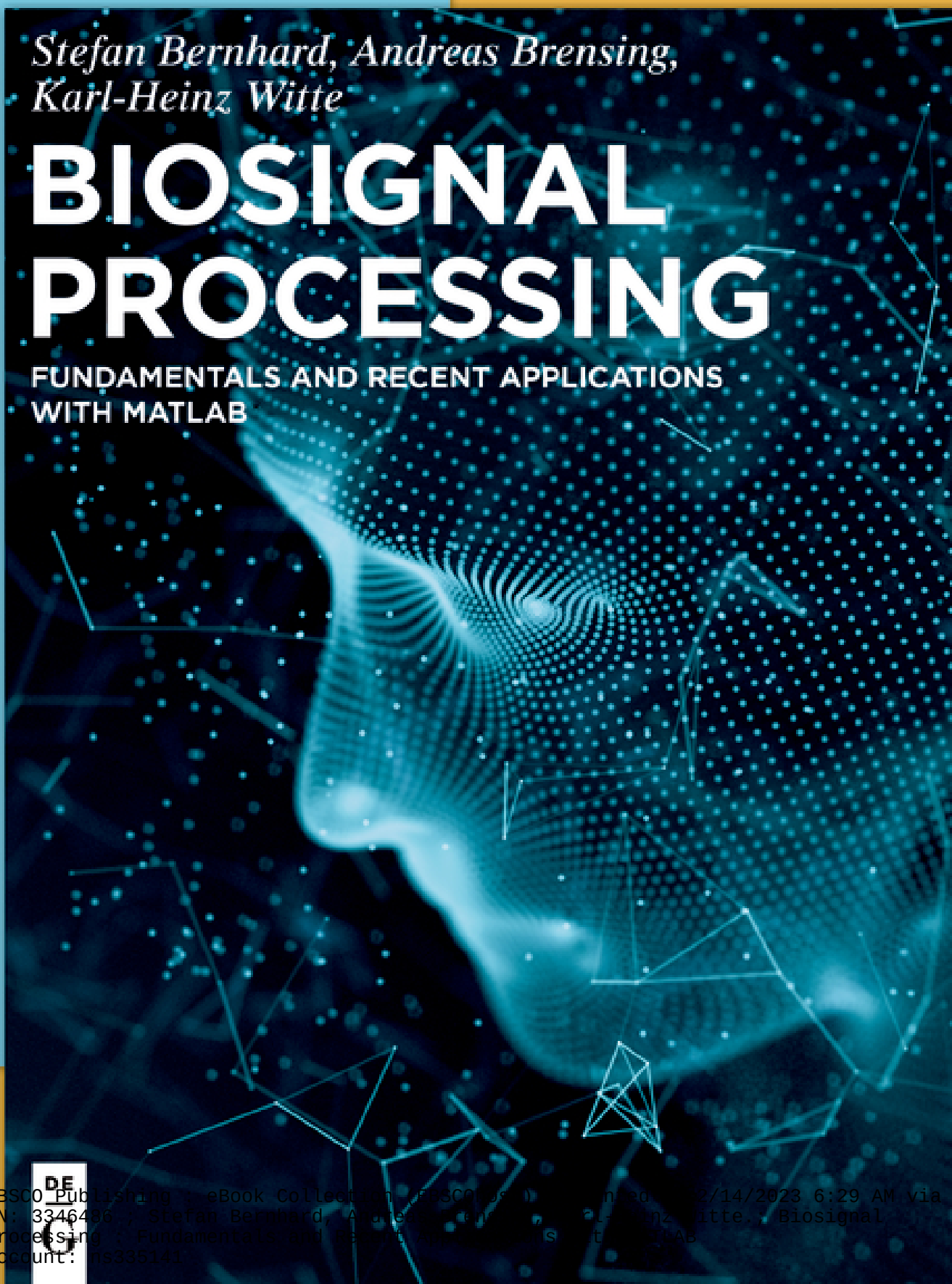
DE GRUYTER
OLDENBOURG

GRADUATE

Stefan Bernhard, Andreas Brensing, Karl-Heinz Witte

BIOSIGNAL PROCESSING

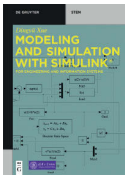
FUNDAMENTALS AND RECENT APPLICATIONS WITH MATLAB



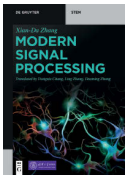
Stefan Bernhard, Andreas Brensing, Karl-Heinz Witte

Biosignal Processing

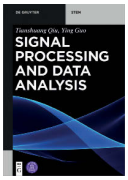
Also of interest



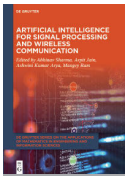
Modeling and Simulation with Simulink®
For Engineering and Information Systems
Dingyü Xue, 2022
ISBN 978-3-11-073904-6, e-ISBN (PDF) 978-3-11-073495-9



Modern Signal Processing
Xian-Da Zhang, 2022
ISBN 978-3-11-047555-5, e-ISBN (PDF) 978-3-11-047556-2

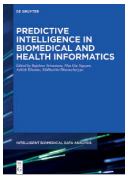


Signal Processing and Data Analysis
Tianshuang Qiu, Ying Guo, 2018
ISBN 978-3-11-046158-9, e-ISBN (PDF) 978-3-11-046508-2



Artificial Intelligence for Signal Processing and
Wireless Communication
Edited by Abhinav Sharma, Arpit Jain, Ashwini Kumar Arya,
Mangey Ram, 2022
ISBN 978-3-11-073882-7, e-ISBN (PDF) 978-3-11-073465-2
De Gruyter Series on the Applications of Mathematics in Engineering
and Information Sciences
ISSN 2626-5427, e-ISSN 2626-5435

Communication, Signal Processing & Information Technology
Edited by Faouzi Derbel, 2020
ISBN 978-3-11-059120-0, e-ISBN (PDF) 978-3-11-059206-1
Advances in Systems, Signals and Devices
ISSN 2364-7493, e-ISSN 2364-7507



Predictive Intelligence in Biomedical and Health Informatics
Edited by Rajshree Srivastava, Nhu Gia Nguyen, Ashish Khanna,
Siddhartha Bhattacharyya, 2020
ISBN 978-3-11-067608-2, e-ISBN (PDF) 978-3-11-067612-9
Intelligent Biomedical Data Analysis
ISSN 2629-7140, e-ISSN 2629-7159

Stefan Bernhard, Andreas Brensing,
Karl-Heinz Witte

Biosignal Processing

Fundamentals and Recent Applications with MATLAB®

DE GRUYTER
OLDENBOURG

Authors
Prof. Dr. rer. nat. Stefan Bernhard
Technische Hochschule Mittelhessen
FB Life Science Engineering
Wiesenstrasse 14
35390 Giessen
stefan.bernhard@lse.thm.de

Prof. Dr. Andreas Brensing
Hochschule Rhein Main
FB Ingenieurwissenschaften
Fachgebiet Medizintechnik
Am Brückweg 26
65428 Rüsselsheim
andreas.brensing@hs-rm.de

Prof. Dr.-Ing. Karl-Heinz Witte
Hochschule RheinMain
FB Ingenieurwesen
Am Brückweg 26
65428 Rüsselsheim
Karl-Heinz.Witte@hs-rm.de

ISBN 978-3-11-073959-6
e-ISBN (PDF) 978-3-11-073629-8
e-ISBN (EPUB) 978-3-11-073967-1

Library of Congress Control Number: 2022941711

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

Cover image: Jian Fan / iStock / Getty Images Plus
Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Foreword

The interdisciplinary field of biosignal processing has become an indispensable part of today's medical technology applications. In everyday clinical practice, large quantities of image data and signals are continuously evaluated for diagnostic purposes. In order to be able to offer patients the most efficient therapy for their diseases, reliable diagnostic tools are the foundations for proper medical decision-making. In many situations, the doctor only has minutes – one reason why the data and signals must be processed in a meaningful way. In addition to desirable real-time processing, this includes above all reliability and robustness as well as a clear and unambiguous presentation of the diagnostic statements. According to these requirements, some fundamental rules must be considered in the analogue and digital processing of biosignals.

This book provides a fundamental work on biosignal processing in which, besides a general introduction to the generation, measurement and analogue/digital processing of biosignals, also recent application examples are discussed. Thus, the book is on the one hand suitable for students and teachers as an introduction to the fundamental methodology and on the other hand a compact reference for experienced users in the field of medical technology and medical informatics.

The subject matter of the book is supplemented by numerous recent research topics in biosignal processing in the form of application-oriented programming examples and thus also offers researchers from the natural and life sciences such as medicine and medical technology diverse incitation for the development of novel methods in the processing of biosignals.

The book introduces the fundamental physiological processes in the formation of biosignals, the authors discuss the basics of analogue signal amplification and processing as well as the digitisation for various biosignals along the measurement chain by means of sensor and measurement technology. The design of important digital filters and basic methods for analysing the signals in the time, frequency and composite domain as well as methods for statistical or model-based evaluation of signals are also discussed in detail.

The content is equally convenient for students of natural sciences, namely mathematics, physics and biology, engineering sciences, especially computer science, medical technology and physical engineering, as well as life sciences, especially medicine. The presentation of the mathematical and experimental methods originated in the authors' long-standing activity as lecturers in biosignal processing.

Compact introductions to the physical and physiological basics make it easier for readers from related disciplines to get started and also serve as a refresher on necessary basic concepts. The sections do not claim to be complete and therefore refer to further literature at the appropriate places.

https://doi.org/10.1515/9783110736298-202

Each chapter is supplemented by an example and exercise section, which can be used by the reader to deepen and prepare for the preparation of the material. The programming examples for Matlab, Scilab/COS and LTSpice are available as additional material and can be downloaded from the publisher's website.

The content and form of the book originates from the authors' didactic experience of lectures held at three different German University of Applied Sciences. The wealth of experience gained over the years thus also includes influences from the indirect collaboration of students, doctoral students and colleagues, who cannot all be mentioned here by name. The authors would therefore like to express our gratitude for all the support we have received in the form of suggestions for improvement, criticism and praise during our time as lecturers at the University! Furthermore, the authors like to thank Prof. Dr. Jörg Subke and Benedict Schneider for their contribution of subsection 6.2.2 and Dr. Urs Hackstein for his contribution in subsection 6.3.4. Finally, we would like to thank Ms Eva Funk and Franziska Marschall for creating, editing or translating numerous illustrations.

For the sake of better readability, the book uses the masculine form throughout, although all readers should of course feel addressed by their respective person.

Stefan Bernhard

Contents

Image credits

Fig. 1.1: Image library from TMSi, The Netherlands

Fig.2.7: MRI and CT angiographies of the heart/brain from http://www.osirix-viewer.com/resources/dicom-image-library/

Fig.2.9: MRI and CT angiographies of the heart/brain from http://www.osirix-viewer.com/resources/dicom-image-library/

Fig.3.4: Cell membrane from © natros – stock.adobe.com, http://stock.adobe.com

Fig.3.12: Cell, axon, muscle from © joshya – stock.adobe.com, http://stock.adobe.com

Fig.3.38: Brain from © bilderzwerg – stock.adobe.com, http://stock.adobe.com

Fig.3.38: Sensory organs from © www.freepik.com

Fig.4.2: Skin from © Neokryuger – stock.adobe.com, http://stock.adobe.com

Fig.6.1: © Dule964 | Dreamstime.com

Fig.6.2: Modified from © Alila07 | Dreamstime.com

Fig.6.4: © Alila07 | Dreamstime.com

Fig.6.6: Motor unit from © Balint Radu – stock.adobe.com, http://stock.adobe.com

Figs.3.1, 3.20, 4.6, 4.13, 4.16, 6.9, 6.10, 6.11, 6.12, 6.13, 6.14: Modified body outline from © Aaltazar – istock.de

Figs. 2.1, 2.2, 2.5, 2.6, 2.7, 2.9, 2.10,2.20, 2.21, 2.24, 3.3, 3.5, 3.6, 3.7, 3.34: Generated using TikZ library and examples [75].

Figs. 3.13, 3.17, 6.11 - 6.13: Adapted from [47].

Post-processing, compilation and creation of visuals © Eva Funk – www.evafunk.com

https://doi.org/10.1515/9783110736298-204

1 Introduction to Biosignal Processing

Biosignal processing is an interdisciplinary field of research that includes medical informatics, signal processing and life sciences. The main objective of the analysis of human biosignals is to support medical *diagnosis* with the help of *mathematical* methods. Intelligent evaluation of the signals should provide the medical practitioner with valuable quantitative information in the diagnosis and effectively support them in medical decision-making. The topics are wide-ranging: On the one hand, they include signal processing, methods for monitoring and controlling vital functions in intensive care medicine, for example through the automatic classification of signals, and on the other hand, the formal description of the relationships between signals and physiological functions in medical research. Another growing discipline is the modelling of physiological phenomena to gain a deeper understanding of the underlying pathological mechanisms to subsequently improve the technology in medical device engineering. Modern simulation techniques often make use of statistical methods like parameter variation and estimation, as well as the quantification of uncertainties in the model approach – just to develop increasingly sophisticated data evaluation methods for diagnostic purposes. In many cases, these methods are the basis for optimal control and regulation or lead to a complete replacement of physiological functions, as for example in prosthetics.

The acquisition of biosignals for diagnostic purposes has a long history. Beginning with the recognition of the electrical activity of nerve and muscle cells, which engaged Luigi Galvani in his famous frog's leg experiment in 1787, a chain of far-reaching findings on the fundamental mechanisms of electrophysiology followed. As early as 1876, E. J. Marey succeeded in graphically depicting these processes for the first time. Willem Einthoven received the Nobel Prize in Medicine in 1924 for the development of the string galvanometer and the physiological interpretation of the electrocardiogram. Further development through the use of improved sensing and measuring technology, such as the tube amplifier, transistors and later integrated circuits as well as the microprocessor technology, led to a considerable improvement in signal quality over time.

Nowadays, the main focus of research is on increasingly sophisticated evaluation algorithms, for example based on large amounts of data[1] and the miniaturisation of measurement and transmission technology as well as data storage on the internet. Today, the evaluation of information and signals from the human body is the basis of almost every medical diagnosis. Electrocardiography (ECG) in particular has developed into one of the most frequently used medical examination methods – millions of ECGs are recorded in the world every day. Especially in the field of long-term

1 Evaluation of extremely large data sets with the help of computer algorithms with the aim of making connections such as trends, correlations or patterns in the behaviour of the data visible.

https://doi.org/10.1515/9783110736298-001

ECGs (over 24 h), modern methods are indispensable, as a heart beats approximately 80,000 to 100,000 times during this time. In this order of magnitude, the search for abnormal heartbeats such as cardiac arrhythmias or unusual changes such as ventricular fibrillation presents doctors with a practically impossible task. In addition to the precious time that would be involved in manual review, there would be a lapse of attention and thus a loss of important events – not to mention the compilation of statistics.

However, large amounts of data are not only generated in the analysis of intensive care data, but also in telemedicine. The variety of data in the field of commercial sports medicine has grown rapidly through the use of mobile technologies and software on smartphones, as has the continuous monitoring of older people to monitor their health. The list of possible vital signs and health parameters besides ECG is long: blood pressure, heart rate, blood oxygen saturation, body weight and temperature – all require conscientious and contextual evaluation with mathematical methods to guide diagnosis. With their help, the essential parameters are to be found automatically and the physician is to be pointed to relevant contents. Especially when monitoring patients in intensive care units or in telemedicine, threatening conditions should be detected quickly and reliably and the treating staff should be informed by an alarm.

The problem with this form of automatic detection of important events is to develop a reliable algorithm with high accuracy –an almost impossible task given the diversity of patients and the variability in the signals. In other words, algorithms must be applicable to all possible variants of a given signal. This so-called *robustness* of an algorithm is an essential feature in the subsequent approval as a medical device. In addition, the measurement technology and algorithms must be absolutely insensitive to external electromagnetic interference to which a measurement setup may be exposed.

Mathematical methods and the development of software are therefore indispensable components of biosignal analysis. The success of modern monitoring systems therefore comes only to a small extent from the development of electronic hardware – the far greater contribution to the success of an innovative product is now made by intelligent signal evaluation. For example, in addition to exclusive filtering in the frequency or time domain, interfering signals are nowadays increasingly analysed using the wavelet transformation in the time-frequency domain, since it allows the best possible time and frequency resolution for a given signal section. The detection of important signal sections in an ECG signal course, for example, requires the extraction of statistically robust features, which are then fed to a classifier or a neural network for analysis. With this method, a large number of anomalies in ECGs can already be reliably distinguished today.

Model-based techniques such as the "Kalman" filter or "Markov" models are also used to detect abnormal (pathological) states and the associated changes in state. Early detection of trends or random fluctuations in a signal course, such as signal variations shortly before an epileptic seizure in the electroencephalogram (EEG), using

machine learning methods can be used to identify an upcoming seizure. These statistical methods are intended to distinguish stochastic fluctuations from deterministic ones or to recognise significant changes before they occur themselves. The mathematical relationships that can be used to predict the probability of a future change from current readings, in order to detect possible signs of a developing pathology, often lie in the theory of complex systems and non-linear dynamics. For example, in normal heart rhythms one observes stable rhythms (trajectories in phase space), so-called *attractors*, which, however, can sometimes pass into chaos via so-called *bifurcations* and thus into pathological states such as ventricular fibrillation. Perhaps in the not too distant future, this insight can be used for the early detection and diagnosis of such incidents to the benefit of the patient.

This book provides an introduction to the theory and principal methodology of biosignal processing, describes the origins of the most common human biosignals, and teaches the techniques for measurement and modern information processing using LTSpice and Matlab/Simulink[2]. After a brief introduction and historical review to the individual topics of electrophysiology and analogue and digital signal processing, the reader is introduced to the practice of biosignal processing with Matlab/Simulink through selected applications of the methodology learned.

The reader is given an overview of the variety of human biosignals (cf. Figure 1.1) and is introduced to the topic on the basis of selected biosignals, such as muscle activity in the electromyogram (EMG), the activity of the heart muscle in the electrocardiogram (ECG), the activity of the nerve cells of the brain in the electroencephalogram (EEG) or the measurement of the oxygen saturation of the blood in the photoplethysmogram (PPG). In this context, the fundamentals of deriving, preprocessing, recognising and interpreting these signals are taught with the help of the simulation environment LTSpice and the programming language Matlab/Simulink. Carrying out the exercises in Matlab/Simulink also teaches the necessary techniques of practical biosignal processing and offers the opportunity to apply acquired theoretical knowledge in practice. For a better overview of the formula symbols, units and constants used, a table sorted by chapter is provided in chapter 7 for reference.

The current state of research and development is presented in three fields of application from the authors' research topics: (i) mathematical modelling and analysis of signals from the heart / circulatory system, and analysis of the electrical activity of (ii) muscles and (iii) the brain. Each chapter includes a series of examples and exercises, as well as a presentation of the future perspectives of the respective field.

2 The MathWorks, Inc.

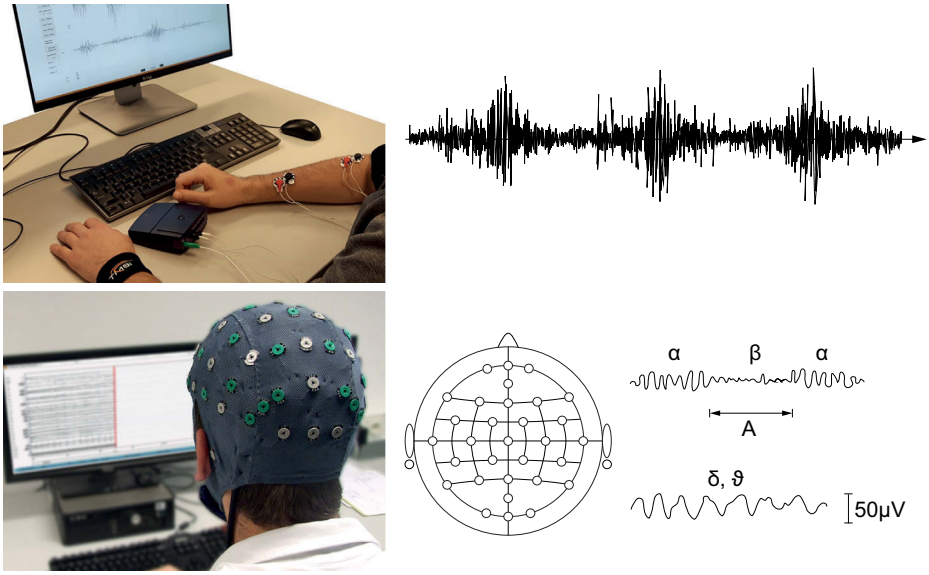


Fig. 1.1: Possible sources from the fields of electroencephalography (EEG) and electromyography (EMG) with illustration of the typical temporal signal sequences of the individual rhythms in the EEG (bottom) and of a repeated muscle contractions in the EMG (top).

2 Fundamentals of Information, Signal and System Theory

In biosignal processing, a variety of methods from different disciplines are applied to problems in medicine. The methods of classical information and signal processing form an important basis. The following chapter introduces methods from these disciplines that are commonly used in biosignal processing and discusses the context along practical examples.

2.1 Information and Information Transmission

The concept of information occupies a crucial place in all scientific disciplines, including biosignal processing. However, the concept of information is not uniformly defined in the various disciplines and has different meanings. The most common is the concept of information coined by C. E. Shannon[1], which is used in communications engineering.

Digital Information

However, the Shannonian definition of *digital information* refers exclusively to statistical aspects of information, i.e. the *probability of occurrence* of *characters* of a character set in a string to be transmitted. The statistical information content I of this sequence of n characters with the respective occurrence probability p_i of the individual characters is:

$$I = \sum_{i=1}^{n} \log_2 \frac{1}{p_i} = \log_2 \frac{1}{p_1} + \log_2 \frac{1}{p_2} + \cdots + \log_2 \frac{1}{p_n} \geq 0 \,, \quad n \in \mathbb{Z} \,. \tag{2.1}$$

According to this equation, the information increases steadily with the number n of characters in the chain, as long as the following relation holds for a character from a character set with N different characters:

$$0 \leq p_j \leq 1 \,, \quad \sum_{j=1}^{N} p_j = 1 \,. \tag{2.2}$$

Equation 2.2 stands for the existence of a character and at the same time guarantees that each character appears only once in the character set. From the two equations it becomes clear that the information content of a character string increases strongly

1 Claude Elwood Shannon (1916–2001), US mathematician and electrical engineer, founder of information theory

https://doi.org/10.1515/9783110736298-002

through rare characters, i.e. for small values of p_j. This connection is justified by the fact that rarely occurring characters of a character set usually have a special meaning in a message and therefore also carry a large information content. If one relates this fact to the German language, less frequently occurring letters have a higher information content and consequently a greater significance in the decoding of a text message. In fact, based on the occurrence probability of letters in an encrypted message, a decryption of the message can be achieved under certain conditions based on the total occurrence probability of the individual letters in the German language.

Data Set

An important result of statistical information theory is the consistent description of information transmission in terms of *symbols*. The information-bearing symbols consist of *one or more information units*, the so-called *digital bits*[2] with a *word width index W* of up to 10 bits per symbol, since larger word widths can usually no longer be clearly distinguished on the receiving side due to noise. Since a symbol has a different number of data bits depending on the coding, a distinction is made between the *data rate* and the *symbol rate* of a data stream.

Data Transmission Rate

Determining the data stream of a digital signal is of great importance, for example, for dimensioning a radio connection for data transmission from sensor networks via Bluetooth. If one first defines the symbol duration T_s as the transmission duration for a symbol, the so-called symbol rate or baud rate index $f_s = 1/T_s$ can be derived by forming the reciprocal value. The baud rate denotes the number of transmitted symbols per second. Common baud rates are between 9600 and 11, 520 baud (1 baud = 1 Bd = 1 symbol per second).[3]

The data transmission rate C is therefore the product of the symbol rate and the word width $C = D \cdot W$ and is given in bits per second. A Bluetooth 3.0 connection theoretically achieves data rates of up to 25 Mbit/s, so at 8 bits per byte a maximum of 3.125 MByte of data can be transmitted in one second.[4] Other data rates for comparison are given in Table 2.1.

2 bit is the short form for *binary digit*. 1 bit is the information content contained in a choice of two equally probable possibilities. The information content can be any real, non-negative value. These smallest information units have only two distinguishable states – zero and one. The number of bits is called *dataset D*. A symbol of D bits therefore has $Z = 2^D$ *distinguishable states*. In transmission technology, it is common to work with 4 to 6 bit/symbol, which leads to $2^4 = 16$ or $2^6 = 64$ different states for a symbol.

3 After Jean-Maurice-Émile Baudot, who invented the Baudot code in 1874.

4 The bit is usually abbreviated with a lower case b and the byte with an upper case B.

Tab. 2.1: Data rates of common signals.

Type	data rate
Eye	100 Mbit/s
Ear	50 kbit/s
Language	64 kbit/s
Audio CD (44.1 kHz, 16 bit)	700 kbit/s
Bluetooth	25 Mbit/s

However, these theoretical data rates are never achieved in practice, the reasons for this are on the one hand the available *bandwidth* of the *transmission channel* and on the other hand the existing interference on the transmission path (cf. Figure 2.1). A

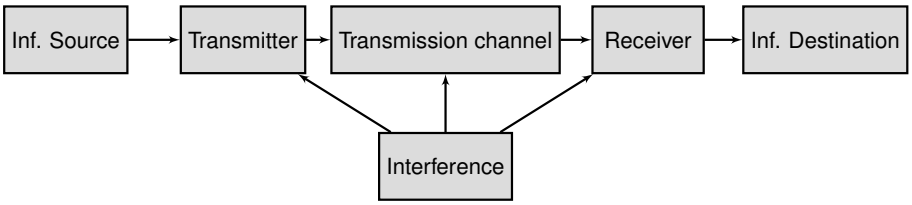


Fig. 2.1: Transmission path of information transmission under the influence of interference.

maximum data transmission rate $C_{\max}$ can be calculated using the *Shannon-Hartley* law as a function of the bandwidth B and the signal to noise ratio or *signal-noise* ratio *SNR* as follows:

$$C_{\max} = B \cdot \log_2(1 + SNR)\,. \tag{2.3}$$

Signal-to-Noise Ratio

In electrical engineering, the signal-noise-ratio is the ratio of the average power of the desired signal P_{signal} to the average noise power of the interfering signal P_{noise}. The average electric power of an AC circuit with resistive load R, instantaneous voltage u and instantaneous current i, is defined as $P = \overline{u \cdot i}$. Using Ohm's law $i = u/R$, the following relationship can be established for averaging over the period $T = t_0, \ldots, t_1$:

$$P = \overline{u \cdot i} = \frac{1}{T} \int_{t_0}^{t_1} u \cdot i \, \mathrm{d}t = \frac{1}{T} \int_{t_0}^{t_1} \frac{1}{R} u^2 \, \mathrm{d}t\,. \tag{2.4}$$

The *effective value* of an alternating quantity is the value of a current or a voltage at which a resistive load converts the same electrical power in a representative time $T = t_1 - t_0$ as a corresponding direct quantity. This means that the converted average power of the instantaneous values in the period $t_0, \ldots, t_1$ corresponds exactly to the

electrical power from the *effective values* of the respective AC voltages U_{eff} and currents I_{eff}. With the associated Ohm's law $I_{\text{eff}} = U_{\text{eff}}/R$, the following relationship holds:

$$P = \frac{1}{T}\int_{t_0}^{t_1} \frac{1}{R}\, u^2 \,\mathrm{d}t = U_{\text{eff}} \cdot I_{\text{eff}} = \frac{(U_{\text{eff}})^2}{R} \,. \tag{2.5}$$

Converting leads to an expression for the root mean square of an AC voltage, which is often referred to as RMS value[5]:

$$U_{\text{eff}} = \sqrt{\frac{1}{T}\int_{t_0}^{t_0+T} u^2 \mathrm{d}t} = \sqrt{\overline{u^2}} \,. \tag{2.6}$$

Corresponding equations apply to the RMS value of the current intensity and generalised to any other periodic or stochastic signal. The signal-noise-distance therefore corresponds exactly to the ratio of the squared RMS values of the effective voltages:

$$SNR = \frac{P_{\text{signal}}}{P_{\text{noise}}} = \frac{U^2_{\text{eff}_{\text{signal}}}}{U^2_{\text{eff}_{\text{noise}}}} \,. \tag{2.7}$$

Due to the large possible range of numbers, the signal-to-noise ratio is often given in logarithmic scale with the unit decibel as follows:

$$SNR = 10\log(SNR)\text{dB} = 10\log\left(\frac{U^2_{\text{eff}_{\text{Signal}}}}{U^2_{\text{eff}_{\text{Rauschen}}}}\right)\text{dB} = 20\log\left(\frac{U_{\text{eff}_{\text{Signal}}}}{U_{\text{eff}_{\text{Rauschen}}}}\right)\text{dB} \geq 0 \,, \tag{2.8}$$

where in the second line of the equation the squares of the RMS voltages have been pulled out of the logarithm for simplicity. According to the equation, the signal-noise-ratio $SNR \geq 0$, and a $SNR = 20\,\text{dB}$, for example, is equivalent to the ratio of the root mean square values of the amplitudes of the signal and the noise signal being 100 to 1, corresponding to a SNR of 100. Regardless of the available bandwidth B, this results in practice in data transmission rates of about a factor of 6.7 compared to the data transmission rate for $SNR = 1$, i.e. $C_{\max} = B \cdot \log_2(101) \approx B \cdot 6.7$.

Bandwidth and Modulation

The bandwidth of a signal describes the frequency interval (e.g. of a transmission channel of a radio link) in which the dominant frequency components of the signal to be transmitted or stored. It is characterised by a lower and an upper cut-off frequency $f_{\text{u}}, f_{\text{o}}$ of a signal. If the lower cut-off frequency is zero, one speaks of a *baseband position*, otherwise of a *bandpass position*. The difference in magnitude of the two cut-off frequency values is called the bandwidth: $B = |f_{\text{o}} - f_{\text{u}}|$. Table 2.2 shows common (bio-) signals and their approximate bandwidths in comparison.

5 The RMS value is the abbreviation for Root Mean Square

Tab. 2.2: Bandwidth of common signals compared to biosignals.

Signal / application	Approximate bandwidth
Nuclear magnetic resonance spectroscopy	0.1 Hz
Electroencephalogram (EEG)	100 Hz
Electrocardiogram (ECG)	300 Hz
Electromyogram (EMG)	5 kHz
Speech	3.6 kHz
Audio CD	22 kHz
Mobile radio (GSM)	200 kHz
FM broadcast signal	300 kHz
DVB-T	7 MHz
WLAN according to IEEE 802.11 a/b	22 MHz
Front Side Bus in the computer	400–800 MHz
Fibre – Ethernet	20–50 GHz

The bandwidth of the transmission channel, or the signals that are transmitted in it, depends on the information content to be transmitted. Since information cannot be transmitted via a single frequency[6], the information signal is modulated onto so-called *carrier signals*. The simplest type of analogue modulation are methods in which a signal parameter such as the amplitude or frequency or the phase position of the carrier signal are modulated by the information signal. In the case of amplitude modulation this is done by multiplying the carrier signal with the information signal (cf. Figure 2.2).

In todays technology, often digital modulation methods are used in addition to analogue amplitude and frequency modulation in order to optimally utilise the channel capacity. Corresponding to the analogue modulation such as amplitude modulation (AM) or frequency modulation (FM), in the digital procedures on the one hand in pulse amplitude modulation (PAM) the amplitude and on the other hand in pulse position modulation (PPM) resp. pulse width modulation (PWM), the frequency and phase position of the digital pulses are modulated with the amplitude values of the information signal (cf. Figure 2.3 and Listing 2.1).

The problem of every type of modulation, however, is the fundamental widening of the bandwidth depending on the information to be represented and the possible (metrologically still just distinguishable) resolution of the transmission path affected

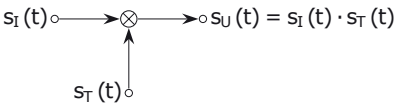


Fig. 2.2: Amplitude modulator, obtained by multiplying the information signal s_I with the carrier signal s_T leading to a transmission signal s_U.

6 A signal with a single frequency does not transmit information, since the measurable quantity does not undergo any change.

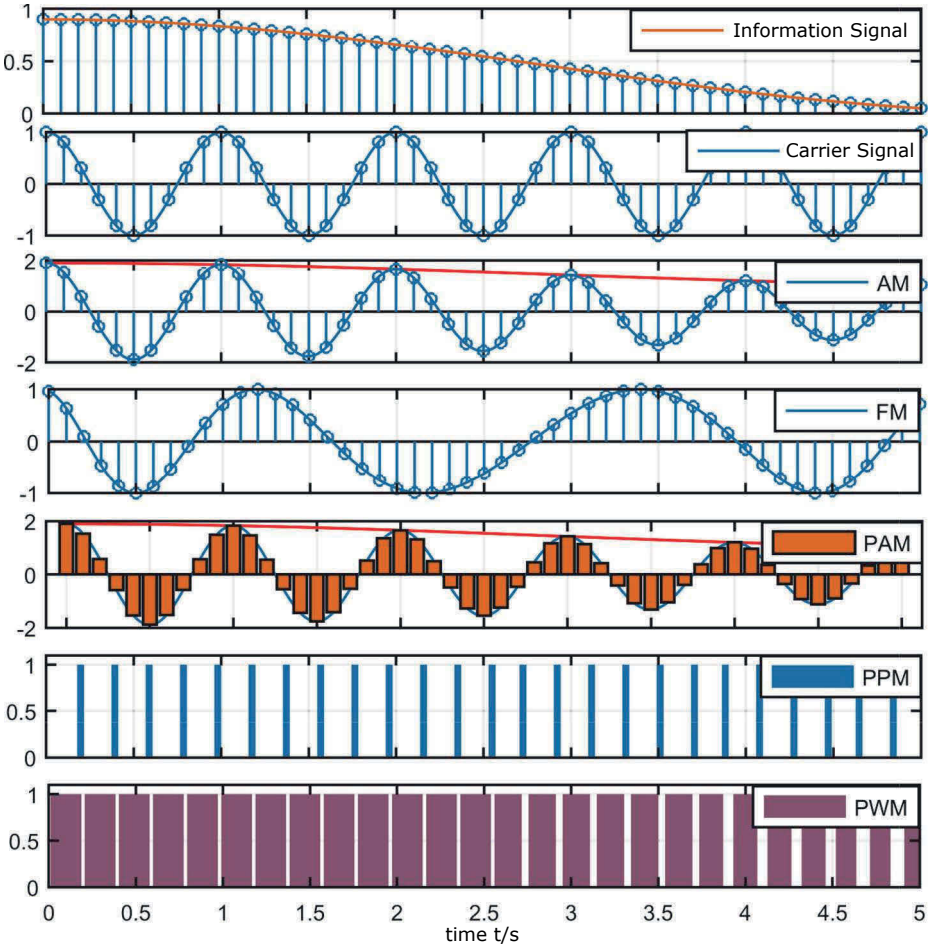


Fig. 2.3: Principle of the basic analogue and digital modulation methods for a given information and carrier signal carrier signal: amplitude modulation (AM), frequency modulation (FM), the frequency modulation (FM), pulse amplitude modulation (PAM), pulse position modulation (PPM) and the pulse width modulation (PWM) in relation to the analogue information signal (top).

by interference. In other words, the information of digital signals is also transmitted analogue in the form of harmonics, i.e. not only the carrier frequency f_T is transmitted, but a whole spectrum in the adjacent frequency ranges. The harmonics of a signal (cf. Figure 2.4) can be described mathematically with the help of the Fourier series:

$$f(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(kt) + b_k \sin(kt)) \,. \tag{2.9}$$

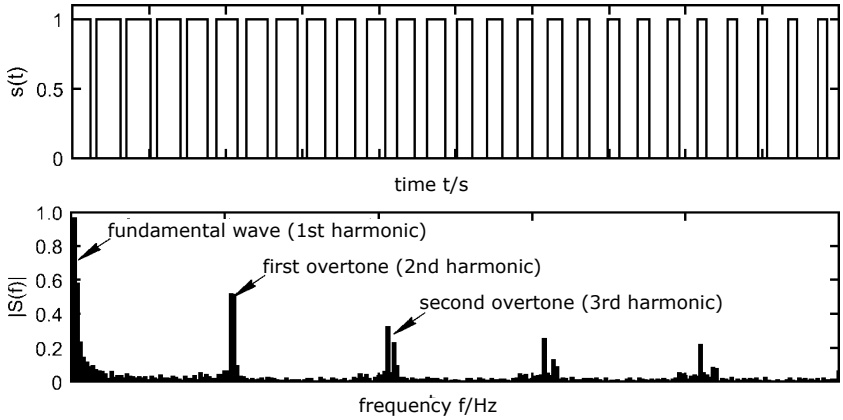


Fig. 2.4: Bandlimited pulse width modulation signal from Figure 2.3 with associated spectrum.

The coefficients a_k, b_k in Equation 2.9 indicate the real amplitudes of the individual sine and cosine harmonics of the k-th harmonics. For band-limited signals k is limited by a value K, $k \leq K < \infty$; they are also numerically only defined on the basis of this.

Listing 2.1: Matlab example for generating different modulation signals.

```
Am1 = 0.35; fm1 = 0.05;      % amplitudes and frequencies
Am2 = 0.25; fm2 = 0.1;       % of the harmonic
Ac = 1; fc = 1;

N=10; fs=N*fc;               % sampling frequency
t = 0:1/fs:0.25/fm1;         % time vector low fs (discrete)
th = 0:1/fs/10:0.25/fm1;     % time vector high fs (cont.)

% Information signal
is = 0.3 + Am1*cos(2*pi*fm1.*t) + Am2*cos(2*pi*fm2.*t);
ish =  0.3 + Am1*cos(2*pi*fm1.*th) + Am2*cos(2*pi*fm2.*th);
subplot(7,1,1)               % subplot for 7 plots among each other
stem(t,is)                   % discrete representation
hold on
plot(th,ish)                 % continuous representation

% Carrier signal
cs = Ac*cos(2*pi*fc.*t);
csh = Ac*cos(2*pi*fc.*th);
subplot(7,1,2)
stem(t,cs)                   % discrete representation
```

```
hold on
plot(th,csh)                % continuous representation

% Amplitude modulation
am = (1+is).*cs;
amh = (1+ish).*csh;
subplot(7,1,3)
plot(t,Ac.*(1+is),'r')
hold on
stem(t,am)                  % discrete representation
plot(th,amh)                % continuous representation

% Frequency modulation
mi = 16;
fm = sin(2*pi*fc*t+(mi.*is));
fmh = sin(2*pi*fc*th+(mi.*ish));
subplot(7,1,4);
stem(t,fm);                 % discrete representation
hold on
plot(th,fmh)                % continuous representation

% Pulse amplitude modulation
subplot(7,1,5)
plot(t,Ac.*(1+is),'r')
hold on
bar(t,am)                   % discrete representation
plot(th,amh)                % continuous representation

% Pulse position modulation
subplot(7,1,6);
ppm = modulate(is,fc,fs,'ppm');
tp = 1:length(ppm);
tp = 10*tp./N./length(is);  % appropriate time base
bar(tp,ppm, 2);             % discrete representation
axis([0 5 0 1.1])

% Pulse width modulation
subplot(7,1,7);
pwm = modulate(is,fc,fs,'pwm');
bar(tp,pwm,1);              % discrete representation
axis([0 5 0 1.1])
```

Acoustic Transmission Channel

The example of the human voice can be used to illustrate the above concept. The human voice consists, in addition to the fundamental frequency, of a large number of overtones, which also make up the timbre. The information of speech, singing, pitch and rhythm is modulated onto the fundamental by the speech apparatus. The transmission signal therefore transports a series of specially arranged vibrations that occupy a certain frequency band in the acoustic transmission channel. Usually, this transmission channel is disturbed by noise and ambient sounds, so that not all of the information always reaches the receiver. The receiver must also be able to understand the pre-agreed code of the speech. In the case of a foreign language, this might not be the case, and consequently neither information is transmitted nor is it followed by an action. In addition to the statistical description according to Shannon, a number of other aspects play a role in the transmission of information that are not yet included in the afore defined information theory. These are included in an expanded concept of information.

Extended Information

The concept of information according to Shannon is of great importance in the transmission and processing of signals and data, but this concept does not make any statements about the meaning of messages, i.e. about their *semantics*. Accordingly, the term information in this representation does not include the meaning of the signs. This is addressed by an extended information concept of natural law information coined by W. Gitt[7]. With its help, the transmission of information can also be described in five different levels shown in Figure 2.5. In the lowest level of this concept, the Shannonian information concept of statistics is found, above it the levels of syntax, semantics, pragmatics and apobetics.

The statistics describe the variety of possible characters, the syntax their arrangement and the semantics their meaning. Pragmatics and apobetics ultimately refer to the understanding, interpretation and execution of an information message; here the message is decoded, interpreted and executed with the agreed code.

According to W. Gitt [19], extended information I^+ only exists if all five hierarchical levels, i.e. statistics, syntax, semantics, pragmatics and apobedtics (according to Figure 2.5), are present in an observable system and the empirical theorems of natural law information theory (NGIT) established by him apply:

1. A material quantity cannot give rise to a non-material quantity.
2. Information is a non-material fundamental quantity.
3. Information is the non-material basis for all program-controlled technical systems and for all biological systems.
4. There is no information without code.

7 Werner Gitt (1937) is a German engineer and founder of natural law information theory

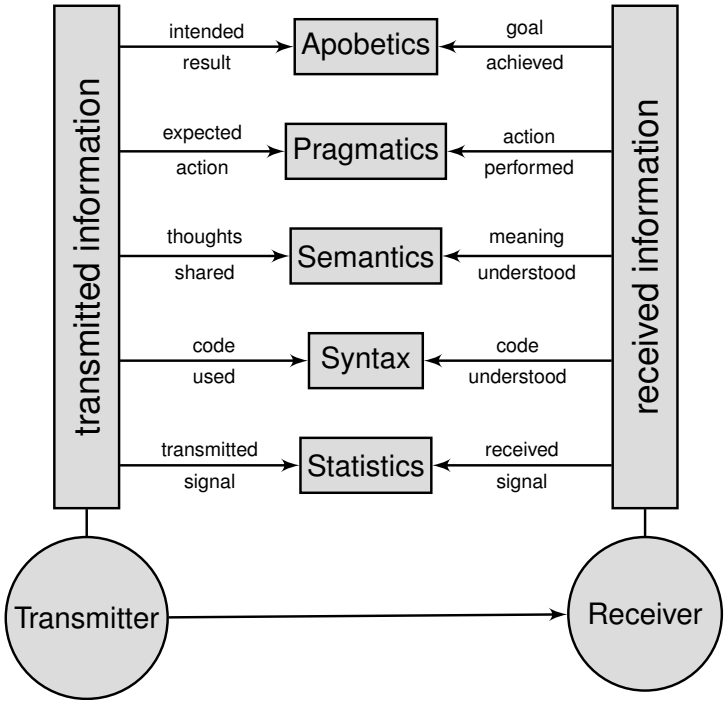


Fig. 2.5: Information transmission according to the concept of W. Gitt's extended natural law information theory [19].

5. Every code is the result of a free will agreement.
6. There is no new information without an intelligent and volitional transmitter.
7. Information at the end of a chain of transmission can be traced back to an intelligent source.
8. Assigning meaning to a set of symbols is a mental process that requires intelligence.
9. No information can be created in statistical processes.
10. Information storage / transmission occurs only on energetic or material carriers.

Definition of the Term Biosignal

With the help of the extended information, the term biosignal can be defined as follows: A biosignal is an energetically-materially-measurable, physical quantity of a living individual in which the diagnostic information of the physiological correlations are encoded in an unknown way. In the case of a biosignal, a further distinction can be made between so-called *evoked* and so-called *autonomous* which cannot/can be influenced by freewill. An important insight that can be drawn from this for biosignal processing is, that the diagnostic information about the function or malfunction of a

physiological system is expressed in the form of signals and the appropriate analysis of the generated signals consequently leads to diagnostically usable system knowledge. If one transfers this knowledge to the process of analogue modulation in Figure 2.2, the information signal $s_I(t)$ in the figure thus becomes the partially unknown influence quantity or the physiological information of a biological system to be decoded and the carrier signal $s_T(t)$ the non-informative part of the biosignal and consequently $s_U(t)$ the measured biosignal. For this reason, the analysis of a biosignal is often done like a decoding of an encrypted signal with incomplete knowledge of the code used.

2.2 Connection between Signals and Systems

In classical signal processing, the processing and transmission of information generally takes place in the form of signals, i.e. the information is encoded by a measurable change in a physical quantity. This quantity can be, for example, the change of an electrical potential difference on the surface of the human body, as in the electrocardiogram, or also the local change of the magnetic field vector of a data bit on a data carrier, e.g. a hard disk. The methods for processing the information are basically the same; they usually take on the character of a *signal processing* system. This can be, for example, a simple shifting of the signal, or as shown in subsubsection 5.3.4.2, a digital filtering or the like (cf. Figure 2.6).

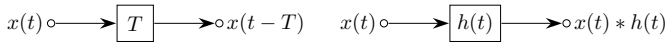


Fig. 2.6: Two different digital systems: a time-shifting system that delays the input signal $x(t)$ by T (left) and a digital convolution filter that convolves the input signal $x(t)$ with the impulse response $h(t)$ of the system (right).

In biosignal processing, signal-processing systems are used in particular for signal conditioning, i.e. freeing the signal from interference. However, in the subsequent analysis of the biosignals for diagnosis finding, it is a matter of identifying important diagnostic parameters of a *signal-generating* system. An example of such a signal generating system is the human heart. This *physical* system has a fundamentally different importance compared to the signal-processing systems. These systems generate important diagnostic information during operation, such as an electrically measurable ECG signal of heart excitation.

To illustrate this fact, let us take an exemplary look at the system character and the biosignals generated by the human heart. Consisting of four blood-filled cavities separated by heart valves and a conduction system of nerve and muscle fibres, the heart is stimulated by nerve impulses to cyclic contractions of the heart muscle. In doing so, it generates a multitude of biosignals which describe the state of the system

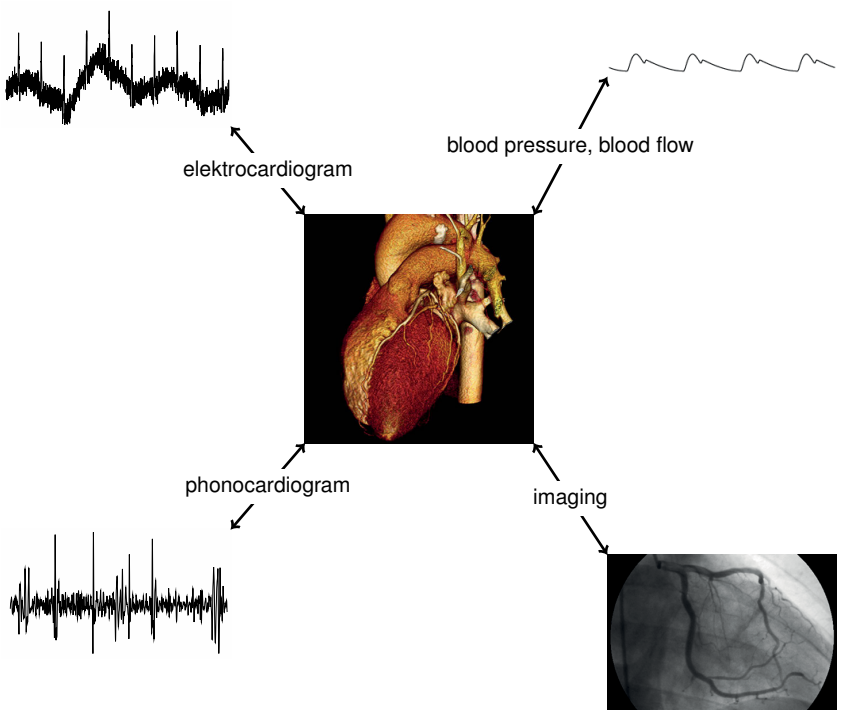


Fig. 2.7: Human heart with the common vital signs from which possible diagnoses can be derived.

"heart" in the respective form (cf. Figure 2.7). These signals are therefore called *state variables* or variables of a system. In addition to the signals of the electrical conduction system of the heart and the muscle contractions themselves, signals such as pressure and flow changes of the blood circulation and the associated flow sounds as well as the sounds of the heart valves can also be measured as acoustic signals. Figure 2.7 shows the most common vital signals of the human heart from which possible diagnoses can be derived. Figure 2.8 shows an overview of the signals of the heart: The pressure curve in the aorta and in the left ventricle as well as the volume of the ventricle indicate flow-related diseases, the electrocardiogram gives hints about the electromechanical functioning of the conduction system and the phonocardiogram about the function of the heart valves.

Each of these signals contains only a part of the total information and is also often only partially known in relation to the biophysical system of the heart. The individual signals thus encode, only very specific partial information of the system state with an only partially known code. For example, if there is an aortic valve defect or a conduction block in the conduction system of the heart, only certain parts in the signal change accordingly. In the first case, one will probably hardly notice a change in the ECG, but the heart tones will sound different. In the second case, the contraction of the vent-

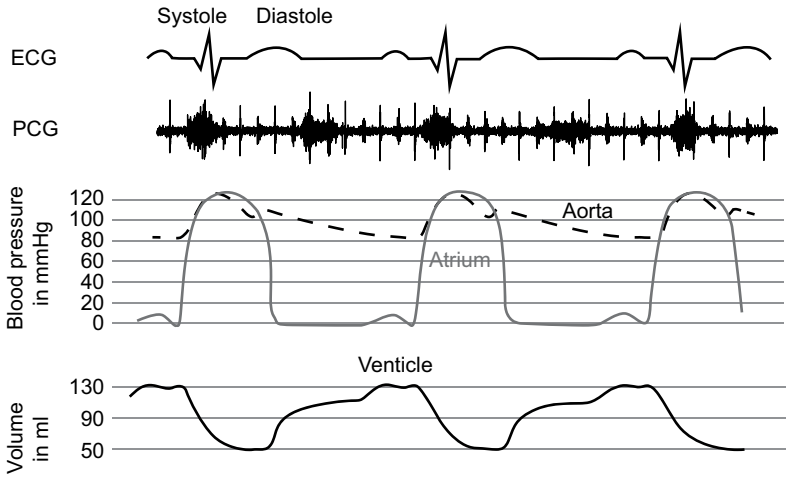


Fig. 2.8: Vital signals of a human heart: Sequence of three heart cycles in the electrocardiogram (ECG), phonocardiogram (PCG) and the corresponding pressure sequence in the left ventricle and the aorta (centre), as well as the ventricular volume (bottom).

ricle will be delayed, resulting in a change in the ECG signals. Thus there exist a more or less pronounced connection between the signal-generating system "heart" and the respective biosignal. In mathematical modelling, this signal-system-relation is called *sensitivity* of a state variable (or measured variable) of the system to the change of a physical system parameter (e.g. the conductivity of the conduction system). In signal processing, such dependencies are made clear with the help of correlation. As will be shown later in the book in subsection 2.3.6, one speaks of a strong/weak correlation of the signal quantities.

One of the main concerns of biosignal processing is the analysis and diagnostic use of these *signal-system* correlations. Due to the complexity and variability of living systems, the correlations are usually not clear, not pronounced enough, incomplete, or covered by artefacts and affected by uncertainties. Mathematical models are helpful in searching, as they greatly improve the basic understanding of the system interrelationships. For example, model systems can be parameterised by exact quantities and their synthetic signals can be analysed possessing the "ground truth". Often, new signal-system-coherences or signal components are found which could not yet be measured on the physiological system for technical reasons (for example, due to unsuitable sensors or measurement locations, too low time resolution, too strong artefacts or filtering of the signals, etc.). A combination of mathematical modelling, sensitivity analysis and adaptation of the measurement technology and evaluation algorithms often proves to be very profitable in the search for the correct parameters or measured variables in the biosignal.

2.3 Definition and Classification of Signals

The methods of biosignal processing refer more often to different classes and definitions of signals. These are also of great importance in continuous signal processing as special test functions in the proof. In the following, we will mathematically define important signals for the course of the book and classify them according to their properties.

2.3.1 Univariate and Multivariate Signals

A common mathematical representation of signals can be achieved through the notions of dependent and independent quantities. According to this definition, a signal is a physical quantity that depends on *one* or *several* independent quantities. If there is dependence on only one variable, signals are referred to in statistics as *univariate* signals, whereas if there is dependence on several independent variables, signals are referred to as *multivariate* signals. Another common definition of multivariate signals concerns the measurement of several signals (e.g. for an EEG) in a measurement arrangement with several sensors. If there is a common dependence on an independent quantity, such as time, when measuring M *univariate* signals $x_1(t), x_2(t), \ldots, x_\mathrm{M}(t)$, the following definition applies to the resulting M-dimensional *multivariate* signal:

$$X_\mathrm{m} = \{x_1(t), x_2(t), \ldots, x_\mathrm{M}(t)\} \quad \forall m \in \mathbb{N}\,. \tag{2.10}$$

In most cases, however, a biosignal is simply the temporal course (time is thus the independent quantity) of a physical (dependent) quantity such as the electrical voltage U – the associated signal thus becomes $U(t)$ or U_t for short. In addition to time as an independent quantity, the location in its three spatial directions (x, y, z) is used in the representation of image signals as two- or three-dimensional spatially resolved intensity signals $I(x, y)$ and $I(x, y, z)$, respectively. If these image signals also have a time dependence, an image signal sequence $I(t, x, y, z)$ is obtained as a function of four independent variables. In principle, the methods of signal processing can also be applied to image or video signals, but in this book we restrict ourselves to signal sequences that depend on only one independent variable, i.e. to univariate or *scalar* signals. The special features of the evaluation of multivariate signals are only dealt with in part in this book in section 6.1.

Figure 2.9 shows on the left a time-dependent raw ECG signal $U(t)$ acquired over a time range of $t = 0, \ldots, 10\,\mathrm{s}$, and on the right a two-dimensional MRI image with a resolution of $x = y = 512$ pixels.

In principle, all physical quantities such as pressure, temperature, voltage, current, etc. can be considered as dependent quantities of a signal. Normally, however, it is an indirect electrical, easily measurable quantity that is assigned to the direct phys-

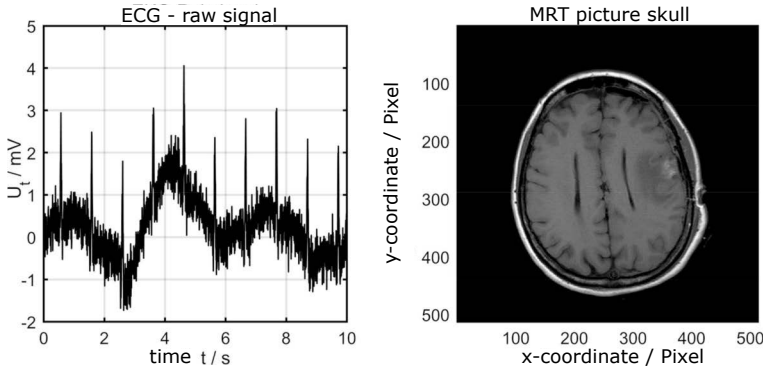


Fig. 2.9: Example of a time-dependent raw ECG signal $U(t)$ (left) and a two-dimensional MRI image $I(x, y)$ showing a section through a skull (right)

ical quantity of the (physiological) process via a signal transformation relationship. A detailed description of this can be found in section 4.3.

2.3.2 Periodic, Quasi-Periodic, Aperiodic and Transient Signals

Periodic processes are frequently encountered in science and technology, whereas *exact-periodic* are rarely found in living nature due to large natural fluctuations. In contrast to very precise oscillators in technology, such as those built into quartz watches, the periodicity of physiological oscillators is often subject to great inaccuracies. However, these so-called *quasi-periodic* processes allow the body to adapt quickly and effectively to changing external conditions and, unlike exact periodic processes, remain insensitive to unwanted perturbations. In section 3.2 it is shown to what extent these physiological fluctuations have an impact in the development of diseases of the heart, and in subsubsection 6.3.1.1 how they can be used for diagnosis.

A mathematical definition for exactly-periodic processes can be found in the harmonic functions and the angular frequency ω. Exact-periodic processes are, as we will see in the next section, always deterministic in nature, i.e. completely predictable, and exist in analytical form. This is also the reason why every exactly-periodic process can be expressed by an *linear combination* of harmonic functions in the form of the Fourier series.

The most important harmonic functions for the analysis of signals are thus trigonometric functions sine and cosine. In science and technology, harmonic functions are ubiquitous for describing oscillations and waves. The time dependent sine function is given by:

$$y(t) = A \sin(\omega_0 t + \varphi_0) = A \sin(\omega_0 (t + t_0)) \,, \tag{2.11}$$

where A is the amplitude, φ_0 is the zero phase angle and ω_0 is the angular frequency of the harmonic function. The zero phase φ_0 can alternatively be written down as the time shift $t_0 = \varphi/\omega_0$. The angular frequency is defined as

$$\omega_0 = \frac{2\pi}{T_0} = 2\pi f_0 \,, \tag{2.12}$$

i.e. the frequency f_0 of the periodic process multiplied by 2π, where f_0 is equal to the reciprocal of the period T_0 of the function $y(t)$:

$$f_0 = \frac{1}{T_0} = \frac{\omega_0}{2\pi} \,. \tag{2.13}$$

Harmonic signals with zero phase angle can be represented as the sum of sine and cosine functions with the help of the addition theorems. A detailed description can be found in [90]. In signal processing, the *pointer representation* in the complex plane has become established for the calculation of harmonically excited systems. Between the two representations can be converted with the help of the Eulerian relation as follows:

$$e^{j\varphi} = \cos\varphi + j\sin\varphi \,. \tag{2.14}$$

Thus a cosine and sine function can be understood as the real and imaginary part of a complex exponential function, respectively:

$$z(t) = Ae^{j(\omega_0 t+\varphi)} = Ae^{j\varphi}e^{j(\omega_0 t)} = A\left(\cos(\omega_0 t+\varphi) + j\sin(\omega_0 t+\varphi)\right) \,. \tag{2.15}$$

This mathematical representation can be represented by a pointer of length A, which rotates in the complex plane around the coordinate origin. The time for a full rotation is the period duration T_0. In this representation, the cosine function $x(t)$ results from the projection of the pointer onto the real axis $\Re\{z\}$, while the sine function $y(t)$ represents the projection onto the imaginary axis $\Im\{z\}$. To illustrate this common form of representation, the projection of the complex pointer onto the imaginary axis in the complex plane is shown in Figure 2.10.

Mathematically, one defines periodic signals by requiring a constant periodic duration $T_0 = \text{const.} \forall t \in \mathbb{R}$. The function $s_{\text{per}}(t)$ repeats exactly after all k multiples of T_0, with $k \in \mathbb{Z}$. This class of signals also includes arbitrary superpositions of harmonic signals as expressed in the real form of the Fourier series:

$$s_{\text{per}}(t) = \frac{a_0}{2} + \sum a_k \cos(k\omega_0 t) + b_k \sin(k\omega_0 t) \,, \quad \omega_0 = \frac{2\pi}{T_0} \,. \tag{2.16}$$

The Fourier coefficients a_k, b_k of the equation indicate the real amplitudes of the sine and cosine oscillations (cf. section 2.4):

$$\begin{aligned} a_0 &= \frac{2}{T_0}\int s_{\text{per}}(t)\mathrm{d}t \,, \\ a_k &= \frac{2}{T_0}\int s_{\text{per}}(t)\cos(k\omega_0 t)\mathrm{d}t \,, \\ b_k &= \frac{2}{T_0}\int s_{\text{per}}(t)\sin(k\omega_0 t)\mathrm{d}t \,, \quad k = 1, 2, 3, \ldots . \end{aligned} \tag{2.17}$$

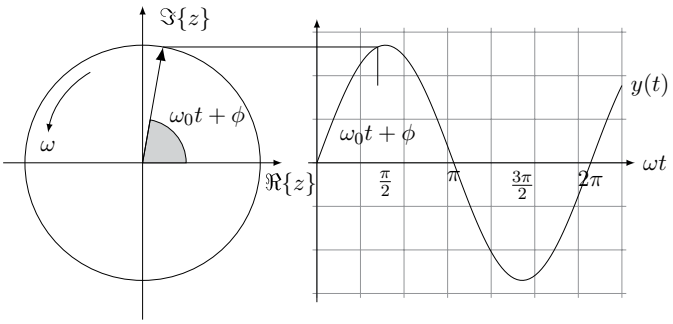


Fig. 2.10: Projection of a complex pointer onto the imaginary axis in the pointer diagram (left) and the plot of the resulting harmonic sine function $y(t)$ (right).

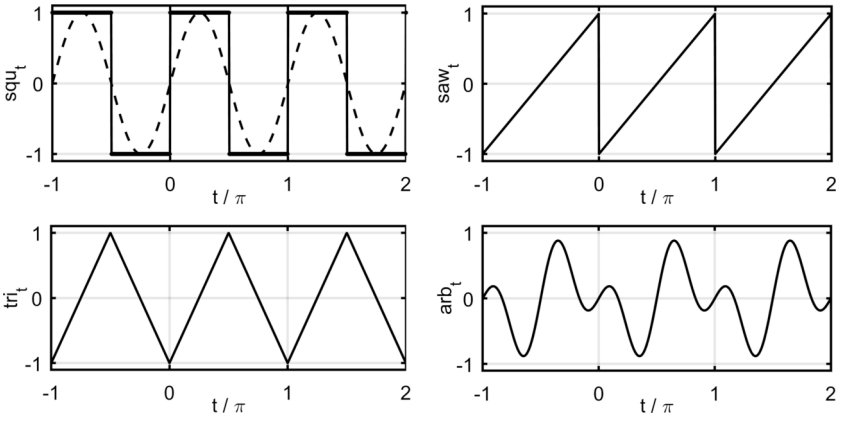


Fig. 2.11: Examples of periodic signals: rectangular signal (top left), sawtooth signal (top right), triangular signal (bottom left) and arbitrary signal from the superposition of two sinusoidal signals (bottom right).

In the limiting case for $k \to \infty$, any periodic signals such as the rectangular, triangular or sawtooth signal can be represented. In practice, however, when representing these signals, one works with a finite Fourier series, i.e. with $k \to N$, $k \leq N$, $N \in \mathbb{N}$.

The plots of the signals in Figure 2.11 were generated with Matlab (cf. Listing 2.3.2). In the following applications with Matlab, we omit labelling elements such as axis labels and titles for reasons of space.

Listing 2.3.2: Matlab example for generating periodic signals with the Fourier series.

```
f = 1;                          % fundamental frequency of signals
N = 1000;                       % number of places of support
t = linspace(-pi,2*f*pi,N);     % time vector between -pi and pi
```

```
% Generation of signals and graphical representation in the subplot
subplot(221)
squ = square(2*t);
hold on
plot(t/pi,squ,'.-',t/pi, sin(2*f*t))
ylabel('squ_t')
xlabel('t / \pi')
title('Rectangular signal')
grid on

subplot(222)
saw = sawtooth(2*f*t);
plot(t/pi,saw)
ylabel('saw_t')
xlabel('t / \pi')
title('Sawtooth signal')
grid on

subplot(223)
tri = triangle(2*f*t);
plot(t/pi,tri)
ylabel('tri_t')
xlabel('t / \pi')
title('Triangle signal')
grid on

subplot(224)
arb = -0.5*sin(2*f*t)+0.5*sin(2*2*f*t);
plot(t/pi,arb)
ylabel('arb_t')
xlabel('t / \pi')
title('Arbitrary superposition of harmonic signals')
grid on
```

Quasi-periodic signals, such as the ECG signal in Figure 2.12, however, do not exactly fulfil this condition. The period T_0 changes with each current period of the signal by $\pm\Delta T$. This behaviour can be expressed by the so-called *instantaneous* period duration $T_i = \{T_1, T_2, \ldots, T_N\}, i \in \mathbb{N}$:

$$\begin{aligned} s(t) &= s(t + kT_0) \quad \text{(periodic)} \\ s(t) &\approx s(t + kT_0) \quad \text{(quasi-periodic)} \end{aligned} \qquad \forall t\,, \quad k \in \mathbb{N}\,. \tag{2.18}$$

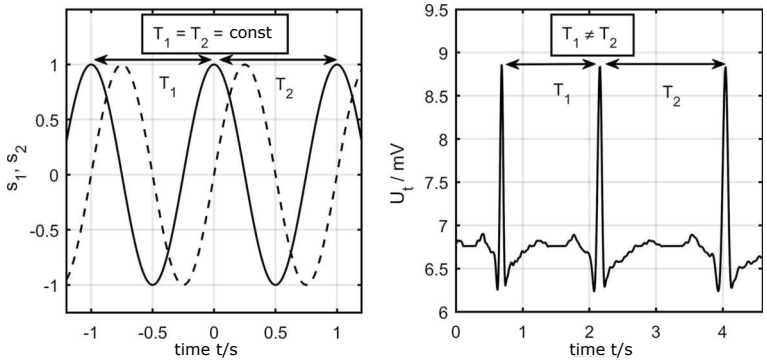


Fig. 2.12: Examples of periodic signals $s_1(t)$ and $s_2(t)$ from exactly-periodic processes (left): the quasi-periodic ECG signal $s_3(t)$, on the other hand (right), comes from a process that is subject to uncertainties. Its period $T_1, T_2, T_3, \ldots$ changes with time.

Aperiodic signals are defined by the absence of a period, i.e. by the condition $T_0 \to \infty$. To this category belong, on the one hand, arbitrary *transient* signals and, on the other hand, special functionals of system theory such as impulses and step functions, which model the on,/off switching or redirection of processes. These signals include both the purely monotonic increasing exponential functions $s_1(t) = e^{-0,25t}$ or their products with harmonic functions $s_2(t) = \sin(2\pi t)\, e^{-0,25t}$ as well as the density function of the "standard" normal distribution (cf. Figure 2.13)

$$\mathcal{N}(\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} . \tag{2.19}$$

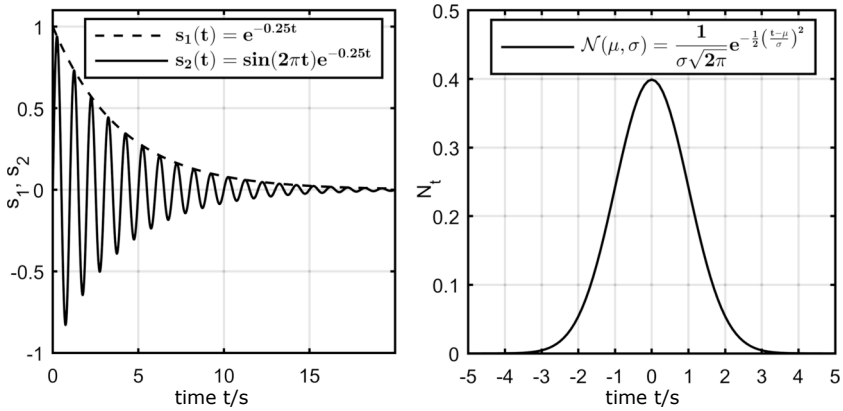


Fig. 2.13: The two transient signals $s_1(t)$ and $s_2(t)$ (left) and the pulse signal of the normal distribution $\mathcal{N}(\mu = 0, \sigma = 1)$ (right).

Another important signal form in signal processing is the so-called delta-distribution (also called δ-function, Dirac-function, Dirac-pulse[8] and unit impulse). Although for the function value of this distribution holds: $\delta(t = 0) = \infty$, in discrete signal processing it is defined for numerical reasons as follows:

$$\delta(t) = \begin{cases} 1 & \text{if } t = 0 \\ 0 & \text{otherwise } t \neq 0 \end{cases}, \quad t \subset \mathbb{R}\,. \tag{2.20}$$

This distribution, like all other distributions, can be understood as the limit of a function series such as the *Dirac series*. In the following, two common approximations for the delta-distribution $\delta_\epsilon(t)$ are given. In the limit $\epsilon \to 0$ the continuously differentiable normal distribution

$$\delta_\epsilon(t) = \frac{1}{\sqrt{2\pi\epsilon}}\, e^{\left(-\frac{t^2}{2\epsilon}\right)} \tag{2.21}$$

produces functions with very narrow and high maxima at $t = 0$. The area under the functions always has the value one which is a conservation variable for all ϵ. Consequently, the mean width $\sqrt{\epsilon} \to 0$ becomes narrower and narrower in the limit transition, while the height $1/\sqrt{\epsilon} \to \infty$ of the function conversely increases strongly. For $\epsilon \to 0$, this results in an infinitely narrow and infinitely high momentum, the so-called Dirac momentum (cf. Figure 2.14). Since the values "infinitely narrow" and "infinitely high" are not usable in discrete signal processing for numerical reasons, one has agreed on the representation of the *weight* . The weight corresponds exactly to the area of the pulse and thus has the value one.

Another common approximation results from the only piecewise continuously differentiable function of the rectangular pulse

$$\delta_\epsilon(t) = \frac{\mathrm{rect}(t/\epsilon)}{\epsilon} = \begin{cases} \frac{1}{\epsilon} & |t| \leq \frac{\epsilon}{2} \\ 0 & \text{otherwise} \end{cases}. \tag{2.22}$$

In this case, the limit value consideration $\epsilon \to 0$ leads to an infinitely narrow and infinitely high impulse with an area of one. Regardless of the Dirac series used, the delta-distribution has special properties that play an important role in the digitisation of signals. To be mentioned is the so-called *equation property* or sieve property

$$\langle \delta, f \rangle = \int_{-\infty}^{\infty} \delta(t)\, f(t)\, \mathrm{d}t = f(0)\,, \tag{2.23}$$

which, when multiplying a function $f(t)$ by the delta-distribution, hides all function values for $t \neq 0$, i.e. only the product at the point $t = 0$ is different from zero and has

8 After the physicist Paul Dirac.

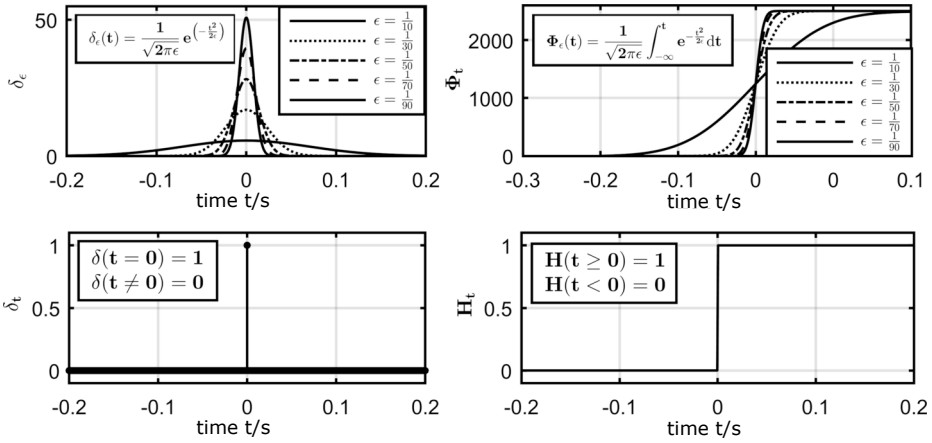


Fig. 2.14: Dirac series of the normal distribution $\delta_\epsilon(t)$ as a boundary transition to the delta-distribution $\delta(t)$: Top left the density distribution of the normal distribution $\delta_\epsilon(t)$ and top right the distribution function of the normal distribution $\Phi_\epsilon(t)$ for $\mu = 0, \epsilon \to 0$. The lower left shows the delta-distribution used in discrete signal distribution $\delta(t)$ applied in discrete signal processing and on the right the Heaviside-function $H(t)$ as the density distribution of the Delta-distribution.

the function value $f(0)$. The shift of the delta-distribution by a leads to

$$\int_{-\infty}^{\infty} f(t)\,\delta(t-a)\,\mathrm{d}t = \int_{-\infty}^{\infty} f(t)\,\delta(a-t)\,\mathrm{d}t = f(a) \tag{2.24}$$

and thus hides all values of the function $f(t)$ at the places $t \neq a$. For the case of the constant function $f(t) = 1$, the weight of the delta-distribution is again obtained:

$$\int_{-\infty}^{\infty} \delta(t-a)\,\mathrm{d}t = 1\,. \tag{2.25}$$

Another important signal form results from the limit value consideration of the cumulative sum of the density function of the normal distribution (cf. Figure 2.14), i.e. the cumulative integral from $-\infty$ to t:

$$\Phi_\epsilon(t) = \frac{1}{\sqrt{2\pi\epsilon}} \int_{-\infty}^{t} e^{-\frac{t^2}{2\epsilon}}\,\mathrm{d}t\,, \tag{2.26}$$

the so-called distribution function of the normal distribution. In the limit value consideration $\epsilon \to 0$ one obtains the so-called Heaviside-function (cf. Figure 2.14)[9], it is also called step or unit step function. It has a function value of zero for arguments less

9 After the British mathematician and physicist Oliver Heaviside

than zero and a function value of one otherwise. The Heaviside-function is continuous everywhere except at $t = 0$:

$$H(t) = \begin{cases} 1 & t \geq 0 \\ 0 & t < 0 \end{cases} . \tag{2.27}$$

In the technical literature, there is also a different nomenclature. In systems theory, the symbol $1(t)$ or $u(t)$ is used after the English term *unit step function*. The Heaviside-unction is the integral of the Dirac-distribution

$$H(t) := \int_{-\infty}^{t} \delta(s)\,\mathrm{d}s , \tag{2.28}$$

and consequently the derivative of the Heaviside-function is the diracian delta-distribution.

2.3.3 Even and Odd Signals

The consideration of symmetries of a signal can be helpful in the decomposition and description of a signal. Periodic signals can basically be decomposed into an even and an odd signal component. A signal is called even if it is axisymmetrical to the ordinate. Odd signals have point symmetry to the origin (cf. Figure 2.15).

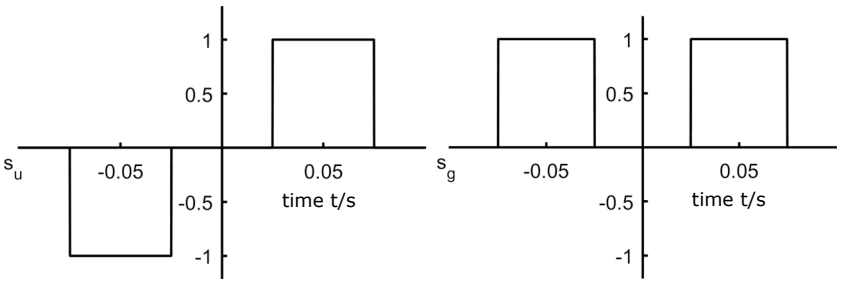


Fig. 2.15: An odd signal $s_u(t)$ of two square pulses (left) and an even signal $s_g(t)$ of two square pulses (right).

Mathematically, this symmetry can be expressed as follows:

$$s(t) = \begin{cases} s(-t) & \text{even} \\ -s(-t) & \text{odd} \end{cases} , \quad \forall t . \tag{2.29}$$

As we will see later in the introduction of the Fourier-transformation in subsection 2.4.1, even signal components can only be expressed by a linear combination of even harmonic functions such as the cosine and odd signal components can only be expressed by a linear combination of odd harmonic functions such as the sine.

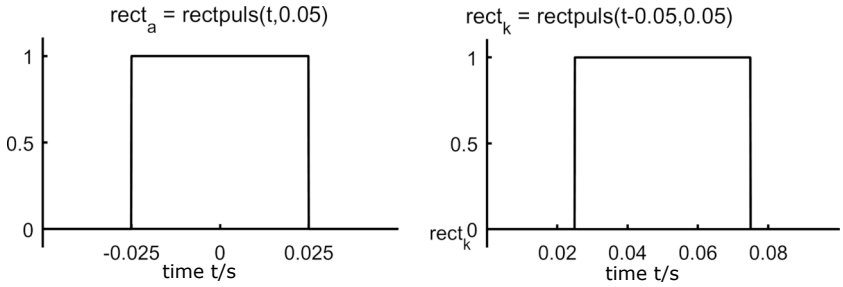


Fig. 2.16: Rectangular pulse in the form of an acausal signal $s_a(t)$ (left) and a causal signal $s_k(t)$ (right).

2.3.4 Causal and Acausal Signals

Causality as a property is often used in the classification of signals and systems to describe switch on/off processes. Figure 2.16 shows a *causal* signal (left), where the switch-on process is in the past, and a *causal* signal (right), which was only switched on at positive time.

Mathematically, a signal $s(t)$ is called causal if it does not exist or is identical to 0 for all times $t < 0$. If this condition is not fulfilled, an acausal signal is present, thus:

$$s(t)_{\text{causal}} = \begin{cases} s(t) & t \geq 0 \\ 0 & t < 0 \end{cases}, \quad s(t)_{\text{acausal}} = \begin{cases} s(t) & t \geq 0 \\ s(t) \neq 0 & t < 0 \end{cases}, \quad \forall t\,. \tag{2.30}$$

The classification causal/acausal also finds its application in the description of systems. In contrast to signals, the causal connection, i.e. that an effect (output values of a system) cannot emerge before the cause (input values of a system), is only fulfilled for systems whose output values are equal to zero for times $t < 0$. Systems that do not fulfil this condition are called acausal and also have no real physical cause-effect relation.

2.3.5 Energy and Power Signals

The concepts of energy and power are fundamental quantities in physics that can be used, for example, to determine the electrical energy required for the displacement of an electron in the electrical field of a capacitor. The assignment of these quantities to a purely analytical signal, on the other hand, is not obvious, since mathematical functions have no physical dimension. In section 2.1 the process of information transfer through the energisation or materialisation of information into information signals is clarified. Signals are thus the energetic or materialised form of information,

consequently their transmission requires energy [10]. However, this refers to the transmission process and not to the signal itself. If one assumes an energised form of the signals, the following relationships can be defined for energy and power signals of signal processing, following electrical engineering: The energy of a complex-valued, time-continuous and dimensionless signal $s(t)$ is determined by

$$E(s_t) = \int_{-\infty}^{\infty} s(t) \cdot s^*(t) \mathrm{d}t = \int_{-\infty}^{\infty} |s(t)|^2 \mathrm{d}t \,. \tag{2.31}$$

The integral is to be calculated section by section over all time intervals in which $s(t)$ is defined. For dimensionally affected quantities, such as the electrical voltage U in volts (V), the signal energy $E(s_t)$ then has the dimension $\mathrm{V}^2 s$ as in electrical engineering.

Based on the signal energy, one distinguishes two classes of signals, the so-called *energy signals* with finite but non-zero energy and the *power signals* with infinite energy but finite mean power. Accordingly, a signal $s(t)$ is an energy signal if it satisfies the following condition:

$$E(s_t) = \int_{-\infty}^{\infty} s(t) \cdot s^*(t) \mathrm{d}t = \int_{-\infty}^{\infty} |s(t)|^2 \mathrm{d}t < \infty \,. \tag{2.32}$$

The energy signal is therefore a square-integrable function and has a non-vanishing energy E. Typical energy signals are all signals with finite signal values that are switched on and off at some point. This class of signals can be assigned to oscillation and decay processes or to time-limited pulse-shaped signals. The following examples are intended to clarify the concept of an energy signal: A square-wave pulse $s(t) = A\,\mathrm{rect}(T)$ with the amplitude A and the width T is, according to the above definition, an energy signal with the energy

$$E = A^2 \int_0^T \mathrm{rect}^2(T) \mathrm{d}t = A^2 T < \infty \,. \tag{2.33}$$

The energy calculation for a delta-distribution,

$$\delta_\epsilon(t) = \frac{\theta}{2\epsilon} \,\mathrm{rect}\left(\frac{2\epsilon}{\theta}\right) \,, \quad \theta \to \infty \,, \tag{2.34}$$

on the other hand, does not lead to an energy signal, since:

$$E(\delta_\epsilon) = \frac{\theta^2}{4\epsilon^2} \frac{2\epsilon}{\theta} = \frac{\theta}{2\epsilon} \to \infty \,, \quad \text{für} \quad \theta \to \infty \,, \;\; 0 < \epsilon < \infty \,. \tag{2.35}$$

The limiting process from the rectangular function to the delta distribution thus leads to a change in the energy properties of the signal. Typical non-energy signals are auto-

10 for example, the transmitting power of the sender and the energy input of the sender/receiver

matically all power signals, because they have an infinite signal energy at a finite average power:

$$0 < P = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} s(t) \cdot s^*(t) \mathrm{d}t < \infty \,. \tag{2.36}$$

The instantaneous power of real signals at time t is:

$$P(t) = \lim_{T\to 0} \frac{1}{2T} \int_{t-T}^{t+T} |s(t)|^2 \mathrm{d}t = |s(t)|^2 \,. \tag{2.37}$$

The class of power signals includes periodically continued energy signals, such as a sine-/cosine signal, or stochastic signals, such as noise with infinite energy. For example, the energy for a DC signal $s(t) = A, \forall t, 0 < A < \infty$ is infinite, and consequently there is no energy signal but a power signal:

$$E = A^2 \int_{-\infty}^{\infty} \mathrm{d}t = A^2 \lim_{\theta\to\infty} \int_{-\theta}^{\theta} \mathrm{d}t \to \infty \,. \tag{2.38}$$

2.3.6 Deterministic and Stochastic Signals

Analytical signals are of great importance in the theory of signal processing. This is due, among other things, to the fact that these signals are mathematically exactly predictable, i.e. completely deterministic. The predictability of signals is therefore an important property. Exactly predictable, so-called *deterministic* signals can be expressed in an analytical mathematical context and predicted for all times and places. *Stochastic* random signals, on the other hand, cannot be fully expressed as an analytic function and consequently cannot be predicted exactly (cf. Figure 2.17). As we will see in subsection 5.3.2, for deterministic signals $x(t)$ there always exists a spectral function $X(f)$ which can be calculated via the Fourier series or the Fourier-transformation. In the case of stochastic signals, however, this cannot be specified, since Fourier series and Fourier-transformation require the exact knowledge of the time function for all times t.

Definition: A deterministic signal can be described exactly in analytical form and predicted at all future times, while a stochastic signal does not fulfil this condition or only imperfectly.

Statistical Moments, Expected Value and Variance

The behaviour of stochastic signals can be characterised using statistical moments such as expected value $E(\bullet)$, variance $\mathrm{Var}(\bullet)$ and standard deviation $\sigma(\bullet)$ as follows: The *first statistical moment* or *expected value* of a random variable X is the value that

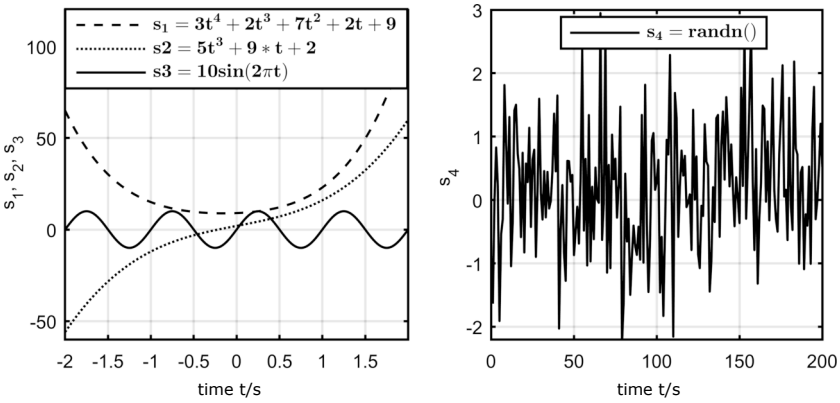


Fig. 2.17: Examples of signals $s_1(t)$ to $s_3(t)$ originate from a deterministic (exactly predetermined) process and can be completely described analytically as a polynomial or harmonic function (left). The random signal $s_4(t)$, on the other hand, comes from a random process and cannot be fully expressed by analytical functions (right).

the random variable is most likely to take. If x_i is a real discrete random variable with the values $(x_i)_{i\in\mathbb{N}}$ and with the respective probabilities $(p_i)_{i\in\mathbb{N}}$ (with $\mathbb{N}$ as the set of natural numbers), the expected value (1. moment) of the time series is given by:

$$E(X) = \mu_X = \sum_{i\in\mathbb{N}} x_i p_i = \sum_{i\in\mathbb{N}} x_i P(X = x_i)\,. \tag{2.39}$$

The *expectation* $E(X)$ is thus the *weighted mean* μ_X of X weighted by the probabilities P and thus the most probable value for a realisation of X (1. moment). With equal probability of N realisations $p_i = p = 1/N$, the expected value is equal to the mean $\mu_X = \mu$ of X. For integrable expected values, i.e. $E(X) = \mu_X < \infty$, the *second moment* or variance is the expected value of the squared deviation of the random variable X from the mean μ_X.

$$\mathrm{Var}(X) = E\big((X-\mu)^2\big) = \sum_{i\in\mathbb{N}} (x_i - \mu_X)^2 P(X = x_i)\,. \tag{2.40}$$

The *variance* is a quadratic quantity that gives the mean squared deviation of a random variable from the expected value of X. It is thus the expected value of the squared deviation (2. moment). The associated non-squared quantity is the *standard deviation* σ_X, which is defined as the square root of the variance:

$$\sigma(X) = \sqrt{\mathrm{var}(X)}\,. \tag{2.41}$$

Both the variance and the standard deviation are positive quantities. Thus, $\mathrm{Var}(\bullet) \geq 0$ and $\sigma(\bullet) \geq 0$ hold.

Covariance and Correlation

Another important quantity in statistics is the covariance of two signals. If X and Y are two real, integrable random variables whose product is also integrable, i.e., the expected values $E(X)$, $E(Y)$ and $E(XY)$ exist, then the covariance of X and Y is defined as follows:

$$\mathrm{Cov}(X, Y) = E\left((X - E(X)) \cdot (Y - E(Y))\right) . \tag{2.42}$$

The value of the covariance makes tendential statements about the values of two random variables X and Y: Positive covariances stand for a monotonic relationship (common tendency of X and Y), negative ones for an inverse monotonic relationship (opposite tendency of X and Y) of the random variables. The random variables X and Y have no monotonic relationship if the covariance is zero.

The covariance thus indicates the direction of a monotonic relationship between two random variables, but the strength of the relationship cannot be read from it. Comparability with other signal pairs is achieved, for example, by normalising the covariance with the standard deviation. As shown in subsection 2.3.6, this leads to the correlation coefficient of the two random variables X and Y.

Signal Components of Biosignals

Normally, real biosignals $U(t)$, like the ECG raw signal in Figure 2.9, basically contain a deterministic signal part $s(t)$ – mostly the biosignal of the physiological process to be measured – and a stochastic signal part consisting of noise and artefacts $r(t) + a(t)$. The measured biosignal is thus:

$$U(t) = s(t) + r(t) + a(t) . \tag{2.43}$$

The relevant information of the biosignals is generally contained in the deterministic part $s(t)$, while the stochastic signal part often only contains noise and artefacts. However, there are also numerous applications that deal with the random deviations in the signals, i.e. the stochastic component, during evaluation. For example, in repeatedly measured motion sequences, the deviations from the mean value curve of the motion can provide important information about the accuracy of the motion sequence. This issue is discussed in detail in section 6.2.

Stationarity of Stochastic Signals

The class of stochastic signals can be further divided into *stationary* and *stationary* stochastic signals. In terms of content, stationarity means that certain stochastic properties of the process X_t, $t = 0, 1, 2, \ldots$ are *time-* or *location-invariant*.. Basically, time or location invariance means that a certain stochastic property is independent of a shift in time or location, i.e. it does not change, for example, if the time origin is shifted by some value τ.

Strictly speaking, stationary signals are divided into strong and weak stationarity. A stochastic signal is covariance stationary or *weakly stationary* if the first two

moments, i.e. the expected value and the variance of the signal are *time invariant*.

$$\begin{aligned} E(X_t) &= \mu_X = \text{ const.} \\ \text{Var}(X_t) &= \sigma_X^2 = \text{ const.} \end{aligned} \tag{2.44}$$

and the covariance between X_t and $X_{t+\tau}$ depends only on τ, and not on time t:

$$\text{Cov}(X_t, X_{t+\tau}) = \text{Cov}(X_0, X_\tau)\,. \tag{2.45}$$

From the property $E(X_t) = \text{const.}$ it can be concluded that a stationary process has no *trend*. A trend in this context is a long-term movement around which the process fluctuates. As shown in Figure 2.18, this can be both linear (V_t) and non-linear (W_t) in nature. The constant variance condition $V(X_t) = \text{const.}$, on the other hand, implies that the signal amplitude of a stationary process does not increase or decrease. In the case where only $E(X_t) = \text{const.}$ is required, it is called mean-stationarity.

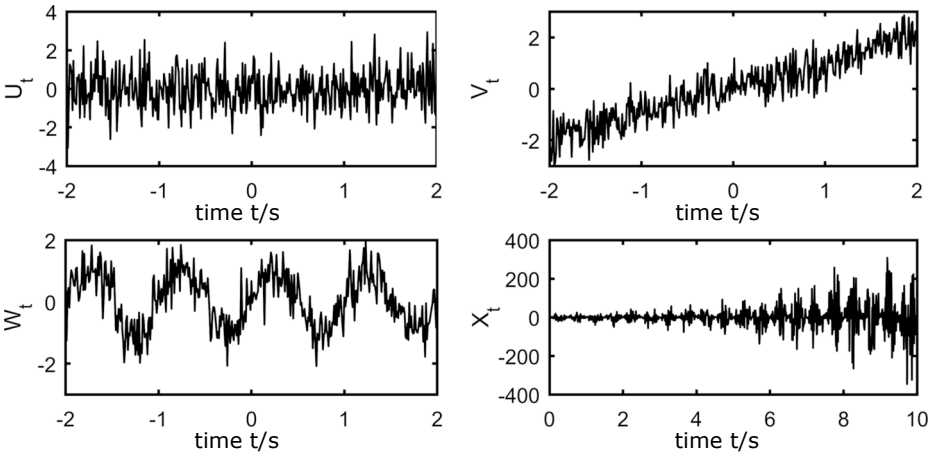


Fig. 2.18: Examples of a stationary signal U_t (top left) and non-stationary signals with linear trend V_t (top right), non-linear trend W_t (bottom left) and increasing variance X_t (bottom right).

The class of *strongly stationary* signals satisfies a more fundamental requirement: the distribution functions themselves must not depend on the shift. An identical, i.e. stationary, distribution of the random variables X_t thus means that in a stationary process all realisations of X_t have the same distribution:

$$P(X_{t_1} \le x_1, \ldots, X_{t_n} \le x_n) = P(X_{t_1+\tau} \le x_1, \ldots, X_{t_n+\tau} \le x_n)\,. \tag{2.46}$$

With this definition, the joint distribution of the random variables $X_{t_1}, \ldots, X_{t_n}$ and $X_{t_1+\tau}, \ldots, X_{t_n+\tau}$ is equal. However, strong stationarity is analytically more difficult to handle than weak stationarity.

2.3.7 Continuous and Discrete Signals

Due to their deterministic nature, purely analytical signals are used for proof in the fundamentals of theoretical signal processing. This class of continuous signals is called *value and time continuous* because both their dependent and independent quantities can take all values of a continuum. This means that the value set of a time-/value interval is in principle infinite, as in the case of the representatives of the real numbers $\mathbb{R}$. Due to their analogue character, these continuous signals are also found in *analogue* signal processing.

In contrast, a digital signal is always discrete in value and time, and the information contained therein consists of a limited number of possible symbols, i.e. the time-value interval of these signals is limited to a countably large number of different values. The number of possible values M is called the *interval number*. Digital signals are named according to the number of intervals, such as binary signal ($M = 2$) or ternary signal ($M = 3$), etc. Only the finite range of values or information content of these signals makes it possible to store them on data carriers for further processing.

Whereas with continuous signals the signal value was defined at any time t, with discrete-time signals this is only the case at discrete times $t(n) = t_n$. These time points are usually chosen *equidistant*, i.e. as a whole multiple $n \in \mathbb{N}$ of a discrete time interval $t_n = nT_s$. In practice, these discrete signals are created by sampling the continuous-time signal. Therefore n is called the sample index, T_s the sampling index or sample interval and its reciprocal $f_s = 1/T_s$ the *sample* or *sample frequency*. A discrete-time signal $x(t_s)$ is thus completely determined by the time sequence of its samples:

$$x_s(t_s) = \{x(n_1 T_s), x(n_2 T_s), x(n_3 T_s), \ldots, x(n_N T_s)\}\,, \quad n \in \mathbb{N}\,. \tag{2.47}$$

These discrete signal values can be both continuous and discrete. With complete digitisation, however, the instantaneous continuous signal value $x(t)$ at the time t_s is rounded to a discrete signal value. This is then held as a signal value in the digital signal for the duration of a sampling interval.

The signals shown in Figure 2.19 were generated with the programming language Matlab, i.e. due to the numerical calculation method, the signals are all discrete in nature. However, the form of representation can also be influenced there by selecting the appropriate function. For the representation of continuous signals, high time resolutions and the command `plot()` are used (top), whereas for the representation of discrete signals low time resolutions and the command `stem()` (bottom) is applicable. In section 5.1, the digitisation of continuous signals by *time discretisation* and *value quantisation* and the associated laws are discussed in detail.

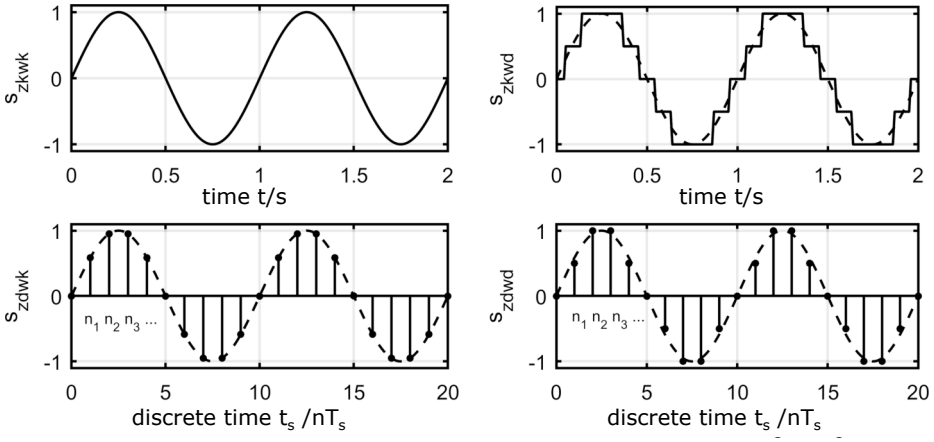


Fig. 2.19: Examples for a continuous-time and continuous-value signal s_{zkwk} (top left), for a continuous-time and discrete-value signal s_{zkwd} (top right), for a discrete-time and discrete-value signal s_{zdwk} (bottom left) and a discrete-time and discrete-value signal s_{zdwd} (bottom right).

2.4 Signal Processing Transformations

In the previous sections, a signal was described by the time-dependent quantity $s(t)$ and a system by the time-dependent impulse response $h(t)$. However, in some cases it is useful to represent the signal and system in a different function space called image domain (or reciprocal space, frequency domain, Fourier domain). For example, transforming the signal $s(t)$ into the image domain can reveal signal information that is hidden in the original function space. The transformation into the image domain is done by means of a mathematical operation T, which transforms $s(t)$ into a new quantity $S(\xi)$ with the new variable of the image domain ξ. In signal processing, however, it also happens that the analysis is performed a priori in the image domain. For example, the effect of filters is usually considered in the frequency domain. Therefore, the mathematical operation T must be designed in such a way that a back transformation T^{-1} exists which uniquely returns the transformed quantity $S(\xi)$ from the image domain back into $s(t)$ of the original function space.

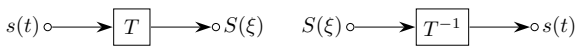


Fig. 2.20: By means of the mathematical operation T the time-dependent quantity $s(t)$ is transformed into $S(\xi)$ (image domain). In signal processing, the requirement for T is that there exists a unique back transformation T^{-1} from $S(\xi)$ back to $s(t)$.

This requirement is fulfilled by the Fourier-transformation and the Laplace-transformation. Both belong to the class of integral transformations[11]. The structure of the transformation operation is the same for all integral transformations. The quantity $s(t)$ to be transformed is multiplied by a function $K(\xi, t)$ (integral kernel) and integrated over the entire range of the variable t. The result is a new function $S(\xi)$:

$$T\{s(t)\} = S(\xi) = \int s(t)K(\xi, t)\mathrm{d}t \,. \tag{2.48}$$

In this section the basics of four important integral transformations are presented, namely the Fourier- and Laplace-transformation, the wavelet-transformation and the convolution. A detailed description can be found, for example, in [30, 45, 82].

2.4.1 Continuous Fourier-Transformation

The best known integral transformation is the Fourier-transformation[12]., where a time-dependent quantity $s(t)$ is transformed into the frequency domain $S(\omega)$. The integral core of the Fourier-transformation $K(\xi, t)$ in Equation 2.48 is the eigenfunction of linear differential equations:

$$K(\xi, t) = K(\omega, t) = e^{-j\omega t} \,. \tag{2.49}$$

Equation 2.49 inserted into Equation 2.48 yields the mathematical operation of the Fourier-transform:

$$T\{s(t)\} = S(j\omega) = \int s(t)e^{-j\omega t}\mathrm{d}t \,. \tag{2.50}$$

According to the Euler formula

$$e^{-j\omega t} = \cos(\omega t) - j\sin(\omega t) \tag{2.51}$$

Equation 2.49 represents a linear combination of a cosine and sine function with the angular frequency ω as a variable in the image domain. In Equation 2.50 the product of the quantity to be transformed $s(t)$ and cosine and sine functions with the angular frequency ω is formed and then integrated.

This is done for all possible values of ω. The result is a complex function in the image domain, $S(j\omega)$. Accordingly, the Fourier-transformation yields large function values for such ω values where the product of $s(t)$ and the corresponding cosine or sine function comes with a large area. This is the case when $s(t)$ has a large similarity to Equation 2.51. Accordingly, the Fourier-transformation $S(j\omega)$ can be interpreted as a measure of the similarity of the transformed quantity $s(t)$ with cosine and sine functions of the respective angular frequency ω. As already mentioned in subsection 2.3.3,

11 Other examples of integral transformations are the wavelet and Radon-transformation.

12 Jean Baptiste Fourier (1768–1830), important French mathematician and politician

the even part is always described by the cosine and the odd part by the sine of the integral core. The largest values in $S(j\omega)$ are obtained when $s(t)$ itself is a periodic function that can be represented by a Fourier series according to Equation 2.16.[13]

The Fourier-transformion has a special significance in the physical-technical context, where the magnitude of $S(j\omega)$ is interpreted as the frequency spectrum of $s(t)$. The frequency spectrum provides an indication of which frequency components are contained in the signal $s(t)$. The analysis in the frequency domain is explained in detail for discrete-time signals in subsection 5.3.2. The frequency spectrum further forms the basis for understanding the transmission behaviour of filters in terms of magnitude and phase frequency response, the application of which is discussed in section 4.4 and subsubsection 5.3.4.2.

In system theory, the transformation into the frequency domain offers an alternative to the consideration in the time domain. As already explained in Figure 2.20, in the time domain the result $y(t)$ of the transmission of a signal through a linear, time-invariant system $h(t)$ is obtained from the mathematical convolution of the input quantity $s(t)$ with the impulse response of the system: $y(t) = s(t) * h(t)$. Transformation into the frequency domain, the convolution is represented by an algebraic multiplication (cf. last row in Table 2.3).

Tab. 2.3: Fourier transform theorems.

Theorem	**Time domain $s(t)$**	**Frequency domain $S(j\omega)$**
Linearity	$a_1 s_1(t) + a_2 s_2(t)$	$a_1 S_1(j\omega) + a_2 S_2(j\omega)$
Similarity	$s(bt)$	$\frac{1}{\lvert b\rvert} S(\frac{j\omega}{b})$
Time shift	$s(t - t_0)$	$S(j\omega) e^{-j\omega t_0}$
Frequency shift	$s(t) e^{j\omega_0 t}$	$S(j(\omega - \omega_0))$
Differentiation	$\frac{\partial^n s(t)}{\partial t^n}$	$(j\omega)^n S(j\omega)$
Integration	$\int s(t) \mathrm{d}t$	$\frac{S(j\omega)}{j\omega} + \frac{1}{2} S(0)\delta(j\omega)$
Multiplication	$s(t)\, h(t)$	$S(j\omega) * H(j\omega)$
Convolution	$s(t) * h(t)$	$S(j\omega)\, H(j\omega)$

Thus, the spectrum of the system output $Y(j\omega)$ is obtained from the multiplication of the Fourier-transformation of input $S(j\omega)$ and the Fourier-transformation of the impulse response $H(j\omega)$:

$$Y(j\omega) = S(j\omega)\, H(j\omega)\ . \tag{2.52}$$

This relationship is illustrated by Figure 2.21. The spectrum of the impulse term $H(j\omega)$ is called the transfer function of the system. In principle, neither of the two ways of looking at linear time-invariant systems (LTI systems for short) is to be preferred to the

13 Every periodic function $s_{\mathrm{per}}(t)$ with the period T_0 can be represented in the form of a Fourier series.

other. In some cases of signal processing, the consideration in the frequency domain is much clearer or simpler. This is particular the case at filters in relation to signals that are already available in the form of their spectrum.

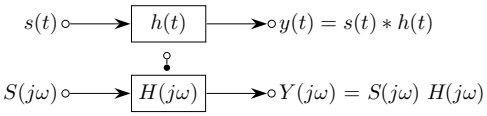


Fig. 2.21: The relationship between the input and output of an LTI system is described by its impulse response $h(t)$ or its transfer function $H(j\omega)$. The output quantity $y(t)$ is calculated by the convolution of the input quantity $s(t)$ with the impulse response $h(t)$, or in the frequency domain by the multiplication of the respective spectra.

The reverse transformation from the frequency to the time domain is done by the inverse Fourier transformation:

$$T^{-1}\{S(j\omega)\} = s(t) = \frac{1}{2\pi}\int S(j\omega)e^{j\omega t}\mathrm{d}\omega\ . \tag{2.53}$$

In the previous considerations it was assumed that the Fourier transform exists. There is no necessary condition for this. A sufficient condition regarding the convergence of the Fourier transform is the absolute integrability of $s(t)$, which is also known as the Dirichlet condition[14]:

$$\int |s(t)|\mathrm{d}t < \infty\ . \tag{2.54}$$

ne This condition is fulfilled by energy signals. The continuous Fourier transforms of common signals are listed in Table 2.4. For signals that do not fulfil Equation 2.54, the

Tab. 2.4: Fourier transforms of important deterministic signals.

Signal	**Time domain $s(t)$**	**Frequency domain $S(j\omega)$**
Rectangular pulse	$\mathrm{rect}(t)$	$\frac{\sin(\omega/2)}{\omega/2} = \mathrm{si}\left(\frac{\omega}{2}\right)$
Si-function	$\frac{1}{2\pi}\,\mathrm{si}\left(\frac{t}{2}\right)$	$\mathrm{rect}(j\omega)$
Dirac pulse	$\delta(t)$	1
Constant	1	$2\pi\delta(j\omega)$
Dirac pulse sequence	$\sum \delta(t - kt_0); k = \ldots -1, 0, 1, \ldots$	$\omega_0 \sum \delta(\omega - k\omega_0); \omega_0 = \frac{2\pi}{t_0}$
Cosine-function	$\cos(\omega_0 t)$	$\pi\left[\delta(j(\omega - \omega_0)) + \delta(j(\omega + \omega_0))\right]$
Sine-function	$\sin(\omega_0 t)$	$j\pi\left[-\delta(j(\omega - \omega_0)) + \delta(j(\omega + \omega_0))\right]$
Step-function	$u(t)$	$\pi\delta(\omega) - j\frac{1}{\omega}\big\vert_{\omega\neq 0}$
Exponential pulse	$\frac{1}{t_0}u(t)\mathrm{e}^{-\frac{t}{t_0}}$	$\frac{1}{1+j\omega t_0}$

14 Dirichlet: German mathematician (1805–1859).

transformability can be achieved by an additional convergence term. This leads us to the Laplace transform.

2.4.2 Continuous Laplace Transform

The Laplace[15] transformation also belongs to the class of integral transformations and is an extension of the Fourier transformation. The integral core is similar to that of the Fourier transform (cf. Equation 2.49):

$$K(\xi, t) = K(p, t) = e^{-pt} \tag{2.55}$$

where p is a complex variable,

$$p = \sigma + j\omega \,. \tag{2.56}$$

Here σ is a real number and j is the imaginary unit. Equation 2.55 inserted into Equation 2.48 yields the mathematical operation of the Laplace transformation:

$$S(p) = \int s(t)e^{-pt}\mathrm{d}t \,. \tag{2.57}$$

If we separate the variable p in Equation 2.57 into its components according to Equation 2.56 the similarity of the Laplace transform with the Fourier transform becomes even clearer:

$$S(\sigma, j\omega) = \int s(t)e^{-j\omega t}e^{-\sigma t}\mathrm{d}t \,. \tag{2.58}$$

The term

$$e^{-\sigma t} \tag{2.59}$$

corresponds to an additional damping term, whereby also signals become transformable which do not satisfy the Dirichlet condition according to Equation 2.54. Thus Equation 2.54 for the Laplace transform can be extended to

$$\int |s(t)e^{-\sigma t}|\mathrm{d}t < \infty \,. \tag{2.60}$$

The damping effect of Equation 2.59 occurs only when the argument of the exponential function is negative overall. Since we are dealing with causal signals in signal processing ($s(t) = 0, \forall t < 0$), so the negative time domain is not considered, σ must be > 0. This region in the complex p-plane is called the convergence region of the Laplace transform. There, polynominals with arbitrary power of t are transformable. For exponentially increasing signals of type $s(t) = u(t)e^{at}$ with $a > 0$, one further restrictions for the convergence region of σ have to be considered and the dampening

15 Pierre-Simon Laplace (1749–1827), french mathematician, physicist and astronomer, Minister of the Interior under Napoleon.

Tab. 2.5: Laplace transformations of various causal signals.

Signal	Time domain $s(t)$ for $t \geq 0$	Frequency domain $S(p)$
Dirac pulse	$\delta(t)$	1
Constant	1	$\frac{1}{p}$
Power-function	t^k	$\frac{k!}{p^{k+1}}$
Exponential-function	e^{-at}	$\frac{1}{p+a}$
Exponential-function	$1-\mathrm{e}^{-at}$	$\frac{a}{p(p+a)}$
Cosine-function	$\cos(\omega_0 t)$	$\frac{p}{p^2+\omega_0^2}$
Sine-function	$\sin(\omega_0 t)$	$\frac{\omega_0}{p^2+\omega_0^2}$

effect of Equation 2.59 to outweigh the exponential increase e^{at}, $\sigma > a$ must hold. The Laplace transform of selected causal signals are listed in Table 2.5.

The interpretation of the Laplace transform is less descriptive than that of the Fourier transform because the image domain has a two-dimensional variable space due to Equation 2.56. For $\sigma = \sigma_0 = \text{const.}$ within the convergence region, the Laplace transform $S(\sigma_0, j\omega)$ is equivalent to the Fourier transform for a signal damped with $e^{-\sigma_0 t}$. The theorems of the Laplace transform are identical to those of the Fourier transform according to Table 2.3 when $j\omega$ is replaced by p.

Analogous to the Fourier transform, the inverse Laplace transform can be written as follows:

$$T^{-1}\{S(p)\} = s(t) = \frac{1}{2\pi j}\int S(p)e^{pt}\mathrm{d}p\,. \tag{2.61}$$

The integration limits in Equation 2.61 are given by $[\sigma_0 - j\infty, \sigma_0 + j\infty]$, where σ_0 must again lie in the convergence region. The back transformation, is usually done by correspondence tables (Table 2.5). This will be presented by means of an example. Let a series resonant circuit with the given electrical components R, L and C (resistance, inductance and capacitance respectively), be excited by the signal $s(t)$ The output quantity sought is the current $i(t)$ of the series resonant circuit. The voltage equation of the mesh through the series resonant circuit yields

$$s(t) = Ri(t) + L\frac{\partial i}{\partial t} + \frac{1}{C}\int i(t)\mathrm{d}t\,. \tag{2.62}$$

Laplace transforming the differential equation 2.62 using the theorems for differentiation and integration, leads to:

$$S(p) = RI(p) + LpI(p) + \frac{1}{pC}I(p) \tag{2.63}$$

$$\Longleftrightarrow\ pS(p) = \left(pR + p^2L + \frac{1}{C}\right)I(p)\,. \tag{2.64}$$

The transfer function of the series resonant circuit $H(p)$ can be given as the quotient of transformed output quantity $I(p)$ and input quantity $S(p)$:

$$H(p) = \frac{I(p)}{S(p)} = \frac{1}{L}\frac{p}{p^2 + \frac{R}{L}p + \frac{1}{LC}} . \tag{2.65}$$

Excitation of the system by a unit step function, $s(t) = u(t)$, leads to the Laplace transform $T\{u(t)\} = S(p) = \frac{1}{p}$. Substitution into Equation 2.65 leads to:

$$I(p) = \frac{1}{L}\frac{1}{p^2 + \frac{R}{L}p + \frac{1}{LC}} . \tag{2.66}$$

Converting the polynomial in Equation 2.66 to zero form yields

$$I(p) = \frac{1}{L}\frac{1}{(p - p_1)(p - p_2)} , \tag{2.67}$$

with

$$p_{1,2} = -\frac{R}{2L} \pm \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}} . \tag{2.68}$$

Using partial fraction decomposition, Equation 2.67 can be written as

$$I(p) = \frac{1}{L}\left[\frac{A}{p - p_1} + \frac{B}{p - p_2}\right] , \tag{2.69}$$

with the coefficients A and B still to be determined. The coefficient comparison of Equation 2.69 with Equation 2.67 yields

$$A = \frac{1}{p_1 - p_2} \tag{2.70}$$

and

$$B = \frac{1}{p_2 - p_1} . \tag{2.71}$$

The partial fraction form in Equation 2.69 can be easily transformed back by using Table 2.5 and the linearity of the Laplace transform thus

$$i(t) = \frac{1}{L}[Ae^{p_1 t} + Be^{p_2 t}]u(t) \tag{2.72}$$

with the previously calculated values for A, B, p_1 and p_2. Note that the excitation by the unit step function $s(t) = u(t)$ can cause oscillations if the root term in Equation 2.68 is negative, thus leading to complex zeros.

2.4.3 Continuous Short-Time Fourier-Transform and Wavelet Transform

The Fourier transform according to Equation 2.50 performs a integration out over all times. of the signal Therefore, the spectrum obtained in this way does not contain any

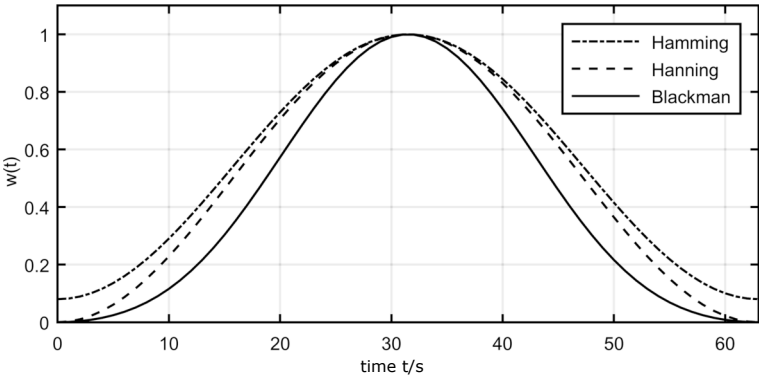


Fig. 2.22: Window functions for the short-time Fourier transform (STFT): The window width corresponds to the time resolution Δt of the STFT, Δt can be varied via the parameterisation in the functional description of the window functions.

time information, i.e. no information about which part of the spectrum originates from which time section of the signal. In order to determine the spectrum of a signal at least in sections, the short-time Fourier transform (STFT) is available. For this purpose, the signal is multiplied by a window function $w(t)$, which sets all signal components out of the window function to zero. With the signal "windowed" in this way, the Fourier transform yields only the spectrum of the section that lies in the window. Then the window is shifted and the spectrum of the next section is calculated. The window width is freely selectable and corresponds to the time resolution Δt of the STFT. Commonly used window functions are known as Hamming, Hanning, Blackman-Harris or Gaussian windows (see Figure 2.22).

An important property of such windows is that the signal is gently guided towards zero at the edges of the window. Otherwise, jumps could occur at the edges, which would result in an infinitely extended spectrum. Therefore, the square wave function is usually ruled out as a window function. Limiting the time resolution to Δt, according to the uncertainty principle of communications engineering[16], a limited frequency resolution results, since the product of time and frequency resolution cannot fall below a certain value. A frequently chosen definition of bandwidth and signal duration leads to the uncertainty condition

$$\Delta t \, \Delta f = 1 \, . \tag{2.73}$$

Accordingly, frequency and time accuracy cannot be high for the same windows parameters. If a high time resolution is required, i.e. Δt is small, the frequency resolution

16 The uncertainty principle of communications engineering was formulated by Karl Küpfmüller (1897 - 1977).

decreases (Δf is large) and vice versa. For example, using a time resolution of 1 ms, the achievable frequency resolution would be 1000 Hz.

An alternative to the STFT is the wavelet transform. The word *wavelet* describes the form of the integral kernel function $\psi(t)$. The transformation rule for the wavelet transformation is

$$S(a,\tau) = \frac{1}{\sqrt{a}} \int s(t)\psi\left(\frac{t-\tau}{a}\right) \mathrm{d}t \,. \tag{2.74}$$

The parameter τ in the argument of the wavelet causes a time shift, the parameter a a compression or stretching of the wavelet. Similar to the Fourier transform, the wavelet transform can be interpreted as a correlation of the signal $s(t)$ with the wavelet $\psi(t)$, where the correlation value $S(a,\tau)$ depends on the stretching a and the time shift τ. The factor $1/\sqrt{a}$ before the integral is necessary for the normalization of the transform to the wavelet width. The functional description of $\psi(t)$ is not fixed. Rather, $\psi(t)$ can be largely freely designed and adapted to the signal $s(t)$, which is an advantage of the wavelet function over transforms with a given integral kernel. For the wavelet synthesis only two conditions must be fulfilled. First, the area fractions of the function above and below the zero line must be equal. Thus applies

$$\int \psi(t)\mathrm{d}t = 0 \,. \tag{2.75}$$

Equation 2.75 thus describes the wave character of the wavelet. The second requirement for $\psi(t)$ is formulated as an admissibility condition:

$$\int \frac{\Psi^2(\omega)}{\omega} \mathrm{d}\omega < \infty \,, \tag{2.76}$$

with $\Psi(\omega)$ as the Fourier transform of $\psi(t)$. A consequence of Equation 2.76 is

$$\lim_{\omega \to 0} \Psi(\omega) = 0 \,. \tag{2.77}$$

In signal processing wavelets based on the Gaussian function

$$e^{-\frac{t^2}{2}} \tag{2.78}$$

are often used. Equation 2.78 forms the envelope when multiplied by a second function, which corresponds to the effect of the Gaussian window in STFT. Examples are the *Morlet*[17] wavelet and the *Mexican-Hat* wavelet. The Morlet wavelet has the general structure

$$\psi(t) = e^{-\frac{t^2}{2\sigma}} e^{-jct} \,. \tag{2.79}$$

Here, σ is a scaling parameter used to determine the width of the wavelet, and c is a modulation parameter for determining the frequency of the oscillation described by

17 Jean Morlet (1931–2007): French geophysicist and one of the founders of the wavelet transform.

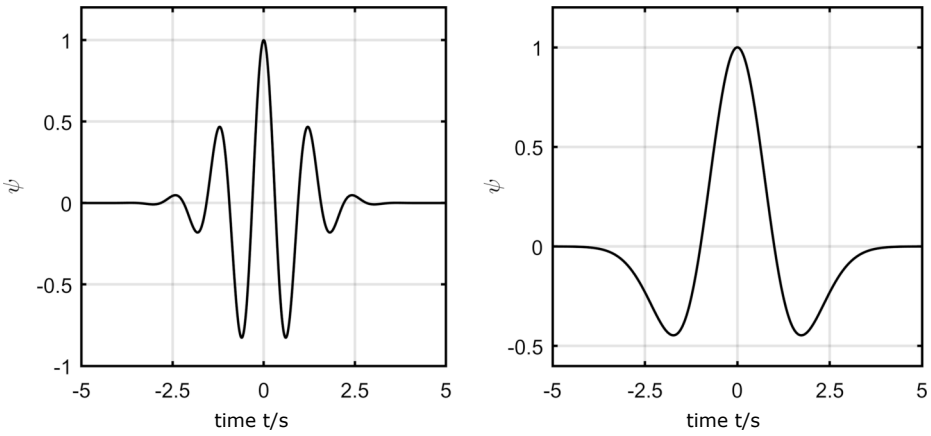


Fig. 2.23: Morlet wavelet (left) according to Equation 2.80 and Mexican Hat wavelet (right) according to Equation 2.81.

the complex exponential function. A possible realisation of a Morlet wavelet is e.g. the function

$$\psi(t) = e^{-\frac{t^2}{2}} \cos(5t) \,. \tag{2.80}$$

Figure 2.23 (left) shows the composition of the Morlet wavelet according to Equation 2.80 from harmonic oscillation multiplied by a Gaussian envelope function. The Mexican Hat wavelet (cf. Figure 2.23, right) has the mathematical form

$$\psi(t) = e^{-\frac{t^2}{2\sigma}} \left(1 - t^2\right) . \tag{2.81}$$

Unlike the Morlet wavelet, the Mexican Hat wavelet does not contain a harmonic function, which leads to differences in the interpretation of the transformation, as will be discussed later.

The graphical representation of $S(a, \tau)$ is either in a three-dimensional figure with a and τ as x- and y-axis and S as the z-axis, or a two-dimensional figure in which a is plotted downward in ascending order over τ. The transformation result is then rendered as a color or brightness in the two-dimensional (a, τ) plane.

The benefit of the wavelet transform for signal analysis lies in the variable wavelet width. If, for example, a high time resolution is required because very short signal events occur in the signal that are to be analysed spectrally, the width of the wavelet can be reduced by means of the scaling value a so that the required time resolution is achieved. This narrow wavelet then passes through the entire signal by means of the displacement parameter τ, yielding high S values whenever the wavelet encounters the short signal events. However, if the same signal also contains periodic events with long period durations, such as may be caused by respiration in biosignals, these are captured in the same signal analysis for large a values. This makes the wavelet transform particularly well suited for the analysis of signals that are com-

posed of sequences with different bandwidths and are sometimes of short duration. Thus, the wavelet transform finds wide application in the analysis of electroencephalograms, for example, to detect short epileptic episodes in the low-frequency fundamental waves of electrical neuron activity.

Finally, the filter property of wavelets should be pointed out. The Equation 2.77 $\psi(t)$ can also be understood as a bandpass filter. The bandwidth of the bandpass is directly related to the scaling value a. The smaller a is, the larger is the bandwidth of the bandpass filter. This suggests building a filterbank from a single wavelet where only the scaling value a is changed. The Morlet wavelet is particularly suitable as a bandpass filter because it is based on a harmonic function (cf. Equation 2.80). Varying the scaling value a, the frequency of the harmonic function of the Morlet wavelet is modulated. The signal $s(t)$ to be filtered is correlated with the harmonic function of the Morlet wavelet during the transformation, which means a spectral decomposition of $s(t)$ according to the frequencies determined by the scaling value a.

2.4.4 Continuous Linear Convolution

In the section on the Fourier transform, it was already mentioned that for a linear time-invariant system, in the time domain the output signal $y(t)$ is calculated from the mathematical convolution of the input signal $s(t)$ with the impulse response of the system $h(t)$ (cf. Figure 2.21). Equation 2.82 gives the mathematical operation for this:

$$y(t) = \int s(\tau)h(t-\tau)\mathrm{d}\tau \,. \tag{2.82}$$

Here, the variable t is renamed to τ: $s(t) \rightarrow s(\tau)$, $h(t) \rightarrow h(\tau)$. The second function (here h) is mirrored at the ordinate by the negative sign of the variable τ. The value t in the argument of the second function is on the one hand the displacement parameter of the function h and at the same time the variable of the output function $y(t)$. Thus, the function h is mirrored, shifted by t, then multiplied by s and the result is integrated over the time interval τ. This process is repeated for all displacement values t, from which the new function $y(t)$ is formed.

The convolution integral, given by Equation 2.82, is identical to the correlation function except for the negative sign in the argument of the second function, which thus also belongs to the class of integral transformations. The correlation function provides a measure for the similarity of two signals or functions. In this sense, in the Fourier transform the signal $s(t)$ was correlated with the complex exponential function, and the convolution integral provides a measure for the similarity of $s(\tau)$ and $h(-\tau)$ at the respective shift values t. The symbol for the notation of the convolution is a star between the functions to be convolved:

$$y(t) = s(t) * h(t) \,. \tag{2.83}$$

An overview of the common convolution theorems can be found in Table 2.6. Applications of integral transformations are given in connection with discrete transformations in section 5.2 and in practical applications in chapter 6.

Tab. 2.6: Convolution theorems.

Theorem	Relation
Commutative law	$a(t) * b(t) = b(t) * a(t)$
Associative law	$a(t) * b(t) * c(t) = [a(t) * b(t)] * c(t) = a(t) * [b(t) * c(t)]$
Distributive law	$a(t) * (b(t) + c(t)) = a(t) * b(t) + a(t) * c(t)$
Identity	$a(t) * \delta(t) = a(t)$
Differentiation	$a(t) * \delta'(t) = a'(t)$
Integration	$a(t) * u(t) = \int a(\tau)\mathrm{d}\tau$
Displacement	$a(t - t_0) = a(t) * \delta(t - t_0)$
Dilation	$a(t) * \delta(bt) = \frac{1}{\lvert b\rvert} a(t)$

2.5 Biosignal Processing and the Derivation of Diagnostic Information

The definition of important properties and transformations of signal processing in the previous sections allow a mathematical description of the analysis of biosignals. A fundamentally important idea in the analysis of biosignals, as with signals in general, is the direct connection between the signal generator, in our case a physiological system, and the signal itself, already discussed in section 2.2. Ideally, this direct connection allows the system state (state of health) to be determined from the signal in the form of a system diagnosis. The basic prerequisite for this is that the *signal measurand* has a sufficiently high *sensitivity* to the physiological variable under consideration and is also not obscured by other artefacts such as movements or noise. In most cases, for this reason, the primary task is to separate the important from the unimportant signal components and to generate a suitable measurement and signal processing chain for the signal to be acquired.

The signal processing chain results from the measurement task and, as shown in Figure 2.24, always contains a physiological process (and a physiological signal) which takes place in the patient and which is measured (in the form of the signal measurand) with the aid of suitable sensor technology and measuring equipment and is digitised, stored and evaluated in the computer or microcontroller for further processing. It should be noted that in most cases the physiological signal does not correspond exactly to the signal measurand, either because of the measurement method used or the type of signal processing that follows.

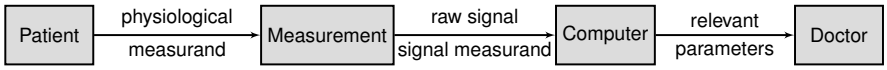


Fig. 2.24: An example of a signal processing chain in biosignal processing.

In the further course of the signal analysis, the aim is often to reduce the amount of data to a few, but meaningful parameters that can be used for medical decision-making. For this purpose, the raw signal is usually first freed from disturbing artefacts and noise by filtering and only then evaluated with regard to diagnostically relevant parameters such as heart rate and respiratory rate. A mostly graphical representation of the evaluation is forwarded to the attending physician for a decision on how to proceed with the patient's treatment. In this scenario, the final evaluation is carried out by an experienced doctor, but medical expert systems can also provide important services for further evaluation. Here, on the basis of statistically processed expert knowledge, a probability-based decision aid is generated with the help of an algorithm, which is intended to support the doctor in his decision-making.

When creating the signal processing chain, the relevant and unimportant signal components as well as the methodology for separating them are often still unclear. The procedure for finding them is more like searching for a needle in a haystack. If the signals are heavily distorted by filtering, for example, it can happen that the relevant signal components have also become unusable for diagnostic evaluation. For example, the heart rate and respiratory rate can be determined as diagnostic parameters from an ECG signal under optimal conditions. On the other hand, after a correction of the *Baseline* in the ECG by a steep high-pass filter above 0.5 Hz, the respiratory rate will no longer be present in the filtered signal and an evaluation in this respect will also not provide reliable values. Performing an in-depth spectral analysis of the signals (cf. subsection 5.3.2) before filtering, on the other hand, allows, for example, to determine the frequency content of the raw signal in order to identify a possible unwanted influence on the useful signal. For this and similar reasons, care must always be taken during the measurement and processing of the signals to ensure that the methods used are optimal and *signal-preserving*. The methods of the above elaborated processing steps in the signal processing chain are presented in detail in chapter 4, in chapter 5 and in chapter 6 and discussed on the basis of the selection criteria and practical application possibilities.

2.6 Post-Reading and Exercises

Information and Information Transmission

1. What characterises information? What is meant by extended information and what is it needed for?

2. What conditions must be fulfilled in the transmission of information? Explain with a simple example.
3. How is information transmitted technically or generally? Explain with one example each.
4. What is the signal-to-noise ratio used for in technology, what forms of representation do you know? To which *SNR* does 3 dB, −3 dB, 10 dB, 120 dB correspond?
5. Explain the terms data quantity, symbol and data transmission rate. What does the maximum data transmission rate depend on for an information transmission? What data transmission rates do you expect for common biosignals? Calculate the maximum data transmission rate for a bandwidth of 5 kHz and a signal-noise-ratio of 40 dB.
6. What is meant by bandwidth, what is meant by a band-limited signal? State the bandwidth of different bio-signals compared to technical signals.
7. Give a definition for biosignals, including the generalised concept of information. In your explanations, distinguish between evoked and autonomous biosignals and give an example of each. What conclusions can be drawn from the newly obtained definition?

Signals and Systems

1. According to DIN 44300, how is the term signal defined?
2. Name at least five reasons why one does signal processing.
3. What is the difference between a continuous and a discrete-time signal? What is a digital signal?
4. Explain the relationship between signal and information by way of example.
5. What is the relationship between signals and systems, on the one hand in signal-processing systems from technology and on the other hand in signal-generating systems from biology? Name the significance of each in the analysis of biosignals.
6. Name four common biosignals of the heart. What diagnoses or conclusions can be drawn from them?
7. Explain the difference between univariate and multivariate signals, use the terms un/dependent variable, give one example each.

Definition and Classification of Signals

1. Explain the meaning and general use of the terms: expected value, mean, variance and standard deviation. In each case, draw a distribution function that you would expect to see for N subjects and M subjects when measuring height repeatedly. Compare the expected values and standard deviations and justify the differences.
2. Classify the term signal and give the appropriate definitions (a subclass is e.g. deterministic signals).
3. What is the characteristic of a periodic, or a transient signal?

4. Draw an arbitrarily shaped, causal, aperiodic time signal in a labelled Cartesian coordinate system.
5. What is the relationship between the period and the frequency of a sinusoidal signal?
6. Interpret the formula of the Fourier series with regard to the postulate mentioned earlier that every signal is built up from a sum of harmonic signals of different amplitude and frequency.
7. Explain the term linear superposition using a sawtooth oscillation.
8. Use the superposition of two sine functions to produce a signal similar to the one in Figure 2.11 (bottom right). Alternatively, use the Fourier series. What is the effect of changing the amplitude and phase of each harmonic?
9. What special properties does the delta-distribution have, how can these be expressed mathematically?
10. How can a DC-function in the time domain 1(t) also be described mathematically? To do this, use a suitable limit analysis for the period of a complex exponential function.
11. Explain clearly, using a limit value consideration of the period, why the sampling function (Dirac-pulse train) again has a sampling function as a Fourier-transform, although the Fourier-transform of the delta-function is a DC-function.
12. Analogous to Listing 2.3.2 create a Matlab-script for the graphical representation of the two functions from Figure 2.13.
    (a) Change the sign in the exponent of the exponential function to create a transient.
    (b) Gradually decrease the value of the standard deviation of the normal distribution and perform the limit transition to the Dirac-distribution to some extent. To do this, use Equation 2.21 and compare the output with Figure 2.14.
13. Describe the sieve property of the Dirac-distribution and how it acts on a signal $f(t)$. What is the effect of shifting the Dirac-distribution and how is this done in the positive time direction?
14. What is the relationship between the delta-distribution and the Heaviside-function? How do you explain the connection and the discontinuity with the help of the derivative?
15. Name the properties for energy and power signals with an example for each. Check by calculation whether the square wave pulse, the Dirac pulse, the sine function, the exponential function and the DC signal are energy or power signals.
16. Explain the different terms of stationarity of a stochastic signal and give examples for the respective category.
17. Generate a mixed signal from a deterministic sine function and a noise signal in an additive way, analogous to the representation in Figure 2.17 (right). To do this, use the Matlab-functions `mean()`, `var()` and `std()` to calculate the respective magnitudes of the first two moments.

18. Generate another signal analogous to the previous task by using the cosine function. What covariance of the two signals do you expect? Check your assumption with the help of the Matlab-function `cov()`. Now replace the cosine function with a sine function and discuss the result.
19. Explain the difference between analogue and digital signals using the technical terms. What is the main advantage of processing digital signals? Using Matlab, plot a sine function as a continuous or discrete signal following Figure 2.19 using the functions provided.

Transformations of Signal Processing

1. What is the mathematical form of the Fourier series and what is it used for? How is it different from the Fourier transform?
2. What are the coefficients of the Fourier series? What is their significance in frequency analysis?
3. What is the Gibbs phenomenon, when does it occur in the representation of signals using the Fourier series? Can it be avoided?
4. Describe the consequences of the individual theorems of the Fourier-transformation, how do the Fourier pairs of linearity, time/frequency shift, multiplication and convolution have an effect in the time and frequency domain respectively? Interpret.
5. What fundamental postulate underlies the process of spectral analysis? Without this assumption, spectral analysis would be meaningless.
6. Describe the behaviour of a linear time-invariant system in the time and frequency domain with the help of the Fourier-transformation. What are the computational advantages in the individual forms of representation with the aid of the Fourier theorems?
7. Describe the integral transformations in general, what is the significance of the integral kernel? Name the most important integral transformations in signal processing and their integral kernels. What are the differences? Interpret.
8. What form does a sine or cosine transformation take? What results do you expect in comparison to the Fourier transformation for exclusively odd or even signals?
9. Under which conditions does a solution of the Fourier-transformation exist, in which cases does it not converge? Do you know a practicable solution to transform non-converging signals in another way?
10. Describe the significance of convolution in relation to linear, time-invariant systems of signal processing.
11. Explain the mathematical approach to convolution, how can its result be interpreted? Carry out the convolution of two rectangular functions as an example, interpret the convolution product. What does the convolution of a function with the Dirac momentum look like?

12. Using the table of common Fourier transformations, name characteristic features of the spectrum of a continuous periodic time signal. On the other hand, what are the characteristics of a pulse-shaped signal? Interpret the results.

3 Fundamentals of the Formation of Biosignals

This chapter introduces the history and the interesting topic and the concepts of the formation of biosignals. Following the historical explanations in [47, 71, 86], the discovery of bioelectric phenomena reaches back to the year 2750 B.C., where at first bioelectrical activity was observed in the accidental and painful contact with electric fish. Besides the catfish, which generates voltages on the order of 350 volts, the electric ray[1] is also known for this characteristic. The latter, despite considerable potentials between 60 and 230 volts, was used for medical treatments as early as the 1st century AD.

The mode of action became clearer in the 18th century, based on the work about electricity by the physicist Georg Christof Lichtenberg[2], the Briton John Walsh was the first to make the discharge of the organ visible by a flash of light during his investigations [80]. At about the same time, the Italian physician and anatomist Luigi Galvani[3] discovered by chance the contraction of prepared frog legs when they came into contact with a bimetallic strip of copper and iron [5] . These findings raised the fundamental question of the significance of electricity for living organisms – it was even regarded in those days as the central distinguishing feature between animate and inanimate matter.

According to the historical records in [5, 12], Luigi Galvani and the physicist Alessandro Volta[4], who at the time were studying the phenomena of electricity, also considered electricity to be the basis of the activity of the nerve and muscle cells of living beings. Galvani himself called the phenomenon "animal electricity" based on his findings about the electrical activity of fish and claimed to have discovered a new form of electricity in his experiment. He assumed that animal electrical energy present in the tissue was discharged by contact with the metals and that this triggered the contraction of the muscle. Volta, on the other hand, assumed that there was only one form of electrical energy and claimed that the movements caused by contact with the two metals were due to an electrostatic potential difference between the metals, which was transferred to the animal's muscle. To argue against independent animal electricity, he presented the voltaic pile[5] as a model of an electrical organ, thus he demonstrated the similarity to the natural electrical organ of the electric ray [12]. Even if Volta was

1 Torpedinidae from Lat. *torpere* = to be paralysed.

2 Georg Chr. Lichtenberg (1742–1799), German physicist and one of the co-founders of the theory of electricity; he became known for the Lichtenberg-figures named after him.

3 Luigi Galvani (1737–1798), Italian physician, physicist, biologist and philosopher, was considered a pioneer of bioelectromagnetics.

4 Alessandro G. A. A. Earl of Volta (1745–1827), Italian physicist, inventor of the battery and co-founder of the theory of electricity.

5 The voltaic pile consisted of thin zinc and copper discs separated by cardboard discs soaked in saline solution.

https://doi.org/10.1515/9783110736298-003

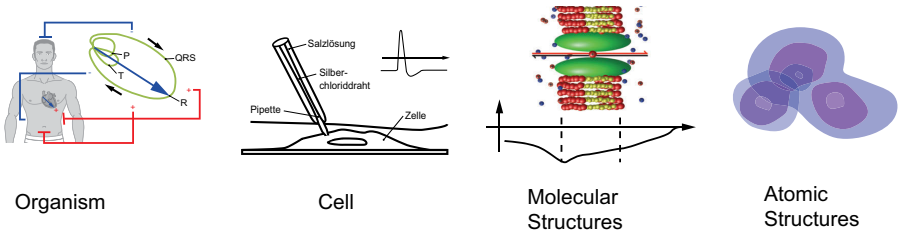


Fig. 3.1: Biophysical approach and model descriptions of the electrophysiologic mechanisms at the different scales.

right in his statements, the foundations of electrophysiology that are still valid today were made much later.

A number of approaches have been established over time to describe the underlying processes, Figure 3.1 considers them according to their scaling. In this, the viewpoint differs in the richness of detail of the models, the associated questions to be answered as well as their validity.

The mode of action of the electrical activity of nerve and muscle cells inside the body and outside a living organism is described by the laws of electrophysiology and electrostatics/-dynamics by the Maxwell-equations.[6] The latter is applied, for example, in the measurements of electrical potentials on the surface of the body or by needle or microelectrodes on the scale of the living organism down to individual cell clusters. Diffusive processes and movements of individual ions at the cell scale, on the other hand, are addressed by the Nernst-Planck-equation[7] of phenomenological thermodynamics. Here, the processes in the cell, for example the electrodiffusive movement of ionic currents, are considered down to individual ions and their effects on the scale above. The arrangement of supermolecular structures, as they occur, for example, in the formation of a lipid membrane or the arrangement of membrane proteins in a cell wall, requires the methods of statistical thermodynamics, the laws of which are described by the Poisson-Boltzmann-equation[8]. Interactions of atomic and molecular processes are ultimately quantum mechanically explained by the Schrödinger-equation[9] or simplified molecular dynamics described by New-

6 Fundamental equations of electricity and magnetism, named after the Scottish physicist James C. Maxwell (1831–1879).

7 Fundamental equation of the movement of ions taking into account the electric field, named after the Nobel Prize winners Walter H. Nernst (1864–1941) and Max Planck (1858–1947).

8 Fundamental equation of electrostatic interactions between molecules in liquids with ions dissolved in them, named after Simeon D. Poisson and Ludwig Boltzmann.

9 Fundamental equation of quantum mechanics, named after the Austrian physicist and Nobel laureate Erwin R.J. Schrödinger (1887–1961).

ton's equations of motion[10] in addition to empirical force fields. At this scale chemical bonds and the interaction of individual ions with, for example, functional proteins such as ion channels can be explained. The latter methods probably offer the most accurate description of the processes, but due to the limited computing power of today's mainframe computers, they are only able to describe small systems over very short periods of time (ps to ns).

In biosignal processing, the description of measurable phenomena on the cellular scale and their effects on the scale of the entire organism are primarily applied. Starting from the cellular electrophysiological processes of individual ions, the far-field view is elaborated in the course of the chapter. Due to the analogy to AC technology we refer to the basic work by Harriehausen and Schwarzenau [26]. In the further course of the chapter the most important facts from electrical theory and physiology are introduced in section 3.1. The emergence and propagation of action potentials is considered in subsection 3.1.3 and exemplified on the electrical excitation of the heart muscle in section 3.2. Finally the chapter extends this view to biosignals in general and concludes by a taxonomy of biosignals in section 3.3 and a section consisting of post-reading and exercises in section 3.4.

3.1 Physiology and Electrical Activity of Muscle and Nerve Cells

Electrophysiology deals with the electrochemical signal transmission in the nervous system, which is of great importance both in the field of clinical electrophysiology, such as cardiology and neurology, and in experimental electrophysiology, such as neurophysiology and muscle physiology. Electrophysiological measurements are used in these disciplines, for example, to check nerve and muscle activity. Ion flows in or on biological tissue are measured and evaluated. Well-known procedures are electroencephalography, electroneurography, electromyography and electrocardiography. Cardiologists, for example, examine the electrical potential courses at the heart muscle by measuring potentials at the body surface (non-invasive) or within the framework of a special heart catheter examination (invasive) before implanting a cardiac pacemaker. In addition to measuring signals, methods for stimulating these electrophysiological systems for therapeutic and diagnostic purposes are being developed in parallel. These include, for example, the stimulation of nerve tissue in the brain by provoking sensory stimuli, the stimulation of muscle contractions in cardiac or gastric pacemakers, or stimulation/inhibition by pharmaceuticals. In the latter case, the electrophysiological measurements provide important quantitative statements about their effectiveness and applicability.

10 Equations of motion of classical mechanics, named after the English physicist and mathematician Isaac Newton (1642-1726).

To analyse the electrodynamic potential changes on the surface of the body, an understanding of the electrophysiological relationships of individual nerve or muscle cells or the associations of such cells is required. These are investigated within the framework of experimental neuro- and muscle physiology using invasive microelectrodes. Although the findings emerged from the interplay of theoretical and experimental research, this section will introduce the physiological and relationships in an electrical analogy. Experimental methods for measurement and common models used are presented in section 3.2.

3.1.1 Formation and Function of Biomembranes

The cell membrane or biomembrane consists of a closed double layer of phospholipids (cf. Figure 3.2), i.e. a chemical structure consisting of a phosphate group and one or more hydrocarbon chains. Due to the *amphiphilic*[11] properties of the polar, *hydrophilic* (water-attracting) head group and *hydrophobic* (fat-attracting) ends of phospholipids, they form energetically favourable molecular aggregates (-associates), so-called association colloids in aqueous solutions, depending on concentration and temperature.

The phase separation is caused by the energetically favourable arrangements between the negatively charged hydrophilic head groups and the polar water, as well as by the contact avoidance of the water with the hydrophobic fatty acid-containing ends by repulsive interaction. This process is called self-assembly.

Lipid bilayers, i.e. a parallel and facing arrangement of the hydrophobic ends and a resulting membranous arrangement of the head groups on the two opposite sides, occur at high lipid concentrations (cf. Figure 3.3). At a lower mass concentration, the critical micellar concentration (CMC) on the other hand, so-called micelles[12]

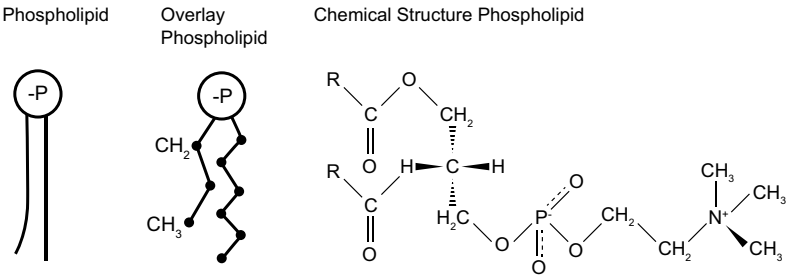


Fig. 3.2: Phospholipid in simplified representation (left), with overlay of chemical structure (middle) and as chemical structural formula (right).

11 From Greek *amphis* = both and *philia* = love, friendship; composite with polar (water-loving) head groups and lipophilic (fat-loving) ends.

12 From Latin, *mica*. *mica* = lump, small morsel.

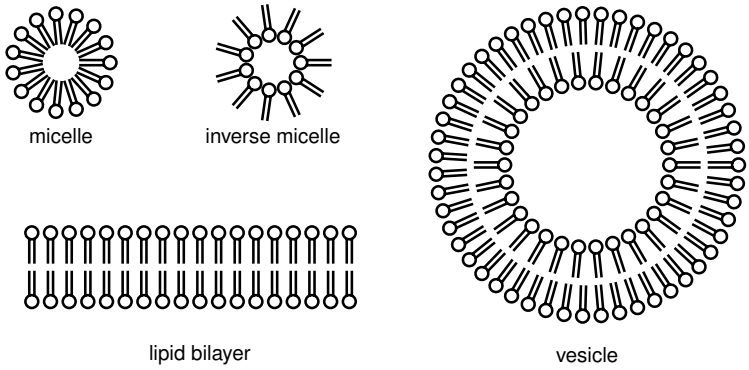


Fig. 3.3: Behaviour of the self-assembly of lipids in aqueous solution: depending on the concentration and temperature of the lipid-water-mixture, different liquid crystalline structures are formed. At a low lipid concentration, micelles (top left) are energetically most favourable. If the concentration is increased, lipid bilayers (bottom left) and closed vesicles (right) form. Inverse micelles are usually found only in organic solvents (top centre).

are formed. During phase separation, the hydrophilic head groups align with the adjacent water molecules, whereas the hydrophobic ends cluster together to form their own internal lipid phase. Inverse micelles with reversed arrangement, on the other hand, only form in organic solvents, because there the interactions behave in exactly the opposite way compared to water.

If the lipid concentration is further increased, so-called vesicles are formed.[13] These are about one micrometre in size and are responsible for the transport of many substances into and between cells. Substances stored in the vesicles are released, for example, by fusion of the vesicles with the cell membrane. These synaptic vesicles are involved in the transmission and processing of nerve impulses through the release of neurotransmitters in the synaptic cleft. The liposomal effect is also used for target-oriented drug release, corresponding migration processes are observed by so-called tracers.[14]

Above stated properties site phospholipids a central role in biological systems such as the cell and their signal interactions. In the cell membrane they form a lipid-double layer in which numerous functional molecules such as proteins or glycolipids[15] are embedded (cf. Figure 3.4).

13 *vesicula* = vesicle; roundish small intracellular (located in the cell) vesicles surrounded by a single or double membrane.

14 A labelled substance (radioactive or fluorescent) being introduced into the living body to participate in the metabolism.

15 Glycolipids (Greek *glykys* = sweet, *lipos* = fat) are phosphorus-free membrane lipids on the outside of the lipid double layer.

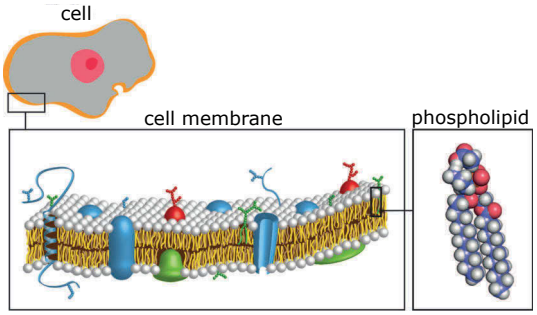


Fig. 3.4: Schematic representation of a cell (top), a cell membrane as a lipid bilayer with embedded membrane proteins (bottom left) and a phospholipid (bottom right).

In electrical analogy, the membrane wall itself is a well insulating layer with electrical resistance, $R_{LDS} \sim 10^9\,\Omega$, which, due to its impermeability for ions, acts like an energy storage of a capacitance C_{LDS}. Figure 3.5 shows a lipid bilayer (left) with corresponding Na^+, K^+ and Cl^- ion distribution between intra- and extracellular space. The corresponding ion concentrations for one ion species $A^{\pm}$, in the intracellular space $c^{\mathrm{i}}_{A^{\pm}}$ and $c^{\mathrm{e}}_{A^{\pm}}$ in the extracellular space, are shown in the equivalent circuit on the right (concentrations found in Figure 3.9).

However, the functional properties of the biomembrane result from the behaviour of the embedded proteins, which, for example, act as *receptors* for certain substances or as *transporter proteins* for the transport of certain ion types or metabolic products across the cell membrane. A distinction is made between *diffusion-driven passive channel proteins* in the form of a simple pore and so-called *active transport proteins* in the form of a molecular pump (cf. Figure 3.6). Passive transport through an open channel (pore) occurs through diffusion along a gradient of the electric field or the substance concentration of certain ions. Active metabolic transport requires energy

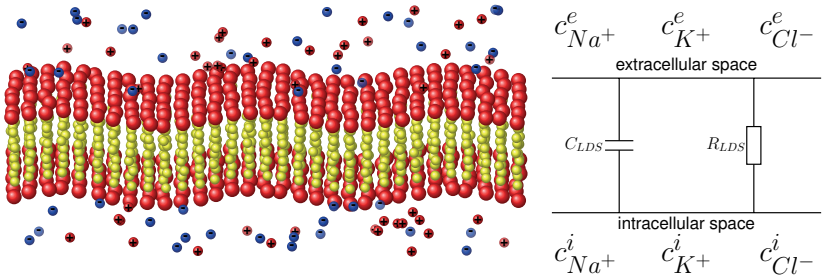


Fig. 3.5: Schematic representation of a nearly impermeable cell membrane as a lipid bilayer (phosphate-head group – red, hydrocarbon-lipids – yellow): in the intra- and extracellular space there are positive (red) and negative (blue) ions. The corresponding equivalent circuit of the membrane wall built by an electrical capacitance C_{LDS} and the insulation resistance R_{LDS}.

in the form of adenosine triphosphate (ATP) and is thus able to transport substances even against these existing gradients. Both forms are caused by the *conformational change* of the proteins, which in the case of the pore causes the opening/closing and in the case of the transporters a folding over. The conformational change of the proteins can have different causes: For example, there are channel proteins which open or close due to the change in electric field gradients; others react to mechanical stimuli or docking-mechanisms of messenger substances. The passive sodium or potassium channels of the nerve cell, for example, can be activated electrically, while the potassium-sodium pump is concentration-controlled. Channels usually have a signalling function due to their high transport rates, in contrast the much smaller transport rates of the transporters are counted as the housekeeping function of the cell [47, 67].

However, both types allow the passage of ions across the membrane wall, inducing a conductance like an ohmic resistor, which can change with time depending on the actual state (open/closed). The current flow of charged molecules or ions across the biomembrane (*electrogenic transport*), is the basis for most cellular signals such as the emergence and propagation of the *action potential* or local changes of field gradients etc.. Since the electrical conductivity of the cell membrane is very low, by Ohm's law $U_{\mathrm{LDS}} = R_{\mathrm{LDS}} I_{\mathrm{LDS}}$ even small ionic currents lead to reasonable high potential differences of $U_{\mathrm{LDS}} \sim$ mV across the cell membrane [47, 67]. Besides the *ion selective* conductivities or permeabilities of the membrane wall, the electrical activity of nerve and muscle cells is also based on *asymmetric* ion distributions (for Na^+, K^+, Ca^{2+} and Cl^-) between the intra- and extra-cellular space. Assuming the cell in thermodynamic equilibrium, ion concentrations and permeabilities can be considered as constant quantities and the resting potential is given by the Goldmann-Hodgkin-Katz-equation (detailed derivation in [13, 47])

$$U_{\mathrm{RP}} = \frac{RT}{F} \ln \left(\frac{P_{\mathrm{Na^+}} c^{\mathrm{e}}_{\mathrm{Na^+}} + P_{\mathrm{K^+}} c^{\mathrm{e}}_{\mathrm{K^+}} + P_{\mathrm{Cl^-}} c^{\mathrm{i}}_{\mathrm{Cl^-}}}{P_{\mathrm{Na^+}} c^{\mathrm{i}}_{\mathrm{Na^+}} + P_{\mathrm{K^+}} c^{\mathrm{i}}_{\mathrm{K^+}} + P_{\mathrm{Cl^-}} c^{\mathrm{e}}_{\mathrm{Cl^-}}} \right) . \tag{3.1}$$

R and T are the ideal gas constant and the absolute temperature respectively. Given the concentration gradients between the intracellular i and extracellular space e of the individual ion species $c_{\mathrm{Na^+}}$, $c_{\mathrm{K^+}}$ and $c_{\mathrm{Cl^-}}$ and the permeabilities $P_{\mathrm{Na^+}}$, $P_{\mathrm{K^+}}$ and $P_{\mathrm{Cl^-}}$ the resting potential can be calculated, concentrations of the mammals given in Figure 3.9 lead to $U_{RP} \sim -70\,\mathrm{mV}$. The generation of nerve impulses (action potentials) is limited by a temporal exchange of these charges according to time varying membrane conductivities (cf. Figure 3.9).

3.1.2 Analogy to Electrical Circuits

The basic principles of the electrical activity of cells in the previous section led to a description via the electrical membrane capacitance and the membrane resistance. Even

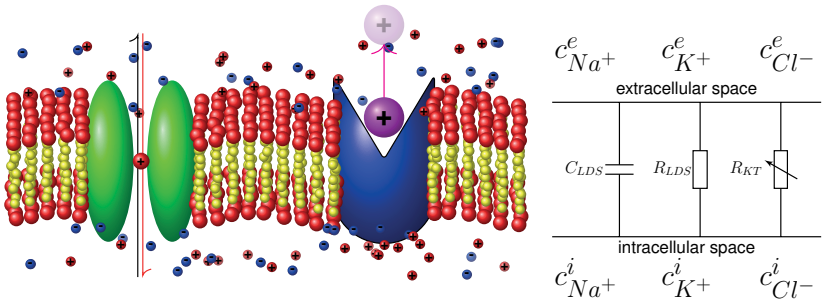


Fig. 3.6: Schematic representation of a cell membrane with an embedded, open channel protein (green) and an outward facing transport protein (blue): compared to closed channels, the variable sum resistance due to the embedded channel and transport proteins R_{KT} is much lower than the membrane resistance R_{LDS}.

though not exactly identical to the analogue components of electrical engineering, they deliver quite good predictions about the temporal course of the currents and potentials in the cell. Thus for simplicity, electrical equivalent circuit diagrams are used as models. They consist of simple electrical components, such as resistors, batteries and capacitors and thus allow a prediction of the measurement results of an electrophysiological experiment. According to the model of Hodkin and Huxley[16] electrical properties of the cell membrane can be represented by the electrical equivalent circuit (cf. Figure 3.7).

As shown in Figure 3.5 ff. the cell membrane is represented by a combination of the individual conduction channels for Na^+, K^+, Cl^- (for the nerve cell) and additionally Ca^{2+} (for the cardiac muscle cell), the membrane capacity C_{LDS} and the respective

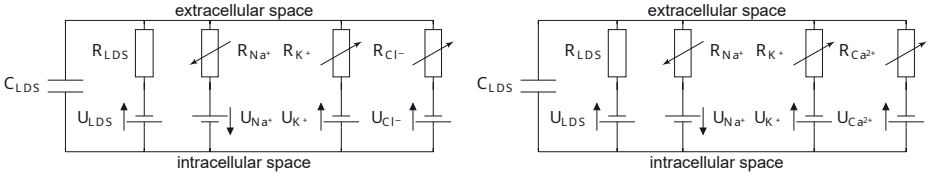


Fig. 3.7: Electrical equivalent circuit based on Hodkin and Huxley's cell membrane model [29] for the nerve cell (left) and the cardiac muscle cell (right): Ion channels are represented by the variable resistances R_{Na^+}, R_{K^+} and R_{Cl^-} (nerve cell, left) and additionally $R_{Ca^{2+}}$ (cardiac muscle cell, right). The membrane wall is represented by a capacitance C_{LDS} and the leakage current resistance R_{LDS}. The electrochemical grandient of an ion species A (cf. Figure 3.9) generates a corresponding current flow I_A through the respective ion channels.

16 Alan Lloyd Hodgkin (1914-1998) and Andrew Fielding Huxley (1917-2012), Nobel prize winners of physiology for their discoveries about the ionic mechanism involved in excitation/inhibition of the nerve cell membrane.

proportions of the membrane voltage. The Hodkin and Huxley model includes only the ion species Na^+ and K^+ and their conduction across the ohmic resistances R_{Na^+} and R_{K^+} and a leakage current that can flow across the ohmic resistance R_{LDS}. The respective fractions of membrane voltages $U_{A^\pm}$ result from the ion distributions between intra- and extracellular space (cf. Figure 3.9); these are counter-directional in the case of Na^+ and K^+, consequently the voltage sources are also counter-directional. In subsection 3.1.3, the emergence and propagation of action potentials of the nerve cell is described in more detail. There, an action potentials is triggered by a stimulus (coming from outside) under certain conditions, which propagate along the axons. During this process, the permeabilities or the conductances $g_{A^\pm} = R_{A^\pm}^{-1}$ change for individual ion species with time. In the electrical equivalent circuit, this behaviour is represented by variable resistances. The instantaneous total current I_m through the cell membrane is obtained analogous to electrical engineering:

$$I_m = I_{Na^+} + I_{K^+} + I_{Cl^-/Ca^{2+}} + I_{LDS} + C_{LDS}\frac{dU_{LDS}}{dt} . \tag{3.2}$$

In the course of the book, further equivalent circuit diagrams, for example for nerve conduction in the axon (cf. subsection 3.1.3), and basic models of heart excitation (cf. section 3.2) are modelled in detail using Scilab/COS. In chapter 4, their significance in the measurement of potentials at the body surface becomes apparent.

3.1.3 Emergence and Propagation of Action Potentials

In the section 3.1 the electrophysiological basics of the resting potential were clarified under the assumption of a thermodynamic equilibrium and the temporal processes in the emergence of *nerve impulses*, the so-called *action potentials*, were deliberately neglected. This section will now deal with the mechanisms of the emergence and propagation of nerve impulses in the body.

Physiology of the Nervous System

The basic building blocks of the central nervous system (CNS) and the peripheral nervous system (PNS) and the sensory and motor systems are the *neuron* or nerve cells. A neuron has several synapsessynapses (signal inputs) and an axonal (signal output) and is therefore often compared to a multi-input single-output-system (MISO system) of signal processing (cf. Figure 3.8). Signal processing in neuron networks takes place on an electrochemical basis. Thus, through the release of chemical messengers in the synaptic cleft of a neuron and internal processing (decision) by the nerve cell, an electrical nerve signal is generated and transmitted through the axon to other nerve cells.

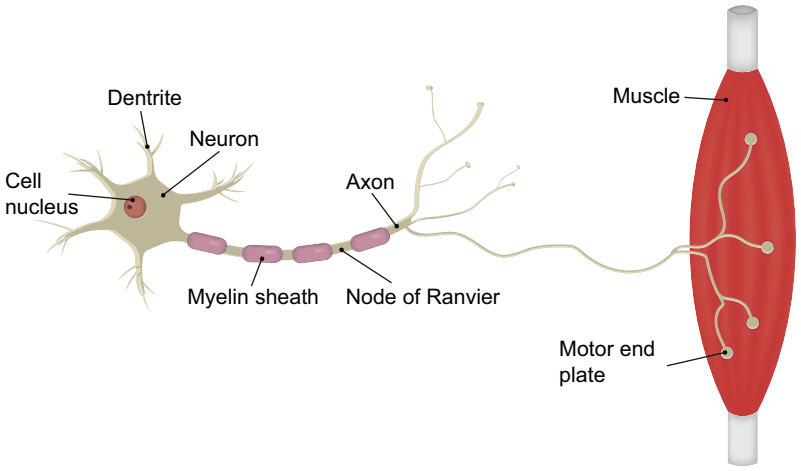


Fig. 3.8: Schematic representation of a nerve cell (neuron) with its dentrides and the synapses arriving there (left), an outgoing axon (nerve conduction) with myelin sheath and branches (middle) and the muscle as a motor system with its end plates (right)

Formation of the Action Potential

In the resting state of the nerve cell, the ion concentrations inside and outside the cell are different. This state is maintained by the sodium-potassium pump and leads to a potential difference between –50 mV and –100 mV (inside negative, outside positive). During the generation of the action potential, this condition changes fundamentally due to a temporary opening-/closing of the Na^+- and K^+- channels. In this process, the permeabilities for the individual ion species change in a very specific temporal sequence. The reason for such a change can be, for example, a brief change in the electric field (potential) in the vicinity of a channel, which then causes an opening and thus a further change in the potential. Such a short-term change may have been triggered by a stimulation of the nerve cell by other nerve cells (via the release of neurotransmitters in the synaptic cleft), as in peripheral somatic stimulation, or, for example, by an optical stimulus in the visual cells. The Na^+- and K^+- ion fluxes are, from an electrical point of view, a charge shift across the membrane wall (cf. Figure 3.9, left) and lead to a change in the electrical potential. The sequence of this voltage reversal is different for nerve and muscle cells. In the case of the nerve cell, the principle sequence (cf. Figure 3.9) of this so-called *depolarisation* is as follows:

1. *depolarisation*: Voltage-gated Na^+-channels are opened due to a short-time exceeding of a threshold potential (membrane threshold), so that Na^+ ions flow into the cell due to the concentration gradient. Afterwards, the channels close again. The membrane potential changes from the resting potential to a positive maximum value of about +20 mV (overshoot). Due to the absolute refractory period,

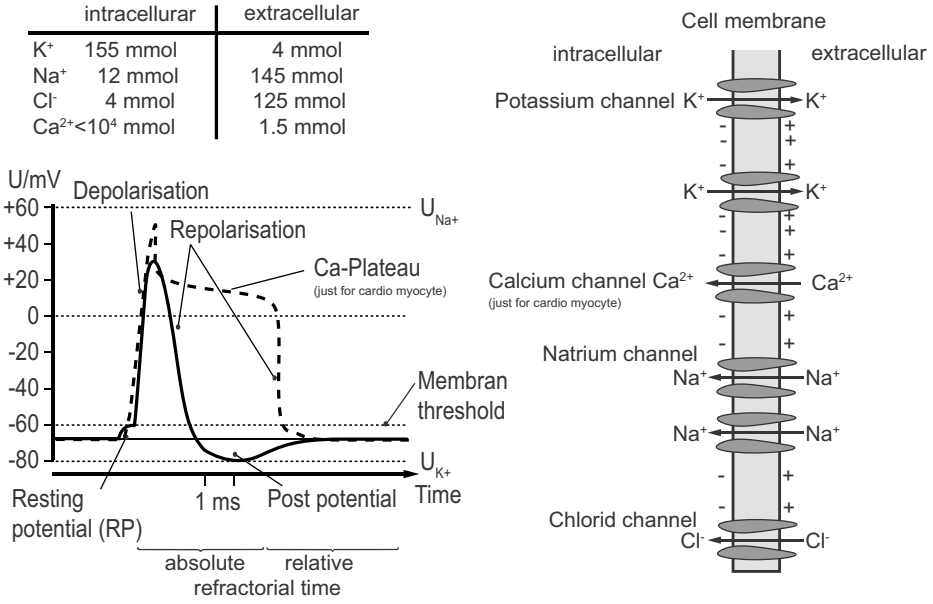


Fig. 3.9: Potential course of an action potential and the Na^+, Cl^- - and K^+ Ion currents through the membrane wall of a nerve cell (mammal intra-/extracellular concentrations from [67]) and the sum of the ionic currents as action potential (bottom left): due to the negative post potential, renewed excitations of action potentials are only possible during the relative refractory period. In the case of cardiac myocytes, there is still a Ca^{2+} ionic current, which is responsible for the Ca plateau. The membrane wall (right) shows a temporary state during the post-potential phase with the corresponding active channels.

no re-triggering is possible over a short period of time after the action potential has been triggered.

2. *repolarisation*: During this period, the Na^+- channels close again, whereas the K^+-channels open with a delay.
3. *hyperpolarisation*: Due to the concentration gradient, there is a flux of K^+ ions out of the cell and a lowering of the membrane potential. The delayed closure of the K^+- channels with respect to the Na^+- channels leads to an overshoot called hyperpolarisation. During this time, the potential threshold is increased for the renewed triggering of an action potential (relative refractory period). Afterwards, the equilibrium potential or resting potential of the cell is restored.

In the case of a muscle cell or cardiac muscle cell, the Ca^{2+}-ionic currents must also be taken into account, which, in contrast to the nerve cell, produce a pronounced plateau in the potential curve of the action potential (cf. Figure 3.9).

Propagation of the Action Potential

The electrical communication of the nerve cells with each other in the brain, with sensory cells or to stimulate the contraction of distant muscle cells occurs by means of the action potential. Accordingly, there are nerve pathways in the body on which the action potentials can spread, the so-called *axon*. Basically, a distinction is made between two forms of conduction, the *continuous*, i.e. the conduction from one point of the axon to the directly adjacent one, and the so-called *saltatory* conduction, in which the action potential spreads abruptly along the axon.

Saltatory excitation conduction is mainly found in vertebrates due to the faster propagation of the action potential, as here usually larger distances have to be overcome and delays in the transmission of nerve stimuli would be intolerable. An example of this is the whale, in which a nerve stimulus may have to travel over 50 m before it reaches the brain. The basic structure of saltatory nervous system differs from the continuous form by an additional insulation of the axon in the form of the so-called *myelin sheath* (cf. Figure 3.8). The saving of the myelin sheath in continuous nerve conduction is due to space problems, which is why it is mainly absent in smaller creatures such as insects and on the last centimetres of a vertebrate nerve conduction. Besides the lower propagation speed, another disadvantage of continuous excitation conduction is that action potentials are subject to strong attenuation. In contrast to saltatory excitation conduction, the action potential is not re-formed at the approximately five millimetre long *node of Ranvier* between the sections of the insulating myelin sheath. Because the action potential is constantly regenerated in saltatory excitation conduction, attenuation along even long pathways is negligible. This means it doesn't just arrive faster, it is also transmitted with *constant* amplitude. At the same time, it can be concluded that the amplitude of an action potential cannot stand for its stimulus intensity, i.e. stronger stimuli produce constant amplitudes, but an increased frequency of action potentials. However, there is also an upper stimulus limit for this, which can be explained with the help of the previously discussed refractory period of the nerve cell. When the upper stimulus limit is exceeded, the frequency of the action potentials no longer increases, although the stimulus continues to increase.

Continuous Excitation Conduction

In continuous excitation conduction, the action potential is transmitted by depolarisation of directly adjacent nerve cells in the axon. This means that after stimulation of a nerve cell, e.g. by incoming signal stimuli in the synaptic cleft between the dendrites and the nerve cell, an action potential spreads along the axon by depolarising directly adjacent nerve cells in the axon due to the increase in potential and these in turn depolarise the adjacent ones and so on. Thus, each nerve cell must depolarise in turn in the axon before the signal arrives at the other end of the nerve. The propagation velocities achieved in this process range from 1 to 5 m/s. The nerve conduction velocity is

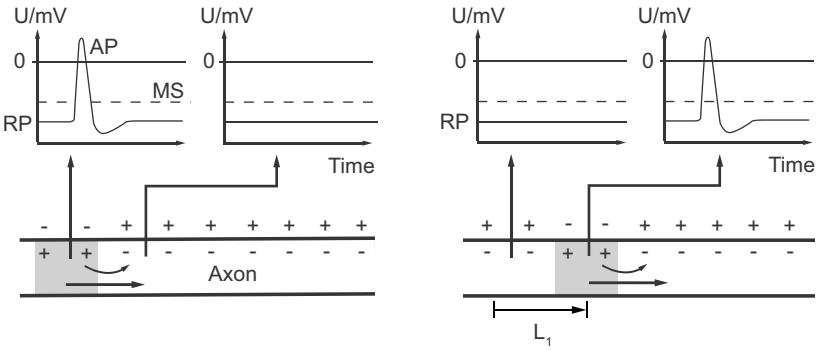


Fig. 3.10: Propagation of the action potential in continuous excitation conduction: the action potential propagates as a wave along the axon with a nerve conduction velocity of 1 to 5 m/s; the partial figures on the left and right refer to different times.

indeed a diagnostic parameter that is used in the daily practice of neurologists for the diagnosis of numerous nerve diseases such as multiple sclerosis or amyloidosis.

In contrast to the invasive measurement with the help of microelectrodes as shown in Figure 3.10, the doctor usually measures with surface electrodes (cf. chapter 4) and determines the speed of propagation from the running time and the distance between the measuring points. If this is reduced, this indicates a pathological situation of the measured nerve. Figure 3.10 also shows the depolarisation and charge shift of the ions associated with the conduction. Analogous to the electrical resistance of a conductor, a conductance and an attenuation or a decay interval of the action potential can also be determined in the case of the axon. The specific resistance g_a of an axon moves around 30 Ωm, the associated decay distance l, at which the input potential has decayed to half, is a few millimetres.

Saltatory Excitation Conduction

The saltatory excitation conduction, which only occurs in vertebrates, has been optimised by evolution with regard to the paths and response times to be overcome with it by a periodic isolation from myelin sheath and a node of Ranvier. With this excitation conduction, the action potential is not transmitted to the directly adjacent nerve cells in the axon by depolarisation, but due to the field propagation in the axon, it jumps across the sections of the myelin sheath and is regenerated in the next node of Ranvier. This type of propagation is advantageous not only with regard to the speed of propagation, which differs by a factor of about ten, but also because of the preservation of the amplitude of the action potential. However, nerve conductors require more space due to their isolation, which is why they only occur in larger organisms. This is also the reason why continuous nerve conduction is unavoidable in the last millimetres from the nerve end, even in vertebrates. However, the associated attenuation is

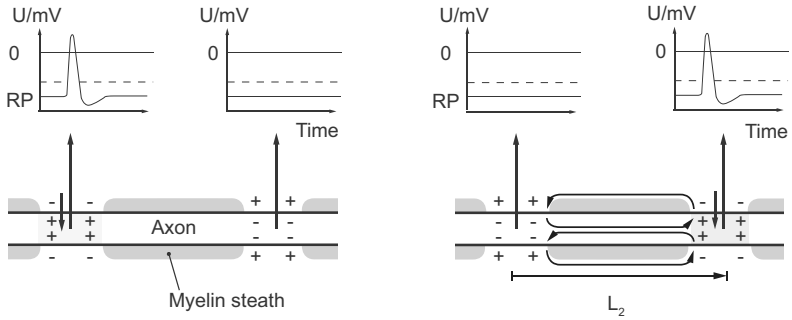


Fig. 3.11: Propagation of the action potential in saltatory excitatory conduction. The action potential propagates by regeneration at the node of Ranvier abruptly with a nerve conduction velocity of 10, . . . , 50 m/s; the partial illustrations on the left and right refer to different times.

not a hindrance to function. Nerve diseases such as multiple sclerosis result from the dissolution of the myelin sheath, i.e. from the body's autoimmune reaction to destroy its own cells; on the other hand, in diseases such as amyloidosis or Alzheimer's disease, misfolded proteins accumulate in the node of Ranvier, which damage the erratic conduction.

The propagation of the action potential along a saltatory excitation conduction is shown in Figure 3.11. Depolarisation here occurs abruptly from node of Ranvier to node of Ranvier, resulting in propagation velocities between 10 and 50 m/s. The depolarisation occurs practically instantaneously due to the propagation of the electric field across the sections of the myelin sheath, i.e. the potential in the adjacent node of Ranvier rises faster than in the area of the myelin sheath due to the much lower electric capacity there and thus exceeds the membrane threshold, so to speak, before the action potential reaches the node of Ranvier.

3.2 Electrophysiology of the Heart

The transfer of electrophysiological processes to the level of the organism is exemplified in this important section on the heart physiology using biosignal processing. However, as explained in section 3.3, the knowledge gained in this process can easily be used for other parts of the body such as muscle and nerve activity. The Respective applications are explained in more detail in chapter 6 using a practical example.

3.2.1 General Excitation of Muscle Cells

The heart muscle pumps blood through the heart-circulatory-system, which is distributed throughout the body in pulse waves to supply the cells with the required nutrients and oxygen for metabolism. Without oxygen, the cells could not fulfill their tasks in the body. Because this transport is so important, the heart contains autonomous pacemaker cells that control the excitation of the heart muscles on their own, without the need for a connection to the central nervous system in the brain. For example, if the heart were removed from the body and stored in a suitable nutrient fluid, it could continue to beat without external stimulation from outside. Two types of heart cells are distinguished:

1. The *cardiac muscle cell*, like other muscle cells, can generate an action potential when they are in turn excited by an external action potential. When excited, they contract and reduce the size of the anterior and main chamber of the heart. The resulting increase in internal pressure leads to the opening of the heart valves and to the ejection of blood. An example of such an action potential is shown by Figure 3.12. Unlike other muscle cells such as from skeletal muscles, the force with which they contract does not depend on the strength of the excitation. A heart muscle therefore does not contract more forcefully when the electrical stimulus increases. So the all-or-nothing-principle [49] applies here, in contrast to the arm muscle, whose strength can vary depending on the physical exertion.
2. The *pacemaker cells* can also generate action potentials on their own *without external stimulation*, as they do not have a constant resting potential. Rather, after the relaxation phase, the potential grows independently until the excitation threshold is reached and a new action potential is triggered. These cells are very common:
   - in the *sinus node (SA node)* – in the right upper atrium,
   - in the *atrioventricular node (AV node)* – just before the transition from the atrium to the main ventricle,
   - in the *His bundle* – in the ventricular limb after the AV node between the main chambers and
   - in the *Purkinje* fibres – in the left and right walls of the main chamber after the end of the ventricular thigh.

Because of their specific tasks, the action potentials of the heart cells show considerable differences from the action potentials of the other nerve cells in the body, the description of which is possible by the Hodgkin-Huxley-equations.

In the group of pacemaker cells, the cells in the SA node generate the clock-determining action potentials for the entire heart. The other pacemaker cells synchronise to this. After a delay, their action potential reaches the cells in the AV node and from there to the cells in the His bundle and the Purkinje-fibres (HP complex). The action potential coming from the sinus node is the quickest (see Figure 3.12).

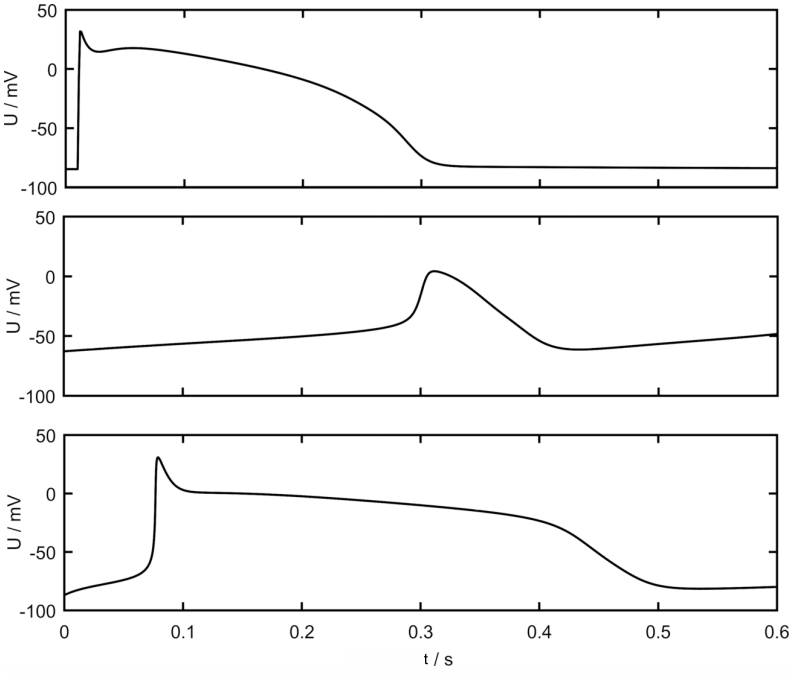


Fig. 3.12: Action potential after external excitation of a muscle cell in the ventricle according to the model of Beeler and Reuter (top) [2], action potential according to the model of Yanagihara for the sinus node (SA) (middle) [87], and action potential according to the model of Noble for the Purkinje-fibres (bottom) [54].

The action potential of the cells in the Purkinje-fibres as well as in the muscle fibres takes much longer: 300–400 ms compared to 3 ms in the large nerve cell of an octopus, which can be described according to the Hodgkin-Huxley-equations. In other words, the cells of the sinus node produce an oscillation with the highest frequency, on average at 60 to 80 oscillations periods per minute. The oscillation of the AV node is slightly lower (40 to 50 per minute), and the cells in the His bundle and in the Purkinje-fibres have the lowest frequency of about 39 to 40 oscillations per minute. The associated action potentials are shown in Figure 3.12.

Since the heart does not beat at a combination of these oscillatory frequencies in the non-pathological case, the sinus node synchronises the AV node and the latter synchronises the excitation by the His bundle and the Purkinje-fibres. There exists a certain redundancy; if, for example, the sinus node fails, the heart can still continue to beat, but at the lower frequency of the AV node. If the AV node also fails, it can still continue to work with the excitation from the His bundle and the Purkinje fibres, although the heartbeat is then even lower. The interaction of the individual pacemaker centres (SA node, AV node, HP complex) can be described by three oscillators, whereby coup-

lings from the SA node to the AV node and from the AV node to the HP complex must be taken into account in any case (see Figure 3.13).

3.2.2 Measurement of Electrical Potentials at the Body Surface

The ionic charges present in a body generate an electric field and, in the case of ionic movement, also a magnetic field. These fields usually cannot be measured directly at the point of origin, e.g. at the membrane of a pacemaker cell of the heart, where the migration of sodium and potassium ions occur. However, there is the possibility of a non-invasive measurement via electrodes attached to the surface of the body. This is because the fields generated at the point of origin spread to the surface of the body.

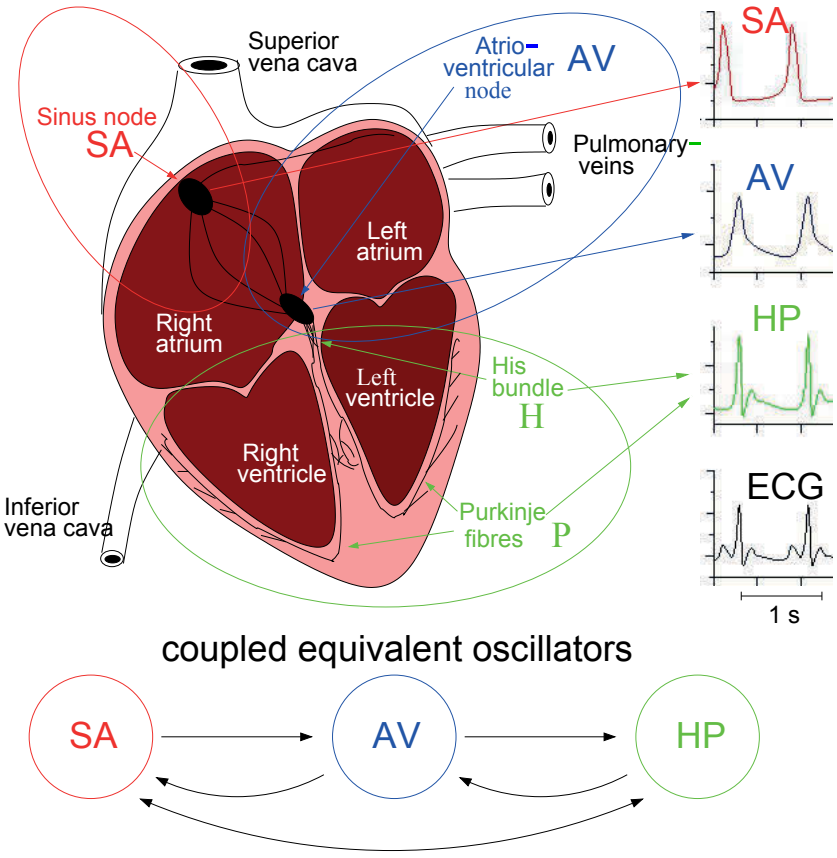


Fig. 3.13: Heart with excitation guidance system (top left); illustration of the equivalent pacemaker-oscillators (SA, AV and HP oscillators) with coupling (bottom); associated action potentials measured at the body surface by electrocardiogram (ECG) (top right).

However, they are attenuated on their way there and also overlap with the fields of ions from other points of origin.

In general, Maxwell's equations [18, 47, 61] also apply here:

$$\nabla \times H = \epsilon \frac{\partial E}{\partial t} + J \,, \tag{3.3}$$

$$\nabla \times E = -\mu \frac{\partial H}{\partial t} \,, \tag{3.4}$$

$$\nabla \cdot (\epsilon E) = \rho_{\mathrm{v}} \,, \tag{3.5}$$

$$\nabla \cdot J = -\frac{\partial \rho_{\mathrm{v}}}{\partial t} \,, \tag{3.6}$$

$$\nabla \cdot (\mu H) = 0 \,. \tag{3.7}$$

In Equation 3.3 to Equation 3.7, E and H are the electric and magnetic field strengths, J is the current density, ϵ and μ are the electric and magnetic permeabilities, ρ_{v} is the space charge density and ∇ is the Nabla operator ($\nabla := \frac{\partial}{\partial s} = e_x \frac{\partial}{\partial x} + e_y \frac{\partial}{\partial y} + e_z \frac{\partial}{\partial z}$ for Cartesian coordinates). These equations can be simplified and reformulated for medical considerations:

- The electrical and magnetic permeabilities are the same as those of the vacuum, i.e. $\epsilon = \epsilon_0 = 8.854 \cdot 10^{-12}\mathrm{F/m}$ and $\mu = \mu_0 = 4\pi \cdot 10^{-7}\mathrm{H/m}$.
- Since the static space charge density ρ_{v} is negligible in a conducting system such as a body, Equation 3.5 need not be considered. Although there are charged ions on a cell membrane, they balance each other on the different sides of the membrane, such as in a plate capacitor.
- The current density J can be divided into a current passing through the electric field $J_{\mathrm{E}} = \kappa E$ (κ for electrical conductivity), and an ionic current that flows between the cell membranes. It can be described by an internal current source J_{i}, ie. h. $J = J_{\mathrm{E}} + J_{\mathrm{i}} = \kappa E + J_{\mathrm{i}}$.

A further simplification results if the Maxwell equations are subjected to a Fourier transformation and used in complex form in the image domain. With

$$E(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \underline{E}(\omega)\, e^{j\omega t}\, \mathrm{d}\omega \,,$$

$$H(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \underline{H}(\omega)\, e^{j\omega t} \mathrm{d}\,\omega$$

results because of

$$\frac{\partial E(t)}{\partial t} \circ\!\!-\!\!\bullet\; j\omega\, \underline{E}(\omega) \quad \text{but.} \quad \frac{\partial H(t)}{\partial t} \circ\!\!-\!\!\bullet\; j\omega\, \underline{H}(\omega) \tag{3.8}$$

after insertion into Equation 3.3, Equation 3.4 and Equation 3.7 with the following simplifications:

$$\nabla \times \underline{H} = (\kappa + j\omega\epsilon_0)\,\underline{E} + \underline{J}_\mathrm{i}\,, \tag{3.9}$$

$$\nabla \times \underline{E} = -j\omega\mu_0\,\underline{H}\,, \tag{3.10}$$

$$\nabla\cdot \underline{H} = 0\,. \tag{3.11}$$

Since according to Equation 3.11 the divergence of $\underline{H}$ vanishes, $\underline{H}$ can be expressed by the rotation of any scalar vector field $\underline{A}$ for easier determination of the solution of these Maxwell equations, after the divergence of a rotation of any vector field always vanishes. Here one chooses, for example.

$$\mu_0\,\underline{H} := \nabla \times \underline{A}\,. \tag{3.12}$$

After inserting in Equation 3.10 it follows:

$$\nabla \times (\underline{E} + j\omega\underline{A}) = 0\,. \tag{3.13}$$

Since the rotation of $\underline{E} + j\omega\underline{A}$ also vanishes, it can now be expressed by any scalar function ϕ as follows:

$$\underline{E} + j\omega\underline{A} = -\nabla\phi\,. \tag{3.14}$$

According to the Helmholtz theorem, a vector field is uniquely described by specifying its rotation *and* divergence [60]. Since for the vector field $\underline{A}$ according to Equation 3.12 only the rotation has been defined so far, the divergence would have to be specified additionally. For this purpose

$$\nabla \cdot \underline{A} := -\kappa\mu_0\Phi \tag{3.15}$$

can be defined. With the help of this definition, the Maxwell equations can now be reduced to the solution of an equation for the vector potential $\underline{A}$. If Equation 3.12, Equation 3.14 and Equation 3.15 are substituted in Equation 3.10 and if one additionally considers the Graßman development theorem

$$\nabla \times \nabla \times \underline{A} = \nabla(\nabla \cdot \underline{A}) - \nabla^2\underline{A}\,, \tag{3.16}$$

then we obtain the vectorial Helmholtz-equation

$$\nabla^2\underline{A} - j\omega\mu_0\kappa\underline{A} = -\mu_0\underline{J}_\mathrm{i}\,, \tag{3.17}$$

whose solution is well known in classical electromagnetic theory and is given by

$$\underline{A} = \frac{\mu_0}{4\pi}\int \frac{\underline{J}_\mathrm{i}e^{-kr}}{r}\,\mathrm{d}v \tag{3.18}$$

$k^2 = j\omega\kappa(1 + j\omega\epsilon_0/\kappa)$ (k: wave vector)

$r^2 = (x - x')^2 + (y - y')^2 + (z - z')^2$ (r: distance current source to measuring point)

can be specified. The location of the current source is described by the coordinates x, y, z, the location of the measurement outside the volume in which the current sources are located by the coordinates x', y', z'.

According to Equation 3.18, the vector potential $\underline{A}$ propagates in a body due to the e-function e^{-kr} like a harmonic wave whose propagation constant k describes the magnitude of the wave vector or the local wavelength. To estimate the maximum value of the product $kr_{\max}$ in the exponent of the e-function, for a maximum distance between the location of the current source and the measurement location of $r_{\max} = 100\,\text{cm}$, a frequency of 1 kHz and an average conductivity κ of 4 mS/cm one obtains the result $kr_{\max} = 0.04$ resp. $e^{-kr_{\max}} = e^{-0.04} = 0.96$. Since this value is very close to 1, this e-function in Equation 3.18 can be approximately neglected for the measurement in a body and we obtain:

$$\underline{A} \approx \frac{\mu_0}{4\pi} \int \frac{\underline{J}_\text{i}}{r}\,\mathrm{d}v\,. \tag{3.19}$$

If this result is substituted into Equation 3.15, it follows for the potential Φ:

$$\Phi = \frac{-1}{4\pi\kappa} \int \underline{J}_\text{i} \cdot \nabla' \left(\frac{1}{r}\right) \mathrm{d}v\,. \tag{3.20}$$

The Nabla operator with the apostrophe ∇' is meant to indicate that the local derivatives are to be performed according to the coordinates x', y', z' of the measurement location, to which Equation 3.15 refers, while in Equation 3.19 it is integrated over the coordinates of the source locations x, y, z. However, because of $\nabla'(\frac{1}{r}) = -\nabla(\frac{1}{r})$, the potential equation can also be described with a nabla operator related to the coordinates of the measurement location:

$$\Phi = \frac{1}{4\pi\kappa} \int \underline{J}_\text{i} \cdot \nabla \left(\frac{1}{r}\right) \mathrm{d}v\,. \tag{3.21}$$

The result is also obtained by assuming that a flow of charge carriers in the body does not cause self-induction and that it is an almost static problem which can be solved by means of Poisson's equation [18]. It also allows the interpretation that the current density multiplied by a volume element $\underline{J}_\text{i} \cdot \mathrm{d}v$ is a *current dipole* and summed over all inner current dipoles with the weight $\nabla(\frac{1}{r})$. A current dipole is a current source with an associated current sink multiplied by the distance between the source and the sink.

Probably the strongest current sources in the body are generated by the action potentials of the pacemaker cells in the heart and can be measured particularly well as a potential on the thorax (see Figure 3.14).

To simplify the measurement of the electrical activity of the heart, all the current dipoles of the heart in Equation 3.21 can be combined into a single current dipole by vectorial summation and its effects on the body studied. If, for example, measurements are made on the left and right arms and on the foot, the measured potential differences can be interpreted as projectionsindexprojection of the heart vector onto the respective individual sections between the arms and a foot (see Figure 3.15).

Using the potentials at the left arm (Φ_L), right arm (Φ_R) and foot (Φ_F), the following potential differences important for an ECG can be determined:

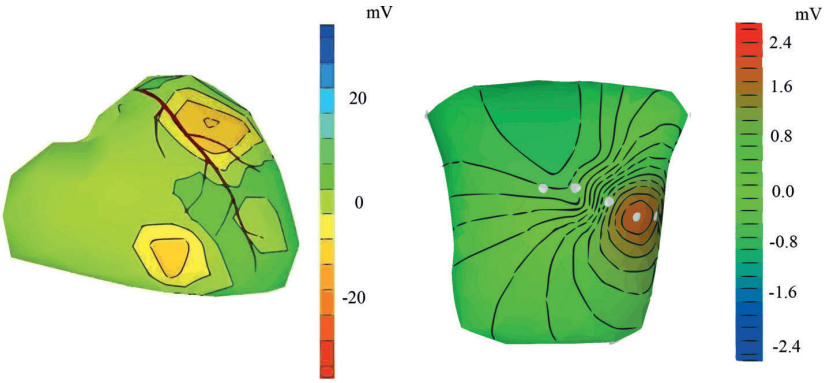


Fig. 3.14: Example potentials at heart (left) and thorax (right) as generated by the freeware programme ECGSIM [55] of the University of Nijmegen; in the thorax representation the locations of the recording electrodes according to Wilson are marked by grey circles

Einthoven Potentials

For these potentials, one takes a foot as the reference potential and obtains the voltages U_{I}, U_{II}, U_{III} from the potential differences:

$$U_{\mathrm{I}} = \Phi_{\mathrm{L}} - \Phi_{\mathrm{R}}$$
$$U_{\mathrm{II}} = \Phi_{\mathrm{F}} - \Phi_{\mathrm{R}}$$
$$U_{\mathrm{III}} = \Phi_{\mathrm{F}} - \Phi_{\mathrm{L}}$$

Goldberger Potentials

Here the voltages aV_{L}, aV_{R}, aV_{F} are not related to the potential at the foot, but to a virtual ground point, which is the average of the voltages derived according to Einthoven as follows:

$$aV_{\mathrm{L}} = U_{\mathrm{II}}/2 - U_{\mathrm{III}}$$
$$aV_{\mathrm{R}} = U_{\mathrm{III}}/2 - U_{\mathrm{II}}$$
$$aV_{\mathrm{F}} = U_{\mathrm{I}}/2 - U_{\mathrm{II}}$$

Wilson Potentials

Further electrodes are applied by adding six (to nine) further measurement points on the thorax near the heart in a ring-shaped arrangement along the chest wall and relating them to the indifferent collecting electrode, which is obtained by merging the extremity leads according to Goldberger (see Figure 3.16).

In total, twelve ECG potential differences (see Figure 3.16) can thus be obtained from all three measurement arrangements (according to Einthoven, Goldberger and Wilson), which can be used for the standard assessment of electrical cardiac activity.

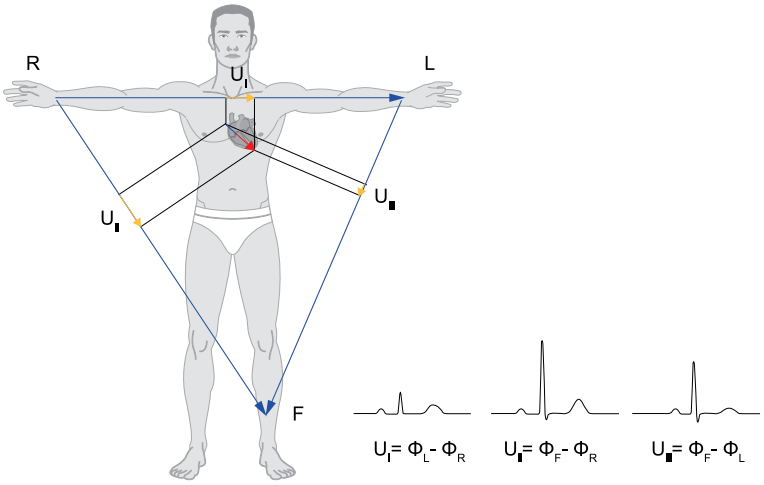


Fig. 3.15: Potential differences between the left (point L) and right (point R) arms and a foot (point F) as projections of the heart vector to form the ECG electrodes according to Einthoven; where the signal waveforms U_I, U_{II} and U_{III} are obtained by forming the difference between the potentials Φ_L, Φ_R and Φ_F.

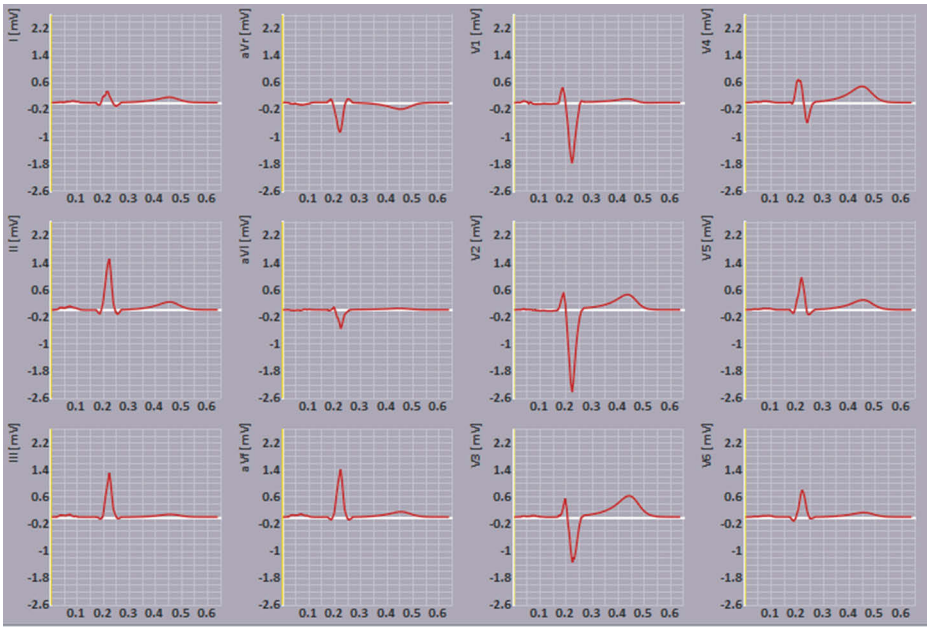


Fig. 3.16: Standard ECG with 12 leads for a young man generated with the freeware ECGSIM [55] from Nijmegen University (V_I, V_{II}, V_{III} after Einthoven, aV_L, aV_R, aV_F after Goldberger and V_1 to V_6 after Wilson).

In particular, the ECG can also provide information about whether, for example, a heart attack or other heart diseases is present.

As an alternative to the determination of a single heart vector, whose projection onto the body surface yields the respective potential differences, the effects of the electrical potential distributed over the closed surface of the heart on the body surface can also be investigated. The *equivalent double layer model* (EDL, see [8, 57]) used for this purpose can be formulated as a matrix equation:

$$\mathbf{\Phi} = \mathbf{A} \cdot \mathbf{S} \,. \tag{3.22}$$

S here is a $\mathbf{N} \times \mathbf{T}$ potential matrix of the heart surface and **Φ** is the corresponding $\mathbf{L} \times \mathbf{T}$-potentiometer matrix$\mathbf{L} \times \mathbf{T}$-potential matrix of the body surface for **N** or **L** measurement locations and **T** time points. **A** is the so-called $\mathbf{L} \times \mathbf{N}$ transfer matrix. Thereby, according to Equation 3.22, given the potential on the heart surface, the potential on the body surface can be determined by matrix multiplication. This is called the *forward problem*. Often, however, because of the non-invasive measurement on the body surface, only the potential at this location is known. If one wants to determine from this the potential at the heart surface, the so-called *inverse problem* has to be solved, which presupposes an invertibility of the transfer matrix **A**:

$$\mathbf{S} = \mathbf{A}^{-1} \cdot \mathbf{\Phi} \,. \tag{3.23}$$

3.2.3 Process of Excitation Propagation during a Heart Beat

As already described, the excitation of the heart muscle cells takes place through a conduction system, whereby the pacemaker cells in the sinus node excite further pacemaker cells in the AV node and these in turn excite pacemaker cells in the His bundle and the Purkinje-fibres. The latter finally cause the heart muscle cells to periodically contract and relax again. The sequential course of a heartbeat is shown by Figure 3.17 in eight phases A to H. The course of this excitation can be measured with the help of electrodes on the surface of the body and is known as an electrocardiogram. For the sections shown in Figure 3.17, the individual phases (A to H) can be assigned to a wave or stretch of the ECG (P, Q, R, S or T).

Section A: The cardiac excitation cycle is initiated by an action potential from the pacemaker cells in the sinus node.

Section B: The excitation spreads across both atria, and the cell membranes of the atrial cells become depolarised, which is evident on the ECG as a P wave. In this section, the atria contract and fill the ventricles with blood, causing the leaflet valves (mitral and tricuspid valves) to open between the atria and the main chamber and the semilunar valve indexes (aortic and pulmonary valves) to open between the atria and the main chamber; a (aortic and pulmonary) valves are closed.

Section C: The excitation through the sinus node spreads to the AV node. It can only spread further into the wall between the left and right main chamber ("tawara" =

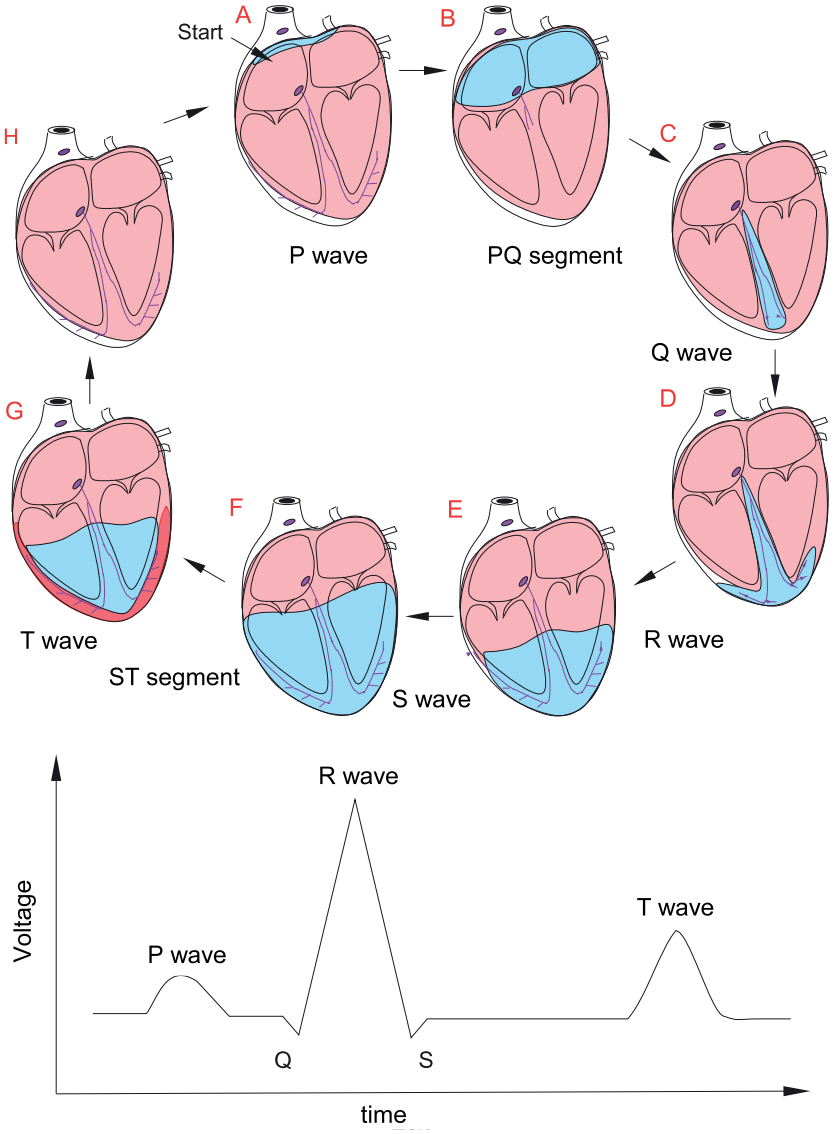


Fig. 3.17: Excitation of the heart muscles over one heartbeat in eight steps (A to H) with labelling of the individual sections (top); associated ECG with assigned waves (bottom)

thigh) via this node and the adjoining His bundle. This section can be seen as a Q wave in the ECG.

Section D: At the end of the wall between the left and right main chambers, the excitation can travel further into the outer wall of the main chambers via the Purkinje-fibres, which generates the central R-wave in the ECG.

Section E: The cardiac muscle cells of the main chambers are now excited via the Purkinje-fibres, which can be seen as the S-wave in the ECG.

Section F: Once all the cardiac muscle cells of the main chambers have been excited, the main chambers contract, which increases the pressure there, the valves open and the blood is forced out of the chambers into the aorta or the pulmonary artery. At the same time, the atria are also filled with blood during this expulsion phase. As an ST segment it is found in the ECG between the S wave and the T wave.

Section G: Since in the meantime all heart cells have been excited, the conduction of impulses between the atria and the main chambers is normally only possible via the His bundle and the AV node, and the cells also need a certain time (refractory period) until they can be excited again by an action potential; the excitation from the main chambers can no longer spread back into the atria. The relaxation phase or diastole now begins, recognisable as a T-wave in the ECG.

Section H: Finally, all cells of the main chamber return to their resting state by repolarising. Now a new cardiac cycle can begin again by stimulating an action potential from the sinus node.

3.2.4 Modelling the Excitation System

As already explained in subsection 3.2.1, a simple system of three coupled oscillators can be used to model cardiac excitation, which generates the rhythm of the sinus node, the AV node and the His bundle with the Purkinje-fibres. In this system, the sinus node oscillator has the highest natural frequency and controls the other oscillators (AV node and HP complex) from the outside (see Figure 3.13). The coupling is particularly strong from the sinus node to the AV node and from the AV node to the His bundle with the Purkinje fibres. Other couplings are also present, but much weaker, so they do not need to be considered in the modelling.

Oscillating Nerve Cells Potentials in the Heart

In modelling the nerve cells with three oscillators for SA node, AV node and His bundle with the Purkinje-fibres, one can in principle start from the investigation carried out by Hodgkin and Huxley on the giant axon of the squid [28]. Although the equations they set up correctly describe the potentials produced, they are very complicated. Fitzhugh [15] succeeded in simplifying them considerably without substantially distorting the representation of the correct potential. Independently, Nagumo [53] also succeeded in doing so, which eventually led to the common FitzHugh-Nagumo-model:

$$\begin{aligned} \dot{v} &= v - \tfrac{1}{3}v^3 - w + I_{\text{ext}} \\ \tau \dot{w} &= v - a - bw \,. \end{aligned} \tag{3.24}$$

Here v is the membrane potential, w and τ are auxiliary variables, and I_{ext} is an external current. Fitzhugh also calls this model the Bonhoeffer-Van-der-Pol oscillator,

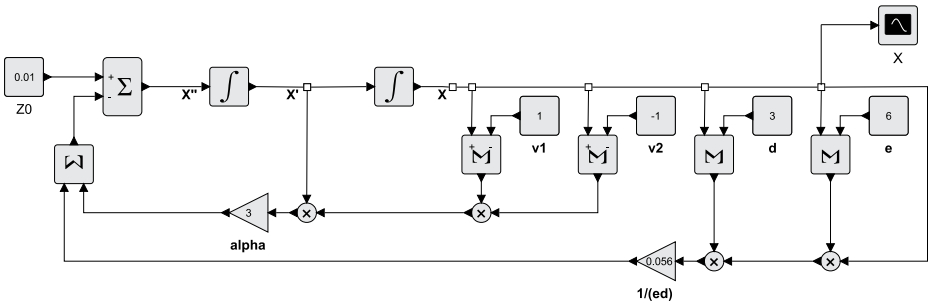


Fig. 3.18: Modified Van der Pol oscillator according to Grudzinski and Zebrowski [23] according to Equation 3.25 to describe the action potential of a vibration-generating nerve cell or a whole bundle of similar nerve cells such as the SA or AV node.

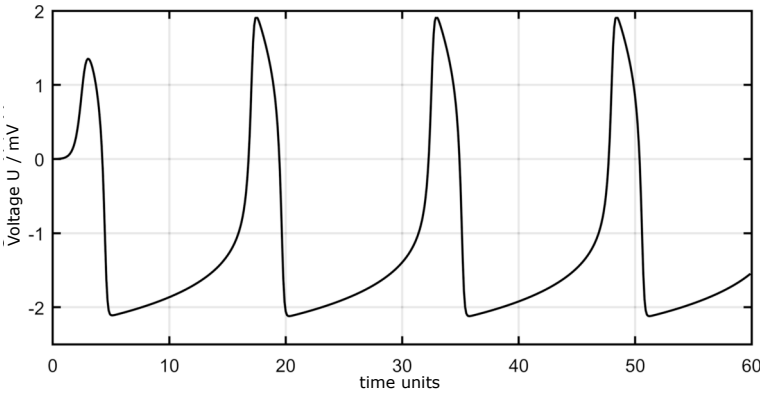


Fig. 3.19: Action potential $x(t)$ to the modified Van der Pol oscillator after [23] in Figure 3.18 with parameters $v_1 = 1,\ v_2 = -1,\ d = 3,\ e = 6$ and $\alpha = 3$.

since these equations for $a = b = 0$ describe the Van-der-Pol oscillator as a special case. Such an oscillator can generate an oscillation without external excitation, which simulates the potential of a nerve cell in generating oscillations. In the general case, however, instead of the parameters a and b the original Van der Pol oscillator can be extended in other ways to describe important properties of the action potential and to influence the frequency and oscillation stability in a simpler way without changing the signal form significantly [23]. This led to the modified equation:

$$\ddot{x} + \alpha(x - v_1)(x - v_2)\dot{x} + x(x + d)(x + e)/ed = 0\,, \quad d, e, \alpha > 0\,. \tag{3.25}$$

The corresponding model is shown in Figure 3.18. This can be used to describe de facto not only the action potential of a single nerve cell, but also that of a whole cluster of similar nerve cells, such as in the sinus node. The corresponding action potential shows Figure 3.19.

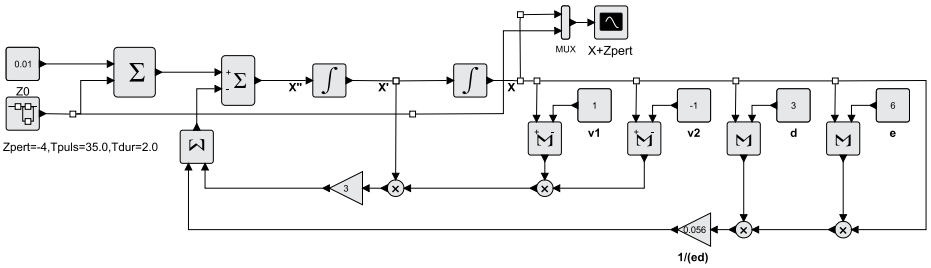


Fig. 3.20: Modified Van der Pol oscillator according to Figure 3.18 perturbed by an externally applied square pulse at time unit 35 with an amplitude of 4 and a width of 2 time units.

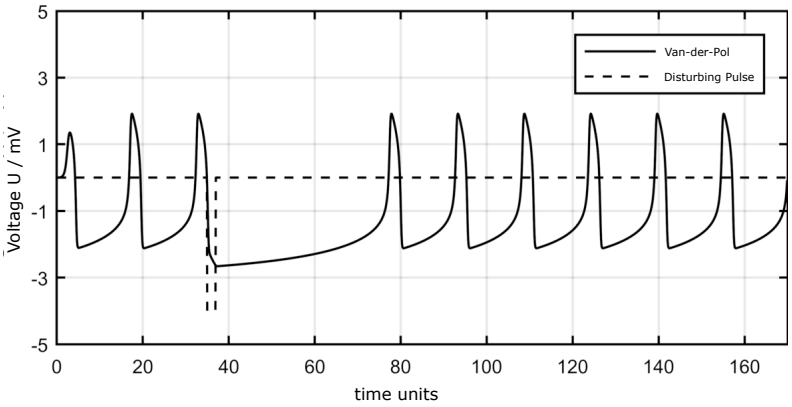


Fig. 3.21: Action potential $x(t)$ belonging to the perturbed modified Van der Pol oscillator Figure 3.20 with suspension of two normal oscillations, where the perturbation is modelled as an externally applied square pulse at time unit 35 with an amplitude of 4 and a width of 2 time units.

However, the modified van der Pol oscillator is not stable against external influences. If a potential is additionally applied to the input, this can cause the oscillations to cease or even stop altogether, depending on the time and amplitude [14]. The associated model is shown in Figure 3.20. A square wave pulse to the input immediately causes a suspension of two oscillations. The associated action potential is shown in Figure 3.21. If the pulse height is doubled, the oscillation subsequently stops completely (see action potential $x(t)$ in Figure 3.22).

Coupling of the SA node with the AV node

The sinus node of the heart consists of several similar heart cells that can independently generate an oscillation that corresponds to the oscillation of a single heart cell. Thus the model of Grudzinski and Zebrowski [89] according to Figure 3.18 is also suitable as a model for the whole sinus node. The sinus node influences the AV node, which also consists of similar oscillation-generating heart cells, but which have a

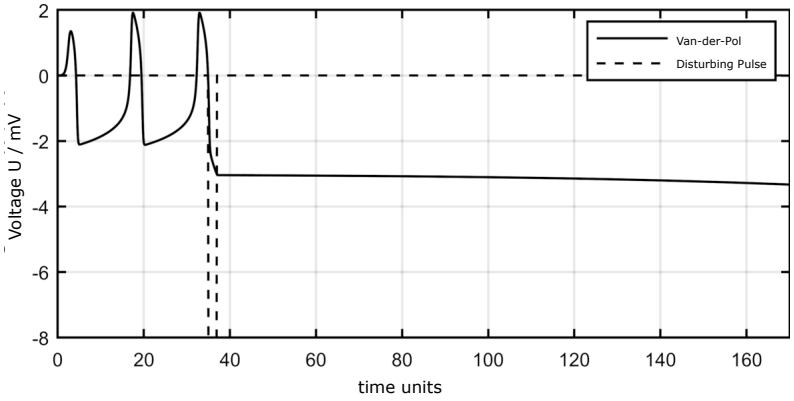


Fig. 3.22: Action potential $x(t)$ belonging to the perturbed modified Van der Pol oscillator in Figure 3.20 leading to the complete cessation of oscillation; an externally applied square pulse at time unit 35 with an amplitude of 8 and a width of 2 time units acts as the perturbation.

lower natural frequency. Due to the coupling, the sinus node works as a pacemaker node of the AV node and causes the heart cells of the AV node to also generate action potentials with the frequency of the sinus node. Both oscillations –that of the sinus node and the delayed one of the AV node –generate an altered signal by weighted superimposition, and by the shape of which pathological situations can also be recognised in a real ECG signal.

For the mathematical description, the equations (analogous to equation 3.25) for *two* mutually coupled modified van der Pol oscillators are used. Initially, only one coupling is considered, namely that from the sinus node to the AV node [89]:

$$\begin{aligned} x_{\mathrm{SA}} &= -\alpha_{\mathrm{SA}}(x_{\mathrm{SA}} - v_{\mathrm{SA}_1})(x_{\mathrm{SA}} - v_{\mathrm{SA}_2})x_{\mathrm{SA}} \\ &\quad - x_{\mathrm{SA}} f_{\mathrm{SA}}(x_{\mathrm{SA}} + d_{\mathrm{SA}})(x_{\mathrm{SA}} - e_{\mathrm{SA}}) \\ x_{\mathrm{AV}} &= -\alpha_{\mathrm{AV}}(x_{\mathrm{AV}} - v_{\mathrm{AV}_1})(x_{\mathrm{AV}} - v_{\mathrm{AV}_2})x_{\mathrm{AV}} \\ &\quad - x_{\mathrm{AV}} f_{\mathrm{AV}}(x_{\mathrm{AV}} + d_{\mathrm{AV}})(x_{\mathrm{AV}} - e_{\mathrm{SA}}) \\ &\quad + k_{\mathrm{SA-AV}} x_{\mathrm{SA}}^{\tau_{\mathrm{SA-AV}}} - k_{\mathrm{AV-AV}} x_{\mathrm{AV}} \,. \end{aligned} \tag{3.26}$$

Here, the lower index *SA* denotes the signals and parameters of the sinus node and the lower index *AV* denotes the signals and parameters of the atrioventricular node. As for the coupling, $k_{\mathrm{SA-AV}}$ is the attenuation factor from the sinus node to the AV node and $k_{\mathrm{AV-AV}}$ is the feedback factor from the output to the input of the AV node. Furthermore, $x_{\mathrm{SA}}^{\tau_{\mathrm{SA-AV}}}$ is the output signal of the sine node at the input of the AV node delayed by the transit time $\tau_{\mathrm{SA-AV}}$, since the output signal of the sine node is not directly applied to the AV node, but must reach it via other nerve cells. In general, let $x^{\tau}(t) = x(t - \tau)$ apply. The corresponding model is shown in Figure 3.23 and Figure 3.24.

If the coupling factor $k_{\mathrm{SA-AV}}$ from the sine node to the AV node is not too large compared to the feedback factor $k_{\mathrm{AV-AV}}$ of the AV oscillator, the two oscillators can

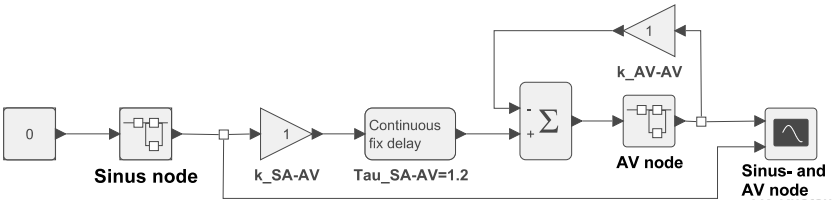


Fig. 3.23: Oscillator model with SA and AV nodes of the heart according to Equation 3.26 analog [89]: The model includes a delay $\tau_{SA-AV} = 1.2$ from the SA to the AV node, which satisfies conduction. The factors k_{SA-AV} and k_{AV-AV} describe the attenuation from the sinusoid to the AV node and the feedback from the output to the input of the AV node. Both have the value 1 here. Both the sinus node and the AV node are represented in the model as blocks symbolising separate networks. The network for the AV node is shown in Figure 3.24. The sinus node has a network with an identical structure but modified parameters.

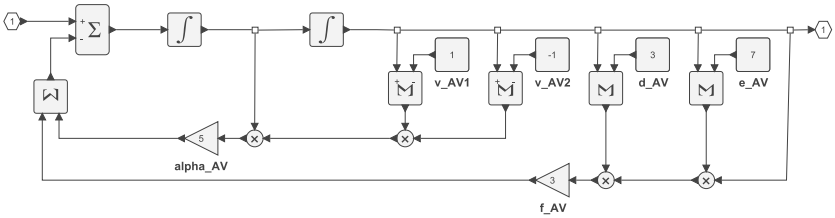


Fig. 3.24: Network for the AV node in Figure 3.29: It corresponds in structure to that of the sine node, but has a different value for the parameter e. According to the information in [89], the parameters $v_{AV_1} = 1$, $v_{AV_2} = -1$, $d_{AV} = 3$, $e_{AV} = 7$, $f_{AV} = 3$ were used for the AV node, $\alpha_{AV} = 5$ and for the sinus node the parameters $v_{SA_1} = 1$, $v_{SA_2} = -1$, $d_{SA} = 3$, $e_{SA} = 12$, $f_{SA} = 3$, $\alpha_{SA} = 5$ were chosen.

oscillate with their natural frequencies (see Figure 3.25). In this case, each of these two oscillators can react to external influences like a single oscillator, i.e. also with partial or complete suspension of the oscillation (cf. Figure 3.21 and Figure 3.22). However, if the coupling factor k_{SA-AV} from the sine node to the AV node is stronger, the signal from the sine node will have a large influence on the AV node:

- If the natural frequency of the AV node is not much lower than that of the sine node, the signal of the sine node can "trigger" the AV node so that both nodes oscillate at the same frequency (see Figure 3.26).
- If the natural frequency of the AV node is too low to follow the signal of the sine node, dropouts will occur (see Figure 3.26).
- In the worst case, if the natural frequencies of the SA and AV nodes are too different, it is particularly dramatic. Then, after a few arrhythmic oscillations of the AV node, excitation soon stops completely; the *heart thus stops beating* (see Figure 3.28).

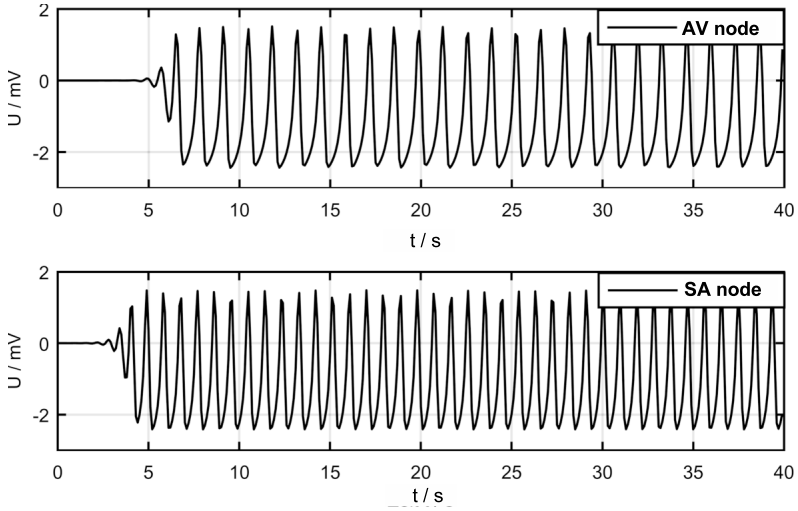


Fig. 3.25: Decoupled oscillations of the sine and AV node according to Figure 3.23 and Figure 3.24 mit $k_{SA-AV} = k_{AV-AV} = 1$, $e_{SA} = 12$ und $e_{AV} = 7$.

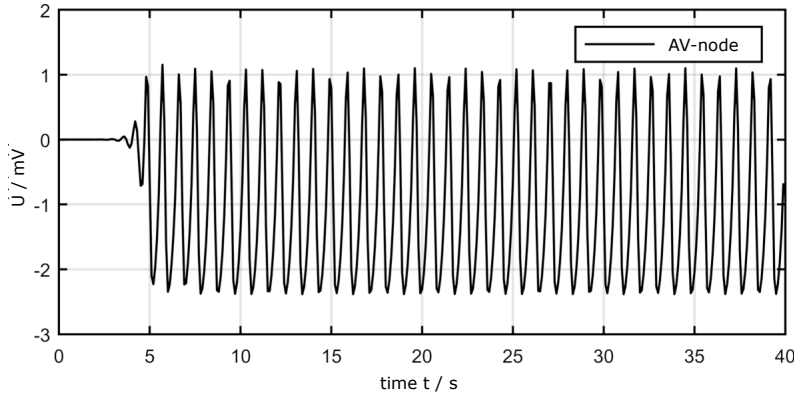


Fig. 3.26: Synchronised oscillations of the sine node according to Figure 3.23 and Figure 3.24 with stronger coupling of the sinus node to the AV node, where the natural frequencies do not differ that much ($k_{SA-AV} = 12$, $k_{AV-AV} = 1$, $e_{SA} = 12$ and $e_{AV} = 9$).

Couplings of SA-node, AV-node and HP-complex

With the two oscillators mentioned in the previous section, which describe the coupling of the sinus node with the AV node, many properties of the generation of oscillations in the heart can already be explained. However, the signals obtained in this way do not resemble those measured at the body surface in the form of an ECG signal $x_{EKG}(t)$. It is important to note that the His bundle and the Purkinje fibres also represent their own oscillator, which affects the ECG signal (see Figure 3.13). The output signals of the sinus node (SA) $x_{SA}(t)$, the atrioventricular node (AV) $x_{AV}(t)$ and the His-

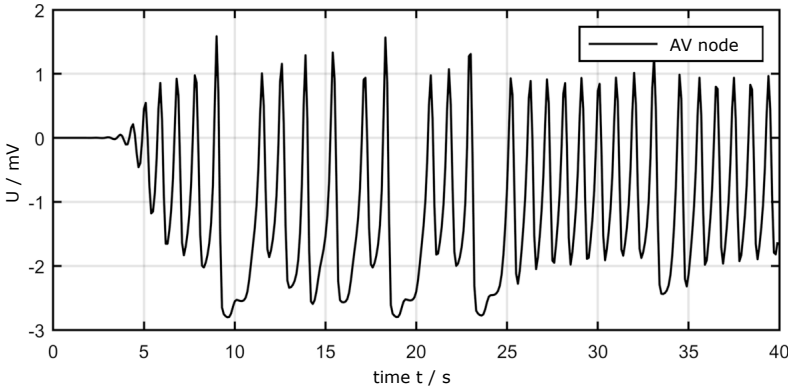


Fig. 3.27: Partially synchronised oscillations of the AV node according to Figure 3.23 and Figure 3.24 with stronger coupling of the sinus node to the AV node, where the AV node can no longer follow the signal of the sinus node due to a too low natural frequency ($k_{SA-AV} = 12$, $k_{AV-AV} = 1$, $e_{SA} = 12$ and $e_{AV} = 9$).

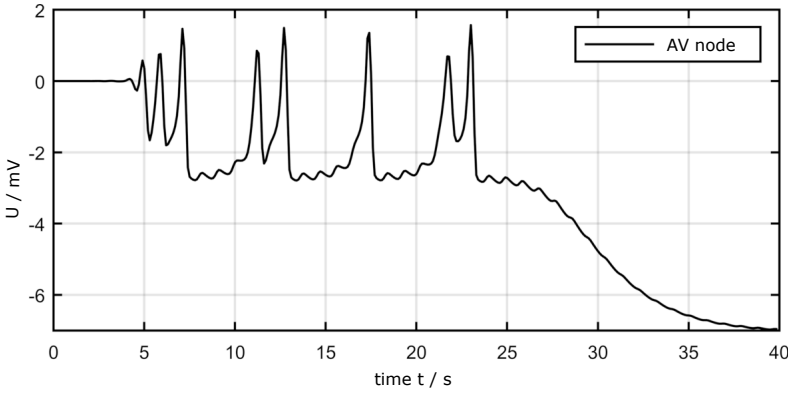


Fig. 3.28: Non-synchronisable oscillations of the AV node according to Figure 3.23 and Figure 3.24 when the sinus node is too strongly coupled to the AV node, whereby the oscillation of the AV node stops completely after irregular dropouts, i.e. the point of singularity has been reached ($k_{SA-AV} = 12$, $k_{AV-AV} = 1$, $e_{SA} = 12$ '$e_{AV} = 7$).

Purkinje-complex (HP) $x_{HP}(t)$ overlap and have different effects on the body surface. The ECG signal can therefore be described approximately linearly by an addition of these signals weighted with the attenuation factors a_{SA}, a_{AV}, and a_{HP} [20]:

$$x_{ECG}(t) = a_{SA}x_{SA}(t) + a_{AV}x_{AV}(t) + a_{HP}x_{HP}(t)\ . \tag{3.27}$$

Since the cardiac excitation travels from the sinus node via the atrial node and then via the AV node to the main chamber, the coupling of the three oscillators should be

approximately in one direction, i.e. from the SA to the AV and then to the HP oscillator, as shown in the lower part of Figure 3.13.

A mathematical description of the three oscillators is obtained by extending the two-oscillator system according to Equation 3.26 with an equation for the third oscillator and adding couplings:

$$\begin{aligned}
\dot{x}_{SA} &= -\alpha_{SA}(x_{SA} - v_{SA_1})(x_{SA} - v_{SA_2})x_{SA} \\
&\quad - x_{SA} f_{SA}(x_{SA} + d_{SA})(x_{SA} - e_{SA}) \\
\dot{x}_{AV} &= -\alpha_{AV}(x_{AV} - v_{AV_1})(x_{AV} - v_{AV_2})x_{AV} \\
&\quad - x_{AV} f_{AV}(x_{AV} + d_{AV})(x_{AV} - e_{SA}) \\
&\quad + k_{SA-AV} x_{SA}^{\tau_{SA-AV}} - k_{AV-AV} x_{AV} \\
\dot{x}_{HP} &= -\alpha_{HP}(x_{HP} - v_{HP_1})(x_{HP} - v_{HP_2})x_{HP} \\
&\quad - x_{HP} f_{HP}(x_{HP} + d_{HP})(x_{HP} - e_{HP}) \\
&\quad + k_{AV-HP} x_{SA}^{\tau_{AV-HP}} - k_{HP-HP} x_{HP}.
\end{aligned} \tag{3.28}$$

The corresponding model shows Figure 3.29. Disturbances caused by fluctuations in the ECG baseline, e.g. due to changes in skin contact, and interspersed noise, e.g. due to 50 Hz mains hum, neon tube influences or radio waves, were taken into account. An ECG signal disturbed by a noise signal and a fluctuating baseline is shown in Figure 3.31.

Figure 3.30 shows one below the other the signals generated by this system for the SA node, the AV node and the HP complex, and the resulting ECG signal (without interference). It is easy to see where the individual components in the ECG signal come from. The P wave comes from the signal of the sinus node, and the QRS complex from the interaction of the AV node and the HP complex. The T wave, on the other hand, is predominantly generated by the HP complex.

3.3 Taxonomy of Biosignals

In the previous section on cardiac excitation, the origin and modelling of the electrical biosignals was described in detail. In conclusion to this chapter, this view will now be extended again to arbitrary biosignals and the different types will be classified according to their properties. Life forms – from the single cell to complex organisms such as mammals – generate biosignals with different properties, function and meaning. Basically, these signals can be divided into:

1. *autonomous* signals, i.e. signals of the body produced by the life form without external influences, such as the electrical activity of the heart and

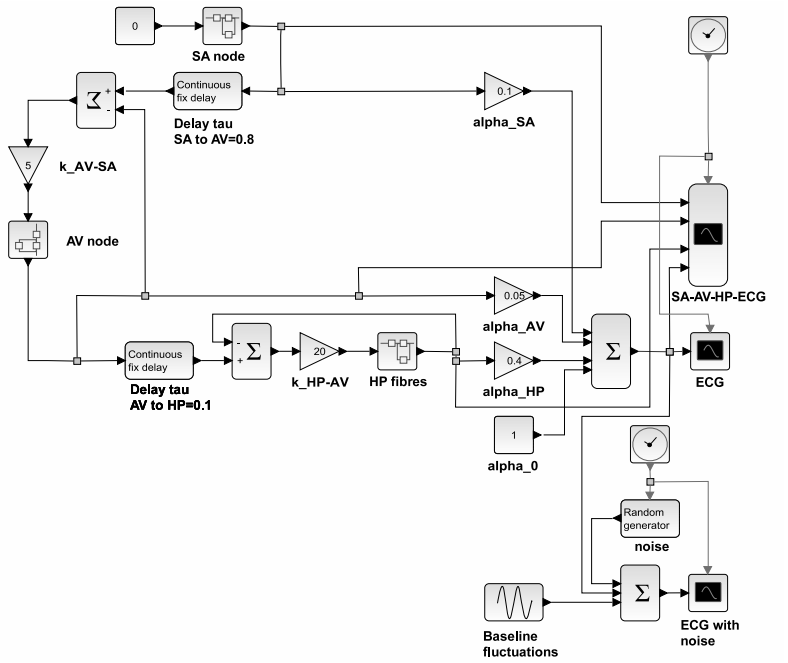


Fig. 3.29: Three-Oscillator-Model of the heart with SA, AV and HP nodes according to Figure 3.13 and Equation 3.28 after [20]: The model includes a delay from the SA node to the AV node by $\tau_{SA-AV} = 0.8$ and from the AV node to the HP complex by $\tau_{AV-HP} = 0.1$. Before the respective signals arrive at the control input of the AV node or the HP nodes, they are influenced by the coupling factors $k_{SA-AV} = k_{AV-AV}$: = $k_{SAAV-AV} = 5$ and $k_{AV-HP} = k_{HP-HP}$:= $k_{AVHP-HP} = 20$ weighted. The oscillator nodes have the same internal structure as in the two-oscillator-model comprising only the SA and AV nodes (see Figure 3.24). The parameters were chosen as in [20], for the SA node: $v_{SA_1} = 0.2$, $v_{SA_2} = -1.9$, $d_{SA} = 3$, $e_{SA} = 4.9$, $f_{SA} = 1$, $\alpha_{SA} = 3$, for the AV node: $v_{AV_1} = 0.1$, $v_{AV_2} = -0.1$, $d_{AV} = 3$, $e_{AV} = 3$, $f_{AV} = 1$, $\alpha_{AV} = 3$ and for the HP-complex: $v_{HP_1} = 1$, $v_{HP_2} = -1$, $d_{HP} = 3$, $e_{HP} = 7$, $f_{SA} = 1$, $\alpha_{HP} = 5$.

2. *evoked*[17] signals, i.e. signals excited by external stimuli, such as the Berger-effect in visual stimulation of the brain.

Thus, in the Berger-effect, the mental and physical state of a test person influences whether so-called α-wave can arise in the brain or whether these frequency components are suppressed by the waking state. If the subject is in physical and mental rest and has closed his eyes, α-waves in a frequency range of 7 to 10 Hz can be detected in the brain waves. As soon as the subject opens his eyes, these disappear as an effect of the *visual provocation*. The disappearance of the frequency components can therefore

17 from lat. *evocare*: to call forth, to cause.

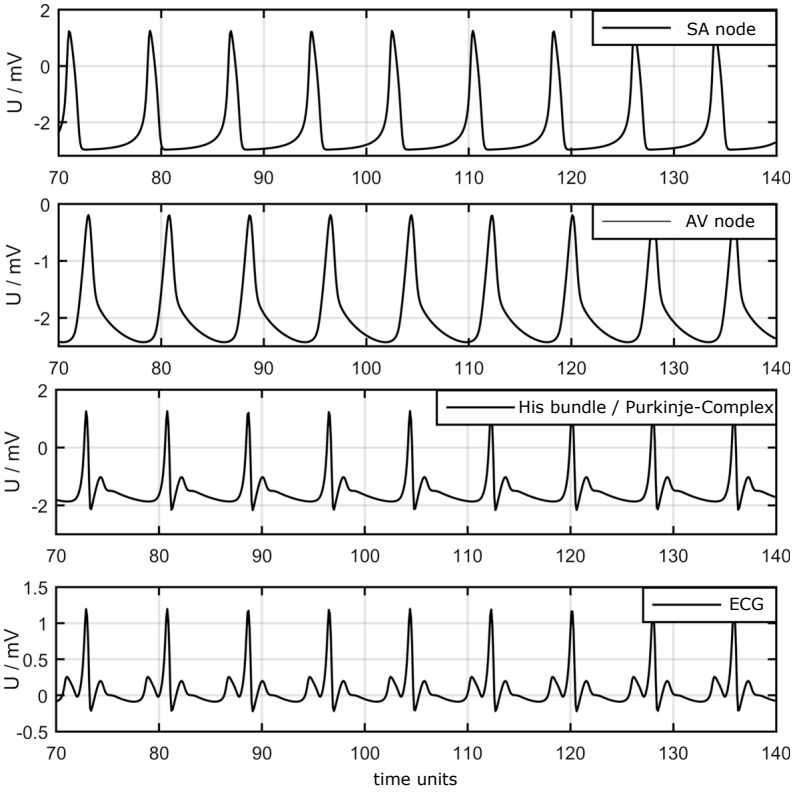


Fig. 3.30: Signals during ECG generation of the three-oscillator-model in Figure 3.29.

be causally linked to the external event of "opening the eyes". The resulting signals are called *visually evoked* signals.

In most cases, however, such a connection is not so easy to detect and requires the use of special biosignal processing methods for quantification. One such method is brainstem audiometry to perform hearing tests on newborns. Since the newborn itself is not capable of providing information about the perception of acoustic stimuli, brainstem audiometry evaluates the electrical signals of the brain immediately after an acoustic provocation in order to gain information about the infant's hearing ability. Such signals are called *acoustically evoked* signals.

Evoked signals occur mainly in connection with sensory stimuli. These stimuli trigger electrical potential changes in the sensory areas of the cerebral cortex. Depending on the stimulation, a distinction is made:

1. Acoustically evoked potentials (AEP), which allow assessment of the auditory pathway: from the cochlea, via the auditory nerve to the brainstem (early AEP,

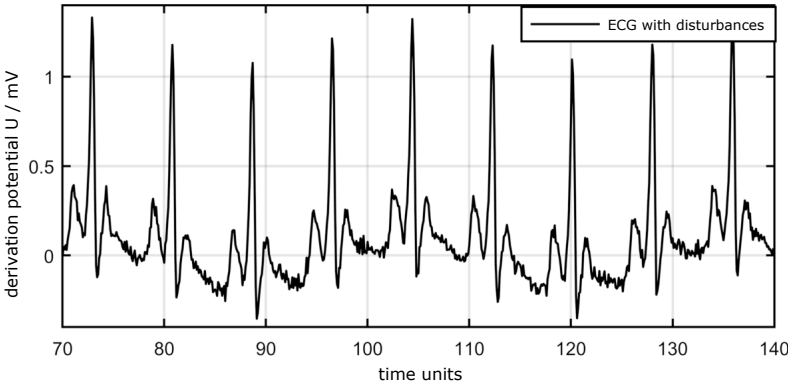


Fig. 3.31: ECG generated with the three-oscillator-model from Figure 3.29 with additional disturbances due to a sinusoidal fluctuation of the baseline with a period length of 32 time units and an amplitude of 0.12 as well as a normally distributed noise signal (mean of 0, dispersion of 0.024).

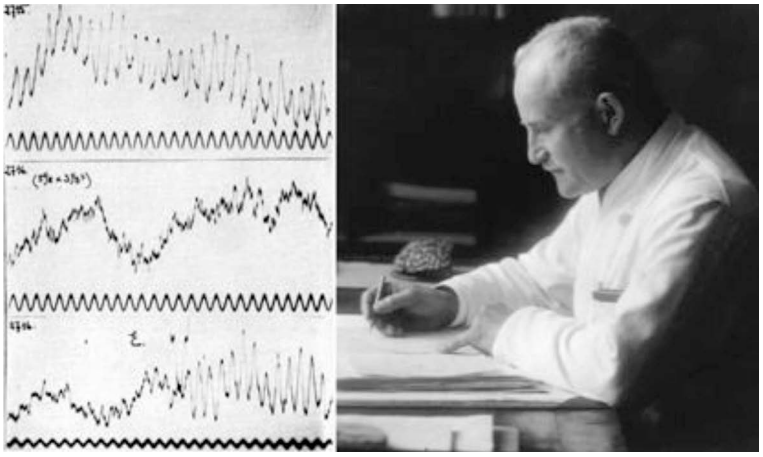


Fig. 3.32: The illustration shows Hans Berger (1873–1941) in 1927 (right) and EEG signals from his daughter Ilse (left). From top to bottom: Ilse at rest (alpha waves), Ilse calculating a sum (beta waves) and Ilse announcing the result of the sum (mixed-waves) [88].

FAEP), further via the midbrain (middle AEP, MAEP), to the auditory cortex (late AEP, SAEP).

2. Visual evoked potentials (VEP), which allow assessment of the visual nerve and visual pathways based on the electrical potentials triggered in the visual cortex (occipital cortex).
3. Somatic evoked potentials (SEP), which can be used to assess the central and peripheral somatosensitive pathways and nerves. This is done by inducing electrical stimuli near a sensitive nerve using a stimulating electrode.

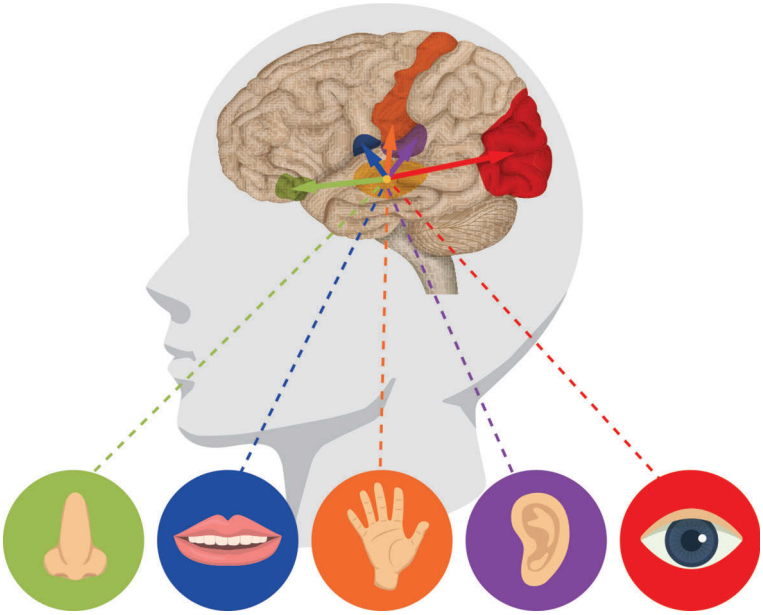


Fig. 3.33: Different forms of sensory stimulation and their relationship to specific brain regions, as well as the associated evoked potentials.

4. Motor evoked potentials (MEP) used in diagnostics primarily to determine the functional state of the nerve pathways between the primary motor cortex and the motor neurons in the spinal cord as well as the peripheral motor nerves up to the muscle.

Another categorisation often used for biosignals, as shown in Figure 3.34, can be derived from the origin, i.e. the place of origin in the body, as well as the physical nature of the signals. Regardless of the type of signal indicated in Figure 3.34, such signals can still be assigned to the categories "autonomous" or "evoked" listed above, as is the case, for example, with electrical heart activity and electrical brain activity, respectively.

The physical categorisation directly results in possible measurement principles for measuring the individual types of signals. A selection of measurement methods for measuring electrical and non-electrical biosignals is discussed in section 4.3 using the example of pulse measurement electrical by conductive electrodes and optically by photoplethysmography. The aim of the conversion of a physical measurand (e.g. temperature, pressure, etc.) into an electrical output variable is often the measurement of the temporal change of the original physiological measurand as a signal curve $s(t)$ as well as its representation and subsequent evaluation. A wide-ranging landscape of applications in medical technology can be found here, which include screening and

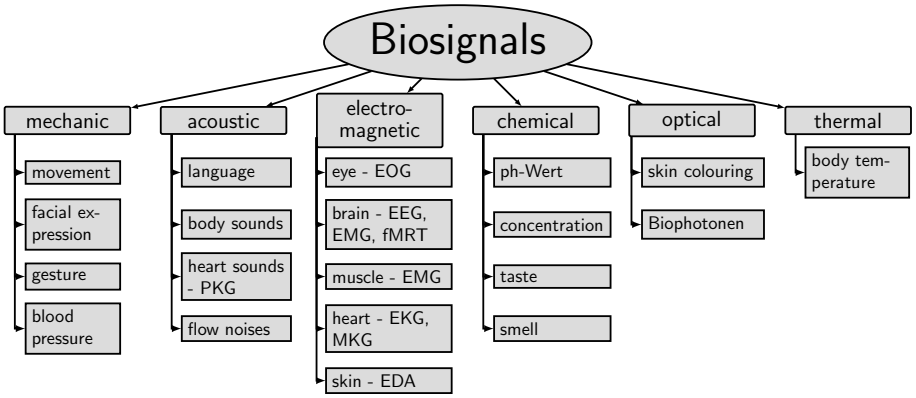


Fig. 3.34: Taxonomy of biosignals.

therapy, real-time analysis in the clinical environment and basic research in addition to functional analysis.

Screening attempts to identify diseases in large parts of the population at an early stage by using methods of biosignal processing through a functional analysis, e.g. measuring the electrical activity of the heart, in order to be able to take therapeutic measures at an early stage. In real-time analysis, data streams generated during clinical monitoring of intensive care patients are analysed online for crisis intervention, for example, to trigger an alarm. Basic research is more concerned with the modelling and simulation of measured variables such as blood pressure or the emergence and propagation of action potentials in cells and nerve conduits. The methods required for evaluation are presented in chapter 5 and deepened in selected applications in chapter 6. The application of mathematical models of physiological processes in the development of novel evaluation algorithms is steadily increasing in importance. An introduction to simulation methods has already been given in section 3.2 on the basis of the development of action potentials and autonomic heart activity.

3.4 Post-Reading and Exercises

Physiology and Electrical Activity of Nerve and Muscle Cells

1. Describe possible diagnostic scenarios using an example from neurology. What conclusions can the neurologist draw, for example, about the transmission of stimuli?
2. What special properties do phospholipids have? How do they manifest themselves in aqueous solution? Describe the formation of a cell membrane and explain the term "self-organisation".

3. What electrical properties do biomembranes have? Give orders of magnitude and draw the equivalent circuit.
4. Explain the term electrogenic transporter. What forms do you know, how do they differ and what are their functions in the cell?
5. In what form do biomembranes obtain their functionality? Describe them in relation to electrogenic transporters.
6. Name the possible conformational changes of channel and transport proteins. What is the function of each of these?
7. What forms of excitation of channel proteins do you know? Explain the excitation on the basis of the Na^+ channel.
8. What is meant by a resting membrane potential and how can it be explained?
9. How does an action potential develop? Describe the exact process and the associated electrochemical processes in the cell.
10. Describe the excitation of a nerve cell across the synaptic cleft. What role does the membrane threshold play in this process?
11. What is meant by the relative and absolute refractory period?
12. In what way does an increase in a stimulus affect an action potential? Within what limits is this possible?
13. Describe the two forms of nerve conduction in the body. What are the differences and in which animal species do the different forms occur primarily?
14. How do you explain the need for active conduction of the action potential?
15. What methods do you know to measure an action potential? What conclusions can be drawn from such measurements? Name their applications in science and technology.
16. Describe the central and peripheral nervous system. For example, in what form does muscle excitation take place?

Electrophysiology of the Heart

1. Describe the function of a pacemaker cell.
2. Which pacemaker cells are present in the heart and what is their function?
3. Can a pacemaker complex consisting of a cluster of these cells fail? If so, what happens?
4. How do the action potentials of a muscle cell and a pacemaker cell of the heart differ? How do you explain these differences?
5. How is the electric field created during cardiac excitation, is it like the electric potentials on the surface of the body? What is meant by a current dipole, what by a projection? How does the current dipole behave during the heartbeat?
6. Give the time course of the potential on the body surface during a heartbeat. Into which sections can the ECG signal be divided, and what meaning and association with cardiac function can be established?

7. What diseases can be identified from a patient's ECG? Why does this inference from the signal to the heart disease work?
8. Why do you always need a reference mass to measure potentials? What is meant by a virtual reference potential?
9. Describe the differences in measuring potentials at the body surface according to Einthoven, Goldberger and Wilson.
10. Describe the mechanical process in the heart during a heartbeat. When do the heart valves open or close and how does the blood flow into the left and right ventricle respectively?
11. Give a model for simulating the time course of an ECG. What differences in the parameterisation do you know, what are the consequences?
12. Why is it dangerous if the heart is exposed to a short potential jump?

Taxonomy of Biosignals

1. Which forms of biosignals do you know, how can they be classified? Give examples.
2. Describe the formation of electrical potentials on the body surface for an EMG or an EEG signal. Are there similarities in terms of the formation of the ECG?
3. What is meant by an evoked potential? Explain it on the basis of somatic evoked potentials. Do you know other evoked potentials and possible fields of application?
4. Explain the function of the autonomic nervous system on the basis of the conduction of impulses in the heart. What fundamental advantage do you see in the autonomic form?
5. Describe Berger's experiment. Into which other frequency bands can the EEG be divided, what is their significance?

4 Measurement of Biosignals and Analog Signal Processing

After the electrophysiological processes in cells as the origin of biosignals were discussed in chapter 3 this chapter deals with the measurement of biosignals and their processing using analog techniques. Signal processing or signal conditioning is an essential step before digital signal processing (chapter 5), because biosignals, captured at the body surface are, on the one hand, too weak for direct analog-to-digital conversion and, on the other hand, usually overlaid by disturbances such as the 50 Hz mains hum. First, the measurement of electrical biosignals (e.g. electrocardiogram, -encephalogram, -myogram) will be discussed in detail, before the measurement of non-electrical biosignals is discussed in the back part of this chapter.

4.1 Measurement of Electrical Biosignals

Electrical biosignals occur as a potential difference (electrical voltage) between two points on the body surface and can be derived there with electrodes. However, signal capture within the body is also possible. For example, in the electrocorticogram electrodes are placed directly on the brain tissue at the open skull, in order to be able to investigate the dynamics of bioelectrical processes in certain diseases such as epilepsy with a comparatively large signal-to-noise ratio and high spatial resolution. Electrodes made of platinum-iridium in a matrix arrangement on silicone or latex are used for this purpose. However, this type of invasive measurement is associated with a high level of stress for the patient and an enormous medical and metrological effort, which is why it is only used in exceptional cases. For a more detailed discussion of this topic, please refer to the relevant literature [27]. In the vast majority of cases, biosignals are derived as potential difference at the body surface via skin electrodes. The starting point for the potential difference is an electrically excited cell area within the body. As explained in chapter 3, in the excited cell area, the ion concentration in the cytoplasm (intracellular space) and in the surrounding interstitium (extracellular space) is different from that in cell areas with resting potential. As a result, a potential difference is formed within the body, which can be regarded as an internal electrical voltage source. Since this voltage source is surrounded by more or less conductive tissue (cf. Table 4.1), *ionic compensating currents* occur within the body. The current paths penetrate large areas of the body and reach the surface of the skin. An electric potential can be attributed to each point on the current paths. The path of the equipotential surfaces is complex because of the different conductivity of the various tissue types within the body. Figure 4.1 shows a simulation result for the potential distribution at the body surface starting from the cardiac electrical activity. Thus, two randomly selected points usually have different potential. The potential difference of these two points is related

https://doi.org/10.1515/9783110736298-004

Tab. 4.1: Electrical conductivity of selected tissue types and physiological saline.

Tissue/Substance	Conductivity in S/m
Physiological saline	2
Body fluid	1,5
Blood	0,67
Heart muscle	0,1–0,2
Brain	0,17
Kidney	0,16
Skeletal muscle	0,08–0,25
Lung	0,07–0,1
Fat	0,02–0,1
Bone	0,006–0,02

to the biosignal source in time and space. In practice, one selects the leakage points according to a standardized arrangement on the body surface, if possible.

Electrical contact with the skin is necessary for biosignal measurement. The skin is composed of the three layers epidermis, dermis and subcutis (cf. Figure 4.2). The total thickness is between 1.5 and 4 mm. Among other things, the skin serves as a fluid barrier for the body. The horny layer, the outermost layer of the epidermis, plays a decisive role in this. The stratum corneum consists of dead epithelial cells joined together to form firm and compact layers with relatively high mechanical and chemical resistance. Its function as a fluid barrier also provides an electrically insulating effect, since electrolytes can hardly penetrate the epidermis. Nevertheless, an electric current can flow through the skin. This is primarily due to sweat pores and hair ducts that begin in the lower part of the dermis and subcutis and penetrate the epidermis. Ion flow is possible along the pores and channels.

For electrical modeling, the skin can be understood as a parallel circuit of capacitor and resistor, where the largely insulating epidermis accounts for most of the capacitive part and the sweat pores and hair ducts for the resistive part. The total impedance of the skin Z_H is of the order of a few kΩ for low-frequency currents up to 100 Hz, although it is clear from what has been said previously, that this value depends strongly on the sweat gland activity or the degree of hairiness. Furthermore, it varies locally and is frequency dependent because of the capacitive property. For simulation purposes, in the equivalent circuit diagram according to Figure 4.2 (left), for example, a capacitance value C_H of 200 nF and a resistance value R_H of 2 kΩ can be used. For the electrical modeling of the lower skin region, an additional resistor of the order of 100 Ω is inserted in series with the parallel circuit from Figure 4.2.

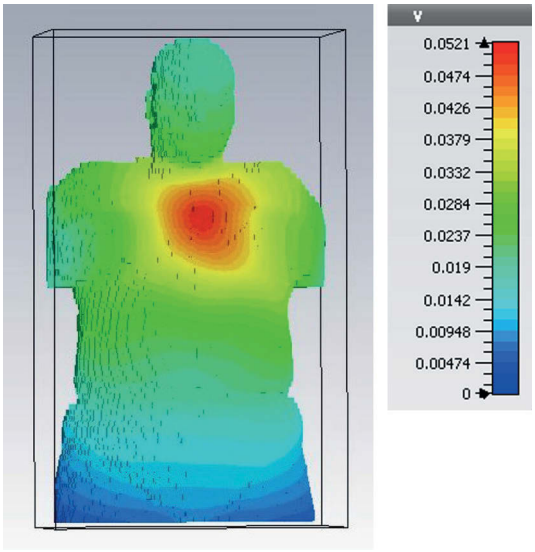


Fig. 4.1: Simulation result for potential distribution on the body surface during ECG; the coloring indicates the potential in volts: The simulation was performed using CST Studio Suite software. The cardiac vector was simulated by two point charges inside the heart with a static potential difference of $100\,\mathrm{mV}$. The numerical body model contains all real tissue components with a spatial resolution of approximately $1\,\mathrm{mm}^3$.

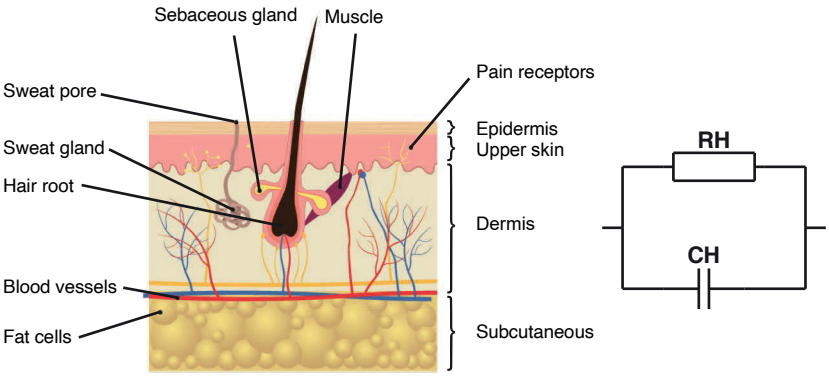


Fig. 4.2: Structure of the skin (left, after [4]): The DC component of the current flows along the hair ducts and sweat glands, which are located within the lower skin layers and penetrate the uppermost skin layer. Electrical equivalent circuit of the skin (right). For simulation purposes, $C_H = 200\,\mathrm{nF}$ and a resistance value $R_H = 2\,\mathrm{k\Omega}$ can be assumed. In fact, the values depend on several factors and can therefore vary considerably.

4.1.1 Electrodes

The next link in the electrical measuring chain between the biosignal source and the measuring amplifier are the electrodes. They represent the coupling element between

the body and the measurement electronics. Thus, electrodes have an important role in the electrical measurement circuit, which is why a basic understanding of the electrochemical processes is required. A more detailed description can be found in [4]. At the electrode the change of the charge carrier type takes place; from dissolved ions (electrolyte) to free electrons in the metallic contact surface of the electrode. First the situation in the thermodynamic equilibrium is considered, i.e. without measuring current. The transition between electrolyte and metal represents a *electrochemical phase boundary*. The two phases usually have a different chemical potential, which is balanced by charge exchange through the *phase boundary*. In a metal-electrolyte transition, the atoms of the metal surface donate electrons to the electrolyte. Positively charged atoms then remain on the metallic side of the interface. As a result, polarized water molecules accumulate on the electrolyte side, which, due to their spatial charge distribution, have an electric dipole moment and are therefore polar. The alignment of the water molecules causes, that dissolved ions with hydrate shells[1] can attach to the ion. This arrangement is called *Helmholtz layer*, an extension of this model Gouy-Chapman layer. The extent of the boundary layer is only one atomic monolayer on the metallic side, i.e. approximately 0.1 nm, and on the electrolyte side a few nm up to several hundred nm at very low electrolyte concentration.

The measurement of electrical biosignals requires charge carrier exchange at the phase boundary. However, direct contact of the electrode metal with the skin can lead to problems with electrical contact for several reasons. For example, wetting of the metal surface with the superficial body fluid is often uneven and highly dependent on patient movement, or body hair prevents direct contact between metal and skin. Therefore, an additional electrolytic *electrode gel* or spray is used as a contact agent. In disposable adhesive electrodes, the gel is in a cup between the internal electrode metal and the skin contact area in the center of the adhesive ring (see Figure 4.3, left). For reusable electrodes such as suction electrodes (cf. Figure 4.3, right), either gel is first applied to the metal surface or spray is applied to the skin before use. Silver is very often used as the electrode metal. With external voltage, chemical reactions begin that allow current to flow across this phase boundary. In one current direction, electrons from the silver enter the silver chloride phase and dissociate AgCl to form Ag^+ and Cl^- (cathodic reaction), in the other current direction, Ag^+ and Cl^- ions associate to form AgCl (anodic reaction). The current is thus carried in the electrolyte by Cl^- ions. The electrode gel or spray has a correspondingly high Cl^-- concentration.

The electrodes have another metrological significance. The charge exchange at the phase boundary to establish thermodynamic equilibrium causes charging of the electrode. The electrode potential depends on the electron affinity of the metal and the chemical potential of the electrolyte. In principle, both measuring electrodes have the

1 The hydrate shell is the attachment of water molecules around an ion. Due to the polarity of the water molecules, they arrange themselves directionally around the ion in the form of a shell.

same structure and should therefore have the same electrode potentials, which then compensate each other in the measuring situation. In practice, however, a differential voltage of a few mV often remains, which overlays the measurement signal as an offset. This can be caused by a different ion concentration at the phase boundary of the two electrodes. As will be shown, this offset can be eliminated by a high-pass filter in the measurement electronics.

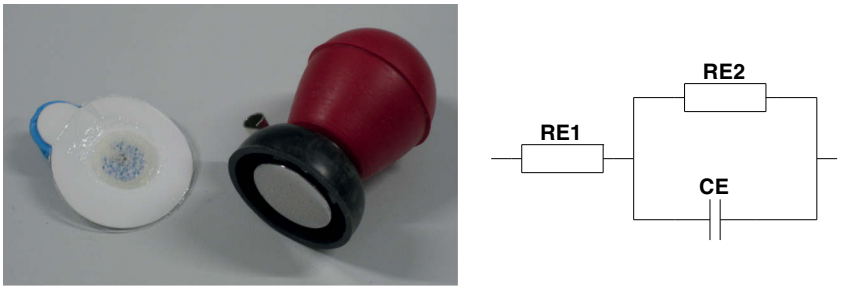


Fig. 4.3: Disposable adhesive electrode and suction electrode (left). Electrical equivalent circuit of the electrode (right): R_{E1} represents the lead resistance in the contact wire and electrolyte. R_{E2} and C_E describe the electrical behavior of the phase boundary between metal and electrolyte (Helmholtz layer).

Electrically, the electrode can be represented by a series connection (cf. Figure 4.3, right), where R_{E1} represents the lead resistance of the contact wire and the electrode gel. The parallel circuit of R_{E2} and C_E describes the phase boundary between electrolyte and metal. The capacitive behavior is explained by the layer sequence metal, Helmholtz layer, electrolyte, which corresponds to a capacitor arrangement (conductor-insulator-conductor). The resistance R_{E2} represents the adsorption and desorption processes that take place between Ag and AgCl. As in the electrical description of the skin, the magnitudes of the individual equivalent circuit components depend on various factors. For simulation purposes, $R_{E1} = 30\,\Omega$, $R_{E1} = 100\,\Omega$, and $C_E = 30\,\mu F$ can be assumed. Overall, the equivalent circuit for skin and electrode is given in Figure 4.4.

At low frequencies below 100 Hz, the impedance of the capacitors C_H and C_E is much larger than the parallel resistors R_H and R_{E2}. In that case, the two capacitors can be neglected in the equivalent circuit. What remains are the three resistors, which are now in series and can be combined into one resistor. Since R_{E1} and R_{E2} are smaller than R_H by more than an order of magnitude, the total impedance of the skin and electrode in this approximation is resistive and determined by the skin resistance R_H. This resistance will be called *transition resistance* in the following.

In electromyography (EMG), *needle electrode* are used as an alternative to flat conduction electrodes. These enable the contacting of individual motor units of the

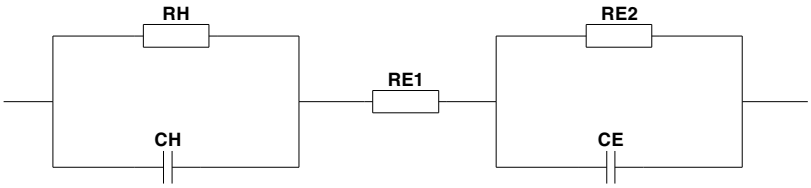


Fig. 4.4: Electrical equivalent circuit of skin and electrode, composed of Figure 4.2 (right) and Figure 4.3 (right).

muscle up to the detection of the action potential of individual muscle fibers, which is impossible with surface electrodes due to the significantly lower spatial resolution. Surface electrodes in the form of electrode arrays, on the other hand, are better suited for studies of the spatial and temporal propagation of the action potential along individual motor units. For in vivo measurements, the needle electrode is usually used to contact the extracellular area where the outer action potential is measured. Controlled puncture of the inner cellular area of a muscle fiber and thus measurements of the inner action potential are usually only possible in vitro. The finer the needle tip, the higher the spatial resolution and more selective the measurement with respect to the action potential of individual muscle fibers. With increasing distance of the electrode to the muscle fiber, however, the EMG amplitude decreases.

In practice, both monopolar and bipolar needles are used (see Figure 4.5). The monopolar version consists of a coated stainless steel needle with a diameter of 300 to 500 µm. The coating electrically insulates the needle body from the punctured tissue. Only the tip contacting the area of interest is exposed. Teflon or medical grade silicone is used as the coating material. The counter electrode can be realized by means of a surface electrode near the puncture site, but still outside the area of the motor unit to be examined. Bipolar electrodes consist of a stainless steel cannula (counter electrode) and a metal core (electrode) surrounded by an insulating layer. The outer diameter is in the range of 300 to 700 µm, the diameter of the core is about 100 µm. The material of the core is usually silver or platinum. In the monopolar electrode, the active area of the needle tip is larger than in the bipolar electrode, which means that a larger area of the motor unit is covered by the electrode, producing a stronger EMG signal. In contrast, a higher spatial resolution is achievable with the bipolar electrode. In addition, with this type the superimposed common mode signal[2] is much weaker, since the electrode and counter electrode are very close to each other, which means that the common-mode signal is present at the input of the measuring amplifier with almost the same amplitude and phase and is largely eliminated by the high common-mode rejection.

2 The term common mode signal is discussed in detail in the following sections.

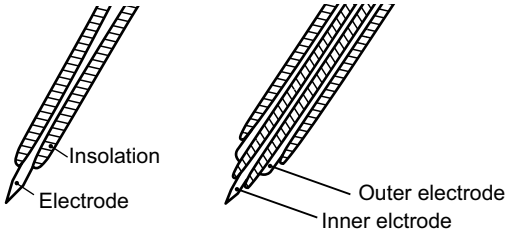


Fig. 4.5: Schematic representation of monopolar (left) and bipolar (right) electrode needle (modified from [50]).

Tab. 4.2: Signal amplitude and bandwidth of selected electrical biosignals.

Biosignal	Electrode type	Signal amplitude	Bandwidth in Hz
Electrocardiogram (ECG)	Skin electrode	0,1–4 mV	0,01–250
Electroencephalogram (EEG	Skin electrode	1–100 µV	0,01–100
Electromyogram (EMG)	Needle electrode	0,1–5 µV	1–5.000
Electrooculogram (EOG)	Skin electrode	0,05–4 mV	0,01–100
Electroneurogram (ENG)	Needle electrode	0,01–3 mV	0–10.000

4.1.2 Electrical Amplifier

Depending on the biosignal and the electrode position, the differential voltage between the measuring electrodes at the body surface ranges from a few microvolts for the EEG to several millivolts for the R-wave in the ECG. The bandwidth of electrical biosignals ranges from 0 Hz to a maximum of 10 kHz. In Table 4.2 values for signal amplitude and bandwidth of some important biosignals are listed.

The amplifier has the task of raising the biosignal to a voltage level in the volt range with a constant gain over the entire bandwidth. As a rule, the biosignal (useful signal) is superimposed by interference signals such as the common-mode signal or noise. A common-mode interference is caused, for example, by power lines located in the vicinity of the measurement setup. The origin of such network disturbances is treated in section 4.2. At this point it is assumed that the common-mode signal is a harmonic alternating voltage whose amplitude may be a few volts and, in the case of the mains disturbance, has the mains frequency (50 Hz or 60 Hz). The common-mode signal also enters the measurement electronics. Figure 4.6 clarifies the measurement situation. The amplitude of the common-mode signal (interference signal) often exceeds that of the biosignal by several orders of magnitude. From this, the necessity of a very high common-mode rejection is directly derived as an important requirement for the amplifier (*Common Mode Rejection*, in short: CMR). Eliminating the common-mode signal with analog filters is not an option for most applications because the filter impedance is small and the rejection frequency band of the filter overlaps the frequency band of the biosignal and would lead to distortion of the biosignal. This aspect is discussed in detail in section 4.4.

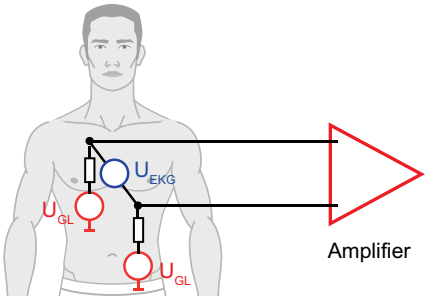


Fig. 4.6: Measurement situation with consideration of the common mode signal U_{GL}: U_{ECG} represents the biosignal applied between the electrodes.

Another requirement for the measurement amplifier is a high input impedance. In the following equivalent circuit diagrams, the biosignal source is assumed to be in series with the signal source assumed to be ideal by an internal resistance of $R_I = 2\,k\Omega$. A current load on this source would cause the source voltage to collapse. This can be avoided by using a high input impedance, several orders of magnitude above the internal resistance of the signal source. In practice, therefore, the input impedance of the measurement amplifier should be at least $Z_E = 1\,M\Omega$.

The next consideration is directed to the thermal noise of the amplifier circuit, expressed by the effective noise voltage U_{eff}. The cause of thermal noise is fluctuations in the spatial distribution of free charge carriers within a conductor or semiconductor due to thermal motion. According to Equation 4.1, U_{eff} depends on the ohmic resistance R, the absolute temperature T and the bandwidth f_B. k_B is the Boltzmann constant.

$$U_{eff} = \sqrt{4k_B T R f_B} \quad . \tag{4.1}$$

The strength of the noise in relation to the strength of the wanted signal is often expressed by the *signal-to-noise-ratio* (*SNR*: signal-noise-ratio) in dB. Related to the thermal noise , the following applies

$$SNR = 20 \log_{10} \frac{U_{biosignal}}{U_{eff}} dB \, . \tag{4.2}$$

Here U_{use} is the RMS voltage of the biosignal and U_{eff} is again the effective noise voltage. Now a calculation example: If the ohmic input resistance of the amplifier is 10 kΩ and the bandwidth is limited to 800 Hz, this results in an effective noise voltage at the input of the measurement amplifier of 0.36 µV at room temperature. Let us assume a useful signal strength of 36 µV, this results in an *SNR* of 40 dB.

A further consequence of the consideration of the thermal noise in Equation 4.1 is to limit the amplifier to high frequencies by a low-pass filter due to the dependence of the effective noise voltage on the bandwidth. The cutoff frequency of the low-pass filter must be designed according to the bandwidth of the biosignal to be measured in order to ensure both distortion-free transmission of the biosignal and to obtain a value for f_B that is as small as possible.

Another noise source, the so-called 1/f noise, makes the limitation of the transmission band towards low frequencies important. The 1/f noise is particularly strong in semiconductor devices and is due to the temporal fluctuation of the charge carrier concentration caused by generation and recombination processes. As the name implies, the noise power spectral density[3] is not uniformly distributed, but increases towards low frequencies. Therefore, high-pass filters are used for suppression. The cutoff frequency of the high-pass filter must again be matched to the bandwidth of the biosignal to avoid signal distortion. Therefore, the cutoff frequency must be well below 1 Hz. A second positive effect of the high-pass filter is the elimination of a possible DC component in the biosignal. The DC component[4] can be caused, for example, by unequal electrical charging at the phase boundaries of the electrodes (cf. subsection 4.1.1).

In summary, the amplifier for electrical biosignals must be able to uniformly amplify a differential signal with a bandwidth of up to 800 Hz, as well as have high input impedance and high common-mode rejection. These requirements can best be met with a circuit based on an Instrumentation amplifier. *Instrumentation amplifiers* are available as integrated devices or can be be built with individual *operational amplifiers* (OPV). The advantage of integrated circuits is the very low component tolerance, which provides a high common mode rejection. In this book, however, the circuit design with lumped OPVs will be dealt with in detail in order to gain a deeper understanding of the individual circuit elements. The following equivalent circuit diagrams were created and simulated with the freely available software LTspice. The exercises for this chapter assume application knowledge in LTspice.

In Figure 4.7 a two-stage instrumentation amplifier consisting of three operational amplifiers is shown. $\mathrm{OPV}_{1,2}$ together with resistors $R_{1,2,3}$ form the first stage. The biosignal is fed as a differential voltage is supplied via the terminals U_{E1} and U_{E2}. The required high input impedance is realized by the high input resistance of the two of the two OPVs. To calculate the gain of this stage, recall that for negative feedback OPVs, the potential difference between the two inputs of an OPV is exactly 0 V. Accordingly, the voltage drop across R_3 is

$$U_{\mathrm{R3}} = U_{e1} - U_{e2} \,. \tag{4.3}$$

Furthermore, let $U_{a1.2}$ be the output voltage of OPV_1 and OPV_2, respectively, referenced to ground. Then, according to the voltage divider rule:

$$U_{\mathrm{R3}} = \frac{R_3}{R_1 + R_2 + R_3}(U_{a1} - U_{a2}) \,. \tag{4.4}$$

3 The noise power spectral density describes, how much noise power is contained per frequency interval. The calculation of the power spectral density is presented in Equation 6.1.

4 The DC component is a DC voltage superimposed on the useful signal. The DC component is not to be confused with the common-mode signal.

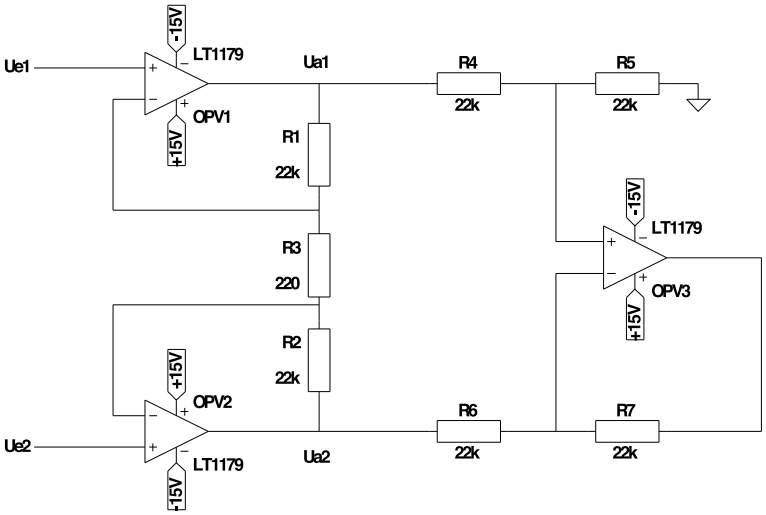


Fig. 4.7: Two-stage instrumentation amplifier with the gain 201.

If in Equation 4.4 U_{R3} is replaced by Equation 4.3 and furthermore it is considered that R_1 and R_2 are equal for symmetry reasons, then the following holds for the difference of the two output voltages:

$$U_{a1} - U_{a2} = \left(1 + \frac{2R_1}{R_3}\right)(U_{e1} - U_{e2})\,. \tag{4.5}$$

With Equation 4.5 and the values $R_1 = 22\,\mathrm{k\Omega}$ and $R_3 = 220\,\Omega$ the gain for the first stage is $V = 201$. The two outputs of OPV_1 and OPV_2 are symmetrically connected via $R_{4,6}$ to OPV_3. The output of OPV_3 is the potential difference $U_{e1} - U_{e2}$ with respect to ground. By increasing the relation between $R_{5,7}$ and $R_{4,6}$ an additional amplification of the second stage can be achieved which was not done here.

Let us now consider common-mode rejection. For this purpose, in the circuit after Figure 4.7 the biosignal U_{EKG} and the common mode signal U_{GL} are added to the circuit input. $R_{\mathrm{u}1,2}$ in Figure 4.8 represent the contact resistances at the electrodes, which are assumed to be equivalent here.

According to the superposition principle, the common mode signal can be considered independently of the differential voltage. Due to the circuit symmetry, it is present with the same amplitude and phase at the input of the operational amplifiers OPV_1 and OPV_2 and consequently also above and below R_3. So there is no current flowing through R_3, which is why this resistor can be omitted for the common mode consideration. If R_3 is replaced by open terminals, OPV_1 and OPV_2 form in conjunction with R_1 and R_2 an impedance transformer. Thus, U_{GL} passes unamplified (gain is 1) the first stage of the instrumentation amplifier and is applied to the positive and negative inputs of OPV_3. The common mode rejection is entirely determined by this operational amplifier. This consideration assumes that the resistors $R_{4,5,6,7}$ have ex-

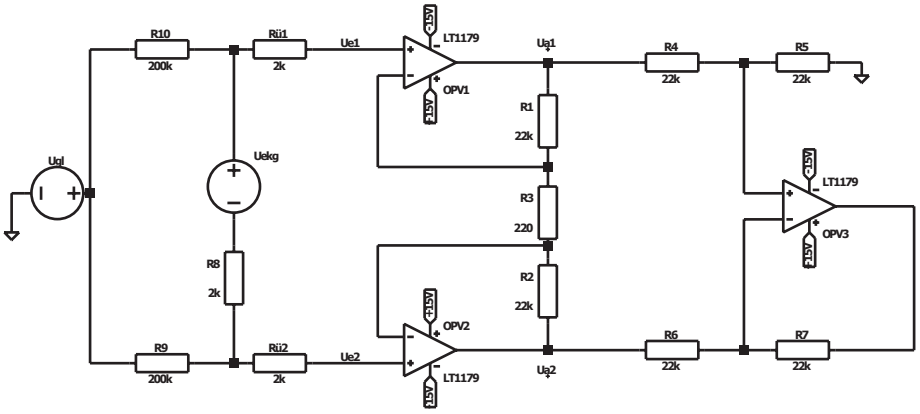


Fig. 4.8: Two-stage instrumentation amplifier with consideration of the biosignal U_{EKG}, the common mode signal U_{GL} and the contact resistances $R_{\mathrm{u1,2}}$.

actly the specified value. Such a condition is usually well satisfied in integrated circuits. However, if resistance tolerances have to be taken into account, as is the case with lumped resistors $R_{4,5,6,7}$, the common-mode rejection is massively degraded. The resistance ratios $\frac{R_5}{R_4}$ and $\frac{R_7}{R_6}$ in the circuit according to Figure 4.8 determine the gain at which the common mode signal is applied to the inputs of the subtractor. For $\frac{R_5}{R_4} \neq \frac{R_7}{R_6}$ the signal amplitudes at the two inputs are different, resulting in a residual signal at the subtractor output. To come as close as possible to the requirement $\frac{R_5}{R_4} = \frac{R_7}{R_6}$, precision resistors must be selected for $R_{4,5,6,7}$. The remaining difference in the resistor ratio can be eliminated by a subsequent circuit adjustment. For this purpose, the resistor R_5 in Figure 4.8 must be replaced by the potentiometer circuit according to Figure 4.9.

A measure of the common mode rejection capability is the *CMRR* value in dB (*CMRR*: Common Mode Rejection Ratio). *CMRR* relates the differential gain V_{diff} to the common mode gain V_{CM}:

$$CMRR = 20 \log_{10} \left(\frac{V_{\mathrm{diff}}}{V_{\mathrm{CM}}} \right) \mathrm{dB} \,. \tag{4.6}$$

Commercially available operational amplifiers have a *CMRR* of more than 90 dB. For example, if the amplitude of the common mode signal is 5 V, then *CMRR* = 100 dB leaves 50 µV at the output of OPV_3, because for the second stage of the instrumentation amplifier after Figure 4.8 $V_{\mathrm{diff}} = 1$. The biosignal, on the other hand, is amplified by the first stage by a factor of 201. Thus, an ECG with an amplitude of 1 mV at the input of the instrumentation amplifier is approximately 200 mV at the output, which is 4000 times than the common mode signal with an input amplitude of 5 V, which is quite sufficient for most applications.

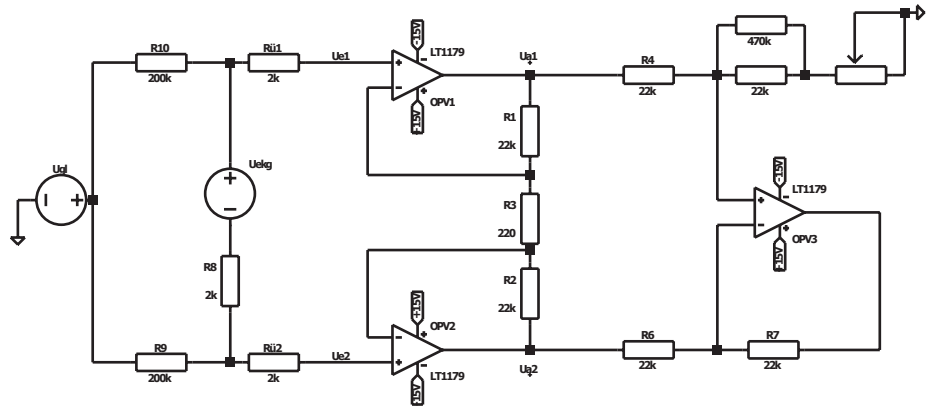


Fig. 4.9: Potentiometer circuit for adjusting $R_5/R_4 = R_7/R_6$ in the instrumentation amplifier according to Figure 4.8: The potentiometer circuit replaces resistor R_5 in Figure 4.8.

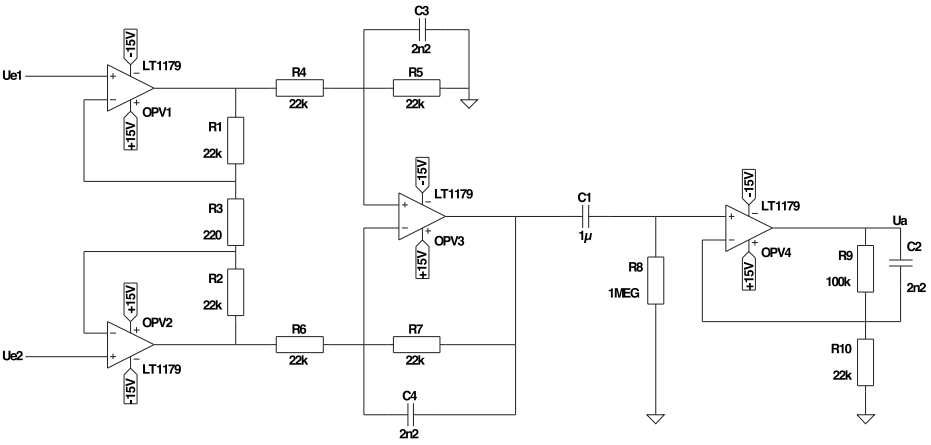


Fig. 4.10: Extension of the instrument amplifier from Figure 4.7 by a further amplifier stage as well as a high-pass and two low-pass filters.

The circuit from Figure 4.7 is now supplemented with regard to the previously described requirements. A further amplifier as well as a high-pass and two low-pass filters are connected downstream. Figure 4.10 shows the extended circuit.

OPV_4 together with $R_{9,10}$ forms a non-inverting amplifier with gain 5.5. This gives a total gain of 1100 for both amplifier stages. C_1 and R_8 form a high-pass filter with cutoff frequency 0.16 Hz. The first low-pass filter is realized by the capacitors $C_{3,4}$. At high signal frequency, the parallel resistors $R_{5,7}$ are almost short-circuited by $C_{3,4}$, whereby the gain of the subtractor falls far below one. The cutoff frequency is approximately 3.3 kHz. The second low-pass filter is produced by the capacitor C_2 in parallel with R_9,

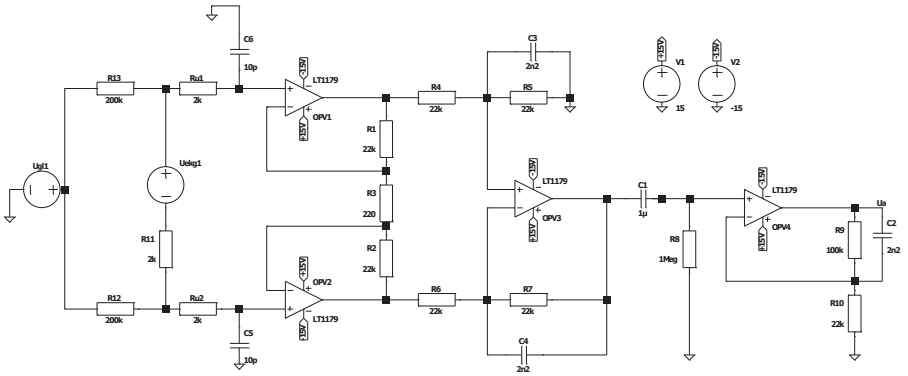


Fig. 4.11: Amplifier considering the biosignal U_{ekg1}, the common mode signal U_{GL} and the resistors $R_{u1,2}$ representing the interface to skin. The capacitors $C_{5,6}$ represent the stray capacitance of the measurement cables to ground.

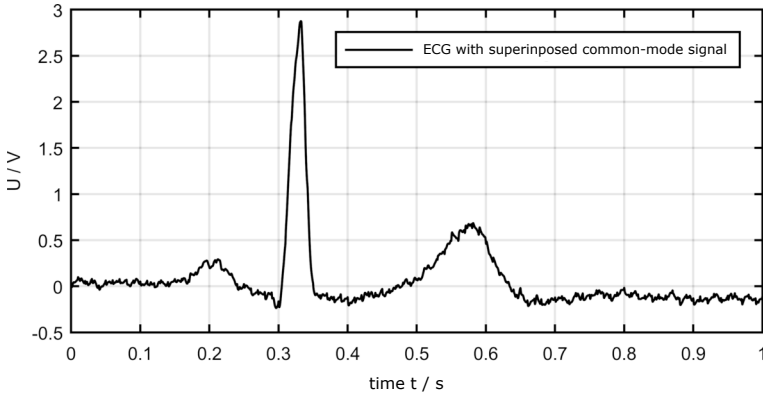


Fig. 4.12: Output signal during simulation of the measuring amplifier after Figure 4.11 with a ECG and superimposed common-mode signal (50Hz hum with 5 V amplitude voltage at the circuit input). The common mode disturbance is hardly visible in the output signal because of the high common mode rejection of the circuit.

which in turn shorts the resistor R_9 at high frequency and thereby reducing the gain of the non-inverting amplifier down to one. The cutoff frequency of the second low-pass filter is 720 Hz. This measurement amplifier is very suitable for recording ECG, EOG or the low frequency part of an EMG.

Figure 4.12 shows the output signal of the circuit simulation from Figure 4.11, where an ECG of amplitude 2.5 mV with a superimposed common-mode signal was applied. Due to the high common-mode rejection, the output signal hardly shows a

common-mode signal (50 Hz mains hum). For an EEG measurement a larger gain is necessary, which can be achieved e.g. by increasing R_9. It should be noted that the cutoff frequency of the second low-pass filter is then reduced accordingly.

4.2 Signal Interference

Biosignals are generally superimposed by interfering signals. In the worst case, a diagnostic evaluation of the biosignal is impossible. In addition, the interfering signal often lies partly or completely within the spectral range of the biosignal and therefore cannot be easily removed with filters without changing the biosignal at the same time. For this reason, precise knowledge of the causes of interfering signals is indispensable, in order to be able to take countermeasures from the outset.

4.2.1 Network Disturbances

In the normal measurement situation, the electronics as well as the test person itself are surrounded by electrical power lines. Even if these are not in direct contact with each other, an interference signal can couple into the signal from the power lines. This is also called mains hum. The mains hum is like the causal power line a harmonic alternating voltage with the same frequency, i.e. 50 Hz in Europe resp. 60 Hz in North America. It can be very effectively eleminated by a narrowband digital bandstop filter (notch filter). However, since the mains hum lies within the bandwidth of biosignals, this is accompanied by signal distortion.

Capacitive Coupling

From an electrical point of view, the human body can be regarded as an inhomogeneous conductive electrolyte container. Between the conductor of the electric network (L1) and the body, a stray capacitance is formed, as well as between the body and the neutral conductor N. Figure 4.13 represents the situation schematically.

The capacitance value of the stray capacitances depends strongly on the distance between body and L1 or N. Following the plate capacitor formula, the capacitance increases with decreasing distance. For further consideration, values between 1 pF and 100 pF can be assumed. Furthermore, stray capacitances also occur between the measuring cables and the power lines. As long as the mains interference couples into the electronics via the two electrodes with the same amplitude and phase, i.e. as a common-mode signal, and the two lead lines up to the subtractor of the amplifier are completely symmetrical to each other, this situation does not pose a problem. As already explained in the previous section, only the differential signal between its two inputs is amplified by the instrumentation amplifier. The common mode signal is sup-

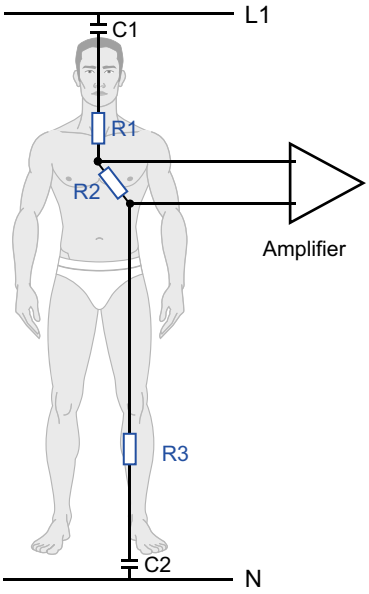


Fig. 4.13: Schematic representation of the capacitive coupling of mains interference into the body by stray capacitances.

pressed by a sufficiently high *CMRR* value to such an extent that it has practically no significance for the output signal. This situation was clarified in Figure 4.8. In fact, the conditions are more complex. In reality, the two electrode junction resistances $R_{u1,2}$ are never exactly equal. For example, if R_{u1} is replaced in the circuit according to Figure 4.11 by 2 kΩ to 4 kΩ, this is reflected in the simulation result shown in Figure 4.14. In it, the ECG is clearly overlaid by the network disturbance.

Usually the amplitude of the mains disturbance at the two electrodes is also not identical. This can be explained in principle with the help of of Figure 4.13. Between the potential of L1 and N the upper stray capacitance C_1, the body and the lower stray capacitance C_2 form a voltage divider. Within the body, the electrical behavior can be described by ohmic resistances: as R_1 between the fictitious coupling point at the head and the first electrode, as R_2 between the two electrodes, as well as as R_3 between the second electrode and the junction to the lower stray capacitance. Now the differential voltage between the both electrodes is to be estimated. For simplification, it is assumed that the two stray capacitances have the same value $C_1 = C_2 = 10\,\text{pF}$ and the resistances R_1 and R_3 can be neglected. The impedance of $C_{1,2}$ at 50 Hz is about 320 MΩ, respectively. For R_2 2 kΩ is assumed as before, between L1 and N 230 V AC voltage is applied. According to the voltage divider rule, this results in a differential voltage between the two electrodes of about 0.7 mV. An ECG signal has an amplitude of about 1 mV at this point. Accordingly, the biosignal and the superimposed network disturbance have the same magnitude. Both are transmitted equally amplified to the output of the amplifier circuit, which makes a medical evaluation of the biosignal impossible.

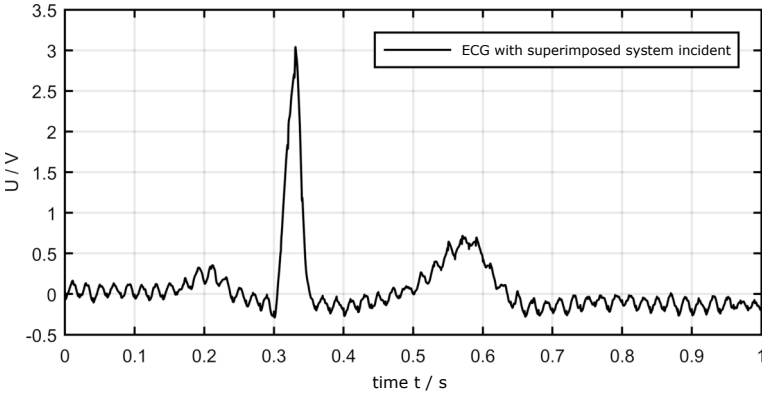


Fig. 4.14: ECG with superimposed mains disturbance: simulation of the output signal of the circuit according to Figure 4.11 with $R_{u1} = 4\,k\Omega$ and $R_{u2} = 2\,k\Omega$.

One way to reduce the network disturbance is the principle of reference potential control, also called *driven right leg*. Here, the common mode signal is tapped behind the first stage of the instrumentation amplifier, inverted (phase shift by 180°), and returned to the body via a third electrode. By superimposing the common-mode signal with the opposite-phase common-mode signal from the third electrode, the two ideally cancel each other out. In practice, however, the two amplitudes of common mode signal and the opposite-phase common-mode signal are different, which means that the common-mode signal is not completely eliminated. Nevertheless, the disturbance can be significantly reduced in this way. Figure 4.15 shows the extension of the circuit according to Figure 4.10 by the reference potential control.

The common mode signal is tapped between the two resistors $R_{11,12}$ and fed to the impedance converter OPV_5. Behind it an inverting amplifier follows, formed by OPV_6 and $R_{13,14}$. There the common mode signal is phase shifted by 180° and amplified by a factor of 22. The amplification is intended to compensate for the voltage drop from the injection point to the electrodes.

Inductive coupling

According to the Biot-Savart law, the current through a power line generates a magnetic field that surrounds the conductor in a circle. For a straight conductor surrounded by air, the magnetic flux density B is calculated to be

$$B(t) = \frac{\mu_0}{2\pi r} I(t) e_\varphi \tag{4.7}$$

with the vacuum permeability μ_0, the perpendicular distance to the conductor r and the current $I(t)$. Since I is an alternating current, B must also be an alternating quantity. According to the law of induction, an alternating magnetic field induces a voltage

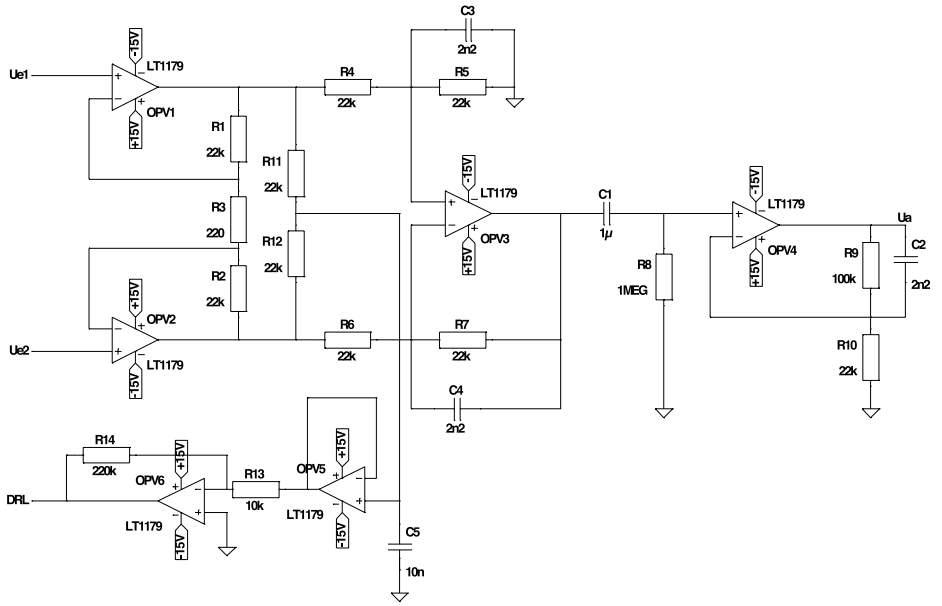


Fig. 4.15: Amplifier circuit for biosignals with reference potential control.

in a conductor loop according to

$$U_{\text{ind}} = -\frac{\partial \Phi}{\partial t} \,, \tag{4.8}$$

where the magnetic flux is defined as follows:

$$\Phi = \iint B \mathrm{d}A \,. \tag{4.9}$$

A is here the area enclosed by a conductor loop. The two measuring cables, the body and the measuring amplifier form a conductor loop (cf. Figure 4.16). If this conductor loop is perpendicular to the alternating magnetic field an induced alternating voltage is generated at the input of the measuring amplifier, which in turn superimposes the biosignal. Overall, the strength of the inductive coupling is determined by four quantities, namely, the current I, the frequency (due to the derivative in Equation 4.8), the distance r between the conductor loop and the mains cable L1 and the area A of the conductor loop. While it is usually not possible to influence the current I in the power cable and the frequency with 50 Hz (60 Hz in North America), the remaining ways to reduce the effect are: i) to keep a large distance r to the current-carrying power cable, and ii) keep the area A of the conductor loop as small as possible. The latter can be achieved by twisting the measuring cables together (*twisted pair*) or at least run close together. With these two methods, inductive coupling can be effectively avoided.

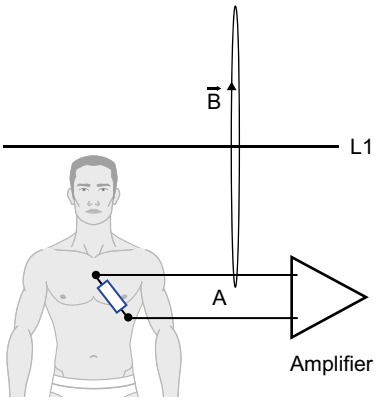


Fig. 4.16: Schematic representation of inductive mains coupling: Starting from the current in the mains cable L1, a magnetic field is generated, which penetrates the conductor loop consisting of the two measuring cables, the body and the measuring amplifier and induces an AC voltage at the input of the measuring amplifier.

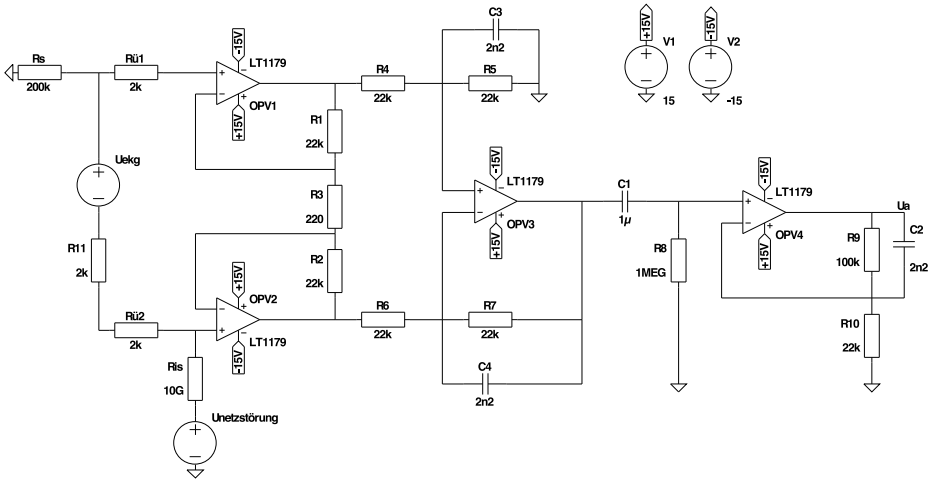


Fig. 4.17: LTspice simulation of the galvanic network disturbance with network coupling via the insulation resistance $R_{is} = 10\,G\Omega$; the resistor R_s is needed for simulation purposes to give a reference potential to the source U_{ECG}.

Galvanic Coupling

The galvanic coupling of mains disturbance requires a limited resistance between the measurement electronics and the mains. At first glance, this does not seem to be a realistic situation, since both –also for reasons of electrical safety –must be sufficiently isolated from each other. However, even at large insulation resistance, a significant network disturbance can occur, as shown by the simulation result in Figure 4.18. The insulation resistance is sometimes lowered by creepage paths. For simulation purpose, the insulation resistance R_{is} and a line voltage were inserted in the electronics circuit (Figure 4.17). An effective countermeasure is to encapsulate the electronic components that carry the mains current (e.g. transformer).

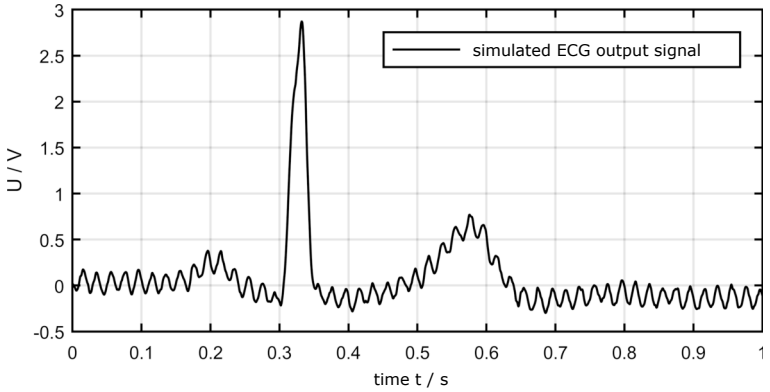


Fig. 4.18: Result of the simulation of the circuit from Figure 4.17.

4.2.2 Transient Disturbances

In the previously discussed circuit according to Figure 4.10, an RC high-pass filter (C_1, R_8) with a cutoff frequency of 0.16 Hz was incorporated to suppress any DC components in the biosignal and to limit the transmission band to low frequencies. This very low cutoff frequency is accompanied by a correspondingly large time constant ($\tau = 1\,\mathrm{s}$), which causes very slow balancing processes . A step-wise DC component can occur, for example, if one of the electrodes is pressed during the measurement to improve the contact to the body. This can change the charge of the electrode. The resulting DC voltage overlaps the biosignal and decays exponentially with the time constant of the RC high-pass filter. Countermeasures in the form of filters do not exist, because in this case a filter of the electronics itself is the cause of the disturbance. As a rule, it is necessary to wait until the transient disturbance has decayed before evaluating the biosignal.

4.2.3 High-Frequency Interference due to Electromagnetic Radiation

In many measurement situations, the electronic circuit is exposed to high-frequency electromagnetic radiation. Sources of radiation are e.g. mobile phone networks, cordless phones or WLAN. The electromagnetic radiation can couple to the electronic circuit by an unshielded measuring cable which acts as an antenna. The antenna is particularly effective if the cable length is half the wavelength of the radiation or a multiple thereof. For a 1.2 m long measuring cable, this would correspond in free space to a transmission frequency of 125 MHz. This frequency is far above the bandwidth of biosignals.

But the high-frequency radiation can couple in by amplitude modulation. Amplitude modulation is a method of communications engineering, which is, however,

only used in a few transmission techniques anymore. However, amplitude modulation can also occur by chance. Let us continue to assume the measuring cable as an antenna. If the position of the measuring cable to the transmitter changes, for example because the patient moves, then the amplitude of the radiation captured by the antenna also changes, which corresponds to amplitude modulation. The electronics consists of components needed for demodulating the high-frequency radiation. For example, a simple envelope demodulator consists of only a diode, a capacitor, and a resistor. Here, the pn junction of a bipolar transistor or the metal-semiconductor junction of a field-effect transistor can act as a diode. It is therefore possible that the amplitude-modulated high-frequency radiation is coupled into the electronics and demodulated. The result is the superposition of the modulation signal with the biosignal. Therefore, the greatest possible distance from transmitters must be ensured.

4.3 Transducer for Non-Electrical Biosignals

In section 4.1 the measurement of electrical biosignals was discussed in detail, where the biosignal is already present as an electrical voltage and can be fed directly to the electronics via electrodes. In addition, there are a number of non-electrical biosignals, as can be seen from the overview in Figure 3.34. Their electronic detection requires the conversion of the physical quantity in which the biosignal occurs into a voltage or a current in order to be able to process it further. In this section, selected transducers for the conversion of non-electrical biosignals are presented.

4.3.1 Sound Transducer

Listening to the body, called *auscultation*, is one of the oldest diagnostic procedures in medicine. Especially the sounds of the heart and lungs provide information about their function and possible diseases. The standard instrument used for auscultation is the stethoscope. It consists of a transducer with a membrane for the conversion of solid-borne sound[5] in airborne sound and a branched sound tube, at the ends of which are two ear olives, which provide the hermetic seal with the auricles of the examiner. Sound recording found its way into clinical practice in the 1950s in the form of phonocardiography, i.e. the recording of heart murmurs. At that time, the devices consisted of a solid-borne sound measuring head, analog measuring electronics and a paper recorder for recording. Since the phonocardiogram is evaluated both in its temporal course and in its spectral composition, the signal was fed to five different analog

5 Solid-borne sound is a term from acoustics. Rather, the term is used to describe the propagation of sound in solids and, in some cases, liquids, which differs from propagation in gases (e.g., airborne sound).

Fig. 4.19: Self-construction of an electronic stethoscope for connection to the PC sound card.

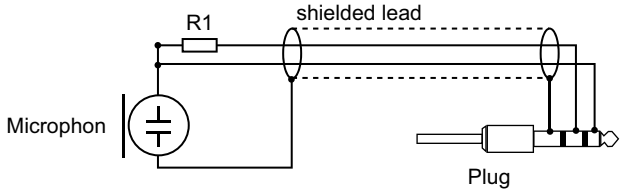


Fig. 4.20: Wiring diagram for an electret microphone.

bandpass filters in parallel and, filtered in this way, output to the paper recorder as individual curves. In parallel, the ECG was registered as an independent marker for cardiac phase. With the advent of imaging techniques, especially echocardiography using ultrasound, phonocardiography was increasingly displaced as a standard cardiological diagnostic technique. Today, electronic stethoscopes are commercially available. A microphone integrated in the sound tube converts the sound, which is then transmitted wirelessly by radio to a PC.

In Figure 4.19 a simple self-construction for the wired connection to a PC sound card is shown. The sound tube of the stethoscope was cut open and an electret microphone was sunk into it. Electret microphones have a very small size and have uniform sensitivity over a wide frequency range, which covers the bandwidth of heart sounds of max. 1 kHz. The microphone capsule has an integrated impedance converter. This is supplied via an upstream working resistor of 2.2 kΩ. The supply voltage is provided by the PC sound card via the microphone input (phantom power). Furthermore, a three-pole cable with jack plug is required. Figure 4.20 shows the connection diagram.

Another form of phonocardiographic self-monitoring by the patient shows Figure 4.21. This is the functional monitoring of an implanted mechanical prosthetic heart valve. Such implants have an increased risk of thrombotic formation, which in turn

may be the cause of reduced valve function or for a thromboembolic event[6] can be. Close monitoring by the patient opens up the possibility of detecting such accumulation processes at an early stage and treating them with medication.

Fig. 4.21: Patient with handheld device for self-monitoring of implanted mechanical prosthetic heart valve.

The method is again based on sound recording and analysis. Mechanical prosthetic heart valves produce characteristic sound in the frequency range up to 20 kHz. In the case of thrombus attachment, the high-frequency component of the valve sounds changes in particular. Stethoscopes cannot be used as sound recorders at high frequencies [7]. The device shown in Figure 4.21 uses an electret microphone that registers the airborne sound emitted from the body. The measurement position is in the sternal axis and not above the heart, because there the high-frequency sound components are largely absorbed by the lungs. The valve sounds are amplified, filtered and fed to a digital signal processor for analysis. The analysis result is displayed on a small screen. The subsection 6.3.2 presents the necessary analysis methods using the Matlab programming language.

6 Thromboembolism: occlusion of a vessel by a thrombus (blood clot).

7 Good stethoscopes have a bandwidth up to approx. 600 Hz. The bandwidth is limited in particular by the inertia of the diaphragm.

4.3.2 Optical Sensors for Plethysmography and Determination of Oxygen Saturation

Plethysmography (Gr. *plethys* = fullness) means the recording of volume variations within the body. In the context of this book, we restrict ourselves to the volume change in vessels due to pulsatile blood flow. The palpation of the pulses belongs to the oldest diagnostic procedures and plays a major role especially in traditional Chinese medicine. On surface (subcutaneous) vessels, the change in volume can also be measured optically. In this case it is called photoplethysmography. The measuring principle is that light is irradiated into the body and the transmitted or reflected radiation is measured with a light sensor. The sensor signal is variable in time as a result of the blood flow. When the pulse wave passes through the irradiated vessel section, the vessel dilates due to the larger blood volume. The larger blood volume then ensures greater light absorption compared to the phase when there is only a small volume of blood in the vessel section. As a result, the light sensor detects less light. In this way, continuous pulse curves can be recorded. The pulse curve contains information about the condition of the vessels and the cardiovascular system. For a differentiated evaluation, pulse curves usually have to be measured at several points. It is particularly easy to determine the heart rate with photoplethysmography. To do this, the duration between the maxima of the pulse curve is determined. The reciprocal of this multiplied by 60 gives the heart rate in the usual unit of beats per minute.

An extension of photoplethysmography is the SpO_2 measurement[8]. It allows the determination of the oxygen saturation of the blood. For this purpose, the different absorption of hemoglobin (Hb) and hemoglobin saturated with oxygen (oxyhemoglobin HbO_2) at different wavelengths is exploited. For the determination of oxygen saturation, measurements at at least two different light wavelengths are necessary, so at least two monochromatic light sources are required. The measurement result is expressed as a percentage and corresponds to the ratio of HbO_2 to total hemoglobin (HbO_2 + Hb). The procedure is presented in more detail in subsection 6.3.2.

Semiconductor light-emitting diodes (LEDs) are predominantly used as the light source. These have a high intensity with comparatively low power consumption, have a small design and are very inexpensive. When measuring SpO_2 with two LEDs, care must be taken to ensure that the emission spectra are far enough apart and do not overlap. LEDs are available in the color spectrum from blue to infrared. Color selection is based on the absorption spectra of Hb and HbO_2. In order to distinguish between the transmitted and reflected radiation of the respective LEDs at the detector, the two LEDs must be operated in pulse mode and multiplexing mode. The accuracy of the SpO_2 determination depends crucially on the adherence to the emission wavelength, as stated in subsection 6.3.2. From Table 4.3, it can be seen that the location of the

8 S denotes saturation, p pulse, O_2 oxygen.

Tab. 4.3: Characteristic values of commercially available LEDs for photoplethysmography and SpO_2measurement; λ_{peak} is the wavelength of maximum emission, TK is the temperature coefficient at λ_{peak}.

LED type	λ_{peak} in nm	Bandwidth in nm	TK in nm/K
infrared	950	42	0,25
red	660	17	0,13
green	525	34	0,03

maximum emission λ_{peak} changes with temperature. From this point of view, short pulse times are preferable because then, at the same power, the absorbed electrical energy remains low and the LED does not heat up much. A positive side effect of short pulse times is the lower power consumption, which is important for battery-powered sensors. On the other hand, shorter pulse times reduce the signal-to-noise ratio, which in turn can be compensated by higher power. In the control scheme, both aspects must always be weighed against each other. The LED is usually controlled by a microcontroller. A typical value for the pulse width is 250 µs and for the repetition rate 500 Hz. A phase without illumination is usually inserted between the LED pulses to determine the ambient light offset. Alternatives to semiconductor LEDs are organic LEDs (OLED) [42] or semiconductor laser diodes. OLEDs are based on flexible polymer films that can conform to the body surface. Semiconductor laser diodes have the advantage of very small bandwidth.

Photodiodes are used as detectors. These convert the incident light into a photocurrent which is proportional to the light intensity. Silicon-based photodiodes have a spectral response of about 400 nm to 1100 nm with a maximum at about 920 nm. Thus, the radiation from the LED listed in Table 4.3 can be detected by a Si photodiode, but with quite different sensitivities in each case. For the SpO_2 determination, however, this circumstance does not cause any problems, since the spectral sensitivity is not included in the calculation of the oxygen saturation index, as will be shown in subsection 6.3.2. A transimpedance converter is connected after the photodiode to convert the photocurrent to a voltage (cf. Figure 4.22, middle). The resistor R_1 also provides amplification, and the capacitor C_1 provides band-limiting and compensation for the photodiode's junction capacitance. When designing the low-pass filter formed by R_1 and C_1, it must be ensured that the short light pulses of the LED are transmitted without distortion.

SpO_2 devices most commonly use finger clip sensors that work with transmission. Light sources and detector are located on opposite sides of the finger (cf. Figure 4.22, left). The finger clip is connected to a small control unit and integrated display, but finger clips with integrated electronics and display are also available. For continuous monitoring of newborns, sensor bands tied around the foot are often found instead of finger clips. In reflection sensors, the LED and photodiode are on one side (cf. Figure 4.22, right). This offers the advantage that areas of the body such as the wrist or

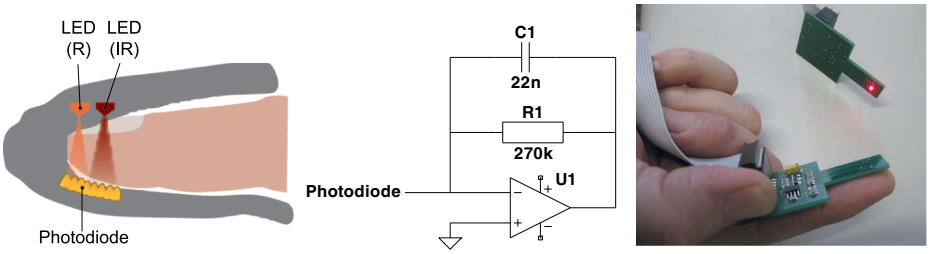


Fig. 4.22: Principle of SpO_2 measurement in transmission (left), exemplary equivalent circuit of a transimpedance amplifier with low-pass filter (center) and pulse curve measurement with a reflection sensor (right): Control and readout are performed via a PC measuring card.

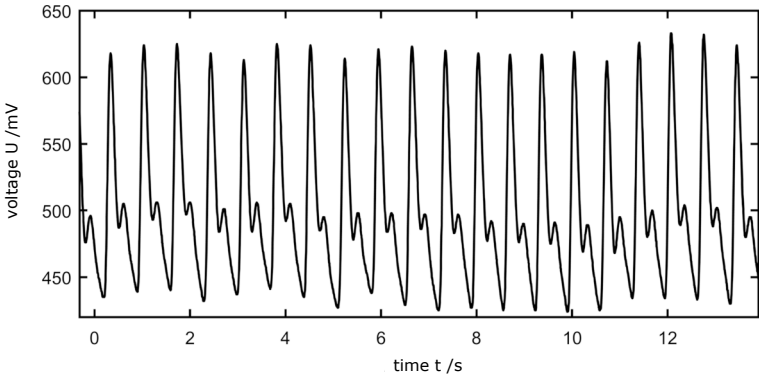


Fig. 4.23: Photoplethysmogram (PPG) from the measurement with the reflectance sensor of Figure 4.22.

temple become accessible to a SpO_2 measurement, which are excluded for a transmission measurement. With regard to mobile applications in the form of wearable functional textiles, this aspect is becoming increasingly important. Figure 4.23 shows a pulse curve recording with a reflectance sensor.

4.4 Interference Suppression and Analog Filtering

The measurement of biosignals is usually overlaid by disturbances. In section 4.2 different causes were presented. In this section, possibilities of interference suppression by means of analog filters are discussed. The essential function of filters is the suppression of defined frequency components. When using filters, it should be noted that the biosignal is distorted if the stopband of the filter overlaps with the biosignal spectrum, and thus the corresponding frequency components of the biosignal are suppressed in addition to the interference signal. In this case it has to be checked whether the resulting signal distortion is tolerable with regard to the medical-diagnostic statement,

especially if the reliability of the diagnosis depends on a original representation of the biosignal. One may imagine the fatal consequences, if, as a result of signal filtering, an ECG change typical of a heart attack would not be apparent, although it is visible in the raw signal. In the following, the different types of interference and their frequency characteristics shall be considered in principle. A simplified representation of the spectra of the respective disturbance can be found in Figure 4.24.

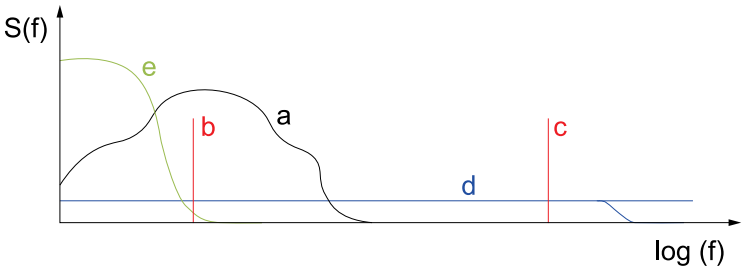


Fig. 4.24: Simplified overview representation of the spectra: (a) arbitrary biosignal, (b) power line hum, (c) high-frequency harmonic disturbance (without sidebands of frequency modulation), (d) noise or impulsive disturbance, and (e) transient disturbance. The decay of the curve (d) would be assumed in the case of $1/f$- or $1/f^2$-noise or a real impulse with temporal expansion. Any harmonics are not shown for clarity.

1. *Periodic disturbances*: The most important periodic disturbances include 50 or 60 Hz mains hum and coupling from high-frequency transmitters. If the interference signal is purely harmonic (sinusoidal), the corresponding spectrum is pulse-shaped at the respective frequency. The 50 or 60 Hz mains hum lies within the spectrum of most biosignals. Carrier signals of transmission channels (e.g. WLAN, mobile radio) lie in their frequency far above the spectrum of biosignals. Frequency-modulated transmission channels produce sidebands in the spectrum near the transmission frequency. In the case of amplitude-modulated signals, the transmission frequency is also above the biosignal spectrum. However, amplitude modulation provides low-frequency spectral components that can overlay the biosignal.
2. *Noise*: Noise is based on stochastic processes. Accordingly, the noise has an arbitrary time course with a very broad spectrum. In the case of thermal noise (cf. subsection 4.1.2) and shot noise[9] the power density spectrum is constant and extended over the entire frequency range. This is referred to as white noise. In the case of $1/f$- or $1/f^2$-noise[10] the power spectral density decreases with increasing

9 Shot noise arises from the random fluctuation in the number of charge carriers passing through a potential barrier (e.g. pn junction).

10 The $1/f^2$-noise arises, e.g., from the recharge of surface and interface states.

frequency. In any case, however, the biosignal spectrum lies within the noise spectrum. By means of a bandpass filter whose passband includes the biosignal spectrum, the noise power can be reduced linearly to the bandwidth of the filtered component.

3. *Transient Interference*: As already described in subsection 4.2.2, transient disturbances can occur during power-on processes due to the charging of the capacitance of the high-pass filter in the measurement amplifier. The charging curve is proportional to $-e^{-t/\tau}$, where τ is the time constant of the high-pass filter. With $R = 1\,\mathrm{M\Omega}$ and $C = 1\,\mu\mathrm{F}$, the time constant is $\tau = 1\,\mathrm{s}$. The spectrum of the charge function is proportional to $\sqrt{1/(1+\omega^2\tau^2)}$. With the calculated time constant, the spectrum of transient disturbances overlaps the lower frequency range of most biosignals.
4. *Pulse type disturbances*: Causes can be artifacts or also pulse-like processes in technical equipment, which couple into the measuring signal. The spectrum of an ideal, infinitesimally narrow pulse is infinitely extended. In the case of a real pulse with temporal expansion, the spectrum falls off towards high frequencies.

The simplest way to realize filters electronically is to use low-pass and high-pass circuits of RC elements. Bandpass and bandstop filters can be created by connecting low-pass and highpass in series and parallel, respectively. The cutoff frequency can be set by selecting the resistor and capacitance values. The cutoff frequency indicates the frequency value at which the transfer function is dropped by 3 dB (corresponding to the factor $1/\sqrt{2}$) with respect to the passband. For a simple RC filter (first-order filter), the cutoff frequency is:

$$f_\mathrm{g} = \frac{1}{2\pi RC}\,. \tag{4.10}$$

The slope is 20 dB/frequencydecade. If steeper edges are required, this can be achieved by connecting the same filter in series several times. Then the slope is $n20$ dB/frequency decade, where n is the number of filters connected in series (filter order). However, when connecting in series, it should be noted that this shifts the cutoff frequency. For example, if two identical filters are connected in series, the cutoff frequency is a factor of 0.37 smaller than the cutoff frequency of a single-stage filter, if the −3 dB- criterion for the cutoff frequency continues to apply.

The calculation of the cutoff frequency in Equation 4.10 assumes that the internal resistance of the signal source at the filter input is very small and the load at the filter output is very large, both related to the impedance of the RC circuit. If this is not the case, internal resistance and load must be considered when calculating the transfer function of the filter and the resulting cutoff frequency. If then still the impedance values of the internal resistance and the load are not known or are known only inaccurately, the filter characteristic, especially with regard to the cut-off frequency, can no longer be controlled.

One way out is the use of operational amplifiers. These have a very high input resistance and a very low output resistance. This electrically decouples the filter output

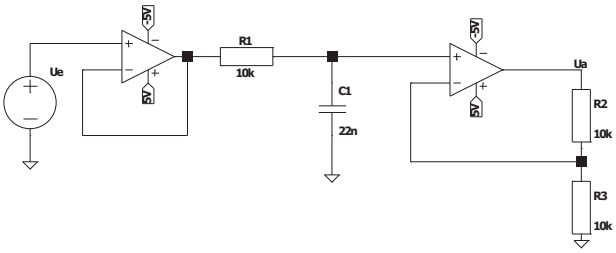


Fig. 4.25: Active low-pass filter of 1st order with cutoff frequency 723 Hz and gain by a factor of 2.

with the load from the actual RC filter element. Figure 4.25 shows the equivalent circuit of an RC low-pass filter with operational amplifier. The first operational amplifier acts as an impedance converter (also called a voltage follower) on the signal source. The second one is connected as a non-inverting amplifier and provides signal amplification, in addition to the impedance conversion[11]. Magnitude and phase frequency response of the circuit from Figure 4.25 are shown in Figure 4.26. Filter circuits with active components such as operational amplifiers or transistors are called active filters. When used in filter circuits, the relatively low bandwidth of operational amplifiers must be taken into account.[12] With regard to high-frequency interference above the bandwidth of the operational amplifiers used, active filters lose their filtering effect. Furthermore, for applications with a very small biosignal amplitude, it should be taken into account that semiconductor components generate more noise (shot noise, $1/f$ noise) than passive components.

With the circuit from Figure 4.25 higher order filters can be realized by series connection. However, the circuit complexity increases rapidly. A more efficient possibility results from the introduction of a positive feedback branch. In Figure 4.27, a 2nd-order low-pass filter with only one operational amplifier is shown[13]. The output voltage is fed back to the positive input by the capacitor C_2.

In practice the resistors $R_1 = R_2 = R$ and the capacitors $C_1 = C_2 = C$ are often chosen equal. For this case, the transfer behavior, i.e. the ratio of output voltage (U_a) to input voltage (U_e) as a function of frequency, is now to be calculated. To do this, we make use of the auxiliary voltages U_1 (voltage between resistors R_1 and R_2) and U_2 (voltage at the positive input of the right OPV), both referenced to ground respectively.

11 The operational amplifier forms a non-inverting amplifier with the resistors R2 and R3. The gain is $1 + R2/R3$.

12 The bandwidth of many operational amplifiers is limited to a range from 10 kHz to 100 kHz. Above this bandwidth, operational amplifiers lose their amplifying effect.

13 Filters of this type are also called Sallen-Key filters.

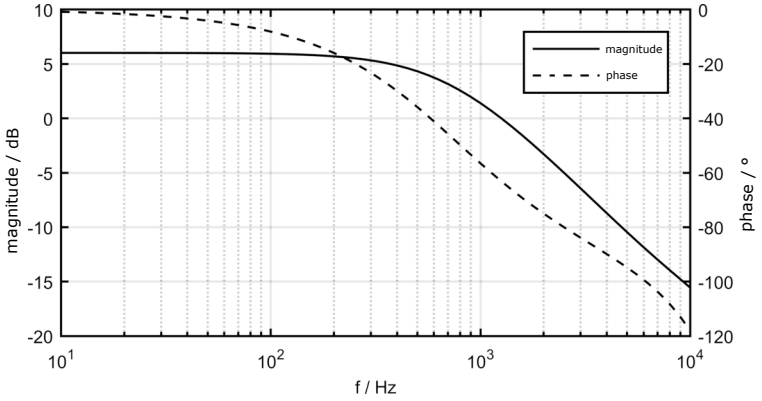


Fig. 4.26: Magnitude and phase frequency response of the low-pass filter from Figure 4.25: The passband is 6 dB because of the gain by a factor of 2. The point for the cutoff frequency intersects the 3-dB line in the segment between 700 and 800 Hz. Mathematically, the cutoff frequency is 723 Hz.

Between U_a and U_2 is a non-inverting amplifier. For the voltage ratio holds:

$$\frac{U_\mathrm{a}}{U_2} = 1 + a. \tag{4.11}$$

Between U_2 and U_1 there is a low pass with the transfer function

$$\frac{U_2}{U_1} = \frac{1}{1 + j\omega RC}\,. \tag{4.12}$$

Another equation is obtained by applying the node rule at U_1. The sum of the three currents flowing to this node must be zero:

$$\frac{U_1 - U_\mathrm{e}}{R} + \frac{U_1 - U_\mathrm{a}}{\frac{1}{j\omega C}} + \frac{U_1 - U_2}{R} = 0. \tag{4.13}$$

Equation 4.11 to Equation 4.13 form a system of equations that can be solved for the transmission ratio we are looking for. After calculation one obtains

$$\frac{U_\mathrm{a}}{U_\mathrm{e}} = \frac{a + 1}{1 + j3\omega RC - \omega^2 R^2 C^2 - j(a + 1)\omega RC}\,. \tag{4.14}$$

The value a corresponds to the internal gain of the filter. This allows to realize different filter types, i.e. filters with different slopes (cf. Table 4.4). In Figure 4.28 for the three a-values from Table 4.4 the magnitude and phase frequency response for a 4th-order filter are shown, which was formed from the series connection of two identical filters according to Figure 4.27 ($R_1 = R_2 = 10\,\mathrm{k\Omega}$, $C_1 = C_2 = 22\,\mathrm{nF}$). The Chebyshev filter exhibits the steepest slope, but it also shows a characteristic ripple in the passband. The Bessel filter has the lowest slope, but has a constant curve in the passband. All

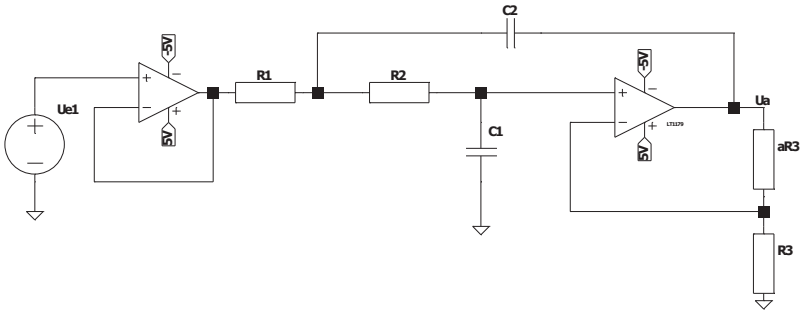


Fig. 4.27: 2nd order low pass filter; the filter type is set via the *a* value (cf. Table 4.4).

Tab. 4.4: Assignment of the value *a* (cf. Figure 4.27) to filter types.

Filter type	Bessel	Butterworth	Chebyshev
a	0,268	0,586	1,234

previous considerations referred to low-pass filters. High-pass filters can be created simply by exchanging R_1 with C_1 in the circuit according to Figure 4.25, or analogously $R_{1,2}$ with $C_{1,2}$ in the circuit according to Figure 4.27.
At the beginning of this section it has already been discussed that parts of the biosignal can also be suppressed by filters. In the unfavorable case, the resulting change in the signal curve can lead to misinterpretations. In some application areas, however, signal distortion is not problematic. For example, in fitness equipment for pulse determination by means of ECG, it is sufficient if the time interval between successive R prongs can be reliably determined. can be reliably determined. In such applications, it is quite reasonable to suppress the 50 or 60 Hz mains hum with very narrowband bandstop filters, so-called notch filters (*notch filter*). A bandstop can be realized from the parallel connection of low-pass and high-pass filters.

Figure 4.29 shows a possible circuit design, Figure 4.30 the corresponding magnitude and phase frequency response. The notch frequency is attenuated here by 45 dB. In Figure 4.31 the effect on an ECG weakly overlaid with 50 Hz mains hum can be seen. The 50 Hz mains hum has completely disappeared due to the filtering. However, the ECG waveform has also changed somewhat: In particular, the R-prong is slightly shifted.

The previous considerations focused on the frequency-selective behavior of analog filters, which is described by the magnitude of the transfer function. However, the filter phase response is also important. Each filter causes a phase shift between input and output signal. This would be unproblematic, if all frequency components of the

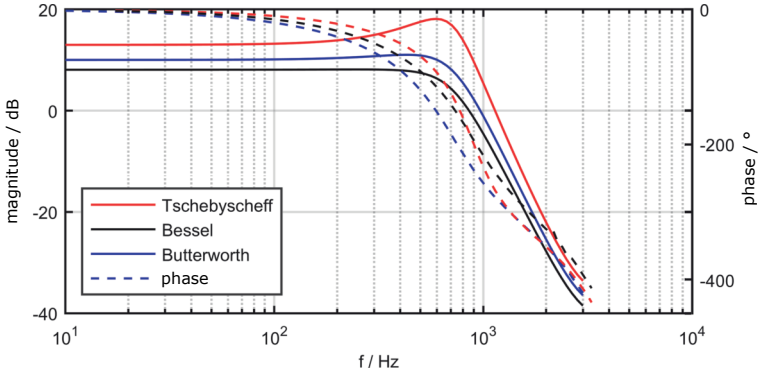


Fig. 4.28: Magnitude and phase frequency response of 4th order Bessel, Butterworth and Chebyshev low-pass filters: For this purpose, the circuit in Figure 4.27 was duplicated and connected in series. The filter type was determined by choosing a after Table 4.4.

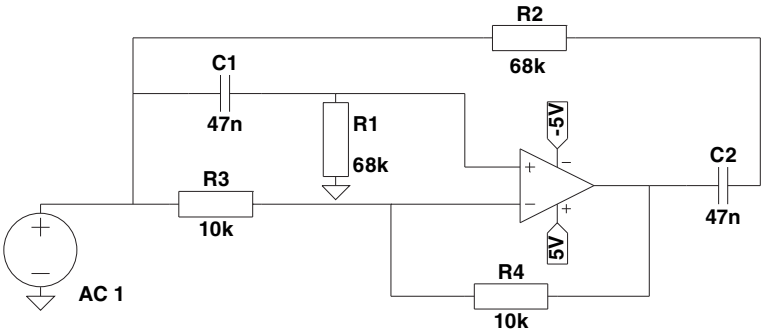


Fig. 4.29: Second order notch filter for 50 Hz suppression.

biosignal had a linear phase frequency response. As the phase responses of the different filter types in Figure 4.28 show, this is generally not the case. There, the phase response at 200 Hz, well into the passband, is $-40°$, while at 10 Hz it still has $0°$. In other words, this means that high-frequency signal components receive a different phase rotation due to filtering than low-frequency signal components. The associated effect on the signal is made even clearer by the introduction of the concept of group delay. This is defined as

$$T_g(\omega) = -\frac{\partial \varphi}{\partial \omega}\,, \tag{4.15}$$

i.e. the negative derivative of the phase φ with respect to the frequency ω. The group delay can be understood as the propagation time for infinitesimally narrowband signal components. In Figure 4.32 the group delay times for a 2nd, 8th and 12th order Bessel filter are shown.

In the example, the group delay times above approx. 100 Hz are no longer constant. Signal components in this frequency range are subject to a different delay dur-

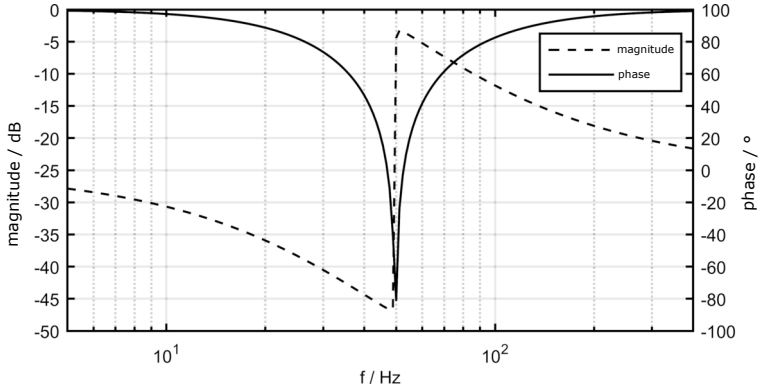


Fig. 4.30: Magnitude and phase frequency response of the notch filter according to the circuit from Figure 4.29: The notch frequency is exactly at 50 Hz and has an attenuation of 45 dB there.

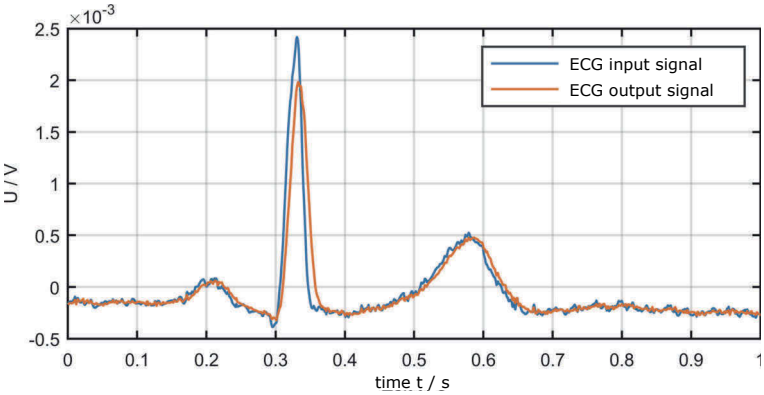


Fig. 4.31: ECG before (blue) and after (red) application of the notch filter from Figure 4.29: The weak 50 Hz mains hum, which is still present in the original signal, was completely eliminated by the notch filter. However, the curve of the filtered signal in the area of the R-prong is changed a little compared to the original curve.

ing transmission through the filter than signal components up to 100 Hz. A steep edge in the signal curve is therefore transmitted differently than signal components with a low gradient. Consequently, the signal is distorted there. In general, the group delay changes with frequency more the higher the filter order and the greater the edge steepness is. In medical measurement technology, Bessel filters are preferably used, because their group delay times vary least with frequency compared to other filter types such as Butterworth, Chebyshev or Cauer filters.

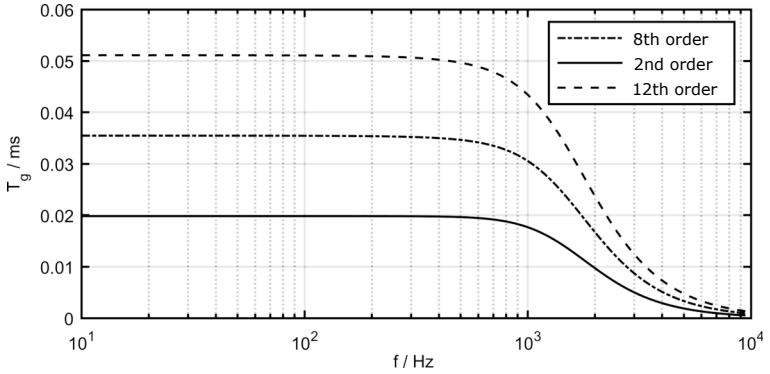


Fig. 4.32: Frequency response of group delay times for 2nd, 8th and 12th order Bessel filters.

4.5 Design of Analogue Filters

Electrical filters are often used in the signal processing of biological measurands to:

- suppress interfering signals such as mains hum, noise, measurement errors or errors caused by transducers when the patient moves,
- select desired useful signals, e.g. in ECG,
- improve detectability, e.g. by applying the raised-cosine filter in the transmission of digital signals,
- limit the highest frequency of the signal spectrum, so that when the conditions for sampling (sampling theorem) when the signal is later sampling conditions (Shannon sampling theorem) are met, and to
- improve the ratio between useful and interfering signal power for later processing (interference ratio increase e.g. by matched- or Kalman-Filter).

Analogue filters can be realised either by purely passive components such as resistors, coils and capacitors or in combination with active components that increase the signal energy, such as transistors or operational amplifiers.

For the application in the biological field, two important design goals will be described in more detail below: i) the design of selective analogue filters to optimise the magnitude frequency response and ii) the design of analogue filters with linear phase response to achieve a constant progression of the group delay, which is particularly important, for example, for the analysis of an ECG.

4.5.1 Selective Filters to Optimise the Magnitude Frequency Response

Selective filters filter out a part of a signal spectrum. In this way, for example, the spectrum of the useful signal can be limited in order to digitise it afterwards (antialiasing-

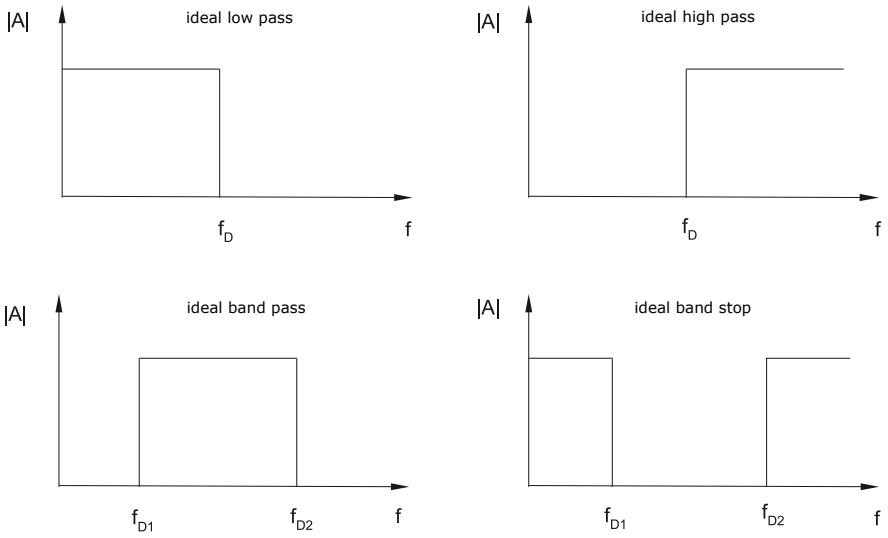


Fig. 4.33: Ideal-selective-standard-filters: $|A|$ denotes the magnitude of the transfer function, f_D the pass frequency of a low or high pass, f_{D1}, f_{D2} the left and right pass frequencies of a bandpass or bandstop.

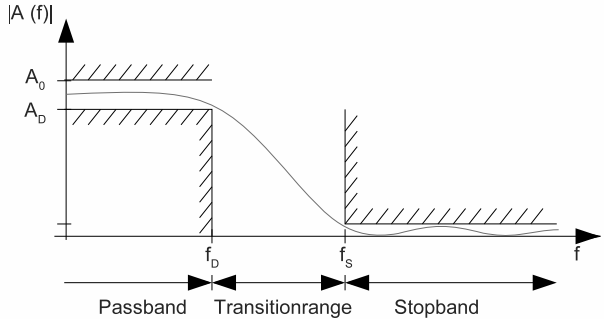


Fig. 4.34: Tolerance-ranges-of-a-low-pass filter, with A_0, A_D as the upper and lower values of the gain in the passband and f_D, f_S as the passband and stopband frequencies.

filter), or certain spectral lines such as the 50 Hz mains hum can be suppressed. Depending on the selectivity requirement, standard filters such as lowpass, highpass, bandpass and bandstop are used (cf. Figure 4.33).

Of course, the properties of ideal filters cannot be realised with real filters, but the tolerance range in which the magnitude frequency response must lie can be specified. This is for example shown in Figure 4.34 for a low-pass filter. The magnitude of the transfer function $|A|$ of the designed filter must lie within the non-shaded area. Furthermore, with the standard filters it can be specified whether the course of the

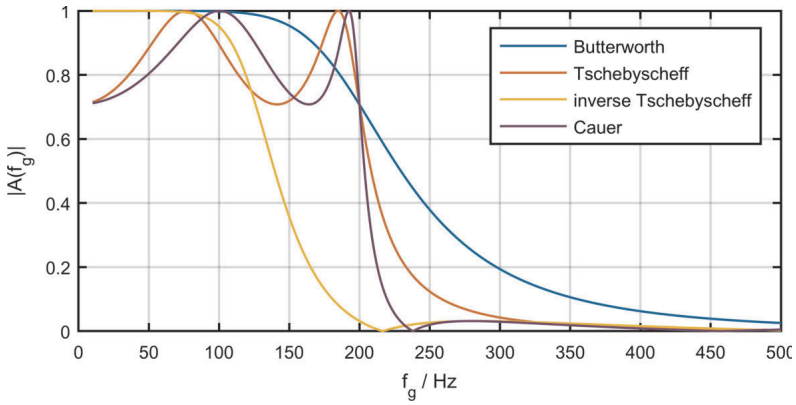


Fig. 4.35: Approximations of the ideal magnitude frequency response for a 4th order low pass filter with a cut-off frequency of 200 Hz using Butterworth, Chebyshev and Cauer filters: The gain in the passband A_D is 1. The passband and stopband tolerances are each 0.3 (corresponding to a passband tolerance of 3 dB), the passband frequency f_D is 200 Hz and the stopband frequency f_S is approximately 210 Hz.

magnitude frequency response in the tolerance range should be flat or wavy. Depending on the specification, these distinctions are made:

- a flat course in the passband and stopband (power- or Butterworth-filter),
- only a wavy course in the passband (Chebyshev-filter),
- only a wavy course in the stopband (inverse Chebyshev-Filter) or
- a wavy course in both the passband/stopband (elliptical filter or Cauer filter).

Figure 4.35 shows an example of a low-pass filter with a cut-off frequency of 200 Hz approximated to the 4th order by the above filter types, and Listing 4.5.1 shows the associated Matlab-script.

Listing 4.5.1: Matlab example of calculating the frequency response of different low-pass filters.

```
n = 4;          % filter order
fg = 200;       % filter cut-off frequency

% Butterworth Filter
[zb,pb,kb] = butter(n,fg,'s');  % filter coefficients
[bb,ab] = zp2tf(zb,pb,kb);      % transfer function
[hb,wb] = freqs(bb,ab,4096);    % frequency response

% Tschebyscheff filter
[z1,p1,k1] = cheby1(n,3,fg,'s');
[b1,a1] = zp2tf(z1,p1,k1);
```

```
[h1,w1] = freqs(b1,a1,4096);

% inverse Tschebyscheff filter
[z2,p2,k2] = cheby2(n,30,fg,'s');
[b2,a2] = zp2tf(z2,p2,k2);
[h2,w2] = freqs(b2,a2,4096);

% Cauer filter
[ze,pe,ke] = ellip(n,3,30,fg,'s');
[be,ae] = zp2tf(ze,pe,ke);
[he,we] = freqs(be,ae,4096);

% graphical representation
plot(wb,abs(hb))
hold on
plot(w1,abs(h1))
plot(w2,abs(h2))
plot(we,abs(he))
grid on
xlabel('f_g / Hz')
ylabel('|A(f_g)|')
legend('Butterworth','Tschebyscheff','inverse Tschebyscheff','Cauer')
```

4.5.1.1 General Procedure for Filter Design

When designing selective filters for the magnitude frequency response, specifications are only made for the magnitude frequency response $|A(f)|$. Requirements for the phase frequency response and the group delay are not defined. In biosignal processing, however, the group delay in the passband should be as flat as possible in order to avoid signal distortion.

To simplify the filter synthesis, first only normalised lowpass filters are designed, which have a dimensionless frequency normalised to a reference frequency f_B. Then, by frequency transformation, a filter can be created which meets the desired requirements in terms of passband and stopband. The values of the tolerance range of the magnitude transfer function remain unchanged. Only the cut-off frequencies are changed. The frequency transformation can be done by the denormalisation. Not only the reconversion from a normalised low pass to a non-normalised low pass, but also a conversion to a high pass, band pass or band stop can be done (cf. Figure 4.36). The conversion must take place in such a way that afterwards a transfer function $A(p)$ (realisable with analogue components) arises again. It can be represented in fractional-rational form with $p := \sigma + j\omega$ and $\omega := 2\pi f$ like

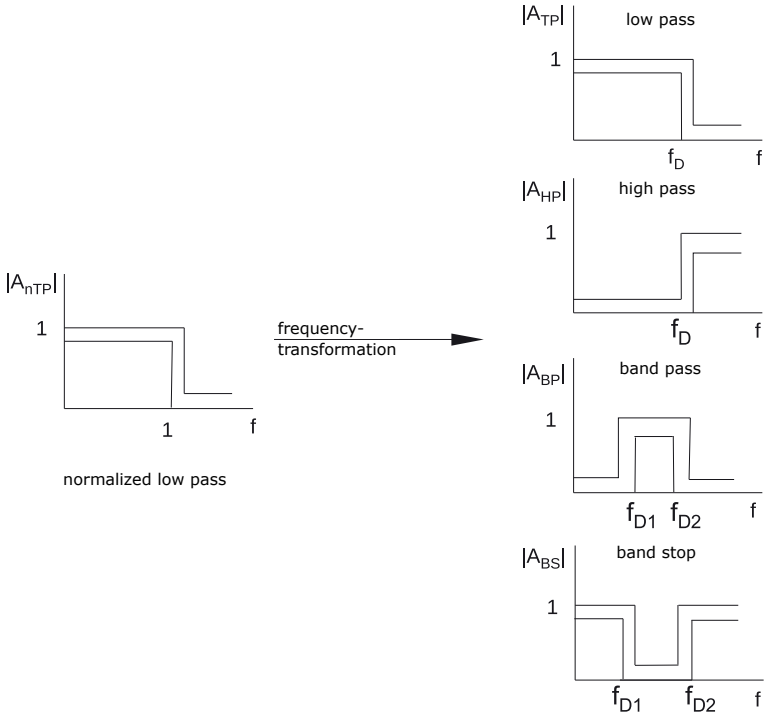


Fig. 4.36: Conversion of normalised low-pass to other standard filters (non-normalised low-, high-, band-pass and band-stop).

$$A(p) = A_0 \frac{1 + b_1 p + b_2 p^2 + \cdots + b_\mathrm{m} p^m}{1 + c_1 p + c_2 p^2 + \cdots + c_n p^n} \tag{4.16}$$

For example, if the impedance of a coil with $j\Omega L$ is replaced with $\Omega := 2\pi F$ by the impedance $\omega_\mathrm{B}/j\omega C$, the transfer function $A(j\omega)$ to Equation 4.16 remains fractionrational in p. This transformation can be achieved by replacing Ω with ω_B/ω and the inductance L with the capacitance $C = -1/L$. However, there is the difference that now, because of the $1/\omega$ function, the frequency axis is now subdivided differently, i.e. a low frequency is mapped to a high frequency and a high frequency to a low frequency. So a low pass becomes a high pass.

A summary of the frequency transformations from a normalised low-pass to a non-normalised low-pass (nTP $\Rightarrow$ TP), a normalised low-pass to a high-pass (nTP $\Rightarrow$ HP), a normalised low-pass into a band-pass (nTP $\Rightarrow$ BP), and a normalised low-pass into a bandstop (nTP $\Rightarrow$ BS) is shown in Table 4.5. Explanatory examples of this are covered in the following sections on power and Chebyshev-filters.

The design of a selective filter can therefore be limited to the design of a normalised low-pass filter. For this, however, only the conditions for the magnitude frequency

response of the normalised low pass with

$$|A_{\mathrm{nTP}}(j\Omega)|^2 = A_{\mathrm{nTP}}(j\Omega) \cdot A^*_{\mathrm{nTP}}(j\Omega) = A_{\mathrm{nTP}}(j\Omega) \cdot A_{\mathrm{nTP}}(-j\Omega) \tag{4.17}$$

and $\Omega := 2\pi F$ are available. In order to obtain a realisable stable transfer function $A_{\mathrm{nTP}}(P)$ analogous to Equation 4.16, this equation is first extended and the normalised angular frequency Ω is replaced by the normalised complex angular frequency $P := \Sigma + j\Omega$. Here the real part Σ is an attenuation part. If one now chooses for the transfer function $A_{\mathrm{nTP}}(P)$ not the polynomial form according to Equation 4.16, but the representation with the help of the poles and zeros (the polynomials of the numerator and denominator in Equation 4.16 can each be described by a product of their zeros), one further obtains:

$$\begin{aligned} A_{\mathrm{nTP}}(P) \cdot A_{\mathrm{nTP}}(-P) = A_0 \frac{(P-P_{n1}) \cdot (P-P_{n2}) \cdot \ldots \cdot (P-P_{nm})}{(P-P_{p1}) \cdot (P-P_{p2}) \cdot \ldots \cdot (P-P_{pn})} \\ \cdot A_0 \frac{(-P-P_{n1}) \cdot (-P-P_{n2}) \cdot \ldots \cdot (-P-P_{nm})}{(-P-P_{p1}) \cdot (-P-P_{p2}) \cdot \ldots \cdot (-P-P_{pn})} \,. \end{aligned} \tag{4.18}$$

The product $G_{\mathrm{nTP}}(P) := A_{\mathrm{nTP}}(P) \cdot A^*_{\mathrm{nTP}}(-P)$ contains twice as many poles and zeros as the complex normalised transfer function $A_{\mathrm{nTP}}(P)$, which is symmetric in respect to the imaginary axis $j\Omega$. This expansion of the magnitude square must therefore, given the magnitude of the normalised low-pass filter, be divided into a symmetrical product with the same number of poles and zeros in each case in such a way that a realisable transfer function $A(P)$ can be determined from it, which describes a *stable* filter. A filter is stable if its impulse response becomes smaller and smaller with time and tends towards zero. To achieve this, all poles must lie in the left P half plane. The zeros of $A(P)$, on the other hand, do not all have to lie in the left P- half plane. Therefore, there are also different subdivision possibilities in this respect. In a minimum phase system, which describes the minimum phase change as a function of the frequency, this is the

Tab. 4.5: Overview of the standard-frequency transformations: Let f_{B} denote the reference frequency, $f_0 := \sqrt{f_{\mathrm{D1}} f_{\mathrm{D2}}}$ the centre frequency, F_{D}, F_{D1} and F_{D2} the passband cut-off frequencies and F_{S}, F_{S1} and F_{S2} the stopband cut-off frequencies.

Transformation	transformation equation	cut-off frequencies
nTP ⇒ TP	$F = \frac{f}{f_{\mathrm{D}}}$	$f_{\mathrm{D}} = f_{\mathrm{B}}$
nTP ⇒ HP	$F = -\frac{f_{\mathrm{D}}}{f}$	$f_{\mathrm{D}} = f_{\mathrm{B}}$
nTP ⇒ BP	$F = \frac{f - f_0^2/f}{f_{\mathrm{B}} - f_0^2/f_{\mathrm{B}}}$	$f_{\mathrm{D1}} = \frac{f_o^2}{f_{\mathrm{B}}}$
	$f_0^2 = f_{\mathrm{D1}} f_{\mathrm{D2}}$	$f_{\mathrm{D2}} = f_{\mathrm{B}}$
nTP ⇒ BS	$F = -\frac{f_{\mathrm{B}} - f_0^2/f_{\mathrm{B}}}{f - f_0^2/f}$	$f_{\mathrm{D1}} = f_{\mathrm{B}}$
	$f_0^2 = f_{\mathrm{D1}} f_{\mathrm{D2}}$	$f_{\mathrm{D2}} = \frac{f_0^2}{f_{\mathrm{B}}}$

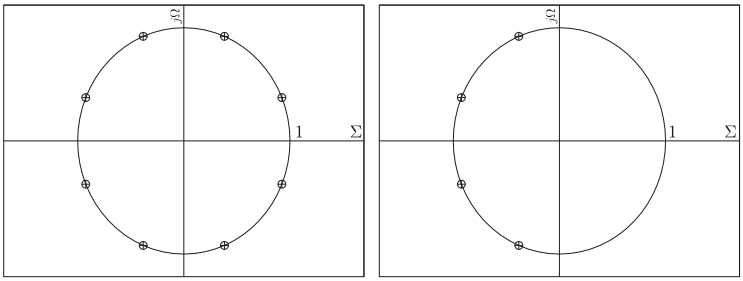


Fig. 4.37: Example of pole selection when the magnitude frequency response of a 4th order selective filter is specified: the pole locations of the extended magnitude frequency response function $A_{nTP}(P) \cdot A^*_{nTP}(-P)$ (left) and the pole locations of the transfer function $A_{nTP}(P)$ (right).

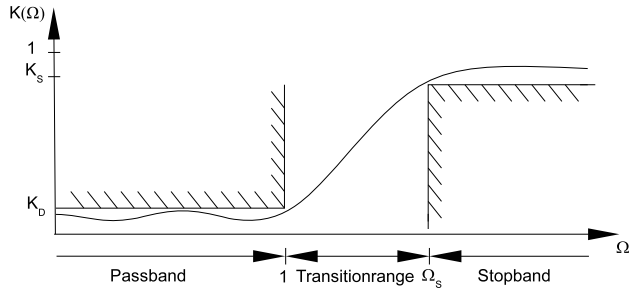


Fig. 4.38: Tolerance scheme of the characteristic function $K(\Omega)$ of a normalised low-pass filter.

case – i.e. all zeros lie in the left P- half plane. An example of the distribution of the pole positions for a 4th order filter is shown in Figure 4.37.

In the further synthesis of normalised low-pass filters, apart from normalisation, it is easier to express the magnitude of the transfer function $|A(j\Omega)|$ by the characteristic function $K(\Omega)$ defined by the following relation:

$$|A(j\Omega)|^2 = \frac{1}{1 + K(\Omega)^2} \,. \tag{4.19}$$

This makes it easier to describe the standard filters, since for these the characteristic function can be described by simple polynomials and the associated tolerance scheme can be simplified (cf. Figure 4.38).

4.5.1.2 Butterworth or Power Filter

In a Butterworth or power filter, only a factor ϵ weighted power of the normalised frequency Ω is chosen for the characteristic function $K(\Omega)$, i.e..

$$K(j\Omega) = \epsilon \cdot \Omega^n \,, \quad n\text{: filtering order} \,, \quad \epsilon\text{: attenuation constant} \tag{4.20}$$

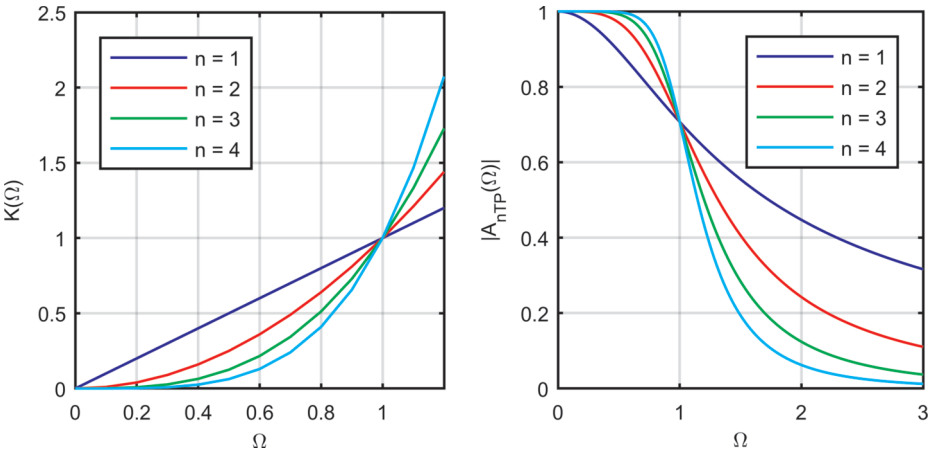


Fig. 4.39: Characteristic functions $K(\Omega) = \epsilon\Omega^n$ (left) and magnitudes of the transfer functions of the associated normalised low-pass filters up to 4th order for $\epsilon = 1$ (right).

and after Equation 4.19 results in:

$$|A(j\Omega)|^2 = \frac{1}{1 + \epsilon^2 \cdot \Omega^{2n}} \,. \tag{4.21}$$

At the normalised passband cut-off frequency $\Omega = 1$, the characteristic function $K(\Omega)$ has the value ϵ and is otherwise always smaller in the passband. The tolerance scheme must therefore have this maximum value. At $\epsilon = 1$ the magnitude of the transfer function $|A_{\mathrm{nTP}}(\Omega = 1)| = 1/\sqrt{2} \approx 0.707$. This corresponds to an attenuation $a = -20 \cdot \log(|A_{\mathrm{nTP}}|)$ of 3 dB (cf. Figure 4.39). If one now extends according to Equation 4.72 the normalised frequency Ω to the complex frequency $P := \Sigma + j\Omega$, then starting from $\Sigma = 0$ for the product $G_{\mathrm{nTP}}(P) := A_{\mathrm{nTP}}(P) \cdot A_{\mathrm{nTP}}(-P)$ we obtain:

$$G_{\mathrm{nTP}}(P) = \frac{1}{1 + K(P/j)^2} = \frac{1}{1 + \epsilon^2 \left(\frac{P}{j}\right)^{2n}} = \frac{1}{1 + \epsilon^2 \left(e^{-j\pi/2} P\right)^{2n}} \,. \tag{4.22}$$

Zeros of $G_{\mathrm{nTP}}(P)$ are not present. The pole places P_k of $G_{\mathrm{nTP}}(P)$ are at the zeros of the denominator, i.e..

$$1 + K(P_{\pm k}/j)^2 = 0 \tag{4.23}$$

respespectively

$$K(P_{\pm k}/j) = \pm j \tag{4.24}$$

and have the values (cf. also [68]):

$$P_{\pm k} = e^{j\pi/2} \left(\frac{-1}{\epsilon^2}\right)^{1/2n} = \frac{e^{j(\pi/2 \pm (\pi + k \cdot 2\pi)/2n)}}{\sqrt[n]{\epsilon}} \,, \quad k = 0, \ldots, n-1 \,. \tag{4.25}$$

Since the amount of these poles always have the same value of $1/\sqrt[n]{\epsilon}$, the poles all lie on a circle (cf. Figure 4.37, for $\epsilon = 1$). Now that the poles and zeros of $G_{\mathrm{nTP}}(P)$ are

known, they still have to be selected for $A_{\mathrm{nTP}}(P)$ according to Equation 4.18 in such a way that they all lie in the left P- half plane. Furthermore, the constant factor A_0 must still be determined that at the passband cut-off frequency $\Omega_{\mathrm{D}} = 1$ the magnitude square of the transfer function $|A_{\mathrm{nTP}}(\Omega = 1)|^2$ drops to the value

$$G_{\mathrm{nTP}}(j\Omega = 1) = |A_{\mathrm{nTP}}(j\Omega = 1)|^2 = \frac{1}{1 + \epsilon^2(\Omega = 1)^{2n}} = \frac{1}{1 + \epsilon^2} \tag{4.26}$$

Considering now the pole and zero position representation of the normalised transfer function $A_{\mathrm{nTP}}(j\Omega)$ analogous to the representation in Equation 4.18, where here there are no zeros and the poles after selection all lie in the left P- half plane,

$$|A_{\mathrm{nTP}}(j\Omega)| = \left| A_0 \frac{1}{(j\Omega - P_{p1}) \cdot (j\Omega - P_{p2}) \cdot \cdots \cdot (j\Omega - P_{pn})} \right|, \tag{4.27}$$

so follows Equation 4.26

$$|A_0| = \left| A_{\mathrm{nTP}}(j\Omega = 1) \cdot \prod_{l=0}^{n-1} (j - P_{pl}) \right| = \frac{1}{\sqrt{1 + \epsilon^2}} \left| \cdot \prod_{l=0}^{n-1} (j - P_{pl}) \right| . \tag{4.28}$$

Substituting the pole places after Equation 4.25 into Equation 4.28 gives:

$$A_0 = \frac{1}{\epsilon} . \tag{4.29}$$

Explanatory Examples

Lowpass 1st order

For filtering an ECG, a passive Butterworth-filter 1st order is to be designed with the aid of an RC element, which has a passband cut-off frequency of 200 Hz and at this an attenuation of 3 dB.

For this purpose, the corresponding normalised low-pass filter is first determined. Since at an attenuation of 3 dB the magnitude square of the frequency response has the value 0.5, it follows from Equation 4.26 that $\epsilon = 1$ must be. From Equation 4.25 it follows for the pole positions

$$P_{\pm p0} = \frac{e^{j(\pi/2 \pm (\pi + 0 \cdot 2\pi)/2 \cdot 1)}}{\sqrt[n]{1}} = \pm 1 \tag{4.30}$$

and for $|A_0|$, when all poles are in the left P- half plane, from Equation 4.29:

$$A_0 = \frac{1}{\epsilon} = 1. \tag{4.31}$$

So the poles are purely real and are at ± 1. The pole in the left P- half plane is therefore $P_0 = -1$ and is selected. The transfer function of the normalised low-pass can now be given:

$$A_{\mathrm{nTP}}(P) = A_0 \frac{1}{P - P_{p0}} = \frac{1}{P + 1}. \tag{4.32}$$

According to Table 4.5, in order to realise the desired low-pass filter with a cut-off frequency f_D of 200 Hz from the normalised low-pass filter, the normalised angular frequency $\Omega = 2\pi F$ must be replaced by a suitable frequency transformation. In this case this would be $F = \frac{f}{f_D}$ resp.

$$\Omega = \frac{\omega}{\omega_D}\,, \quad \text{mit } \omega_D = 2\pi \cdot 200\,\text{Hz}\,. \tag{4.33}$$

For the desired low-pass we then obtain the complex transfer function

$$A_{TP}(j\omega) = A_{nTP}\left(P = j\Omega = j\frac{\omega}{\omega_D}\right) = \frac{1}{\frac{j\omega}{2\pi\cdot 200\,\text{Hz}} + 1}\,. \tag{4.34}$$

A 1st order RC low pass can be realised by a simple voltage divider. The transfer functions of the calculated power low-pass and the RC element must be identical, i.e.

$$A_{TP}(j\omega) = \frac{1}{\frac{j\omega}{2\pi\cdot 200\,\text{Hz}} + 1} = A_{RC}(j\omega) = \frac{1}{1 + j\omega RC}\,. \tag{4.35}$$

From this follows the condition

$$RC = \frac{1}{\omega_D} = \frac{1}{2\pi \cdot 200\,\text{Hz}} = 795.8\,\mu\text{S}\,. \tag{4.36}$$

For example, if one chooses for $C = 1$ nF, one obtains for $R = 800\,\text{k}\Omega$. The RC circuit diagram with the frequency response calculated with LTspice and the corresponding frequency response is shown Figure 4.40.

Bandstop 2nd order

For the suppression of a mains hum interference signal during the measurement of an ECG, a passive Butterworth-bandstop 2nd order with a centre frequency f_0 of 50 Hz at

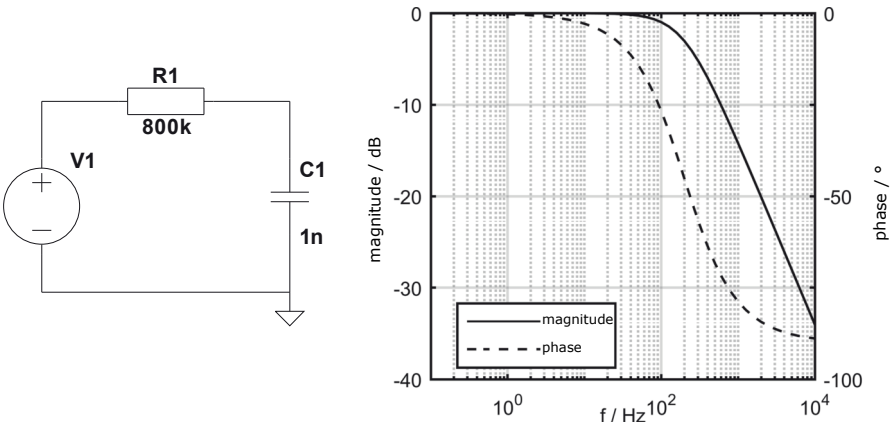


Fig. 4.40: Circuit for a 1st order RC power filter (left) and associated frequency response by magnitude and phase (right).

a bandwidth $\Delta f = f_{D2} - f_{D1}$ of 20 Hz is to be designed using an RLC element. For this purpose, the corresponding normalised low-pass filter from the previous example can be used (cf. Equation 4.32) and then the low-pass-bandstop-transformation according to Table 4.5 can be performed. This results in:

$$A_{nTP}(P) = \frac{1}{P+1} \quad \text{bzw.} \quad A_{nTP}(j\Omega) = \frac{1}{j\Omega + 1} \,. \tag{4.37}$$

If the frequencies are expressed in terms of angular frequencies, the associated frequency transformation on denormalisation gives:

$$\Omega = -\frac{\omega_B - \omega_0^2/\omega_B}{\omega - \omega_0^2/\omega} \,, \quad \text{mit } \omega_B = \omega_{D1} \text{ und } \frac{\omega_0^2}{\omega_B} = \omega_{D2} \,. \tag{4.38}$$

Substituting Equation 4.38 into Equation 4.37, it follows for the bandstop transfer function:

$$A_{BS}(j\omega) = \frac{\omega^2 - \omega_0^2}{\omega^2 - \omega_0^2 - j\omega \underbrace{(\omega_B - \omega_0^2/\omega_B)}_{\Delta\omega}} \,. \tag{4.39}$$

This transfer function must be identical to that of an RLC element, which is obtained by applying the equation for a voltage divider (cf. Figure 4.41,left) it follows for

$$A_{RLC}(j\omega) = \frac{U_2}{U_1} = \frac{\omega^2 - 1/LC}{\omega^2 - 1/LC - j\omega R/L} \,. \tag{4.40}$$

Since both transfer functions must be identical, it follows by comparison:

$$A_{BS}(j\omega) = \frac{\omega^2 - \omega_0^2}{\omega^2 - \omega_0^2 - j\omega \underbrace{(\omega_B - \omega_0^2/\omega_B)}_{\Delta\omega}} = A_{RLC}(j\omega) = \frac{\omega^2 - 1/LC}{\omega^2 - 1/LC - j\omega R/L} \tag{4.41}$$

and hence:

$$\omega_0^2 = \frac{1}{LC} \quad \text{und} \quad \Delta\omega = \frac{R}{L} \,. \tag{4.42}$$

Because of the relationship $\omega_0^2 = \omega_{D1}\omega_{D2} = (2\pi \cdot 50\,\text{Hz})^2$ and $\omega_B - .\omega_0^2/\omega_B = \omega_{D1} - \omega_{D2} = \Delta\omega = 2\pi{\cdot}20$ Hz one obtains e.g. with capacitance choice of the capacitor C of 100 µF:

$$\begin{aligned} L &= \omega_0^2 \cdot C = 101.23\,\text{mH} \\ R &= \Delta\omega \cdot L = 12.73\,\Omega \,. \end{aligned} \tag{4.43}$$

The corresponding frequency response calculated with LTspice according to magnitude and phase is shown in Figure 4.41 (right). The LTspice-simulation of a filtered ECG signal, which was strongly disturbed by a 50 Hz mains hum signal, is shown in Figure 4.42.

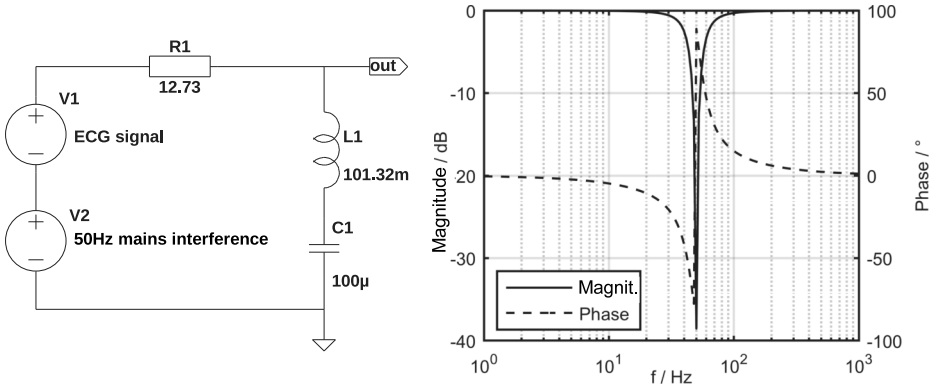


Fig. 4.41: RLC-Potency-Filter-Bandstop 2. orders (left) and associated frequency response (calculation with LTspice, right).

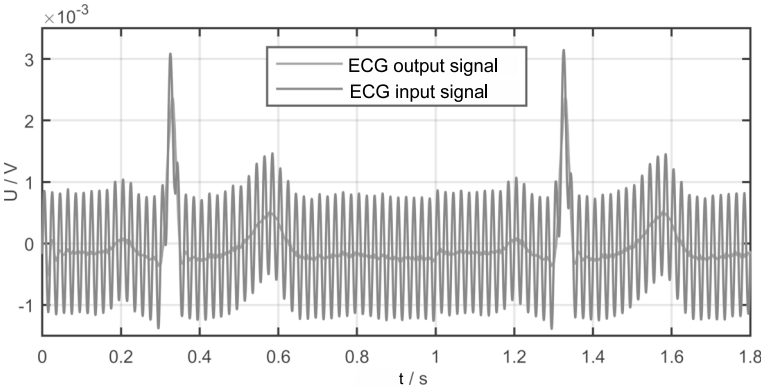


Fig. 4.42: ECG with 50 Hz mains hum before and after filtering with a potency-band block 2nd order.

4.5.1.3 Chebyshev Filter

In a Chebyshev-filter, for the characteristic function $K(\Omega)$ a Chebyshev-polynomial indexChebyshev-polynomial $T_n(\Omega)$ weighted by a factor ϵ is used.

$$K(\Omega) = \epsilon \cdot T_n(\Omega) \tag{4.44}$$

is chosen, with

$$T_n(\Omega) = \cos(n \cdot \arccos(\Omega)) \quad \text{for } |\Omega| < 1 \; T_n(\Omega) = \cosh(n \cdot \operatorname{arccosh}(\Omega)) \quad \text{for } |\Omega| > 1 \,. \tag{4.45}$$

The value n in turn determines the order of filtering. Examples for $T_n(\Omega)$ are:

$$\begin{aligned} &T_1(\Omega) = \Omega \;\; \text{(like power filter)} \,, \\ &T_2(\Omega) = -1 + 2\,\Omega^2 \,, && T_3(\Omega) = -3\,\Omega + 4\,\Omega^3 \,, \\ &T_4(\Omega) = 1 - 8\,\Omega^2 + 8\,\Omega^4 \,, && T_5(\Omega) = 5\,\Omega - 20\,\Omega^3 + 16\,\Omega^5 \,. \end{aligned} \tag{4.46}$$

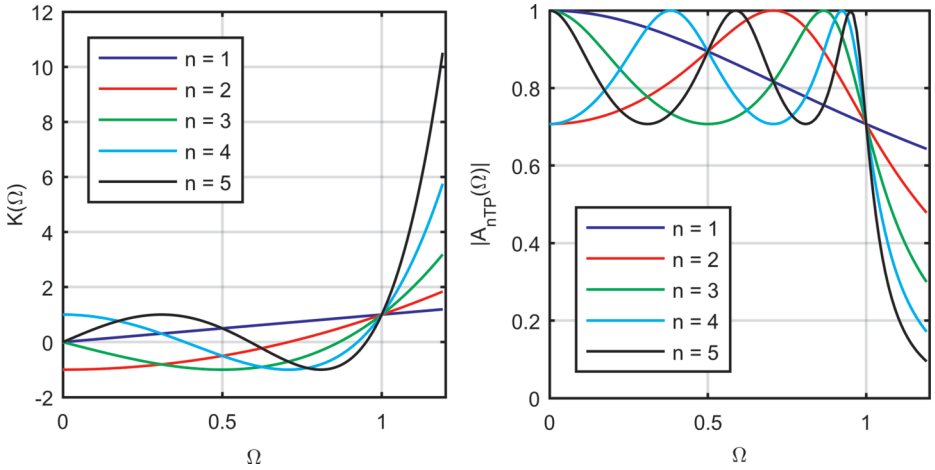


Fig. 4.43: Characteristic functions $K(j\Omega) = \cos(n \cdot \arccos(\Omega))$ of a Chebyshev-filter (left) and magnitudes of the transfer functions of the associated normalised low-pass filters up to the 5th order for $\epsilon = 1$ (right).

The advantage of this choice of approximation of the ideal normalised low-pass filter is that the transition between passband and stopband becomes steeper than with a power filter, allowing a wider bandwidth for the useful signal at a prescribed stopband frequency, as with an antialiasing-filter to satisfy the sampling theorem (cf. Figure 4.43). If one now extends the normalised frequency Ω to the complex frequency $P := \Sigma + j\Omega$ again one obtains, starting from $\Sigma = 0$ for the product

$$G_{\mathrm{nTP}}(P) := A_{\mathrm{nTP}}(P) \cdot A_{\mathrm{nTP}}(-P) :$$
$$G_{\mathrm{nTP}}(P) = \frac{1}{1 + K(P/j)^2} = \frac{1}{1 + \epsilon^2 T_n^2(P/j)} . \tag{4.47}$$

Zeros of $G_{\mathrm{nTP}}(P)$ are again none. The polar places P_k of $G_{\mathrm{nTP}}(P)$ lie at the zeros of the denominator and can be determined by Equation 4.24:

$$K(P_{\pm k}/j) = \epsilon \cdot T_n(P_{\pm k}/j) = \pm j . \tag{4.48}$$

For the derivation of the equation for the pole points, it should first be assumed that the amounts of the pole points are smaller than 1 (which must be checked afterwards). In this case $|\Omega| < 1$ can also be assumed, and one obtains from Equation 4.48 and Equation 4.45:

$$T_n(P_{\pm k}/j) = \cos(n \cdot \arccos(P_{\pm k}/j)) = \pm j/\epsilon . \tag{4.49}$$

Furthermore, in order to improve clarity, the parameter representation of the Chebyshev-polynomials is given with

$$t_{\pm k} = \arccos(P_{\pm k}/j) \text{ or} P_{\pm k} = j \cdot \cos t_{\pm k} \tag{4.50}$$

is chosen. Since in general the pole locations $P_{\pm k}$ are complex, $t_{\pm k}$ is also assumed to be complex, i.e. $t_{\pm k} = t_{R_{\pm k}} + jt_{I_{\pm k}}$. Inserted in Equation 4.50, it follows:

$$\begin{aligned} P_{\pm k} &= \Sigma_{\pm k} + j\Omega_{\pm k} = j \cdot \cos(t_{R_{\pm k}} + jt_{I_{\pm k}}) \\ &= j \cdot \cos t_{R_{\pm 1}} \cos(jt_{I_{\pm k}}) - j \cdot \sin t_{R_{\pm 1}} \sin(jt_{I_{\pm k}}) \\ &= \sin t_{R_{\pm 1}} \sinh t_{I_{\pm k}} + j \cdot \cos t_{R_{\pm 1}} \cosh(jt_{I_{\pm k}}) \,. \end{aligned} \tag{4.51}$$

The pole locations $P_{\pm k}$ can thus be calculated from $t_{\pm k}$ and Equation 4.51. If the parameter $t_{\pm k}$ is substituted into Equation 4.49, it follows:

$$\begin{aligned} T_n(P_{\pm k}/j) &= \cos(n \cdot (t_{R_{\pm k}} + jt_{I_{\pm k}})) \\ &= \cos(nt_{R_{\pm k}}) \cdot \cosh(nt_{I_{\pm k}}) - j \cdot \sin(nt_{R_{\pm k}}) \cdot \sinh(nt_{I_{\pm k}}) \\ &= \pm j/\epsilon \,. \end{aligned} \tag{4.52}$$

This gives rise to two conditions:

$$\begin{aligned} &1.\ \cos(nt_{R_{\pm k}}) \cdot \cosh(nt_{I_{\pm k}}) = 0 \,, \\ &2.\ \sin(nt_{R_{\pm k}}) \cdot \sinh(nt_{I_{\pm k}}) = \mp 1/\epsilon \,. \end{aligned} \tag{4.53}$$

From these conditions it follows:

$$\begin{aligned} n \cdot t_{R_{\pm k}} = (2k+1) \cdot \pi/2 \quad \text{and thus} \quad \sin(nt_{R_{\pm k}}) = 1 \quad \text{and} \\ t_{I_{\pm k}} = -\frac{1}{n} \operatorname{arcsinh}(1/\epsilon) \,. \end{aligned} \tag{4.54}$$

This can now be used to calculate the pole positions from Equation 4.51 , and Equation 4.54 can be calculated:

$$\begin{aligned} &P_{\pm k} = \Sigma_{\pm k} + j\Omega_{\pm k} \qquad \text{with} \\ &\Sigma_{\pm k} = \pm \sin(\pi(2k+1)/2n) \cdot \sinh(t_{I_{\pm k}}) \\ &\Omega_{\pm k} = \cos(\pi(2k+1)/2n) \cdot \cosh(t_{I_{\pm k}}) \\ &t_{I_{\pm k}} = -\frac{1}{n} \operatorname{arcsinh}(1/\epsilon) \,. \end{aligned} \tag{4.55}$$

If the real part is divided by $\sinh(t_{I_{\pm k}})$ and the imaginary part of the pole by $\cosh(t_{I_{\pm k}})$, and if the squares of each are added, it follows:

$$\left(\frac{\Sigma_{\pm k}}{\cosh(t_{I_{\pm k}})}\right)^2 + \left(\frac{\Omega_{\pm k}}{\sinh(t_{I_{\pm k}})}\right)^2 = 1 \,. \tag{4.56}$$

This is the equation of an ellipse, i.e. the pole locations of a Chebyshev-filter all lie on an ellipse, since their semi-axes $\sinh(t_{I_{\pm k}})$ and $\cosh(t_{I_{\pm k}})$ are independent of k and depend only on the filter order n and the damping parameter ϵ. They are also all smaller than unity in magnitude, so our assumption above about using the Chebyshev-polynomials for Equation 4.49 was correct. Substituting the polynomials for Equation 4.55 into Equation 4.28, it follows for the constant factor A_0:

$$A_0 = \frac{1}{\epsilon \cdot 2^{n-1}} \,. \tag{4.57}$$

Explanatory Example

For ECG filtering, design a passive Chebyshev- 2nd order filter using an RLC element which has a passband cut-off frequency of 200 Hz and at this an attenuation of 3 dB.

To do this, the corresponding normalised low-pass is first determined again. Since at an attenuation of 3 dB the magnitude square of the frequency response has the value of 0.5, it follows from Equation 4.26 that here $\epsilon = 1$ must be. From Equation 4.45 it follows for the poles:

$$\begin{aligned} P_{\pm 0} &= \pm \sin(\pi/4) \cdot \sinh(t_{I_{\pm 0}}) + j \cos(\pi/4) \cdot \cosh(t_{I_{\pm 0}}) \\ P_{\pm 1} &= \pm \sin(3\pi/4) \cdot \sinh(t_{I_{\pm 1}}) + j \cos(3\pi/4) \cdot \cosh(t_{I_{\pm 1}}) \\ t_{I_{\pm 0}} &= t_{I_{\pm 1}} = -0.5 \operatorname{arcsinh}(1) = -0.449687 \end{aligned} \tag{4.58}$$

resp.

$$\begin{aligned} P_{\pm 0} &= \mp 0.321787 + j0.776887 \\ P_{\pm 1} &= \mp 0.321787 - j0.776887 \end{aligned} \tag{4.59}$$

and for $|A_0|$, when all poles are in the left P- half plane, from Equation 4.57:

$$A_0 = \frac{1}{\epsilon \cdot 2^{n-1}} = \frac{1}{1 \cdot 2^{2-1}} = \frac{1}{2} \,. \tag{4.60}$$

Thus the poles are conjugate complex. Those lying in the left P half plane are selected, i.e. $P_{+0} = -0.3217871 + j0.776887$ and $P_{+1} = -0.3217871 - j0.776887$. The transfer function of the normalised low-pass filter can now be given:

$$A_{\mathrm{nTP}}(P) = A_0 \frac{1}{(P - P_{+0})(P - P_{+1})} \,. \tag{4.61}$$

In order to realise the desired low pass with a cut-off frequency f_{D} of 200 Hz from the normalised low pass, the normalised angular frequency $\Omega = 2\pi F$ must be replaced by a suitable frequency transformation according to Table 4.5. In this case we get $F = f/f_{\mathrm{D}}$ resp.

$$\Omega = \frac{\omega}{\omega_{\mathrm{D}}} \,, \quad \text{mit } \omega_{\mathrm{D}} = 2\pi \cdot 200\,\mathrm{Hz} \,. \tag{4.62}$$

For the desired low-pass we obtain the complex transfer function

$$A_{\mathrm{nTP}}\left(P = j\Omega = j\frac{\omega}{\omega_{\mathrm{D}}}\right) = \frac{1/\sqrt{2}}{1 + 0{,}9101795\frac{j\omega}{\omega_{\mathrm{D}}} + 1.4142137\left(\frac{j\omega}{\omega_{\mathrm{D}}}\right)^2} \,. \tag{4.63}$$

A 2nd order RLC low pass can be realised by a simple voltage divider. The transfer functions of the calculated Chebyshev-lowpass and the RLC element must be identical for this. This is possible if, for example, the damping factor $1/\sqrt{2}$ is not taken into account, i.e.

$$\begin{aligned} A'_{\mathrm{TP}}(j\omega) &:= A_{\mathrm{TP}}(j\omega) * \sqrt{2} = \frac{1}{1 + 0.9101795\frac{j\omega}{\omega_{\mathrm{D}}} + 1.4142137\left(\frac{j\omega}{\omega_{\mathrm{D}}}\right)^2} \\ &= A_{\mathrm{RLC}}(j\omega) = \frac{1}{1 + j\omega RC + (j\omega)^2 LC} \,. \end{aligned} \tag{4.64}$$

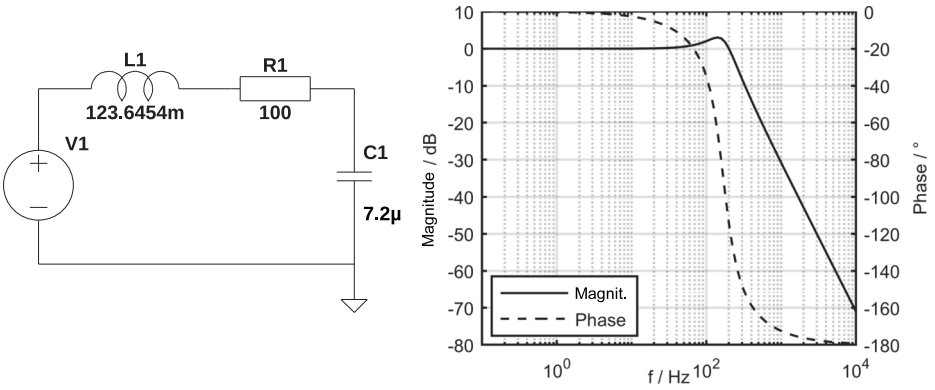


Fig. 4.44: Circuit of a 2nd order RLC Chebyshev-filter with a cut-off frequency of 200 Hz (left) and an associated frequency response by magnitude and phase (right).

From this we obtain the following conditions:

$$\begin{aligned} RC &= 0.9101795/\omega_\mathrm{D}\,, \\ LC &= 1.4142137/\omega_\mathrm{D}^2\,. \end{aligned} \tag{4.65}$$

Choosing $R = 100\,\Omega$, we get $\omega_\mathrm{D} = 2\pi \cdot 200\,\mathrm{Hz}$ for the inductance $L = 123.6454\,\mathrm{mH}$ and for the capacitance $C = 7.2\,\mu\mathrm{F}$. The RLC circuit diagram with the frequency response calculated with LTspice and the corresponding frequency response is shown in Figure 4.44.

Of course, this filter can also be realised actively with the help of an operational amplifier, whereby the constant factor A_0 can be realised exactly. In addition, this filter synthesis can also be done without coils, which makes it cheaper to produce. If one chooses an active filter according to the „Sallen Key"-structure [76], then it follows for its transfer function $\widetilde{A}_\mathrm{TP}(j\omega)$:

$$\widetilde{A}_\mathrm{TP}(j\omega) = \frac{1}{1 + j\omega(R_1 + R_2)C_1 + (j\omega)^2 R_1 R_2 C_1 C_2}\,. \tag{4.66}$$

A comparison with Equation 4.64 gives:

$$(R_1 + R_2) \cdot C_1 = 0.9101795/\omega_\mathrm{D} \quad \text{and} \quad R_1 R_2 C_1 C_2 = 1.4142137/\omega_\mathrm{D}^2\,.$$

For example, choosing $C_1 = 47\,\mu\mathrm{F}$ and $C_2 = 0.33\,\mu\mathrm{F}$ gives $\omega_\mathrm{D} = 2\pi \cdot 200\,\mathrm{Hz}$.

$$R_1 = 6.34\,\mathrm{k\Omega} \quad \text{und} \quad R_2 = 9.09\,\mathrm{k\Omega}\,. \tag{4.67}$$

The circuit with associated transfer function is shown in Figure 4.45.

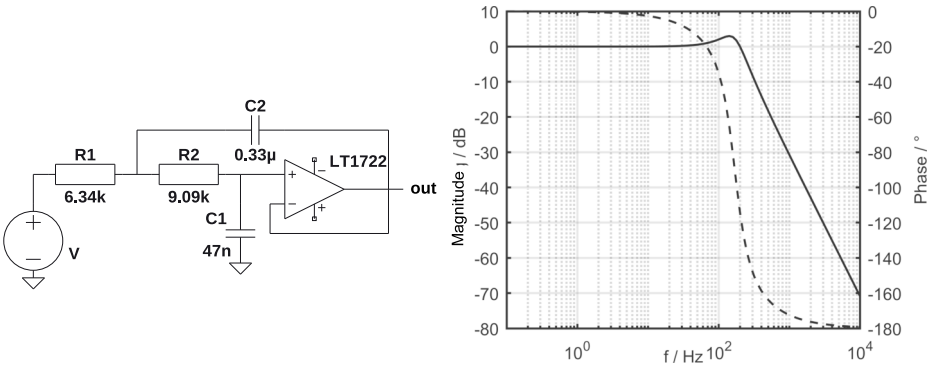


Fig. 4.45: Circuit of an active RC-Chebyscheff-filter of 2nd order with a structure according to „Sallen Key"(left) and associated frequency response according to magnitude and phase (right).

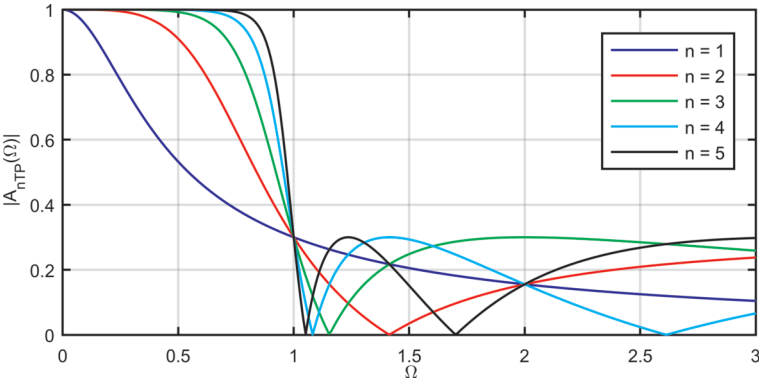


Fig. 4.46: Magnitude frequency response of inverse Chebyshev filters with a normalised cut-off frequency of $\Omega_S = 2$ and $\epsilon = 3.18$.

4.5.1.4 Inverse Chebyshev Filter

With an inverse Chebyshev filter, the tolerance range is not approximated in the passband, but in the stopband. For this purpose, a Chebyshev polynomial is again taken for the characteristic function, but this time with a different argument:

$$K(\Omega) = \epsilon \cdot \frac{1}{T_n\left(\frac{\Omega_S}{\Omega}\right)}, \quad \text{with} \quad \Omega_S\text{: blocking cut-off frequency .} \tag{4.68}$$

However, according to Equation 4.68, the magnitude square of the normalised transfer function $|A_{nTP}(j\Omega)|^2 = 1/(1 + K(\Omega)^2)$ does not have the value $1/(1 + \epsilon)$ at the normalised passband frequency $\Omega = 1$ as before. This value is only reached at the normalised blocking frequency Ω_S. At this frequency , however, a much smaller value of the magnitude of the transfer function should normally be achieved. Therefore ϵ must be chosen correspondingly larger. For example, if the magnitude of the transfer function at the normalised blocking frequency Ω_S is to have the magnitude of 0.3, $\epsilon = 3.18$

must be (cf. Figure 4.46). The pole positions P_k of $G_{\mathrm{nTP}}(P)$ lie at the zeros of the denominator and can be determined with Equation 4.24:

$$K(P_{\pm k}/j) = \epsilon \cdot \frac{1}{T_n\left(\frac{\Omega_S}{\Omega}\right)} = \pm j \,. \tag{4.69}$$

These are not to be derived this time, but as in in [70] only to be stated:

$$\begin{aligned} P_{\pm k} &= \Sigma_{\pm k} + j\Omega_{\pm k} && \text{with} \\ \Sigma_{\pm k} &= \pm\frac{\Omega_S}{Ne}\sin(\pi(2k+1)/2n)\cdot\sinh(\widetilde{t}_{I_{\pm k}}) \,, \\ \Omega_{\pm k} &= \frac{\Omega_S}{Ne}\cos(\pi(2k+1)/2n)\cdot\cosh(\widetilde{t}_{I_{\pm k}}) \,, \\ \widetilde{t}_{I_{\pm k}} &= \frac{1}{n}\operatorname{arcsinh}(\epsilon) \,, \\ Ne &= \left|\cos(\pi(2k+1)/2n) + j\widetilde{t}_{I_{\pm k}}\right|^2 \,. \end{aligned} \tag{4.70}$$

These poles also lie on an ellipse. Zeros, unlike the power and Chebyshev-filter, are present this time and are imaginary. According to [70] they lie at

$$P_{0l} = j\frac{\Omega_S}{\cos\left(\frac{2l-1}{2n}\right)} \,. \tag{4.71}$$

However, to calculate the constant factor A_0 of the transfer function $A_{\mathrm{nTP}}(j\Omega)$ to Equation 4.28, we now have to consider the zeros and obtain

$$|A_0| = \frac{1}{\sqrt{1+\epsilon^2}} \cdot \left|\frac{\prod_{l=0}^{n-1}(j - P_{pl})}{\prod_{l=0}^{m-1}(j - P_{nl})}\right| \,. \tag{4.72}$$

If these are now substituted into this equation for the poles and zeros (cf. Equation 4.70 and Equation 4.71), it follows:
at *even* filter order n:

$$A_0 = \frac{1}{\sqrt{1+\epsilon^2}}, \tag{4.73}$$

at *odd* filter order n:

$$A_0 = (-1)^{(n-1)/2} \cdot \frac{n \cdot \Omega_S}{\epsilon} \,. \tag{4.74}$$

The rest of the filter design is the same as for the Chebyshev-filter. This means the application of suitable frequency transformations, circuit selection (passive or active realisation), determination of the components and a final analysis by simulation.

4.5.1.5 Cauer Filter

With the Cauer filter, both the passband and the stopband are optimally approximated. The transition between passband and stopband is steepest here and thus the distance between passband and stopband cut-off frequency is smallest. This approximation

comes closest to the ideal low-pass filter in terms of its selectivity. However, such filters are most sensitive to parameter fluctuations of the components, so that even small changes can cause a completely different filter behaviour or instability.

For the characteristic function $K(j\Omega)$ a factor ϵ weighted rational elliptic function $R_n(\kappa, \Omega)$ of order n is chosen [70]:

$$K(\Omega) = \epsilon \cdot R_n(\kappa, \Omega)\,, \quad \kappa\colon \text{ module (sensitivity measure)}\,. \tag{4.75}$$

The rational elliptic function $R_n(\kappa, \omega)$ has the property that at the reciprocal of the angular frequency Ω the function is also transformed into its reciprocal, i.e. $R_n(\kappa, 1/\Omega) = 1/R_n(\kappa, \Omega)$. Thus, at the normalised passband frequency $\Omega_\mathrm{D} = 1$, one achieves an optimal approximation of the ideal lowpass in the passband as well as in the stopband. For *even* filter order n is

$$R_n(\kappa, \Omega) = \kappa^{n/2} \cdot \prod_{k=1}^{n/2} \frac{\Omega^2 - \Omega_{0k}}{\Omega^2\kappa^2\Omega_{0k}^2 - 1}\,, \tag{4.76}$$

and for *odd* filter order n is

$$R_n(\kappa, \Omega) = (\sqrt{\kappa})^n \cdot \prod_{k=1}^{(n-1)/2} \frac{\Omega^2 - \Omega_{0k}}{\Omega^2\kappa^2\Omega_{0k}^2 - 1}\,. \tag{4.77}$$

Where

$$\Omega_{0k} = \mathrm{sn}\left(\frac{n-2k+1}{n}K_0\right)\,, \quad k = 1 \ldots \left[\frac{n}{2}\right]\,, \tag{4.78}$$

where $\left[\frac{n}{2}\right]$ is said to be the smallest integer less than $\frac{n}{2}$. sn is the Jacobian elliptic function with modulus $\kappa < 1$, and K_0 is the complete elliptic integral of the first kind also with modulus κ, which is a measure of the cutoff frequency:

$$K_0 := \int \frac{\mathrm{d}x}{\sqrt{(1-x^2)\cdot(1-\kappa^2x^2)}},$$
$$\Omega_\mathrm{S} = \frac{1}{\kappa}\,.$$

An example of the magnitude frequency response of a Cauer filter for filter grades 1 to 5 is shown Figure 4.47. There is a ripple of 0.3 in both the passband and the stopband. Betragsfrequenzgang-Cauer The poles of the transfer function A(P) result in (cf. [74]):

$$P_k = \frac{\sigma_0 F_{0k} \pm j\Omega_{0k}F_0}{1 + \kappa^2\sigma_0\Omega_{0k}^2}\,, \quad k = 1 \ldots \left[\frac{n+1}{2}\right]\,, \tag{4.79}$$

$$\sigma_0 = \frac{\mathrm{sn}(u_0, \kappa')}{\mathrm{cn}(u_0, \kappa')}\,, \quad F_0 = \frac{\mathrm{dn}(u_0, \kappa')}{\mathrm{cn}^2(u_0, \kappa')}\,, \; F_{0k} = \mathrm{cn}(u_{0k}, \kappa)\cdot\mathrm{dn}(u_{0k}, \kappa)\,. \tag{4.80}$$

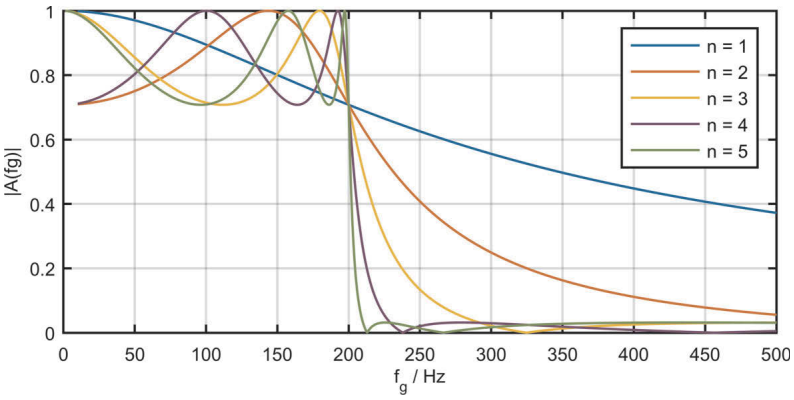


Fig. 4.47: Amount-frequency response of a Cauer filter for filter grades 1 to 5 with a normalised stop-band frequency $\Omega_D = 1$ and a ripple of 0.3.

The still unknown quantities κ', cn, dn, u_0 and u_{0k} in Equation 4.80 have the following meaning:

$$\begin{aligned} \kappa' &= \sqrt{1-\kappa^2}\,, & \mathrm{cn} &= \sqrt{1-\mathrm{sn}^2}\,, \\ \mathrm{dn} &= \sqrt{1-\kappa^2\mathrm{sn}^2}\,, & u_{0k} &= \frac{n-2k+1}{n}K_0\,, \\ u_0 &= \frac{K_0}{n\cdot K_\Delta}\mathrm{sn}^{-1}\left(\frac{1}{\sqrt{1+\epsilon^2\Delta^2}}, k_\Delta\right), & k_\Delta &= \sqrt{1-\Delta}\,. \end{aligned} \tag{4.81}$$

K_Δ is the complete elliptic integral with modulus Δ. The maximum change of the characteristic function $K(j\Omega)$ describes $\epsilon\cdot\Delta$. Δ can now be further determined by the following relations: For *even* filter order n is

$$\Delta = \kappa^{n/2}\prod_{j=1}^{n/2}\Omega_{0j}^2 \tag{4.82}$$

and for *odd* filter order n is

$$\Delta = \left|(\sqrt{\kappa})^n\prod_{j=1}^{(n-1)/2}\frac{1-\Omega_{0j}^2}{\kappa^2\Omega_{0j}^2-1}\right|. \tag{4.83}$$

There are also zeros in the transfer function $A_{\mathrm{nTP}}(P)$ for the Cauer filters. As with the inverse Chebyshev-filters, they all lie on the imaginary axis at

$$P_{0k} = \pm j\frac{1}{\kappa\Omega_{0k}}\,, \quad k = 1, 2, \ldots, \left[\frac{n}{2}\right]. \tag{4.84}$$

To calculate the constant factor A_0 of the transfer function $A_{\mathrm{nTP}}(j\Omega)$ to Equation 4.28, the zeros must now be considered again. From this follows (without derivation):

at *even* filter order n.

$$A_0 = \frac{1}{\sqrt{1 + \frac{\epsilon^2}{\kappa^{n/2} \prod_{j=1}^{n/2} \Omega_{0j}^2}}} , \tag{4.85}$$

at *odd* filtering order n.

$$A_0 = \frac{1}{\epsilon} \left(\sqrt{\kappa}\right)^{n-2} \prod_{j=1}^{(n-1)/2} \Omega_{0j}^2 . \tag{4.86}$$

The further filter design is also done here as for the Chebyshev-filter, i.e. application of suitable frequency transformation, circuit selection, determination of the components and final analysis by simulation.

4.5.2 Selective filters with Group Delay Optimisation

In the previous section, the filters were optimised according to specifications for the magnitude frequency response. However, for some applications, e.g. ECG measurement, it is also important that the waveform of the measurement signal is preserved. To achieve this, all frequency components of the wanted signal must be passed on from the input to the output with the same speed. In this case, the group delay of the filter must have as constant value as possible in the passband. This is possible both by non-recursive digital filtering, which will be discussed in a later section, and by an analogue filter before the analogue-to-digital converter. The Bessel filter will be presented as an example.

4.5.2.1 Bessel Filter

In the Bessel filter, an ideal delay element is first approximated with the normalised delay time T and the transfer function of the normalised low-pass according to $A_{\mathrm{nTP}}(\Omega) = e^{-PT} = 1/e^{PT}$ and then the cut-off frequency is adjusted according to the attenuation specifications. However, since this ideal transfer function is not fractionally rational in P, it cannot be realised directly by analogue components. For this reason, the e- function is approximated in a series with fractional rational members in P, e.g. by a Taylor series n-th order:

$$e^{PT} \approx 1 + \frac{P}{1!} + \frac{P^2}{2!} + \cdots + \frac{P^n}{n!} . \tag{4.87}$$

The disadvantage is that the resulting polynomial does not always have zeros in the left P- half plane. The system can therefore become unstable. The remedy is a Hurwitz-polynomial. Its zeros lie *always* in the left P- half plane [51]. This Hurwitz polynomial must be a Bessel function $B_n(P)$ (hence its name), i.e.

$$A_{\mathrm{nTP}}(P) \approx \frac{B_n(0)}{B_n(PT)} , \quad n\text{: polynomial order} \tag{4.88}$$

$$B_n(P) := \sum_{i=0}^{n} \frac{(2n-i)!P^i}{2^{n-i}i!(n-i)!}\,, \quad i = 0, 1, \ldots, n\,. \tag{4.89}$$

Examples are:

$$\begin{aligned}
B_1(P) &= 1 + P \\
B_2(P) &= 3 + 3P + P^2 \\
B_3(P) &= 15 + 15P + 6P^2 + P^3 \\
B_4(P) &= 105 + 105P + 45P^2 + 10P^3 + P^4 \\
B_5(P) &= 945 + 945P + 420P^2 + 105P^3 + 15P^4 + P^5\,.
\end{aligned}$$

Explanatory Example

For ECG filtering, design a passive Bessel filter of 2nd order using an RLC element which has a passband cut-off frequency of 200 Hz and at this an attenuation of 3 dB. In this case the Bessel polynomial of 2nd order is $B_2(P) = 3 + 3P + P^2$. The delay time T is chosen such that at the normalised passband cut-off frequency $\Omega_D = 1$ the attenuation of 3 dB can be maintained, i.e.

$$|A_{nTP}(P = j\Omega_D)| = |A_{nTP}(P = j)| = \frac{3}{|3 + j3T + (jT)^2|} = \frac{1}{\sqrt{2}}\,. \tag{4.90}$$

One thereby obtains $T = 1.3823$ s. If we now denormalise so that the passband cut-off frequency is $f_D = 200$ Hz and compare this with the transfer function $A_{RLC}(j\omega)$ of an RLC element one obtains because of

$$\begin{aligned}
A(j\omega)_{Bessel} &= A_{nTP}\left(P = j\frac{\omega}{\omega_D}\right) = \frac{3}{3 + 3j\frac{\omega}{\omega_D}T + \left(j\frac{\omega}{\omega_D}T\right)^2} \\
&= A_{RLC}(j\omega) = \frac{1}{1 + j\omega RC + (j\omega)^2 LC}
\end{aligned} \tag{4.91}$$

the relation:

$$RC = T/\omega_D \quad \text{und} \quad LC = \frac{1}{3}\,(T/\omega_D)^2\,. \tag{4.92}$$

Choosing $R = 100\,\Omega$, it follows with the forward frequency $f_D = 200$ Hz and $T = 1.3823$ s for the inductance $L = 36.7$ mH and for the capacitance $C = 11\,\mu$F. The phase response in the passband is linear (cf. Figure 4.48).

4.6 Post-Reading and Exercises

Measurement of Electrical Biosignals

1. Why is analog signal processing needed in the digital age?
2. What would be the skin impedance for a infinitely high frequency current?

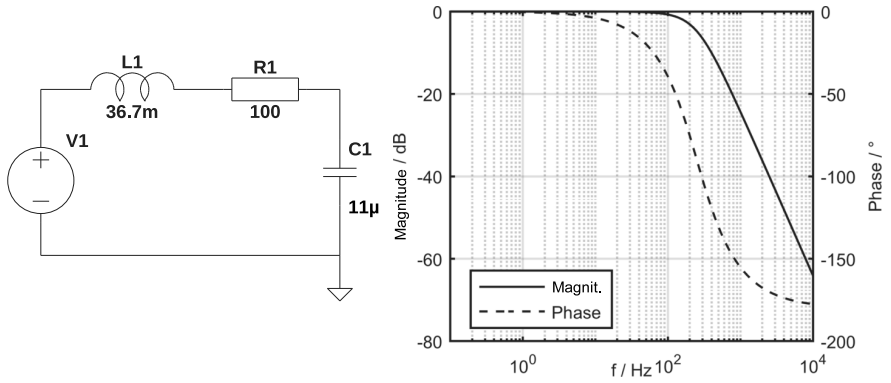


Fig. 4.48: Circuit of a 2nd order RLC Bessel-filter (left) and associated frequency response by magnitude and phase (right).

3. What is a Helmholtz layer?
4. Why is an electrode gel required at the junction between skin and electrode?
5. For what reason does a amplifier for biosignals require a high input impedance?
6. What is a common mode signal and what does common mode rejection mean?
7. What does the signal-to-noise ratio describe and what is it used for?
8. What is the magnitude and phase frequency response of a amplifier?
9. What is the magnitude of the internal resistance of the body and that of the electrodes?
10. What characteristics especially qualify an instrumentation amplifier for amplifying biosignals?
11. Enter the instrumentation amplifier presented in subsection 4.1.2 into LTspice.
12. Check the frequency response of the measurement amplifier after the subtractor, after the high-pass filter and at the circuit output in the frequency range 0.1 Hz to 10 kHz. To do this, apply an AC voltage of amplitude 1 mV between the two inputs, and simulate the frequency range.
13. Which components of the measurement amplifier would have to be changed and how, so that the high pass has a cutoff frequency of 1.6 Hz and the second low pass has a cutoff frequency of 360 Hz?
14. Check the calculated gain of the measurement amplifier after the instrumentation amplifier and at the circuit output.
15. Add a common mode signal source in the LTspice model.
16. At the circuit output of the instrumentation amplifier, calculate the ratio of the useful signal to the common-mode signal when at the circuit input the amplitude of the useful signal is 100 µV and the common-mode signal is 1 V. For this purpose, it should be taken into account that the operational amplifiers have a common-mode rejection of 100 dB.

Signal Interference

1. What forms of interference coupling of mains hum, where do these effects come from and how can they be minimized?
2. What type of shielding would be required to prevent capacitive or inductive coupling of mains hum?
3. Explain why the term 50 Hz noise is nonsensical.
4. What physical quantities determine the magnitude of thermal noise?
5. What is meant by white noise?
6. Draw in a graph the course of a transient disturbance.
7. What influence does the RC high-pass filter in the amplifier have on a transient disturbance?
8. What is the difference between the spectra of transient disturbances and noise?
9. What is amplitude modulation and how is a demodulator for amplitude modulated signals constructed?

Transducer for Non-Electrical Biosignals

1. How is an electret microphone or a solid-borne sound microphone constructed?
2. How does a semiconductor LED work? Why do semiconductor LEDs come in different colors? What does a typical emission spectrum of an LED look like?
3. How does the emission spectrum of an LED differ from that of a laser?
4. Which sensor can be used to detect light? How does it work?
5. What type of sensor can be used to detect magnetic fields?

Interference Suppression and Analog Filtering

1. What advantages do active filters have over passive filters and what disadvantage?
2. Enter a 4th order active Butterworth high pass filter with cutoff frequency 160 Hz into LTspice. Extend the high-pass filter to a band-pass filter with upper cut-off frequency 1.6 kHz. Check the magnitude and phase frequency response.
3. What is meant by group delay?
4. How does a frequency dependence of the group delay affect the transmission of signals?
5. What is the principal difference between a selective filter, such as a power filter, and a Bessel filter?
6. What is the purpose of a frequency transformation?
7. Which frequency transformation can be used to realize a 50 Hz mains hum filter from a low-pass filter?
8. Why is a Bessel filter often used when measuring an ECG?
9. Please determine the pole and zero points and the transfer function of a normalized 2nd order Butterworth low-pass filter and from this, determine the transfer function of a 2nd order Butterworth high-pass filter by a suitable frequency transformation, which has a cutoff frequency of 0.2 Hz with an attenuation of 3 dB.

10. Determine the components of a passive high-pass filter using the transfer function determined in the previous point.
11. What are the advantages and disadvantages of a Cauer filter?
12. What are the properties of Chebyshev functions?
13. What property do elliptic functions have?
14. What important property does a polynomial from a Bessel function have in filter synthesis?

Interference Suppression and Analogue Filtering

1. What advantages do active filters have over passive filters and what disadvantage?
2. Calculate an active Butterworth-high-pass filter of 4th order with cut-off frequency 160 Hz into LTspice. Extend the high pass to a band pass with upper cut-off frequency 1.6 kHz. Check the magnitude and phase frequency response.
3. What is meant by group delay?
4. How does a frequency dependence of the group delay affect the transmission of signals?
5. What is the difference in principle between a selective filter, such as a power filter, and a Bessel filter?
6. What is the purpose of a frequency transformation?
7. With which frequency transformation can a 50 Hz mains hum filter be realised from a low-pass filter?
8. Why is a Bessel filter often used when measuring an ECG?
9. Please determine the poles and zeros and the transfer function of a normalised Butterworth-low-pass filter of 2nd order and, using a suitable frequency transformation, determine the transfer function of a Butterworth-high-pass filter of 2nd order which has a cut-off frequency of 0.2 Hz with an attenuation of 3 dB.
10. Determine the components of a passive high-pass filter with the transfer function determined in the previous point.
11. What are the advantages and disadvantages of a Cauer filter?
12. What are the properties of the Chebyshev functions?
13. What is the property of elliptic functions?
14. Which important property does a polynomial from a Bessel function have in filter synthesis?

5 Methods for Discrete Processing and Analysis of Biosignals

5.1 Discretisation of Continuous Signals

After the analogue measurement of the biosignal, amplification and, if necessary, filtering, it is sampled with pulses at discrete points in time (pulse modulation), and only these measured values are used for further processing. If a time-continuous signal is to be generated again after the discrete-time processing, interpolation must be carried out between the discrete-time values. The way in which this interpolation takes place can best be examined in the spectral range. For this purpose, the spectra before and after the sampling as well as after the interpolation system are considered. The interpolation is done by low-pass filtering.

After the interpolation, the original continuous-time signal must arise again (see Figure 5.1). For this purpose, an equivalent system is considered in which the switch for sampling is replaced by a multiplier that multiplies the input signal with a square-wave pulse train (cf. Figure 5.2).

The multiplication with a square-wave pulse train can also be replaced by a multiplication with a Dirac-pulse train, whereby after the multiplication the now resulting weighted Dirac-pulse train $f_{T_a}(t)$ is still changed back into a square-wave pulse train by a *pulse shaper* (see Figure 5.3). This is because it does not matter whether the input signal is multiplied by a square-wave pulse train or a Dirac-pulse train before the multiplication, whereby the square-wave formation takes place after the multiplication.

The output signal $f_{\Delta T}$ after sampling and pulse shaping is obtained by convolution of the signal sampled with Dirac pulses.

$$f_{T_a}(t) = \sum_{k=-\infty}^{\infty} f(kT_a) \cdot \delta(t - kT_a) \tag{5.1}$$

with the impulse response

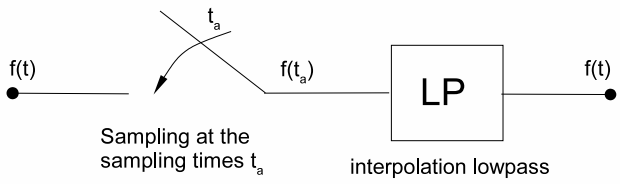


Fig. 5.1: Uniform sampling of a signal in the $t_a = nT_a$, $n = 1, 2, \ldots$ with sampling interval T_a and subsequent interpolation using low-pass filtering.

https://doi.org/10.1515/9783110736298-005

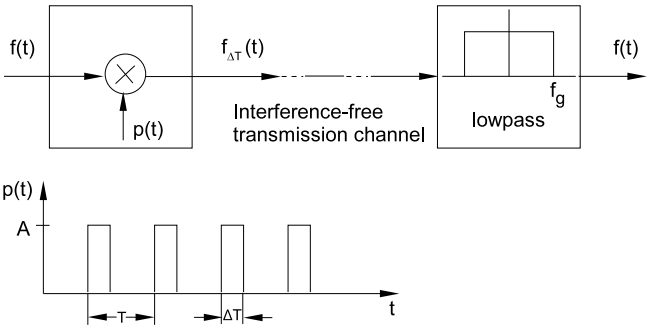


Fig. 5.2: Equivalent uniform sampling of a signal at instances $t_a = n \cdot T_a$ using square pulses (cf. Figure 5.1).

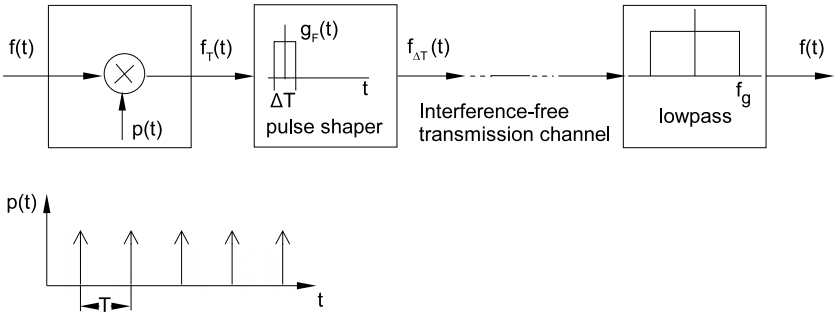


Fig. 5.3: Modified equivalent sampling model with rectangular pulses.

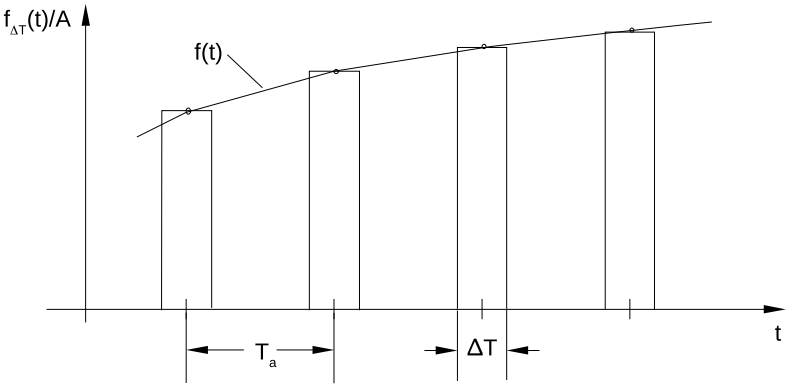


Fig. 5.4: Equivalent uniform sampling of a signal using square pulses (cf. Equation 5.3).

$$g_{\mathrm{F}}(t) = \mathrm{A} \cdot \mathrm{rect}\left(\frac{t}{\Delta T}\right) \tag{5.2}$$

of the pulse shaper, which generates a rectangular pulse $\mathrm{rect}(t)$[1] (see Figure 5.4):

$$\begin{aligned} f_{\Delta T}(t) = f_{T_{\mathrm{a}}}(t) * g_{\mathrm{F}}(t) &= \int_{\tau=-\infty}^{\infty} \sum_{k=-.\infty}^{\infty} f(kT_{\mathrm{a}}) \cdot \delta(\tau - kT_{\mathrm{a}}) \cdot \mathrm{A} \cdot \mathrm{rect}\left(\frac{t-\tau}{\Delta T}\right) \mathrm{d}\tau \\ &= A \sum_{k=-\infty}^{\infty} f(kT_{\mathrm{a}}) \int_{\tau=-\infty}^{\infty} \delta(\tau - kT_{\mathrm{a}}) \cdot \mathrm{rect}\left(\frac{t-\tau}{\Delta T}\right) \mathrm{d}\tau \\ &= A \sum_{k=-\infty}^{\infty} f(kT_{\mathrm{a}}) \cdot \mathrm{rect}\left(\frac{t - kT_{\mathrm{a}}}{\Delta T}\right) . \end{aligned} \tag{5.3}$$

Since a multiplication in the time domain corresponds to a convolution in the frequency domain and, conversely, a convolution in the time domain corresponds to a multiplication in the frequency domain, the associated spectrum can be determined by a convolution of the input signal spectrum $F(f)$ with the spectrum of a Dirac-pulse train and then a multiplication of this spectrum with the spectrum of a square pulse:

$$\begin{aligned} f_{T_{\mathrm{a}}}(t) &= \sum_{k=-\infty}^{\infty} f(kT_{\mathrm{a}}) \cdot \delta(t - kT_{\mathrm{a}}) \,, \\ F_{T_{\mathrm{a}}}(f) &= F(f) * \underbrace{\mathcal{F}\left\{\sum_{\nu=-\infty}^{+\infty} \delta(t - \nu T_{\mathrm{a}})\right\}}_{\sum_{\nu=-\infty}^{\infty} \frac{1}{T_{\mathrm{a}}} \delta\left(f - \frac{\nu}{T_{\mathrm{a}}}\right)} \\ &= \int_{\psi=-\infty}^{+\infty} F(f-\psi) \cdot \left(\frac{1}{T_{\mathrm{a}}} \sum_{\nu=-\infty}^{+\infty} \delta\left(\psi - \frac{\nu}{T_{\mathrm{a}}}\right)\right) \cdot \mathrm{d}\psi \\ &= \frac{1}{T_{\mathrm{a}}} \sum_{\nu=-\infty}^{+\infty} F\left(f - \frac{\nu}{T_{\mathrm{a}}}\right) \end{aligned}$$

and further using the relation for the rectangular function $\mathrm{rect}(t)$ and its Fourier transform $F_{\mathrm{rec}}(f)$:

$$F_{\mathrm{rec}}(f) = \mathcal{F}\{\mathrm{rect}(t)\} = \mathrm{si}(\pi f) \tag{5.4}$$

and the similarity theorem

$$\mathcal{F}\{f(at)\} = \frac{1}{|a|} F\left(\frac{j2\pi f}{a}\right) \quad \text{with} \quad a \neq 0 \,, \quad a = \text{const.} \tag{5.5}$$

1 $\mathrm{rect}(t) := 1$ from a Dirac pulse for $-0.5 \leq t \leq 0.5$; otherwise $\mathrm{rect}(t) := 0$.

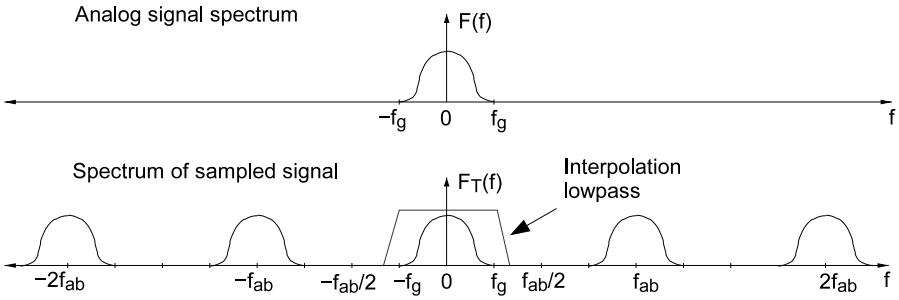


Fig. 5.5: Spectrum of the signal $F_{T_a}(f)$ sampled with Dirac pulses obtained by periodically repeating the spectrum of the original signal $F(f)$.

results for the resulting spectrum $F_{\Delta T}(t)$ of the rectangular sequence $f_{\Delta T}(t)$ after sampling:[2]

$$\begin{aligned} F_{\Delta T}(f) &= F_{T_a}(f) \cdot \mathcal{F}\left\{\mathrm{A} \cdot \mathrm{rect}\left(\frac{t}{\Delta T}\right)\right\} \\ &= \left[\frac{A}{T_a} \sum_{\nu=-\infty}^{\infty} F\left(f - \frac{t}{T_a}\right)\right] \cdot \Delta T \,\mathrm{si}(\pi f \Delta T) \\ &= \frac{A \cdot \Delta T}{T_a} \mathrm{si}(\pi f \Delta T) \cdot \sum_{\nu=-\infty}^{\infty} F\left(f - \frac{t}{T_a}\right) . \end{aligned} \tag{5.6}$$

The result for the spectrum of the signal $F_{T_a}(f)$ sampled with a Dirac-pulse train is a *periodic repetition* of the source signal spectrum $F(f)$ with sampling frequency $f_a = 1/T_a$ (cf. Figure 5.5).

However, this only works if the periodic spectrum repetitions *do not overlap*, as shown in Figure 5.5. In the case of overlapping spectra, a restoration of the original analogue signal by a simple low-pass filtering is no longer possible (cf. e.g. Figure 5.6). For a recovery it is namely necessary that the upper cut-off frequency of the analogue signal is smaller than half the sampling frequency:

$$f_g < \frac{f_a}{2} , \quad \text{Shannon sampling theorem.} \tag{5.7}$$

The spectrum of the analogue signal, and thus the signal itself, can be reconstructed from the spectrum of the sampled signal using a low-pass filter, which filters out only the spectrum component around the frequency zero point. The associated interpolation function between the samples is thus generated by the impulse response of the low-pass filter (see Figure 5.7).

This type can also be interpreted as ideal sampling with Dirac pulses. However, according to Equation 5.6, the signal sampled not ideally but with square wave functions

2 $\mathrm{si}(x) := \frac{\sin(x)}{x}$

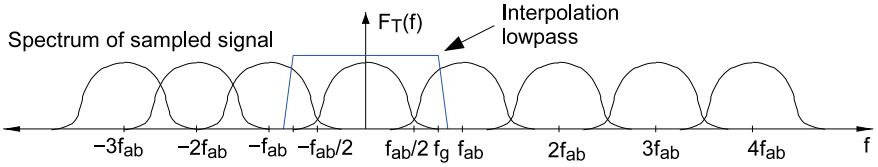


Fig. 5.6: Spectrum of the signal sampled with Dirac pulses with overlapping periodic spectrum of the original signal.

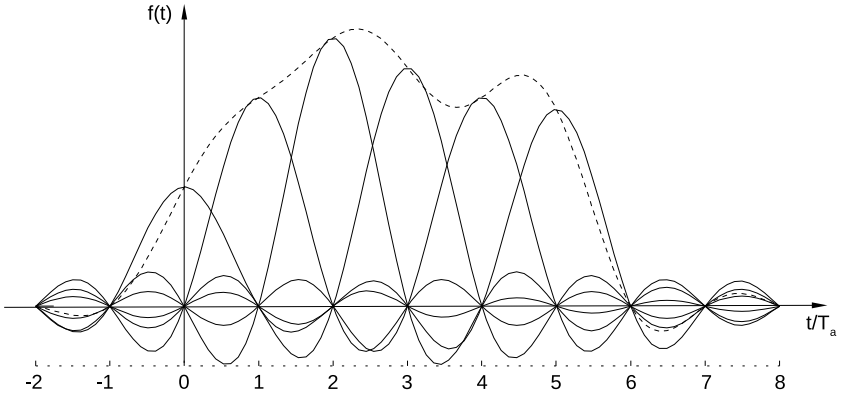


Fig. 5.7: Interpolation between samples using the impulse response of an ideal low-pass filter.

has an attenuation due to the si function. This means that when sampling with square pulses, the spectrum of the sampled signal is not simply repeated with the sampling frequency, but its values are also reduced during the repetition. Furthermore, distortions can also occur during reconstruction, since the si also affects the spectral values of the signal in the passband of the reconstruction low-pass filter at the output.

These distortions are smaller the narrower the pulse width ΔT of the square pulses. However, the spectral components then become smaller and smaller (cf. Equation 5.6). This could be compensated by the fact that the pulse amplitude A becomes larger and overall according to Equation 5.6 the product $A \cdot \Delta T$ is a constant. This is in *ideal way* the case with the Dirac momentum. Its width approaches zero, its height approaches infinity, but its area (product of width times height) is one. Sampling with a Dirac pulse therefore produces no distortions of the spectrum.

5.2 Discrete Transformations of Signal Processing

5.2.1 The Discrete-Time Fourier Transform

A discrete sequence could arise, for example, when sampling a signal with square pulses according to Equation 5.3, if the square width goes $\Delta T \to 0$. If we now calculate the spectrum of this rectangular sequence not according to Equation 5.6, but add up the spectra of these rectangular pulses weighted with the samples and delayed with kT_a, we obtain from Equation 5.3 with the help of the similarity theorem Equation 5.5:

$$\begin{aligned} F_{\Delta T}(f) = \mathcal{F}\left(f_{\Delta T}(t)\right) &= A \sum_{k=-\infty}^{\infty} f(kT_\mathrm{a}) \cdot \mathcal{F}\left(\mathrm{rect}\left(\frac{t - kT_\mathrm{a}}{\Delta T}\right)\right) \\ &= A \sum_{k=-\infty}^{\infty} f(kT_\mathrm{a}) \cdot e^{-j2\pi f k T_\mathrm{a}} \mathcal{F}\left(\mathrm{rect}\left(\frac{t}{\Delta T}\right)\right) \\ &= A \cdot \Delta T \,\mathrm{si}(\pi f \Delta T) \cdot \sum_{k=-\infty}^{\infty} f(kT_\mathrm{a}) \cdot e^{-j2\pi f k T_\mathrm{a}} \,. \end{aligned}$$

The factor $e^{-j\omega k T_\mathrm{a}}$ results from the delay of the square pulses by kT_a, which can be interpreted as passing through a delay element that has this transfer function. The sum

$$F_\mathrm{D}(f) := \sum_{k=-\infty}^{\infty} f(kT_\mathrm{a}) \cdot e^{-j2\pi f k T_\mathrm{a}} \tag{5.8}$$

is the *discrete-time Fourier-transformation* of the samples $f(kT_\mathrm{a})$, where one normally omits the factor T_a in the function $f(kT_\mathrm{a})$ (cf. e.g. [52]). One now obtains altogether for the outward and backward transformation:

$$f(k) = T_\mathrm{a} \int_{-T_\mathrm{a}/2}^{T_\mathrm{a}/2} F_\mathrm{D}(f) e^{j2\pi f k T_\mathrm{a}} \mathrm{d}f, \;\; F_\mathrm{D}(f) = \sum_{k=-\infty}^{\infty} f(k) e^{-j2\pi f k T_\mathrm{a}} \,, \tag{5.9}$$

with the relation

$$F_{\Delta T}(f) = A \cdot \Delta T \,\mathrm{si}(\pi f \Delta T) \cdot F_\mathrm{D}(f) \,. \tag{5.10}$$

If the square pulse $g_\mathrm{F}(t) = A \cdot \mathrm{rect}(t/\Delta T)$ when sampled according to Figure 5.4 goes to infinite amplitude height, i.e. $A \to \infty$ and vanishing width, i.e. i.e. $\Delta T \to 0$, but with $A \cdot \Delta T = 1$ into a dirac pulse $\delta(t)$, it follows from Equation 5.10 with $\lim_{\Delta T \to 0} \mathrm{si}(\pi f \Delta T) = 1$:

$$F_{\Delta T}(f) = F_\mathrm{D}(f) \,, \quad \text{when sampling with } \delta(t) \text{ pulses.} \tag{5.11}$$

The discrete-time Fourier-transform therefore describes in the frequency domain the spectrum of a signal sampled with Dirac pulses.

5.2.2 The Discrete Fourier Transform (DFT)

If the signal to be sampled $f(t)$ (and thus also its samples $f(k)$) is limited in time up to a maximum time duration t_g, the discrete-time Fourier-transformation can be simplified according to Equation 5.9. In this case, the signal to be sampled $f(t)$ is continued periodically, whereby the period duration T_p is greater than the maximum time duration t_g of the original signal $f(t)$ ($t_g < T_\mathrm{p}$). However, the periodic signal $f_\mathrm{p}(t)$ has a discrete spectrum which has a value only for multiples of the fundamental period $f_\mathrm{p} = 1/T_\mathrm{p}$. On the other hand, the signal $f(t)$ is sampled, i.e. it also has a spectrum periodic with the sampling frequency $f_\mathrm{a} = 1/T_\mathrm{a}$. Thus, for a period f_a in the frequency domain only $N = f_\mathrm{a}/f_\mathrm{p} = T_\mathrm{p}/T_\mathrm{a}$ values of the spectrum are needed to describe the periodically continued sampling signal $f_\mathrm{p}(t)$. In the time domain, the original continuous signal $f(t)$ can then be restored by temporally cutting off the periodically continued signal $f_\mathrm{p}(t)$ after one period and subsequent low-pass interpolation according to Figure 5.7 if the sampling theorem (Equation 5.7) is fulfilled.

The fact that a periodically continued signal $f_\mathrm{p}(t)$ can be described by the values of the continuous spectrum $F(f)$ of the original non-periodic signal can be explained by applying the convolution theorem in the time domain. For this purpose, analogous to Figure 5.3, the continuous spectrum $F(f)$ of the original signal $f(t)$ is described by multiplication with a periodic sequence of Dirac pulses $\delta(f)$

$$P(f) := \frac{1}{T_\mathrm{p}} \sum_{\nu=-\infty}^{\infty} \delta\left(f - \frac{\nu}{T_\mathrm{p}}\right) \tag{5.12}$$

as per

$$F_\mathrm{P}(f) = F(f) \cdot P(f) \tag{5.13}$$

"sampled" in the frequency domain. A multiplication in the frequency domain corresponds to a convolution in the time domain. Because of the relations

$$p(t) = \sum_{\nu=-\infty}^{+\infty} \delta(t - \nu T_\mathrm{p}) \circ\!\!-\!\!\bullet P(f) = \frac{1}{T_\mathrm{p}} \sum_{\nu=-\infty}^{\infty} \delta\left(f - \frac{\nu}{T_\mathrm{p}}\right)$$

$$f_\mathrm{p}(t) = \int_{-\infty}^{\infty} f(t-\tau) \cdot p(\tau)\,\mathrm{d}\tau \circ\!\!-\!\!\bullet F_\mathrm{P}(f) = F(f) \cdot P(f)$$

between time and frequency domain is obtained after substituting in Equation 5.13

$$\begin{aligned} f_\mathrm{p}(t) &= \int_{-\infty}^{\infty} f(t-\tau) \cdot \left(\sum_{\nu=-\infty}^{+\infty} \delta(\tau - \nu T_\mathrm{p}) \right) \mathrm{d}\tau \\ &= \sum_{\nu=-\infty}^{+\infty} f(t - \nu T_\mathrm{p}) \end{aligned} \tag{5.14}$$

a periodic time function $f_\mathrm{p}(t)$. We therefore obtain the result that frequency scanning of a spectrum in the time domain results in a periodic signal. Conversely, however, it also follows from this:

> The periodic continuation of a time limited signal results in a discrete spectrum with Dirac pulses whose weighting values, except for the factor $1/T_\mathrm{p}$, correspond to the values at the individual frequency points of the continuous spectrum $F(f)$ of the original non-periodic signal $f(t)$.

In principle, the only requirement for the type of time signal is that it must be that it should be limited in time and integrable. It can therefore also be a square-wave pulse train, as it arises when sampling with square-wave pulses according to Figure 5.4. For this purpose, we investigate the discrete-time Fourier transform according to Equation 5.9

$$F_\mathrm{D}(f) = \sum_{k=-\infty}^{\infty} f(k) e^{-j2\pi f k T_\mathrm{a}} \tag{5.15}$$

of a sampled signal $f(k)$ and determine the spectrum periodic with sampling frequency $f_\mathrm{a} = 1/T_\mathrm{a}$ for one period at N frequency points, i.e. at frequencies $l \cdot f_\mathrm{a}/N = l/(NT_\mathrm{a})$ with $l = 0, 1, \ldots, N-1$, i.e.

$$F_\mathrm{D}\left(\frac{l}{NT_\mathrm{a}}\right) = \sum_{k=-\infty}^{\infty} f(k) e^{-j2\pi k l/N}, \quad l = 0, 1, \ldots, N-1\,. \tag{5.16}$$

The infinite sum in Equation 5.16 is now divided into an infinite sum of these periods, each having N values (cf. [62]), ie:

$$\begin{aligned} F_\mathrm{D}\left(\frac{l}{NT_\mathrm{a}}\right) &= \cdots + \sum_{k=-N}^{-1} f(k) e^{-j2\pi k l/N} + \sum_{k=0}^{N-1} f(k) e^{-j2\pi k l/N} + \sum_{k=N}^{2N-1} f(k) e^{-j2\pi k l/N} + \cdots \\ &= \sum_{m=-\infty}^{\infty} \sum_{k=mN}^{N(m+1)-1} f(k) e^{-j2\pi k l/N}. \end{aligned} \tag{5.17}$$

In the case of a double sum as in the last equation, the order may be reversed. If we further replace the running variable k by $k := i + mN$, it follows from Equation 5.17:

$$F_\mathrm{D}\left(\frac{l}{NT_\mathrm{a}}\right) = \sum_{i=0}^{N-1} \underbrace{\sum_{m=-\infty}^{\infty} f(i+mN)}_{f_\mathrm{p}(i)} e^{-j2\pi i l/N}\,. \tag{5.18}$$

A comparison of the last Equation 5.18 with Equation 5.14 shows that the underclamped part of the periodically repeating signal $f(t)$ is given at the sampling instants $t_\mathrm{i} = i \cdot T_\mathrm{a}$. If the signal $f(t)$ is limited so that it disappears after the period NT_a, the original signal $f(t)$ can be determined from the periodically repeating signal $f_\mathrm{p}(t)$ by truncation after N samples.

However, the values of the spectrum $F_\mathrm{D}(l/NT_\mathrm{a})$ of the discrete-time Fourier-transform of the sampled signal $f(t)$ can also be described by the coefficients of a complex Fourier series of the periodic signal $f_\mathrm{p}(t)$. With the *discrete Fourier-transformation*

the discrete periodic signal can be described like an analogue signal as a sum of e-functions according to

$$f_{\mathrm{p}}(i) = \sum_{l=0}^{N-1} c_l e^{j2\pi il/N}\,, \quad i = 0, 1, \ldots, N-1 \tag{5.19}$$

describe. The Fourier-coefficients c_{i} are then given by

$$c_l = \frac{1}{N}\sum_{i=0}^{N-1} f_{\mathrm{p}}(i) e^{-j2\pi il/N}\,, \quad l = 0, 1, \ldots, N-1\,. \tag{5.20}$$

A comparison with Equation 5.18 shows that the coefficients of the discrete Fourier-transform match the values of the discrete-time Fourier-transform $F_{\mathrm{D}}(l/NT_{\mathrm{a}})$ except for the factor $1/N$:

$$c_l = \frac{1}{N} F_{\mathrm{D}}\left(\frac{l}{NT_{\mathrm{a}}}\right) \quad \text{bzw.} \quad f_{\mathrm{p}}(i) = \frac{1}{N}\sum_{l=0}^{N-1} F_{\mathrm{D}}\left(\frac{l}{NT_{\mathrm{a}}}\right) e^{j2\pi il/N}\,. \tag{5.21}$$

Result

By periodically continuing a time-limited signal sampled with the time interval T_{a} with a period NT_{a} greater than the temporal length $t_g < NT_{\mathrm{a}}$ of the signal, the associated spectrum can be calculated by a discrete Fourier series according to Equation 5.16 and Equation 5.19 at N frequency points within one period of the frequency domain.

The original signal can be realised by cutting out one period of the signal $f_{\mathrm{p}}(i)$ by a multiplication with a temporal rectangular window, and the original spectrum can be interpolated by interpolation in the frequency domain with an si function, which corresponds to a convolution of the discrete spectrum $F(l/NT_{\mathrm{a}})$ with the Fourier-transform of the temporal window. With the help of the abbreviation

$$F(l) := F_{\mathrm{D}}\left(\frac{l}{NT_{\mathrm{a}}}\right) = \sum_{i=0}^{N-1} f_{\mathrm{p}}(i) e^{-j2\pi il/N} \tag{5.22}$$

and Equation 5.19 are finally obtained for the outward and backward transformation of the *discrete Fourier-transformation* (DFT):

$$f_{\mathrm{p}}(i) = \frac{1}{N}\sum_{l=0}^{N-1} F(l) e^{j2\pi il/N}\,, \quad F(l) = \sum_{i=0}^{N-1} f_{\mathrm{p}}(i) e^{-j2\pi \mathrm{fil}/N}\,. \tag{5.23}$$

The DFT can also be given in matrix notation using the Fourier-matrix $\mathbf{W} := \{w_{mn}\}$, with elements $w_{mn} := e^{-j2\pi mn/N}$ as follows: [3]

$$\begin{aligned} f_{\mathrm{p}} &= \mathbf{W}^{-1} \cdot F \\ F &= \mathbf{W} \cdot f_{\mathrm{p}} \end{aligned} \tag{5.24}$$

3 The matrix $\mathbf{A}'$ is the transpose of the matrix $\mathbf{A}$. Given a vector, e.g. f_{p}, this gives a row vector a column vector and a column a row vector.

with equation:

$$\begin{aligned} f_\mathrm{p} &:= (f_\mathrm{p}(0), f_\mathrm{p}(1), \ldots, f_\mathrm{p}(N-1))' \\ F &:= (F(0), F(1), \ldots, F(N-1))' \\ \mathbf{W} &:= \{w_{mn}\}\,, \quad w_{mn} = e^{-j2\pi mn/N} \\ \mathbf{W}^{-1} &= \left\{\frac{1}{N} w_{nm}^{-1}\right\}\,, \quad w_{nm}^{-1} = e^{j2\pi mn/N} \\ m, n &:= 0, 1, \ldots, N-1. \end{aligned} \tag{5.25}$$

The inverse transformation is thus carried out by means of the inverse Fourier-matrix $\mathbf{W}^{-1}$. With the help of computer-algebra-systems (CAS), which can perform matrix operations directly, such as Octave, Scilab or Matlab, the DFT is particularly easy to calculate.

Example

for $N = 3$: Forward transformation:

$$\underbrace{\begin{bmatrix} F(0) \\ F(1) \\ F(2) \end{bmatrix}}_{F} = \underbrace{\begin{bmatrix} 1 & 1 & 1 \\ 1 & e^{-j2\pi/3} & e^{-j4\pi/3} \\ 1 & e^{-j4\pi/3} & e^{-j8\imath/3} \end{bmatrix}}_{\mathbf{W}} \cdot \underbrace{\begin{bmatrix} f_\mathrm{p}(0) \\ f_\mathrm{p}(1) \\ f_\mathrm{p}(2) \end{bmatrix}}_{f}, \tag{5.26}$$

Reverse transformation:

$$\underbrace{\begin{bmatrix} f_\mathrm{p}(0) \\ f_\mathrm{p}(1) \\ f_\mathrm{p}(2) \end{bmatrix}}_{f} = \underbrace{\frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & e^{j2\pi/3} & e^{j4\pi/3} \\ 1 & e^{j4\pi/3} & e^{j8\pi/3} \end{bmatrix}}_{\mathbf{W}^{-1}} \cdot \underbrace{\begin{bmatrix} F(0) \\ F(1) \\ F(2) \end{bmatrix}}_{F}. \tag{5.27}$$

5.2.3 Discrete Laplace Transform and *z*-Transform

The z- transform is particularly well suited for describing linear digital systems consisting only of linear components, since here the relationship between input and output signals in the frequency domain can be described by a simple fractional rational function in the new frequency variable $z := e^{j\omega T_\mathrm{a}}$ (cf. subsubsection 5.3.4.1). For a causal discrete-time signal $f(n)$ (i.e. $f(n) = 0$ for $n < 0$), the z-transform is then described by the new frequency variable as follows:

$$F_\mathrm{D}(f) = F(z = e^{j\omega T_\mathrm{a}})\,, \quad \text{bzw.} \quad F(z) = F_\mathrm{D}\left(j\omega = \frac{1}{T_\mathrm{a}} \ln z\right)\,. \tag{5.28}$$

It then follows according to Equation 5.9 for the z-transformation:

$$f(n) = \frac{1}{2\pi j} \oint F(z)\, z^{n-1} \mathrm{d}z\,, \quad F(z) = \sum_{n=0}^{\infty} f(n)\, z^{-n}\,. \tag{5.29}$$

From the integral with limits from $-\frac{T_a}{2}$ to $+\frac{T_a}{2}$, a closed orbital integral has now emerged, and the imaginary axis in the $p := \sigma + j\omega$-plane is thereby mapped onto the unit circle in the $z = e^{pT_a}$-plane. The z- transformation can be interpreted as a *mapping* between the p- and the z-plane. However, it is *ambiguous* because the imaginary axis in the p-plane is mapped several times to the unit circle in the z-plane. In this case, the integration must be done on a closed path in the mathematically positive direction in the convergence domain of the z-transform. The relation in Equation 5.28 is valid only for absolutely summable causal time signals.

Example
Exponential sequence: $f(n) = e^{-knT}$
Thereby, Equation 5.29

$$F(z) = \sum_{n=0}^{\infty} e^{-knT} z^{-n} = \sum_{n=0}^{\infty} \left(e^{-kT} z^{-1}\right)^n = \frac{1}{1 - e^{-kT} z^{-1}}, \quad |z| > e^{-kT}. \tag{5.30}$$

In the range $|z| > e^{-kT}$, the $z$transform converges. Only for $k < 0$ is the exponential sequence $f(n)$ absolutely summable. Then it also has the discrete-time Fourier-transform

$$F_D(f) = F(z = e^{j\omega T_a}) = \frac{e^{j\omega T_a}}{e^{j\omega T_a} - e^{-kT}}, \quad k > 0. \tag{5.31}$$

5.3 Methods for Analysis and Processing of Discrete Biosignals

As in the continuous-time case, the biosignal can also be studied using its samples in the discrete-time domain, as long as the latter is in good agreement with the continuous-time signal. This requires compliance with the sampling theorem, low sampling distortion, and sufficient accuracy in quantization to avoid excessive rounding noise.[4]

5.3.1 Time Domain Signal Analysis and Matching

The analysis of the biosignal in the time domain, in contrast to the frequency domain, can be clearly illustrated by its sampled values in the course of the curve. In addition to the determination of distinctive points (e.g. minima, inflection points, maxima) in the curve progression by means of a curve discussion, the deviations of individual function values, for example, can also be determined by calculating the statistical

4 Rounding noise arises from the error that occurs in digitization due to truncation or rounding in the course of converting an exact sample to a sample with finite accuracy.

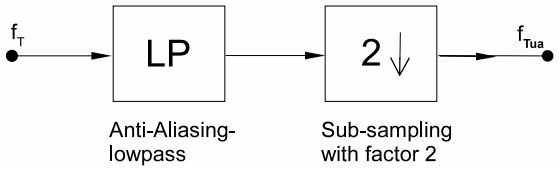


Fig. 5.8: Sub-sampling, omitting every second sample and using a low-pass beforehand to ensure compliance with the sampling theorem.

moments such as the expected value or the variance, as well as the interrelationships of two curve progressions with the help of the covariance or the correlation. Before the signals can be treated in this form, they must often first be synchronized and brought to a common time base.

5.3.1.1 Simplification, Interpolation and Averaging of signals

The samples of signals available for analysis cannot always be processed directly. E.g., there are too many or too few samples corresponding to the sampling frequency for a certain algorithm to be applied, or in the method to be applied, the sampling frequency must be changed, as is the case, for example, with the discrete wavelet transform (see subsubsection 5.3.3.2).

In the case of stochastic signals, it is often not the smaller random fluctuations in the signal waveform, but rather the mean value that changes with time, which is provided by a moving average (MA) filter in a given time interval before processing. Changing the sampling frequency can be done in Matlab with the `resample()` command. However, this can also be done by omitting or adding samples followed by low-pass filtering, which will be shown in the following sections.

Changing the Sampling Frequency

To change the sampling frequency, two cases must be distinguished in principle: Either the sampling frequency must be decreased or increased.

1. If the sampling frequency is to be decreased, the upper cutoff frequency of the signal must not be greater than half the sampling frequency because of the necessary compliance with the sampling theorem Equation 5.7. If, for example, every second sample is omitted to reduce the sampling rate, the sampling frequency is halved, but the maximum permissible cutoff frequency of the signal is also halved, because of the sampling theorem. Before omitting samples, it must therefore be checked whether the sampling theorem is still observed, and if not, the cutoff frequency must be reduced with a low-pass filter (see Figure 5.8).

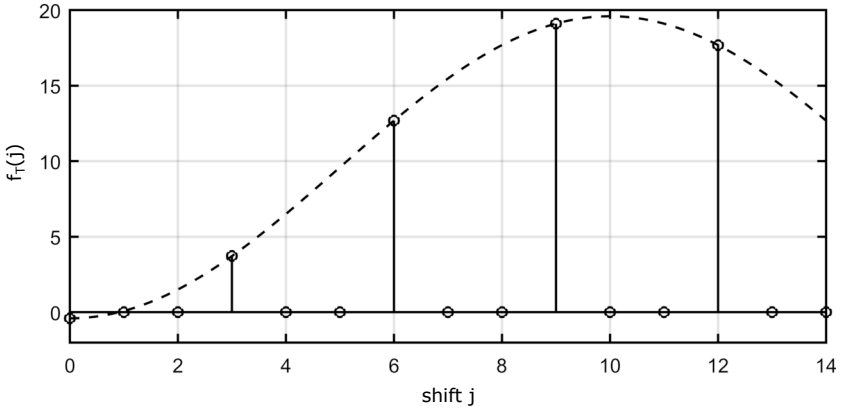


Fig. 5.9: Preparation to increase the sampling frequency by three times by inserting two values between each of the old samples, but setting them all equal to zero.

2. If the sampling frequency is to be increased, this would in principle mean reconstructing the analogue signal and then sampling it again at an increased frequency. An interpolation lowpass according to Figure 5.1 and Figure 5.7 and an additional analogue/digital converter would therefore be necessary. But there is also different:
   First, depending on the requirements $M-1$, additional additional values are inserted between each two samples corresponding to the new sampling frequency, but all of them are set equal to zero (see Figure 5.9), i.e.

$$\begin{aligned}\tilde{f}_{T_a}(j) &= f_{T_a}(j/M) \quad \text{for } j = i \cdot M \\ &= 0 \quad \text{other.}\end{aligned}$$

   The sampling frequency is thus increased by M times and the sampling interval $T_{\text{ш}} = T_a/M$ is decreased by M times. In the frequency domain, according to the discrete-time Fourier-transformation to Equation 5.9, one then obtains for the spectrum of this signal:

$$\begin{aligned}\tilde{F}_{T_a}(f) &= \sum_{j=-\infty}^{\infty} \tilde{f}_{T_a}(j) e^{-j2\pi f j T_{\text{ш}}} = \sum_{i=-\infty}^{\infty} \tilde{f}_{T_a}(i \cdot M) e^{-j2\pi f i M T/M} \\ &= \sum_{i=-\infty}^{\infty} f_{T_a}(i) e^{-j2\pi f i T_a} = F_{T_a}(f).\end{aligned}$$

   Attention: The spectrum does not change after inserting M zeros between every two samples!
   However, since the sampling frequency has now been increased by M times, the new half sampling frequency also changes and thus there are also additional spectral components in the signal that must be suppressed by a low-pass filter (see Figure 5.10).

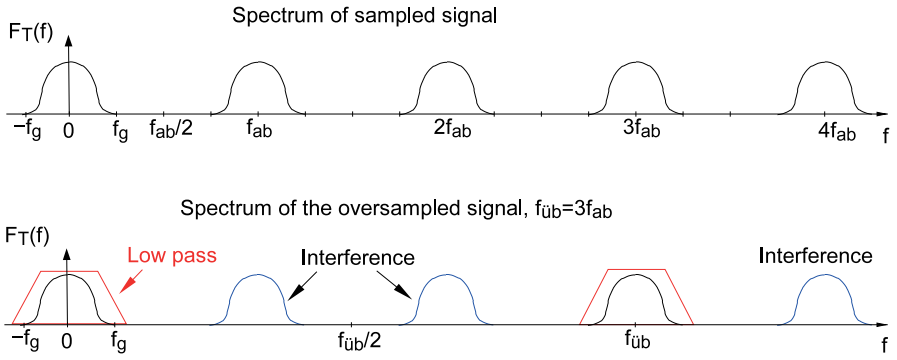


Fig. 5.10: Spectrum of a discrete-time signal at triple oversampling by adding two additional values between each two samples but equal to zero, see Figure 5.9.

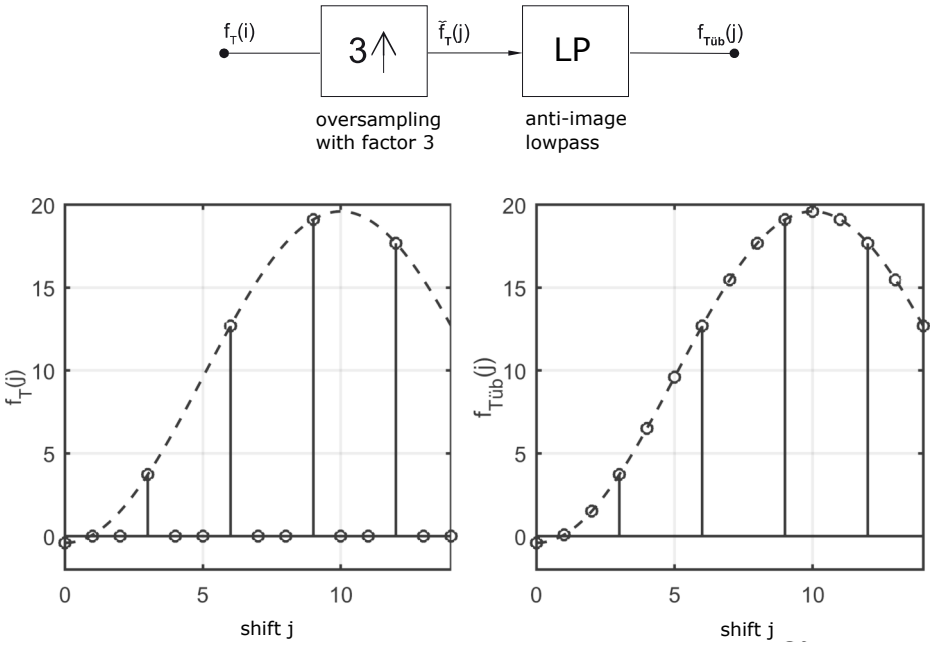


Fig. 5.11: Interpolation example of a discrete-time signal using the insertion of two zeros between each two samples followed by low-pass filtering.

After successful suppression of these "interfering parts" in the spectrum by a low-pass filter (see Figure 5.11), one then obtains new additional samples that interpolate the signal in the same way as if it had previously been converted into a

time-continuous signal and then resampled with a higher sampling frequency. An additional analogue-to-digital converter is therefore unnecessary.

Averaging of signals

The time averaging of signals is used on the one hand for low-pass filtering and on the other hand for separating the deterministic and stochastic signal components of a biosignal. A biosignal can thus be described as

$$s(k) = \mu(k) + \Sigma(k) \tag{5.32}$$

where $\mu(k)$ contains the transient mean, i.e. the deterministic part, and $\Sigma(k)$ contains the stochastic part.

For example, if $s(k)$ is the sampled biosignal, the averaged signal $s_{\mathrm{M}}(k)$ is obtained when averaging over $N + 1$ samples:

$$s_{\mathrm{M}}(k) = \frac{1}{N} \sum_{i=0}^{N} f(k - i)\,, \quad \mathrm{N}\colon \text{filter order.} \tag{5.33}$$

Such averaging filters can be easily realised by a non-recursive digital filter N-th order whose coefficients c_{i}, for $i = 0, \ldots, N$ all have the same value $1/N$ (see Figure 5.35). Due to the averaging over the window width, the sampled signal experiences a time base modified by $1/N$, i.e. the sampling frequency becomes lower.

In addition to this average filtering over several continuous function values through a windowing, there is also the possibility of averaging over the entire signal course of several signals. For example, if one has sampled several periods of a signal, these can be averaged by cutting out coherent. In this case, one obtains a mean value and an associated variance or standard deviation for each sample of the signal. The mean value curve describes the deterministic part of the biosignal, whereby all deviations are expressed in the variance. It is thus possible to find the noise component and outliers and to separate them from the biosignal, since these components behave stochastically and are therefore averaged out during summation.

With periodic signals, a fundamental distinction is made between coherent and non-coherent averaging. In practice, for example, in the analysis of EEG signals, the usually very small evoked potentials are separated from the dominant noise component with the help of a triggered non-coherent averaging of many signals. But the method can also be used to determine a characteristic signal course of an often repeated EMG measurement for a certain body movement or to determine the pulse wave course of a photoplethysmogram.

The Listing 5.3.1.1 shows a Matlab script for a coherent averaging of a N-times measured, noisy signal $S_1(n) = \{s_1(n)\}, \ldots, S_N = \{s_N(n)\}$, in the corresponding Figure 5.12 the associated spectrum and the course of the signal-noise-ratio as a function of N are shown next to the signals.

Listing 5.3.1.1: Matlab example for coherent averaging of signals.

```
T = 100;                    % signal duration
fa = 100;                   % sampling frequency
Ta = 1/fa;                  % sampling period
tn = 0:1/fa:T-1/fa;         % time vector between 0 and T s

%% s, cosine signal with additive noise
f = 1;
x = -0.5*sin(2*pi*f*tn)+0.*sin(2*2*pi*f*tn);% sum oscillation
a = 0.2;                    % noise amplitude
n = randn(size(x));         % random numbers
s = x+(a*n);                % signal with additive noise

subplot(4,2,1);             % graphical representation of signal
plot(tn,x,'LineWidth',2); hold on;
plot(tn,s, '--'); grid on;
legend('without noise', 'with noise')
xlabel('time t /s');
ylabel('amplitude');
axis([0 5 -1.2 1.2]);

subplot(4,2,2);             % graphical representation of spectrum
[orig_Sor,f] = pwelch(x,[],[],[],fa);
[orig_S,f] = pwelch(s,[],[],[],fa);
plot(f,20*log(orig_Sor), f,20*log(orig_S)); grid on;
legend('without noise', 'with noise')
xlabel('frequency f / Hz');
ylabel('amplitude / dB');
axis([0 10 -300 30]);

%% 5-fold coherent averaging of s
N = 5;
average_s=0;

for i=1:N
    n=randn(size(x));
    s=x+(a*n);
    average_s=s/N+average_s;
end

subplot(4,2,3);
```

```
plot(tn,average_s,'LineWidth',2); hold on;
plot(tn,s, '--'); grid on;
legend('averaged', 'with noise')
xlabel('time t /s');
ylabel('amplitude');
axis([0 5 -1.2 1.2]);

subplot(4,2,4);
[S,f] = pwelch(average_s,[],[],[],fa);
plot(f,20*log(S),f,20*log(orig_S)); grid on;
legend('averaged', 'with noise')
xlabel('frequency f / Hz');
ylabel('amplitude / dB');
axis([0 10 -300 30]);

%% 50-fold coherent averaging of s
N = 50;
average_s=0;

for i=1:N
    n=randn(size(x));
    s=x+(a*n);
    average_s=s/N+average_s;
    SNR(i) = snr(average_s,fa,5,'aliased');
end

subplot(4,2,5);
plot(tn,average_s,'LineWidth',2); hold on;
plot(tn,s, '--'); grid on;
legend('averaged', 'with noise')
xlabel('time t /s');
ylabel('amplitude');
axis([0 5 -1.2 1.2]);

subplot(4,2,6);
[S,f]=pwelch(average_s,[],[],[],fa);
plot(f,20*log(S),f,20*log(orig_S)); grid on;
legend('averaged', 'with noise')
xlabel('frequency f / Hz');
ylabel('amplitude / dB');
axis([0 10 -300 30]);
```

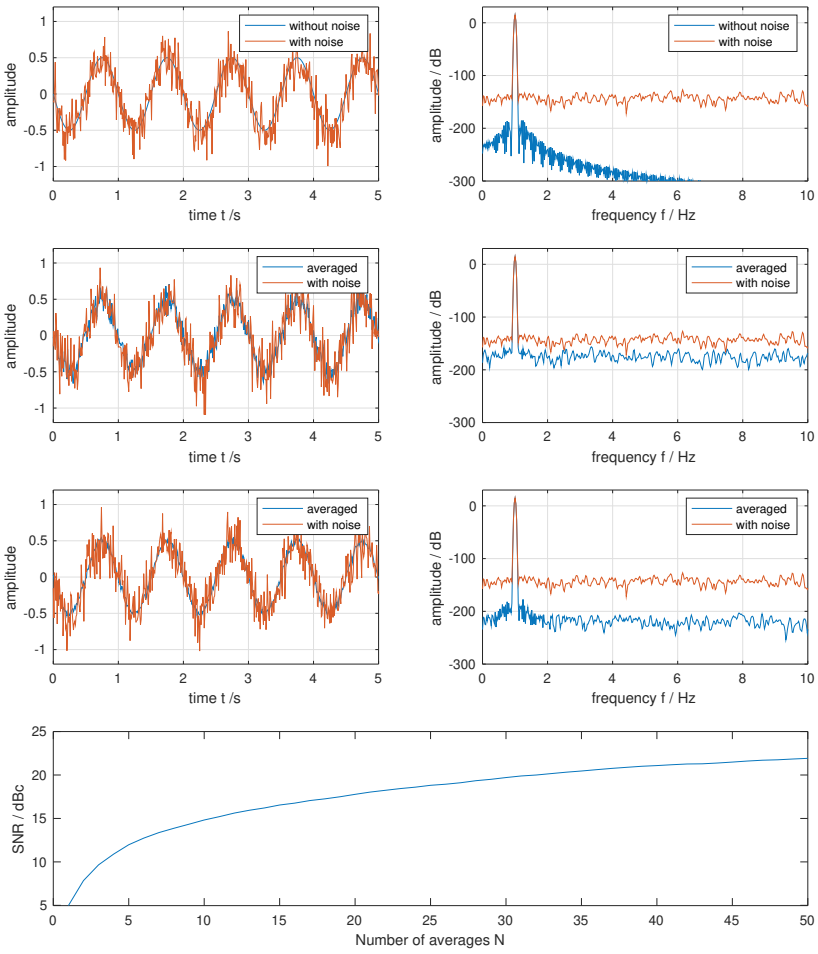


Fig. 5.12: Coherent averaging using the example of a noisy periodic signal (top left): after a 5-fold (centre left) and after a 50-fold coherent averaging (bottom left); the spectra of the signals compared to the non-noisy signal (top right) and the noisy signals (centre right, bottom right) clearly shows the reduction of the signal-noise-distance. The greater the number of averages N, the better the SNR in dBc (bottom).

```
%% graphical representation of the SNR
subplot(4,2,7:8);
plot(1:N,SNR)
xlabel('Number of averages N');
ylabel('SNR / dBc');
```

5.3.1.2 The Auto- and Cross-Correlation

The correlation generally describes the similarity of the random variables X_1 and X_2 and is proportional to the energy of their difference according to [45]

$$E[(X_1 - X_2)^2] = E[X_1^2] - 2\underbrace{E[X_1 X_2]}_{R_{X_1X_2}} + E[X_2^2] \,. \tag{5.34}$$

The operator E forms the expected value of a random variable. Therefore, $E[{X_1}^2]$ and $E[{X_2}^2]$ are the expected values of the squared random variables X_1 and X_2, respectively, and $E[X_1 X_2]$ is the expected value of the product of the random variables. This is also denoted by $R_{X_1X_2}$ and is defined as their associated *correlation*.

The random variables are often signals that depend on time. If they are different, e.g. two different measurements, $R_{X_1X_2}$is called *cross-correlation*. If it is one and the same signal, i.e. $X_1 = X_2 = X$, only at different measuring points, R_{XX} is the corresponding *auto-correlation*.

In the case of random variables, the expected value is also the mean value, which is formed either over a large group (group mean index, e.g. mean value for several identical dice) or over time (time mean index, e.g. mean value for one dice when dice are rolled several times).

In biosignal processing, different time averages of measured values are usually examined, which can arise, for example, from measurements with electrodes at different points on the body, such as in ECG and EEG, or from the measurement of electrical fields in computer or magnetic resonance tomography. For discrete-time signals, one then obtains for the correlation $R_{X_1X_2}$ in general:

$$R_{X_1X_2} = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} X_{i1} X_{j2} P_{X_{i1}X_{i2}}(X_{i1}, X_{j2}) \,. \tag{5.35}$$

Here $P_{X_{i1}X_{j2}}(X_{i1}, X_{j2})$ gives the probability density function for the values X_{i1} and X_{j2} to occur simultaneously. If this function is the same for all combinations X_{i1}, X_{i2}, it is simply given by the inverse of the number of possibilities – e.g., for a die with 6 possible rolls {1 to 6}, the probability of a given number being rolled is $\frac{1}{6}$. Then we get:

$$R_{X_1X_2} = \lim_{L\to\infty} \frac{1}{(2L+1)^2} \sum_{i=-L}^{L} \sum_{j=-N}^{L} X_{i1} X_{j2} \,. \tag{5.36}$$

The probability density function $P_{X_{i1}X_{j2}}(X_{i1}, X_{j2})$ can thereby be expressed by $1/(2L+1)^2$.

In auto-correlation, time dependence is usually investigated, and one obtains with the definitions $X_\mathrm{i} := X(t = t_\mathrm{i})$ and $X_j = X(t = t_\mathrm{i} + \tau_j)$ when summing up all signal values with the same difference τ_j of the measurement time points in an interval with $2N + 1$ values:

$$R_{XX}(t_\mathrm{i}, \tau_j) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{l=-N}^{N} X(t_\mathrm{i} + t_l) X(t_\mathrm{i} + t_l + \tau_j) \,. \tag{5.37}$$

In many cases, the auto-correlation depends only on the difference τ_j of the measurement time points and not on the absolute time t_i at which the measurement took place. Equation 5.37 can therefore be simplified as follows:

$$\mathcal{R}_{XX}(\tau_j) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{k=-N}^{N} X(t_k)X(t_k+\tau_j) \tag{5.38}$$

or even simpler:

$$\mathcal{R}_{XX}(j) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{k=-N}^{N} X(k)X(k+j)\,. \tag{5.39}$$

Mean Influences

When analysing biosignals, it is often important to examine the correlations in the changes of a measurand rather than the measurand itself. For example, the pulse rate of a heart is influenced by respiration. This influence could then be determined by calculating the auto-correlation. This then oscillates around an individual mean value. However, the interesting values can only be recognised in the oscillating values and can often be difficult to determine if a large constant value also comes into play in the correlation. Therefore, the auto-correlation of deviations from the mean is also used, which is called *auto-covariance* $\mathcal{C}_{XX}(j)$, and is defined as follows:

$$\mathcal{C}_{XX}(j) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{k=-N}^{N} \{X(k) - E[X]\}\{X(k+j) - E[X]\}\,, \tag{5.40}$$

with

$$E[X] := \frac{1}{2N+1} \sum_{k=-N}^{N} X(k)\,. \tag{5.41}$$

The relationship between auto-correlation and auto-covariance is obtained by decomposing the random signal X into a mean-free random signal $\widetilde{X}(j)$ and its mean according to $X(\mu) = \widetilde{X}(\mu) + E[X]$. Then one obtains for the auto-correlation $\mathcal{R}_{\widetilde{X}\widetilde{X}}(m)$ of the mean-free random signal or the auto-covariance $\mathcal{C}_{XX}(j)$:

$$\begin{aligned}
\mathcal{C}_{XX}(j) = \overline{R}_{\widetilde{X}\widetilde{X}}(j) &= E[\widetilde{X}(\mu)\widetilde{X}(\mu+j)] \\
&= E[(X(\mu) - E[X(\mu)])(X(\mu+j) - E[X(\mu+j)])] \\
&= E[X(\mu)(X(\mu+j)] - E[X(\mu)]E[X(\mu+j)] \\
&\quad - E[X(\mu)]E[X(\mu+j)] + E[X(\mu)]E[X(\mu+j)] \\
&= \underbrace{E[X(\mu)(X(\mu+j)]}_{\overline{R}_{XX}(j)} - E[X(\mu)]E[X(\mu+j)]\,.
\end{aligned} \tag{5.42}$$

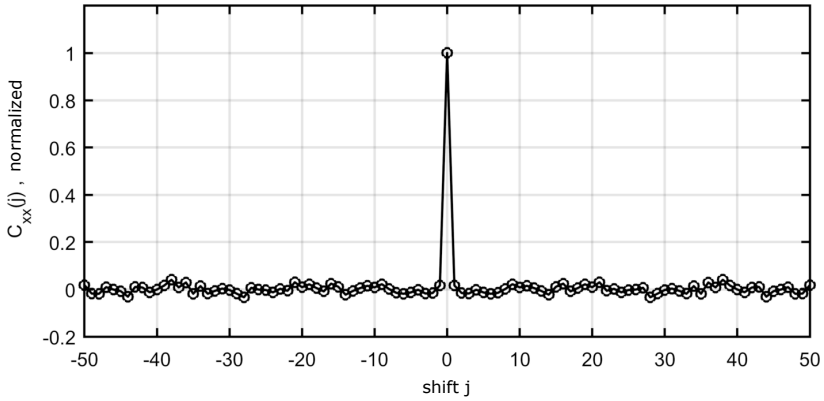


Fig. 5.13: Matlab-Example of a normalized auto-covariance $C_{XX}(j)$ of a Gaussian distributed mean-free random number sequence with 5000 values and a variance of one.

Furthermore, since the random signal is ergodic[5] and thus the expected value is time-independent, it follows because of

$$\begin{aligned} E[X(\mu)] = E[X(\mu + j)] \ &:= E[X]: \\ \mathcal{C}_{XX}(j) = \overline{R}_{XX}(j) - (E[X])^2 \ . \end{aligned} \tag{5.43}$$

As a result, the mean-free random signal $\mathcal{C}_{XX}(j)$ is simply obtained by subtracting the root mean square $(E[X])^2$ from the auto-correlation function.

Redundancy-Free Biosignals

A redundancy-free signal exists if $X(\mu + j)$ for $j \neq 0$ is independent of the preceding measured values $X(\mu)$:[6]

$$\overline{R}_{XX}(j) = E[X(\mu)X(\mu + j)] = E[X(\mu)] \cdot E[X(\mu + j)] = (E[X])^2 \ , \quad j \neq 0 \ .$$

It follows for the auto-covariance $\mathcal{C}_{XX}(j)$ according to Equation 5.43:

$$\begin{aligned} \mathcal{C}_{XX}(j) = . \begin{cases} E[X^2] - (E[X])^2 = \sigma_X^2 \ , & \text{at j = 0} \\ .0 \ , & \text{other} \end{cases} \\ = \sigma_X^2 \cdot \delta(j), \delta(j): \quad \text{discrete unit pulse.} \end{aligned} \tag{5.44}$$

Thus, it follows that for redundancy-free and mean-free signals, the auto-correlation consists only of a discrete unit momentum weighted by the variance σ_X^2 of the random variable X. An example of such a covariance function is shown in Figure 5.13.

5 A random signal is called ergodic if the mean values are equal over the multitude and over time.

6 For independent or redundancy-free random signals, the expected value of the product of two random signals, e.g. A and B, is equal to the product of their expected values, i.e. $E[A \cdot B] = E[A] \cdot E[B]$.

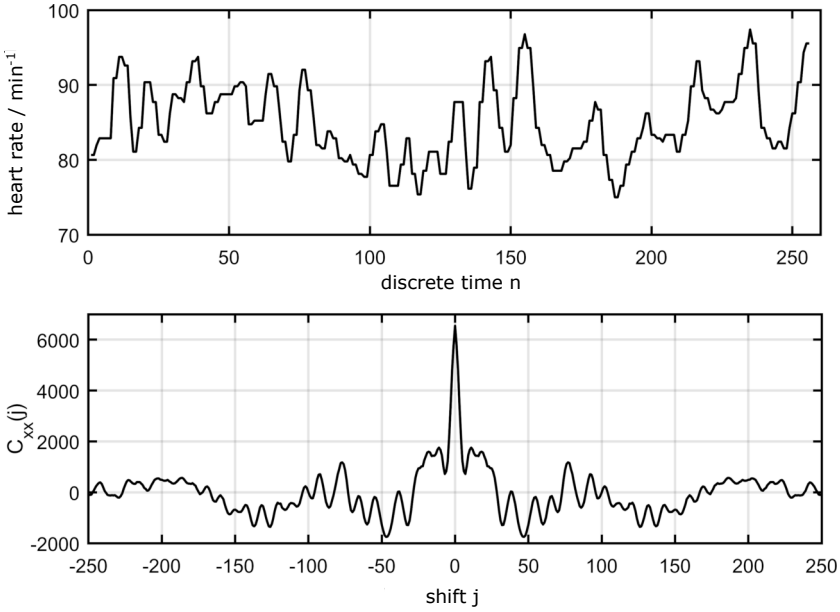


Fig. 5.14: Matlab-Representation of a human heart rate (top) with associated auto-covariance (bottom) using the supplementary material to [6].

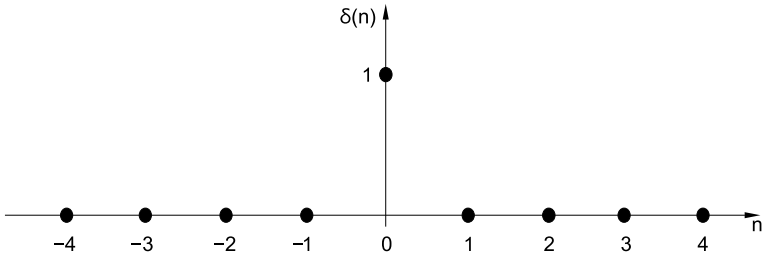


Fig. 5.15: Unit-pulse $\delta(n)$.

Here, a normally distributed mean-free random number sequence with 5000 values was generated with Matlab using the function `randn()` and then its auto-covariance was calculated with `xcov()` and normalised to a maximum value of one. You can see very clearly that this produces the function of a unit impulse (see Figure 5.15), which is approximated the better the more values the random number sequence has.

Another example shows Figure 5.14. Here the auto-covariance of the heart rate of a human is shown. The similarity to white noise disappears and one can see a periodic progression. This can be explained, for example, by the influence of periodic breathing, which also changes the heart rate periodically.

5.3.1.3 Linear and cyclic convolution

Convolution uses the impulse response to determine the output signal of a causal linear system from the input signal, as in the analogue case. In a linear discrete-time system, the *superposition principle* applies as in the analogue case, i.e., if two signals are applied to an input of a linear discrete-time system in succession and the associated output signals are generated and then added, the same output result is also obtained if the two input signals are added beforehand, the sum is applied to the input and the output signal is then determined. Or formulated in mathematical notation for discrete-time systems: If

$$y(n) = T_\mathrm{a}\{x(n)\}\,, \quad T_\mathrm{a}\text{: Operator}\,, \tag{5.45}$$

then

$$T_\mathrm{a}\left\{\sum_{\nu=1}^{N} k_\nu \cdot x_\nu(n)\right\} = \sum_{\nu=1}^{N} k_\nu \cdot T_\mathrm{a}\{x_\nu(n)\}\ . \tag{5.46}$$

For time-invariant systems, their property does not change with time. It does not matter when a certain input signal is given to the system. The same signal then always occurs at the output. This is not the case, for example, with mobile communication using a mobile phone. In the city centre, the reception is worse than in an open field. This transmission system is then time-variant with regard to its transmission properties. Then the spectrum is not only dependent on frequency, but also on time. To plan network coverage and standardisation, various models for the associated transmission channels were developed for this purpose.

However, in the calculation of the convolution, the Dirac momentum $\delta(t)$ is not needed as in the analogue case, but the *unit momentum*. However, the unit momentum is not obtained from the "sampling", but has the value 1 at the point $n = 0$ and is calculated according to

$$\delta(n) = \begin{cases} 0 & \text{for } n \neq 0 \\ 1 & \text{for } n = 0 \end{cases} \tag{5.47}$$

is defined. In contrast to the analogue $\delta(t)$, in this case no transition from a square pulse to an infinite pulse of vanishing width with an area of 1 is required. The digital unit impulse according to Equation 5.47 and Figure 5.15 thus has a value corresponding to the area of the analogue Dirac impulse and therefore has no mathematical peculiarity.

The digital impulse response $g(n)$ serves as the basis for the digital convolution. This is the output signal of a discrete-time system if the unit impulse $\delta(n)$ is present at its input (cf. Figure 5.16).

Because the superposition principle applies to a linear system, each discrete-time input signal $x(n)$ can also be expressed as a weighted sum of impulse responses. In this case, the discrete-time convolution sum is obtained:

$$y(n) = \sum_{\nu=-\infty}^{\infty} x(\nu)g(n-\nu) = \sum_{\mu=-\infty}^{\infty} x(n-\mu)g(\mu)\,, \quad \mu = n - \nu\,. \tag{5.48}$$

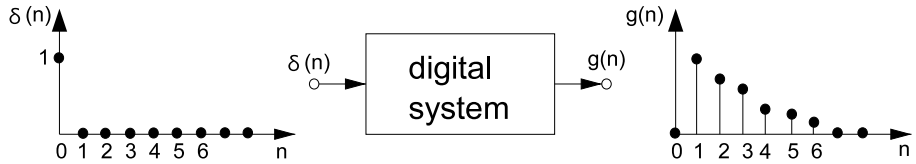


Fig. 5.16: Digital impulse response.

This convolution sum is generally expressed by the operator "$*$" as in analogue signals, i.e. for two discrete signals x_1 and x_2 it is defined as.

$$x_1(n) * x_2(n) := \sum_{\nu=-\infty}^{\infty} x_1(\nu)x_2(n-\nu)\,. \tag{5.49}$$

Explanatory Example
A causal time-invariant system has the finite impulse response

$$g(0) = 1\,; \quad g(1) = 0.75\,; \quad g(2) = 0.5\,; \quad g(3) = 0.25\,; \quad g(i) = 0 \quad \text{for } i > 3\,.$$

If a causal input signal $x(n)$ consists of the values

$$x(0) = -1\,; \quad x(1) = 0.5\,; \quad x(2) = -0.25\,; \quad x(i) = 0 \quad \text{for } i > 2\,,$$

then for the finite output signal $y(n)$ according to Equation 5.48 we obtain:

$$\begin{aligned}
y(0) &= x(0)g(0) &&= -1\\
y(1) &= x(0)g(1) + x(1)g(0) &&= -0.25\\
y(2) &= x(0)g(2) + x(1)g(1) + x(2)g(0) &&= -0.375\\
y(3) &= x(0)g(3) + x(1)g(2) + x(2)g(1) &&= -0.1875\\
y(4) &= x(1)g(3) + x(2)g(2) &&= 0\\
y(5) &= x(2)g(3) &&= -0.0625\,,
\end{aligned}$$

or by a scalar product using matrices:

$$\underbrace{\begin{bmatrix} y(0)\\ y(1)\\ y(2)\\ y(3)\\ y(4)\\ y(5)\end{bmatrix}}_{\vec{y}} = \underbrace{\begin{bmatrix} g(0) & 0 & 0 & & & \\ g(1) & g(0) & 0 & & & \\ g(2) & g(1) & g(0) & & & \\ g(3) & g(2) & g(1) & 0 & g(3) & g(2)\\ 0 & 0 & g(3) & & & \end{bmatrix}}_{\mathbf{Dr}\{g\}} \cdot \underbrace{\begin{bmatrix} x(0)\\ x(1)\\ x(2)\end{bmatrix}}_{\vec{x}} = \begin{bmatrix} -1\\ -0.25\\ -0.375\\ -0.1875\\ 0\\ -0.0625\end{bmatrix}\,.$$

The vectors x and y contain the values for the input and output signals respectively. The matrix $\mathbf{Dr}\{g\}$ is a *strip-triangular matrix* and contains per column the values of the impulse response, whose values are summarised in the vector $g := [g(0), g(1), \ldots, g(N-1)]'$. The values of the matrix $\mathbf{Dr}\{g\}$ in the i-th column are shifted downwards by one index compared to the values in the $i-1$-th column.

The output signal $y(n)$ has a total of $L = M + N - 1$ values for a finite impulse response $g(n)$ with N values and a finite input signal with M values. In our example, $N = 4$ and $M = 3$, so that one obtains the value $L = 3 + 4 - 1 = 6$ for its length. A corresponding graphical representation is shown in Figure 5.17.

In general, for a finite input signal with N values and an impulse response with M values, the discrete-time convolution can be expressed in matrix form by

$$\underbrace{\begin{bmatrix} y(0) \\ y(1) \\ y(2) \\ \vdots \\ \vdots \\ y(L-1) \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} g(0) & 0 & 0 \\ g(1) & g(0) & 0 \\ \vdots & g(1) & g(0) \\ g(N-1) & \vdots & g(1) \\ 0 & g(N-1) & \vdots \\ 0 & 0 & g(N-1) \end{bmatrix}}_{\mathbf{Dr}\{g\}} \cdot \underbrace{\begin{bmatrix} x(0) \\ \vdots \\ x(M-1) \end{bmatrix}}_{x} \tag{5.50}$$

For periodic signals, the convolution requires not the response of a system to *one* unit impulse, but that of an *periodic* impulse sequence of unit impulses with a certain period length N_p. The impulse response then gives rise to the pulse response as shown in Figure 5.18 (centre).

If the period length N_p is greater than the length N of the impulse response $g(n)$, then the impulse response $g_\mathrm{p}(n)$ is simply obtained from the string of impulse responses $g(n)$ repeated after N_p samples. If this is not the case, the individual impulse responses overlap and the impulse response cannot be determined again from the impulse response simply by fading out a period, see Figure 5.18 (below).

With a periodic input signal, the convolution sum (Equation 5.48) does not have to be formed from $n = -\infty$ to ∞. Here, as with the Fourier series, one period is sufficient, i.e.

$$y(n) = \sum_{\nu=0}^{N_\mathrm{p}-1} x(\nu) g_\mathrm{p}(n-\nu) = \sum_{\mu=0}^{N_\mathrm{p}-1} x(n-\mu) g_\mathrm{p}(\mu)\,, \quad \mu = n - \nu \tag{5.51}$$

or in matrix form:

$$\underbrace{\begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(N_\mathrm{p}-1) \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} g_\mathrm{p}(0) & g_\mathrm{p}(N_\mathrm{p}-1) & \cdots\, g_\mathrm{p}(1) \\ g_\mathrm{p}(1) & g(0) \cdots & \cdots\, g_\mathrm{p}(2) \\ \vdots & \ddots & \ddots \\ g_\mathrm{p}(N_\mathrm{p}-1) & g_\mathrm{p}(N_\mathrm{p}-2) & \cdots\, g_\mathrm{p}(0) \end{bmatrix}}_{\mathbf{Zykl}\{g_\mathrm{p}\}} \cdot \underbrace{\begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N_\mathrm{p}-1) \end{bmatrix}}_{x} . \tag{5.52}$$

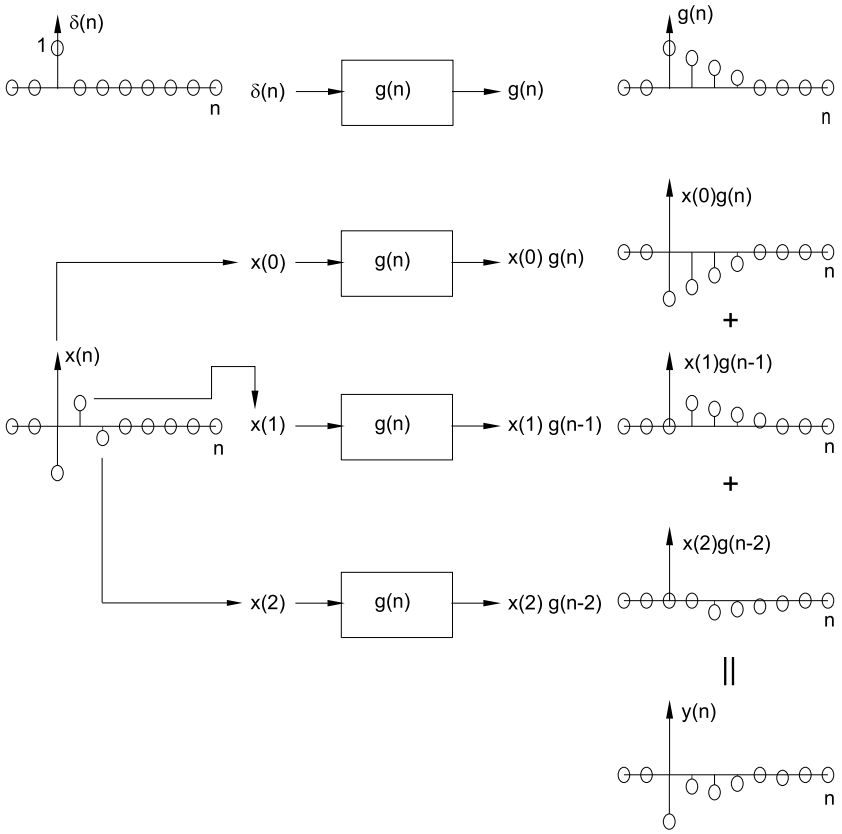


Fig. 5.17: Example of a discrete convolution according to section 5.3.1.3 calculated by a weighted sum of shifted impulse responses.

The matrix **cycl**$\{g_\mathrm{p}\}$ is cyclic. In contrast to linear convolution, the matrix elements of the i-th column are also shifted downwards by one index with respect to the elements of the $i-1$-th column, but the element of the $i-1$-th column that is omitted by the downward shift is added again at the top of the i-th column, because there may only be N_p elements per column.

Explanatory Example

Let a periodically time-invariant system have, at a period length of $N_\mathrm{p} = 3$, the pulse response

$$g_\mathrm{p}(0) = 1\,; \quad g_\mathrm{p}(1) = 0.75\,; \quad g_\mathrm{p}(2) = 0.5$$

and the periodic input signal

$$x(0) = 2\,; \quad x(1) = 1\,; \quad x(2) = 0.5\,.$$

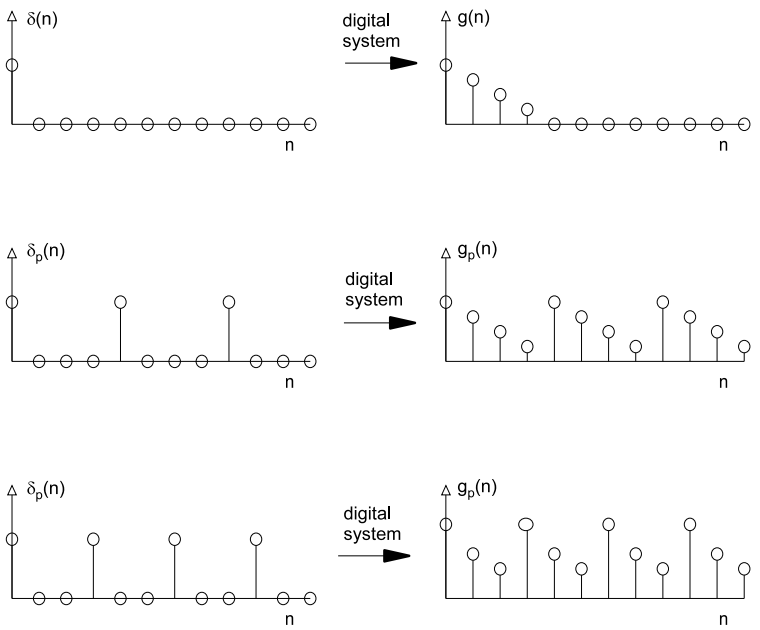


Fig. 5.18: Discrete-time impulse response with a period length N_p: Discrete impulse response $g(n)$ with $N = 4$ values (top), impulse response $g_p(n)$, $N_p = 4$ without $g(n)$ overlap (middle), impulse response $g_p(n)$, $N_p = 3$ with $g(n)$ overlap (bottom).

Then for the finite output signal $y(n)$ according to Equation 5.51 we obtain:

$$\begin{aligned} y(0) &= x(0)g_p(0) + x(1)g_p(2) + x(2)g_p(1) = 3.125 \\ y(1) &= x(0)g_p(1) + x(1)g_p(0) + x(2)g_p(2) = 2.75 \\ y(2) &= x(0)g_p(2) + x(1)g_p(1) + x(2)g_p(0) = 2.25\ , \end{aligned}$$

respectively for the vectors x, y and the matrix $\mathbf{cycl}\{g_p\}$:

$$\underbrace{\begin{bmatrix} y(0) \\ y(1) \\ y(2) \end{bmatrix}}_{\vec{y}} = \underbrace{\begin{bmatrix} g_p(0) & g_p(2) & g_p(1) \\ g_p(1) & g_p(0) & g_p(2) \\ g_p(2) & g_p(1) & g_p(0) \end{bmatrix}}_{\mathbf{Zykl}\{g_p\}} \cdot \underbrace{\begin{bmatrix} x(0) \\ x(1) \\ x(2) \end{bmatrix}}_{\vec{x}} = \begin{bmatrix} 3.125 \\ 2.75 \\ 2.25 \end{bmatrix}\ .$$

5.3.2 Signal Analysis in the Frequency Domain

In section 5.2 various transformations (especially the Fourier-transformations) were described with which a signal can be described not only as a function of time but

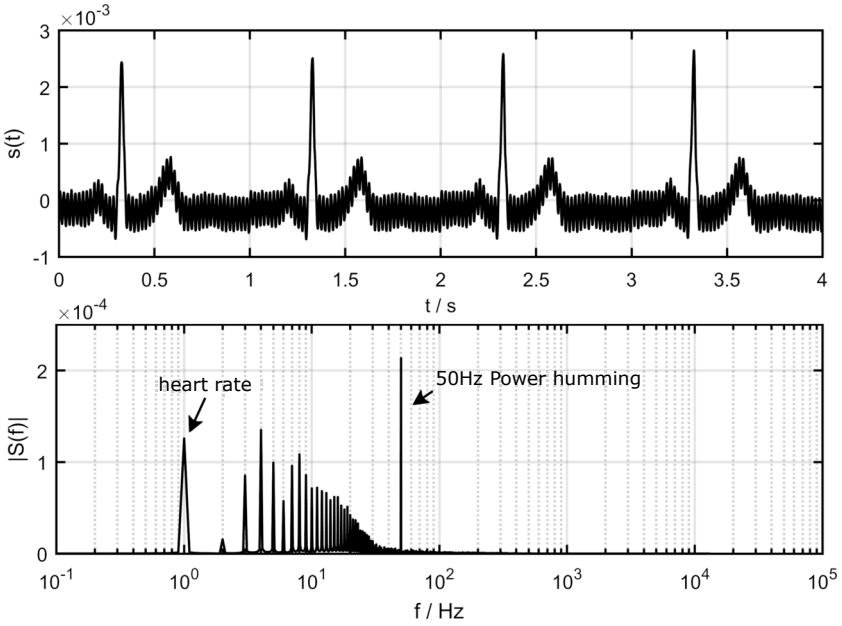


Fig. 5.19: The ECG with a 50 Hz mains hum shows the time domain (top) and the corresponding spectrum (bottom). At 50 Hz, a distinct spectral line of over 200 µV can be seen due to the 50 Hz mains hum (simulation result was generated using LTspice).

also as a function of other variables, e. g. the frequency f or $z := e^{j2\pi fT}$. With this representation in the "image domain", e. B. in f or z, disturbing signal components such as e. B. the 50 Hz mains hum, border areas or a baseline fluctuation as well as shifts are easier to recognise than in the time domain. It can therefore be used as an important basis for filter design.

In Figure 5.19, for example, the section of an ECG is shown above, which is disturbed by a sinusoidal oscillation; the corresponding representation in the frequency domain is shown below. Here it is very clearly visible that the heart rate (first larger spectral line) is at 1 Hz, i.e. the pulse has a value of 60 beats per minute. Furthermore, it can be seen that there is a distinct spectral line at 50 Hz, which is due to the influence of a 50 Hz mains hum. In the time domain (above), this disturbance can also be recognised, but its nature cannot be determined so well, which is better in the frequency domain. Since the spectral components of the 50 Hz interference are relatively far away from the frequency range of the ECG, they could be well suppressed by a bandstop with a centre frequency of 50 Hz, if its bandwidth is narrow enough so that no ECG components are lost.

For a linear time-invariant system, the impulse response can be used in the time domain to determine the output signal using discrete convolution. This is often much easier when determining the output signal for a given input signal in the "image do-

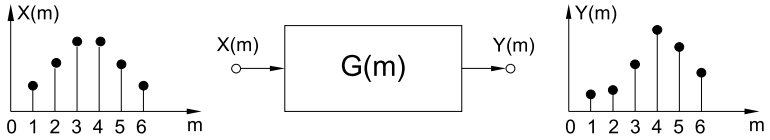


Fig. 5.20: Discrete-time linear system with representation in the image area by its transfer function.

main". E.g. a sum of sine waves is used to represent a periodic signal with the help of the Fourier series. If these are applied to the input of a linear time-invariant system, a sine wave with the same frequency but with different amplitude and phase is also obtained at the output from each individual sine wave of the input signal. These differences are described according to magnitude and phase by the complex transfer function $G(m) = |G(m)|e^{j\varphi(m)}$ (cf. Figure 5.20). This transfer function is also discrete for a periodic input signal in the image domain. The relation between input and output signal can then be described for a period with N_p frequency points by the matrix-equation

$$\underbrace{\begin{bmatrix} Y(1) \\ Y(2) \\ \vdots \\ Y(N_\mathrm{p}) \end{bmatrix}}_{Y} = \underbrace{\begin{bmatrix} G(0) & 0 & \cdots & 0 \\ 0 & G(2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & G(N_\mathrm{p}) \end{bmatrix}}_{\mathbf{G}} \cdot \underbrace{\begin{bmatrix} X(1) \\ X(2) \\ \vdots \\ X(N_\mathrm{p}) \end{bmatrix}}_{X} \tag{5.53}$$

X and Y are vectors and contain the input and output signals, $\mathbf{G}$ is a $\mathbf{N}_\mathrm{p} \times \mathbf{N}_\mathrm{p}$ diagonal matrix whose diagonal elements embody the values of the transfer function. With the help of the matrix form for the Fourier-transformation in Equation 5.24, the back transformation into the time domain can now be carried out simply by applying the matrix calculation, i.e., the transfer function can be transformed back into the time domain. i.e. because of $x = \mathbf{W}^{-1} \cdot X$ and $Y = \mathbf{W} \cdot y$ it follows from

$$Y = \mathbf{W} \cdot y = \mathbf{G} \cdot X = \mathbf{G} \cdot \mathbf{W} \cdot x \tag{5.54}$$

and by left multiplication with the inverse Fourier-matrix $\mathbf{W}^{-1}$:

$$y = \underbrace{\mathbf{W}^{-1} \cdot \mathbf{G} \cdot \mathbf{W}}_{\mathbf{Zykl}\{g_\mathrm{p}\}} \cdot x \,. \tag{5.55}$$

Thus, by matrix calculation, a relation between input and output signal in the time domain is obtained which is identical to Equation 5.52.[7]

Explanatory Example

For a simple illustration of the representation in the frequency domain with filtering of spectral components there, a sinusoidal oscillation with 4 kHz, an amplitude of

7 It can be mathematically proved that by multiplying a diagonal matrix from the left by the inverse Fourier-matrix $\mathbf{W}^{-1}$ from the right by the Fourier-matrix $\mathbf{W}$, an *cyclic* matrix is formed.

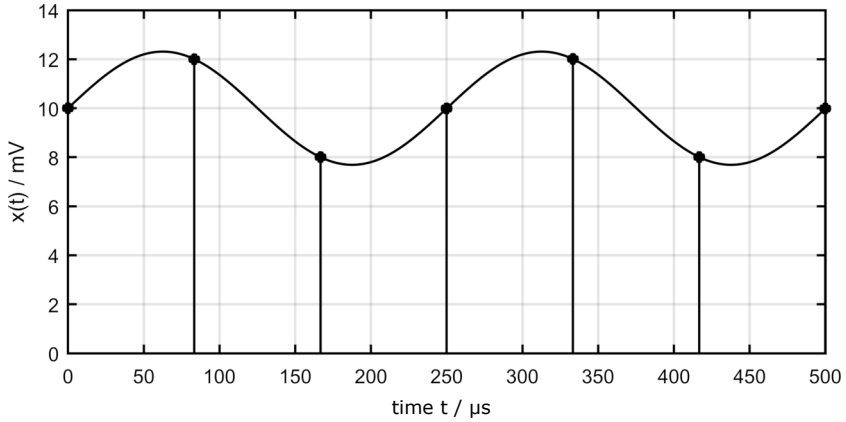


Fig. 5.21: Sine wave of 4 kHz with a DC component of 10 mV sampled at a frequency of 12 kHz.

2.31 mV and a DC component of 10 mV according to

$$x(t) = 10\,\text{mV} + 2.31\,\text{mV} \cdot \sin(2\pi\, 4\,\text{kHz}\, t)$$

with 12 kHz sampled as well as using a *ideal* discrete-time high-pass filter with a purely real transfer function and a cut-off frequency of 2 kHz to suppress the DC component.

With a sampling frequency of 12 kHz one obtains the following three values per period:

$$x(t = 0\,\mu\text{s}) = x(0) = 10$$
$$x(t = 83.33\,\mu\text{s}) = x(1) = 12\; x(t = 166.66\,\mu\text{s}) = x(2) = 8\,.$$

Since this is a periodic signal, the calculation of the spectrum can be done using the Fourier-matrix and we obtain as in the example of Equation 5.26:

$$\underbrace{\begin{bmatrix} X(0) \\ X(1) \\ X(2) \end{bmatrix}}_{X} = \underbrace{\begin{bmatrix} 1 & 1 & 1 \\ 1 & e^{-j2\pi/3} & e^{-j4\pi/3} \\ 1 & e^{-j4\pi/3} & e^{-j8_1/3} \end{bmatrix}}_{\mathbf{W}} \cdot \underbrace{\begin{bmatrix} x(0) = 10 \\ x(1) = 12 \\ x(2) = 8 \end{bmatrix}}_{x} = \begin{bmatrix} 30 \\ -j3.64 \\ j3.64 \end{bmatrix}.$$

The periodic spectrum shows Figure 5.22. It can be seen that the DC component is evident from a real spectral line at frequency $f = 0$ Hz. In order to suppress this, the discrete-time high-pass filter must have a transfer function $\mathbf{G}$ that does not pass a DC component. At the other frequencies a gain of one can be chosen, i.e. with the relation between output and input spectrum according to Equation 5.53 $G(0) = 0$, $G(1) = 1$ and $G(2) = 1$, i.e.

$$\underbrace{\begin{bmatrix} Y(0) \\ Y(1) \\ Y(2) \end{bmatrix}}_{Y} = \mathbf{G} \cdot X = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 30 \\ -j3.64 \\ j3.64 \end{bmatrix} = \begin{bmatrix} 0 \\ -j3.64 \\ j3.64 \end{bmatrix}.$$

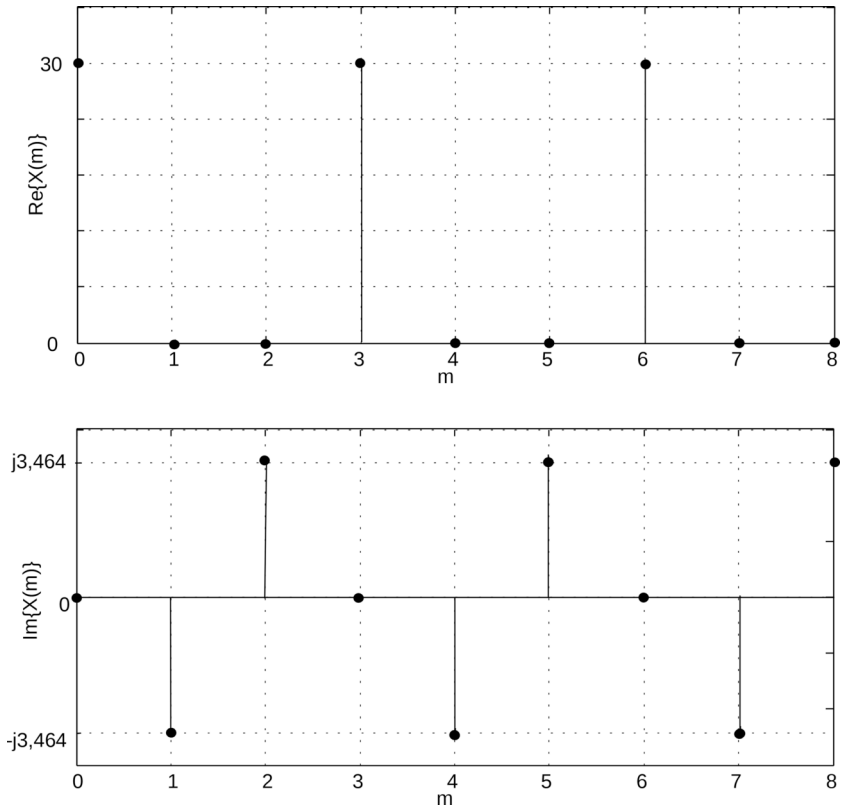


Fig. 5.22: Spectrum of the signal according to Figure 5.21 shows the real part (top) and the imaginary part (bottom). The spectrum is periodic with the sampling frequency of 12 kHz and has three spectral lines within one period, i.e. in the first period at $m = 0$, $f = 0$ kHz, $m = 1$, $f = 4$ kHz and $m = 2$, $f = 8$ kHz.

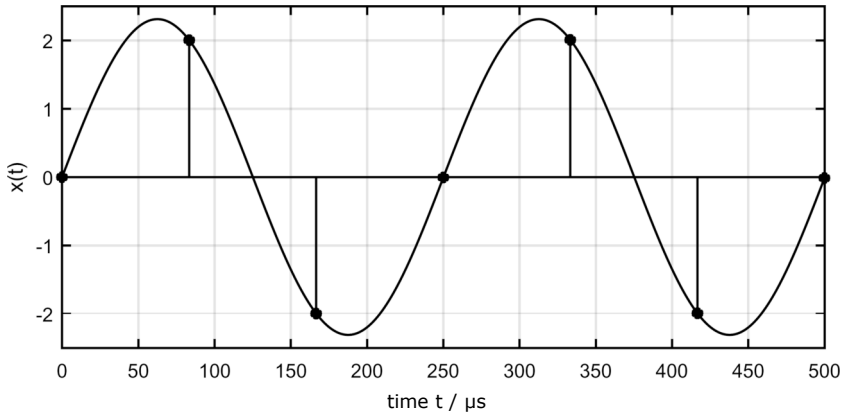


Fig. 5.23: Filtered sinusoidal signal with DC component, cf. Figure 5.21.

Then by applying the inverse Fourier-transform to the spectrum of the output signal Y according to Equation 5.24 in the time domain, it follows:

$$y = \mathbf{W}^{-1} \cdot Y = \underbrace{\frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & e^{j2\pi/3} & e^{j4\pi/3} \\ 1 & e^{j4\pi/3} & e^{j8\imath/3} \end{bmatrix}}_{\mathbf{W}^{-1}} \cdot \underbrace{\begin{bmatrix} 0 \\ -j3.64 \\ j3.64 \end{bmatrix}}_{Y} = \begin{bmatrix} 0 \\ 2 \\ -2 \end{bmatrix}.$$

From the values 10, 12 and 8 of the sampled sinusoidal input signal with a DC component, we have now obtained the values 0, 2 and −2 for the output signal after filtering with an ideal high-pass filter, which corresponds to a sampled sinusoidal signal *without* a DC component, cf. Figure 5.23.

Note

The output signal $y(n)$ can also be determined directly by cyclic convolution. However, for this one needs the pulse response $g(n)$[8], which are obtained analogously to continuous-time signals by Fourier-back-transforming the transmission vector $G := [G(0), G(1), G(2)]^{\mathrm{T}}$:

$$\underbrace{\begin{bmatrix} g(0) \\ g(1) \\ g(2) \end{bmatrix}}_{g} = \underbrace{\frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & e^{j2\pi/3} & e^{j4\pi/3} \\ 1 & e^{j4\pi/3} & e^{j8\imath/3} \end{bmatrix}}_{\mathbf{W}^{-1}} \cdot \underbrace{\begin{bmatrix} G(0) = 0 \\ G(1) = 1 \\ G(2) = 1 \end{bmatrix}}_{G} = \frac{1}{3}\begin{bmatrix} 2 \\ -1 \\ -1 \end{bmatrix}.$$

Now according to Equation 5.55 the cyclic convolution can be performed:

$$\underbrace{\begin{bmatrix} y(0) \\ y(1) \\ y(2) \end{bmatrix}}_{y} = \underbrace{\frac{1}{3}\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}}_{\mathbf{cycle}\{g\}} \cdot \underbrace{\begin{bmatrix} 10 \\ 12 \\ 8 \end{bmatrix}}_{x} = \begin{bmatrix} 0 \\ 2 \\ -2 \end{bmatrix}.$$

As can be seen, the results obtained by Fourier-back transformation and cyclic convolution are identical. In contrast to the continuous-time domain, the cyclic convolution is less complex in the digital representation and much easier to calculate.

In many measured biosignals, the unwanted interference signal does not only consist of a sinusoidal 50 Hz mains hum, as shown in Figure 5.19, which can be easily detected in the frequency range and suppressed with the help of a filter. Noise-like interference signals (e.g. from electromagnetic radiation sources such as neon lamps, TV and radio transmitters or mobile phones) also occur, which cannot be described precisely and of which only the power or statistical quantities such as the correlation can be measured. These quantities can also be investigated in the frequency domain.

8 Since it is known that signal and transfer function are periodic, in this example the index p at g_{p} and $\mathbf{G}_{\mathrm{p}}$ is omitted for simplification.

Between the auto-correlation and the power density[9] (or energy density) there is a close relationship which can be described as follows:

The convolution sum (Equation 5.49) can be generally interpreted, as in the calculation of the output signal by convolution of an input signal with the impulse response of a digital system (cf. Figure 5.16), such that one signal, e.g. x_1, is the input signal and the other, x_2, is the impulse response of a linear digital system. In the frequency domain, this means that to determine the output spectrum X_2, the input spectrum is multiplied by the transfer function **G**, or in abbreviated notation:

$$x(\tilde{n}) := x_1(\tilde{n}) * x_2(\tilde{n}) \circ\!\!-\!\!\bullet X_1(m) \cdot X_2(m) \,, \tag{5.56}$$

with

$$x_1(\tilde{n}) * x_2(\tilde{n}) := \sum_{\tilde{\nu}=-\infty}^{\infty} x_1(\tilde{\nu}) x_2(\tilde{n} - \tilde{\nu}) \,. \tag{5.57}$$

Using the substitution $n := -\tilde{\nu}$ and $\nu := \tilde{n}$ we get:

$$\sum_{n=\infty}^{-\infty} x_1(-n) x_2(\nu + n) = \sum_{n=-\infty}^{\infty} x_1(-n) x_2(\nu + n) \,. \tag{5.58}$$

If $x_1(-n) := x(n)$ and $x_2(n + \nu) := x(n + \nu)$ still hold, it follows.

$$x(\tilde{n}) = \sum_{n=-\infty}^{\infty} x_1(-n) x_2(\nu + n) = \sum_{n=-\infty}^{\infty} x(n) x(\nu + n) = \text{Ŕ}_{XX}(\nu) \,. \tag{5.59}$$

Result

The mean auto-correlation function $\overline{R}_{XX}(\nu)$ can therefore be interpreted as a convolution of the function $x_1(n) = x(-n)$ with the function $x_2(n) = x(n)$.

Insertion into Equation 5.56 further gives the relation between the time and frequency domains:

$$\text{Ŕ}_{XX}(\nu) \circ\!\!-\!\!\bullet X_1(m) \cdot X_2(m) = F(x(-\nu)) \cdot \mathcal{F}(x(\nu)) \,. \tag{5.60}$$

Because of

$$\begin{aligned} x(n) &\circ\!\!-\!\!\bullet X(m) \quad \text{and} \\ x(-n) &\circ\!\!-\!\!\bullet X^*(m) \quad \text{(similarity theorem)} \end{aligned}$$

then follows from this as well:

$$\overline{R}_{XX}(\nu) \circ\!\!-\!\!\bullet X^*(m) X(m) = |X(m)|^2 = S_{XX}(m) \tag{5.61}$$

with $S_{XX}(m)$ as the spectral power or energy density.

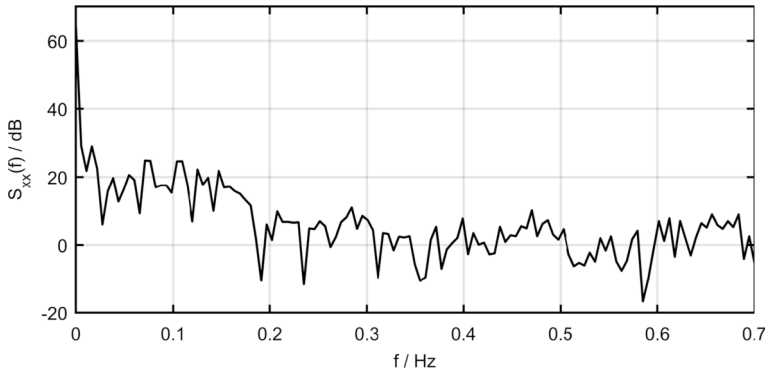


Fig. 5.24: Power-density calculation of heart-rate-variability according to the example given in Figure 5.14 using the magnitude square of the fast Fourier-transformation.

Further result

The Fourier transform of the mean auto-correlation function $\overline{R}_{XX}(\nu)$ is the spectral power or energy density $S_{XX}(m)$, i.e. analogous to Equation 5.23 applies for discrete signals in a range with N values (e.g. periodic signal or clipping by windowing):

$$\overline{R}_{XX}(\nu) = \frac{1}{N}\sum_{m=0}^{N-1} S_{XX}(m)e^{j2\pi\nu m/N}\,, \quad S_{XX}(m) = \sum_{\nu=0}^{N-1} \overline{R}_{XX}(\nu)e^{-j2\pi\nu m/N}\,. \tag{5.62}$$

This means that from the measurement of a statistically distributed interference signal, its power spectral density can be determined by means of auto-correlation or auto-covariance. From this, the magnitude spectrum $|X(m)|$ can be determined due to $S_{XX}(m) = |X(m)|^2$, but not the phase course. For example, Figure 5.24 shows the power density of the changes in the pulse rate shown in Figure 5.14. One can clearly see a high DC component (spectral component at $f = 0$ Hz) resulting from the fact that the pulse rate varies around a certain mean value, about 84 beats per minute, and this variation is mainly due to the fact that the pulse rate varies up to approx. 0.3 Hz is present. To calculate this, one can apply the fast Fourier-transform to the heart rate variability in Matlab using the function `fft()` and then determine the power spectral density according to Equation 5.61 by squaring the magnitude spectrum. Or one can use the function `xcorr()` to first determine the auto-correlation and then use the fast Fourier-transformation and thereby calculate the power spectral density without squaring. A third possibility is to apply the function `periodogram()` to the heart variability, which is the simplest, because the heart rate variability can be passed to it directly as an argument, and one does not need to calculate either an auto-correlation or the square of a Fourier-transformation.

9 The power density gives the power frequency for a signal with finite power (e.g. sinusoidal oscillation), the energy density gives the power frequency for a signal with finite energy (e.g. impulse).

5.3.3 Signal Analysis in the Time-Frequency Domain

$$f(t) = \int_{-\infty}^{\infty} F(f) e^{j2\pi ft} \mathrm{d}f \,, \quad F(f) = \int_{-\infty}^{\infty} f(t) e^{-j2\pi ft} \mathrm{d}t \,. \tag{5.63}$$

This integral quantity does not obtain a statement about when a particular event occurred in the signal waveform, but only a spectral average. E.g. one would have to record a phonocardiogram (PCG) from beginning to end and would not know afterwards when a certain heart sound occurred. This is of course very unfavourable for heart sound recognition, and ideally the doctor would like to know not only the exact frequency but also the time of occurrence. But this is not possible for the following reasons:

In order to examine a certain time interval around a certain time t_0, a signal should only be present in this range. This can be achieved by multiplying the signal by a window function $w(t - t_0)$ which fades out this range. I.e. the faded out signal $f_w(t)$ then results in

$$f_w(t) = w(t) \cdot f(t) \,. \tag{5.64}$$

In general, $w(t)$ is a function that is symmetric around time t_0, e.g. a square or Gaussian function. In the frequency domain, the spectrum must then be convolved with the frequency transformed function $W(t)$ of the window function $w(t)$; i.e.

$$F_w(f) = W(f) * F(f) \,. \tag{5.65}$$

The spectrum now created by convolution naturally also has a width that is influenced by the width of the spectrum of the window function $W(f)$. The spectrum of the window function $W(f)$ can have a similar shape as the window function $w(t)$ in the time domain. If, for example, $w(t)$ can be described by a Gaussian pulse, this can also be done for the associated spectrum $W(f)$ – a Gaussian pulse $i(t)$ in the time domain also produces a Gaussian pulse $I(f)$ in the frequency domain. Unfortunately, the width of the spectrum of the window function in the frequency domain is large when the width of the window function $w(t)$ in the time domain is small, and vice versa. This can be shown by examining a pulse width in the time and frequency domains. If the pulse width is defined according to the width of an equal-area rectangular pulse (cf. Figure 5.25), it follows:

$$T = \frac{1}{i(0)} \cdot \int_{-\infty}^{\infty} i(t) \, \mathrm{d}t \,, \tag{5.66}$$

$$B = \frac{1}{I(0)} \cdot \int_{-\infty}^{\infty} I(f) \, \mathrm{d}f. \tag{5.67}$$

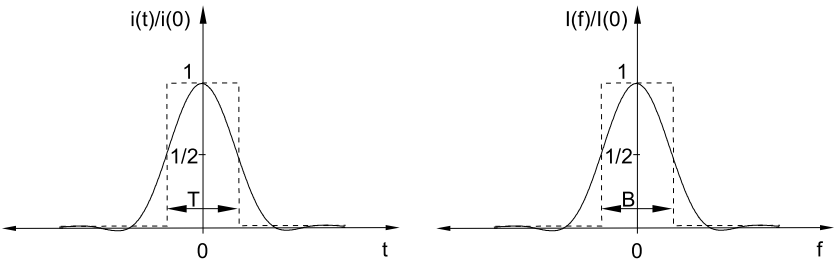


Fig. 5.25: Pulse-width in time and frequency domain defined over equal areas of a rectangular pulse.

By applying the Fourier-transformation (Equation 5.63), for the time $t = 0$ and the frequency $f = 0$, we obtain

$$I(0) = \int_{-\infty}^{\infty} i(t)e^{-j0t}\mathrm{d}t = \int_{-\infty}^{\infty} i(t)\,\mathrm{d}t\,, \tag{5.68}$$

$$i(0) = \int_{-\infty}^{\infty} I(f)e^{j2\pi f0}\mathrm{d}f = \int_{-\infty}^{\infty} I(f)\,\mathrm{d}f\,, \tag{5.69}$$

and you get the result

$$I(0) = T \cdot i(0)\,; \quad i(0) = B \cdot I(0)\,. \tag{5.70}$$

Substituting finally yields the sought relationship between the pulse widths in the time and frequency domain

$$B \cdot T = 1\,. \tag{5.71}$$

Thus the bandwidth B of a pulse is equal to the reciprocal of its temporal pulse width T and vice versa. A narrow window in the time domain therefore produces a wide window in the frequency domain. So if one wants to know exactly when, for example, a heart sound occurs, this can be done by continuously shifting a time window by varying t_0, but the frequency resolution is very small. However, if one wishes to examine the frequency of a heart sound with its frequency components in detail, it is not possible to determine exactly when it occurred.

5.3.3.1 The Short-Time Fourier Transform (STFT)

In the short time-Fourier-transformation (STFT), a uniform time function $w(t-t_0)$ with the same width is multiplied by the signal $f(t)$ to be investigated and the centre of the window function t_0 is continuously shifted until the entire signal waveform has been windowed out and investigated. Independent of t_0, the width of the window function remains constant. In the discrete-time case, the STFT can be calculated analogously to the discrete Fourier-transformation according to Equation 5.23, but here it is not necessary to sum up over the entire period length with N samples, but only over the window width with an odd number of N_F samples. The number of samples N_F within

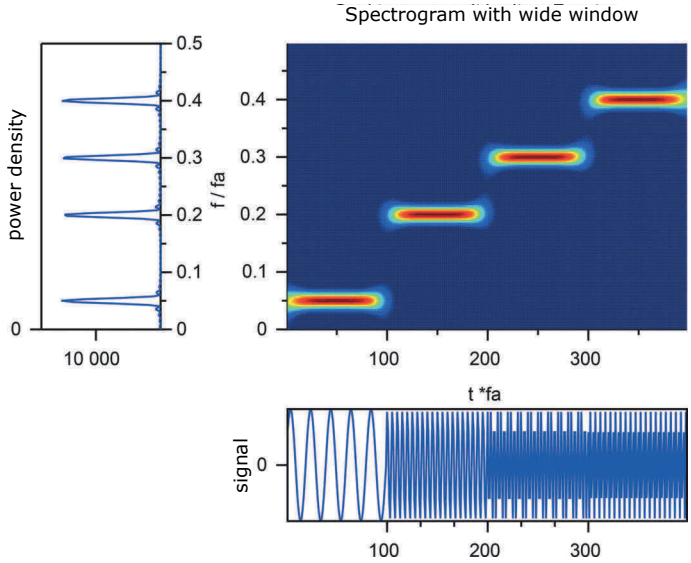


Fig. 5.26: Spectrogram (power density spectrum as a function of window range) of a signal with four sinusoidal oscillations with 1, 4, 6 and 8 Hz in an interval of 20 s and a sliding Gaussian window with a width of 4 s.

the window is therefore chosen to be odd so that the window is symmetrical about its centre $i_M := (N_F + 1)/2$. Within the entire period with N samples, the window centre shall lie at sample $i = i_F$. From this follows for the digital STFT:

$$f_p(i, i_F)w(i - i_F) = \frac{1}{N}\sum_{l=0}^{N-1} \mathrm{STFT}_{f_p}^{(w)}(i_F, l)e^{j2\pi il/N} \tag{5.72}$$

$$\mathrm{STFT}_{f_p}^{(w)}(i_F, l) = \sum_{i=i_F-i_M+1}^{i_F+i_M-1} f_p(i)w(i - i_F)e^{-j2\pi il/N} \tag{5.73}$$

where $i_F = i_M/2$ to $N - 1 - i_M/2$.

In principle, it is now a two-dimensional Fourier transformation; because the location of the window i_F has now been added as an additional variable. The original signal $f_p(i)$ is obtained if the original spectrum is also used in the reverse transformation. Since the Fourier transform of a sum is the sum of the Fourier transforms of the individual summands, the original signal $f_p(i)$ could also be obtained by summing over the Fourier transforms with *all* windows when the individual windows would gradually fade out the *total* spectrum.

Another possibility is that in the frequency domain the spectrum is multiplied by a window function $H(l)$ that cancels the effect of the temporal window $w(i - i_F)$ [33], ie.

$$f_p(i) = \frac{1}{N}\sum_{l=0}^{N-1} STFT_{f_p}^{(w)}(i_F, l)H(l)e^{j2\pi il/N}\ . \tag{5.74}$$

The Fourier-transform of this window function $h(i) := \mathcal{F}\{H(l)\}$ is exactly the inverse of the window function $w(i)$ and must satisfy the condition

$$\sum_{i=-\infty}^{\infty} h(i)w(i) = 1 \tag{5.75}$$

Examples of spectrograms (power density spectra as a function of the position of a time window) obtained using STFT with two Gaussian windows of different widths are shown in Figure 5.26 and Figure 5.27. They were calculated with the function `tfrsp()` from the toolbox `Time Frequency` for Matlab and Octave. This toolbox is free and can be downloaded from http://tftb.nongnu.org/. The Gaussian window function can be calculated in the discrete-time domain by

$$w(n) := e^{-\frac{1}{2}\left(\frac{n-i_M}{\sigma \cdot i_M}\right)^2}, \quad \text{für } n = 0, \ldots, N_F - 1 \tag{5.76}$$

are described [32]. σ is a measure of the width of this window.

Below the spectrogram, the signal sampled at 20 Hz is shown. To the left of the spectrogram the complete spectrum (without window function) is shown, where the spectral lines at the four frequencies (1, 4, 6 and 8 Hz) with the wide Gaussian window of 4 s match well with the horizontal lines of the spectrogram. The frequency can therefore be determined well. However, the width of these lines does not reflect the locations of the power density fluctuations within the duration of a single sinusoidal oscillation, which can be explained by the uncertainty condition according to Equation 5.70. A good frequency resolution means a poor time resolution and vice versa. With the narrow Gaussian window with a width 0.5 s, the frequency of the four sine oscillations can no longer be determined exactly from the spectrogram. However, it can be determined exactly when the minima and maxima of the sinusoidal oscillation with 1 Hz occur with the bar at the bottom left, because the power density is greatest there. Due to the poorer frequency resolution, one now obtains a better time resolution than with the wide window in Figure 5.26.

5.3.3.2 The Discrete Wavelet Transform

A disadvantage of the short-time-Fourier-transform is that the window width in the time domain always has a constant value. If the frequency in a time domain is to be determined with a certain accuracy, enough samples would have to lie within the time window. For low-frequency signals, a sufficiently wide window is needed so that this oscillation can be sampled sufficiently often per period (uncertainty principle of signal processing). A higher-frequency oscillation does not need this wide time window and could also be analysed by a shorter one.

As in the continuous-time case, the discrete-time wavelet transformation (DWT) provides a remedy, whereby the time window depends on the frequency to be investigated. The higher this frequency, the narrower the time window can be and vice versa (see chapter 2).

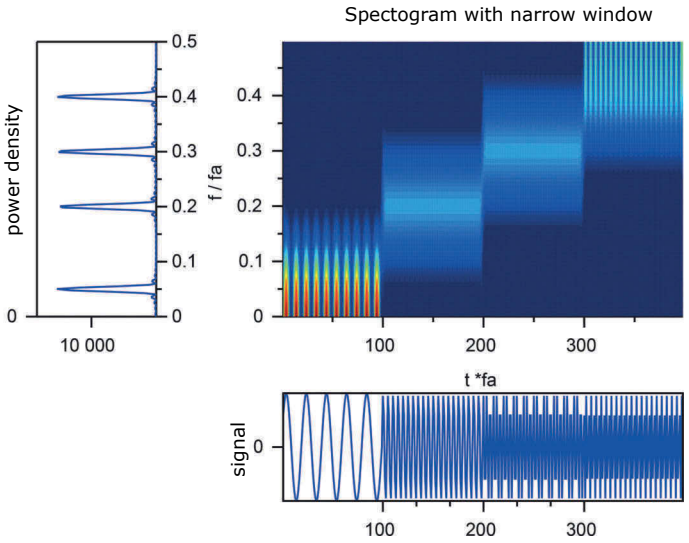


Fig. 5.27: Spectrogram as in Figure 5.26 but with a narrow Gaussian window with a temporal width of 0.5 seconds.

The discrete Fourier transform according to Equation 5.23

$$\begin{aligned} F(l) &= \sum_{i=0}^{N-1} f_{\mathrm{p}}(i) e^{-j2\pi il/N} \\ &= \sum_{i=0}^{N-1} f_{\mathrm{p}}(i) \cos\left(\frac{2\pi il}{N}\right) + j \sum_{i=0}^{N-1} f_{\mathrm{p}}(i) \sin\left(\frac{2\pi il}{N}\right), \end{aligned}$$

can in fact also be understood as a filter bank, where for each frequency to be determined with $f_l = l \cdot f_{\mathrm{a}}/N$ with $f_{\mathrm{a}} = 1/T_{\mathrm{a}}$ (sampling frequency) and $l = 1, \ldots, N-1$ a filter is used. This is because the calculation of the real and imaginary parts corresponds to a cross-correlation (cf. subsubsection 5.3.1.2) between the signal $f_{\mathrm{p}}(i)$ and the functions $\cos(2\pi il/N)$ and $\sin(2\pi il/N)$ with

$$\mathcal{R}_{xy_1}(l, m) := \frac{1}{N} \sum_{i=0}^{N-1} \underbrace{f_{\mathrm{p}}(i)}_{x(i)} \underbrace{\cos\left(\frac{2\pi(i+m)l}{N}\right)}_{y_1(i+m)} \quad \text{und}$$

$$\mathcal{R}_{xy_2}(l, m) := \frac{1}{N} \sum_{i=0}^{N-1} \underbrace{f_{\mathrm{p}}(i)}_{x(i)} \underbrace{\sin\left(\frac{2\pi(i+m)l}{N}\right)}_{y_2(i+m)}, \quad m = 0.$$

The cross-correlation can equally also be interpreted as a filtering of two signals, since a correlation corresponds to a convolution, see subsection 5.3.2. The transformations are therefore "filter banks" that apply a specific filter to the signal under investigation for each pixel (e.g. a specific frequency). Unfortunately, the wavelet-filter", such as

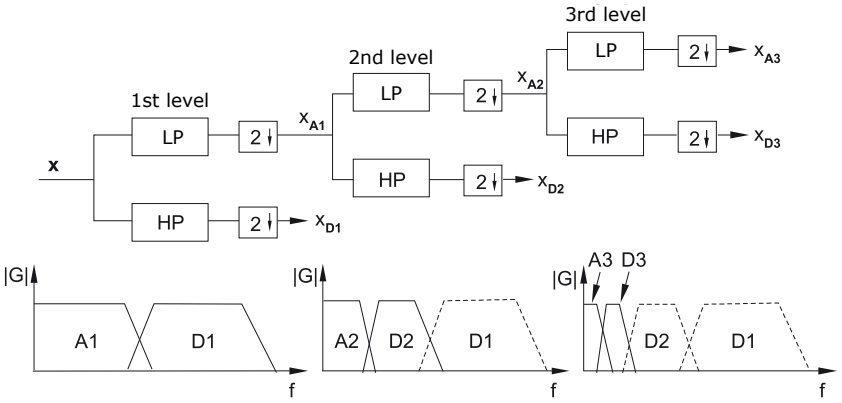


Fig. 5.28: Principle of a digital wavelet-transformation (DWT) up to level 3 using subband coding [46].

the wavelet according to Morlet or the Mexican-Hat have very large overlaps and thus a lot of redundancy [33]. Less expensive wavelets can be obtained by applying subband coding, which has long been used in message transmission in the form of the quadrature-mirror-filter (QMF) [38].

In 1988, Stephane Mallat and Yves Meyer developed a method based on a digital high-pass and low-pass filter pair, in which the algorithm is the same as in the subband coding known in digital signal processing with quadrature-mirror-filter [46]. They showed that continued digital filtering of the low-pass component produces sets of coefficients corresponding to a wavelet-decomposition. The wavelet itself is the impulse response of a high-pass, the scaling function that of a low-pass. The principle of this procedure presented in Figure 5.28 is:

1. The frequency range of the signal $x(i)$ to be investigated by wavelets is divided in the middle in the 1st stage, whereby the low-frequency part is generated by a low-pass filter and the higher-frequency part by a high-pass filter. However, these parts of the signal only have half the bandwidth each and can therefore also be reconstructed with half the samples. This is achieved by discarding every second sample (downsampling). Altogether, the partial signals x_{A1} and x_{D1} are now created. The low-frequency signal x_{A1} is also referred to in the literature as a scaling function, the higher-frequency x_{D1} as a wavelet.
2. The frequency range of the low-frequency signal x_{A1} is now divided further in the middle in the 2nd stage, divided into a low-frequency and a higher-frequency range by the same low-pass and high-pass filters –but this time with half the cut-off frequency –and then further processed with a quarter of the sampling frequency due to the bandwidth being half smaller (after downsampling again with a factor of 2). The signals x_{A2} and x_{D2} are now produced.

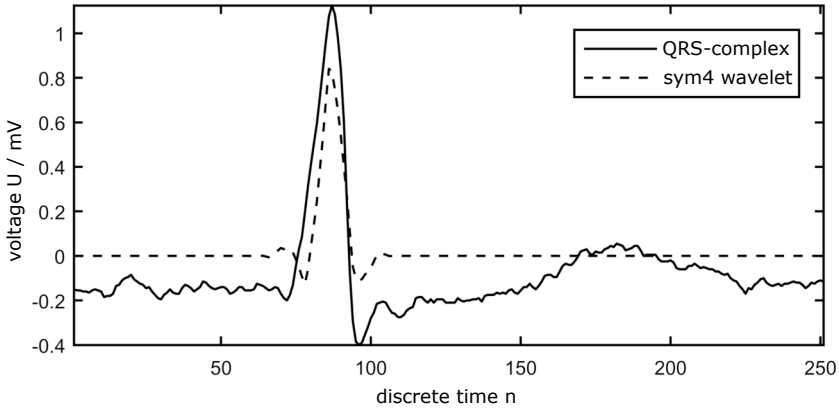


Fig. 5.29: QRS complex compared with a wavelet of the symlet-family.

3. The further decomposition of the low-frequency signal $x_{\mathrm{A}2}$ is done as in the 2nd stage. In further stages the procedure is the same and so many stages are chosen until a sufficiently large accuracy in the reconstruction is achieved. Thus the whole spectrum $X(l)$ of the original signal $x(i)$ can be described by adding the spectra of the last low pass $X_{\mathrm{A}N}$ and all high passes $X_{\mathrm{D}i}$, $i = 1, \ldots, N$, i.e.

$$X(l) = X_{\mathrm{A}N}(l) + \sum_{i=1}^{N} X_{\mathrm{D}i}(l)\,, \quad N\text{: number of steps}\,. \tag{5.77}$$

4. To reconstruct the original signal, start from the low-pass signal of the last stage, reverse the downsampling by adding a new sample between the old samples by interpolation, and apply this to the input of the last low-pass. Proceed in the same way with the higher-frequency signal component. The output signals of the low-pass and high-pass are now added, and one gets again the signal $X_{\mathrm{A}2}$ of the penultimate stage (reverse order as in Figure 5.28). With the following calculation of the low-pass signal $x_{\mathrm{A}2}$, the procedure is repeated for all previous stages up to the first stage, which then completes the reconstruction.

When processing the spectra, they do not necessarily have to be divided with ideal high or low passes. The transfer functions of the high and low pass can also overlap somewhat at one stage and have a smooth transition between passband and stopband, which makes them easier to realise with the help of digital filters. In Matlab, wavelets of the families Daubechies, Coif|lets, Symlets, Fejer-Korovkin Filters, Discrete Meyer, Biorthogonal and Reverse Biorthogonal can also be used. Some wavelets can also be quite similar to the signal under investigation, and thus fewer stages are needed in the decomposition. For example, the symlet `Sym4` resembles a QRS complex in an ECG (see Figure 5.29).

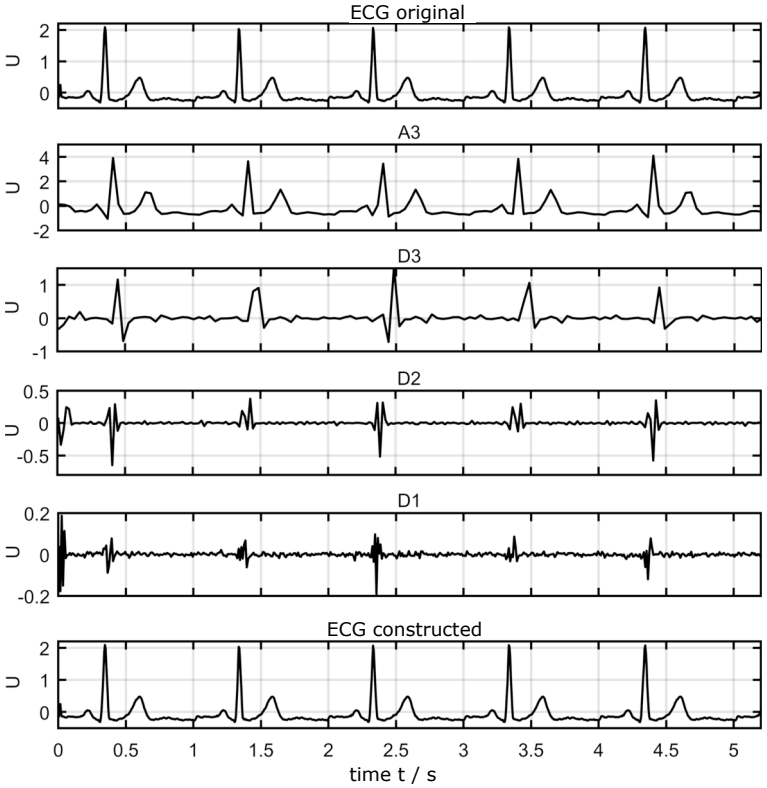


Fig. 5.30: Digital wavelet-transformation and reconstruction of an ECG signal up to the 3rd level: the calculation was done with the help of the Matlab-functions `wavedec()` and `waverec()` using the wavelet `sym4`.

In Matlab, this decomposition can be performed using the function `wavedec()` and the reconstruction using the function `waverec()`. Both are available in the (chargeable) toolbox `wavelet` from MathWorks. Alternatively, you can use the toolbox `wavelet` from Scilab, which is offered as a free software package. As an application example, Figure 5.30 shows the decomposition of an ECG signal using the wavelet `Sym4`.

5.3.4 Discrete Linear Time-Invariant Systems and Digital Filters

In the previous chapters, the concept of system has been used repeatedly to describe methods of signal processing; in the following section, the basics of system theory will be further explored. Since signals are nowadays usually processed in the discrete, we will dispense with an introduction of continuous systems and discuss the discrete systems that are relevant in practice. The following explanations only apply to linear time-invariant systems (LTI systems), i.e. the principle of linear superposition of sig-

nals applies and furthermore the assumption applies that the system properties (or the system parameters or system coefficients) do not change with time. The LTI systems form the basis of the digital filters at the end of the section.

5.3.4.1 Linear time-invariant systems

In the subsection 5.3.1, reference has already been made to some simple digital systems e.g. for changing the sampling or convolution of discrete signals. In the following, these discrete-time systems will be formally described. There are three basic forms of description:

1. If one represents the relations between the input and output quantities in the time domain, one obtains the so-called *difference equations*:

$$y(n) + d_0 y(n-1) = c_1 x(n) + c_0 x(n-1)\,,$$
$$y(n) + d_1 y(n-1) + d_0 (n-2) = c_2 x(n) + c_1 x(n-1) + c_2 x(n-2)\,.$$

2. Analogous to the convolution integral for continuous-time systems, in the discrete-time domain one obtains the *convolution sum* according to

$$y(n) = \sum_{\nu=-\infty}^{\infty} x(\nu) g(n-\nu)\,, \quad g(n)\text{: impulse response}\,. \tag{5.78}$$

3. The *transfer function* describes the behaviour of the discrete-time system N-th order with respect to its input and output quantities in the image domain. For linear systems it is a fractional rational function in z:

$$G(j\omega) = G(z = e^{j\omega T_a}) = \frac{\mathcal{F}\{y(n)\}}{\mathcal{F}\{x(n)\}} = \frac{\mathcal{Z}\{y(n)\}}{\mathcal{Z}\{x(n)\}} = \frac{c_0 + c_1 z + \cdots + c_{N-1} z^{N-1} + c_N z^N}{d_0 + d_1 z + \cdots + d_{N-1} z^{N-1} + z^N}\,. \tag{5.79}$$

The three forms of representation for describing an LTI system are completely equivalent and can be transformed into each other.

5.3.4.2 Digital filter

In general, a digital filter is a linear algorithm that converts a sequence of numbers into another sequence of numbers using linear mathematical operations (cf. Figure 5.31).

As a result, a digital filter also uses only the linear operators adder, subtractor, constant multiplier and delay element (cf. Figure 5.32), i.e. . *not* the saturation element known from control engineering. Discrete-time networks can be realised from these operators such that the coefficients c_i and d_i of the general transfer function (Equation 5.79) occur directly in the values of the multipliers, e.g. in networks of the 1st or 2nd canonical direct form (cf. Figure 5.33 and Figure 5.34). If the coefficients d_i in the denominator are non-zero, i.e. the network contains feedback, the impulse response is infinite (IIR filter), otherwise it is finite and ends after N output values (FIR filter, cf. Figure 5.35).

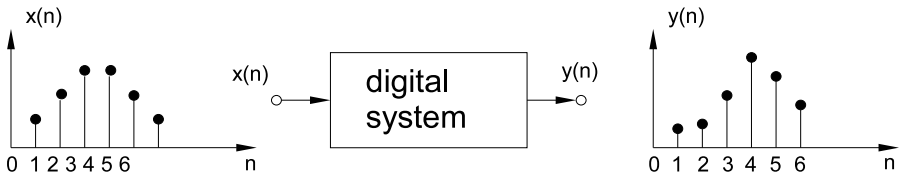


Fig. 5.31: Input and output-Signals of a discrete-time system.

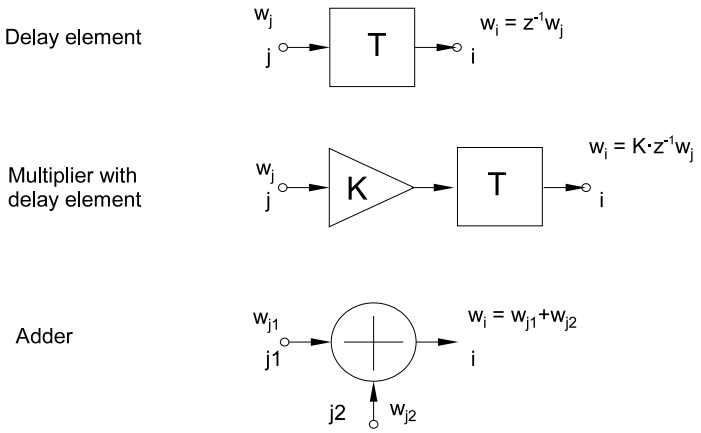


Fig. 5.32: Building elements of a discrete-time linear system.

In general, the input numbers are the encoded samples of an analogue signal. After processing by the mathematical algorithm, the output numbers are then converted back to analogue values and interpolated between sampling times (see Figure 5.36). As explained in section 5.1, the frequency spectrum of the sampled analogue signal is periodic with the sampling frequency (cf. e.g. Figure 5.5). However, this also means that the transfer function of the digital filter, which represents the ratio between the output and input spectrum, must also be periodic. Figure 5.37 shows as an example the periodic spectra of an ideal low and high pass. Next, the spectrum of the high pass can be easily generated from the spectrum of the low pass by shifting the spectrum of the low pass to the left or right by an odd multiple of $f_a/2$, ie:

$$G_{hp}(f) = G_{tp}(f + [2k+1]\, f_a/2)\,, \quad k = 0, \pm 1, \pm 2, \ldots \tag{5.80}$$

Since according to the frequency shift theorem in the Fourier-transformation a frequency shift with the frequency f_V in the spectral domain causes a multiplication by the factor $e^{j2\pi f_V t}$ in the time domain, a frequency shift of $f_a/2 = \frac{1}{2T_a}$ (i.e. i.e. $k = 0$ in

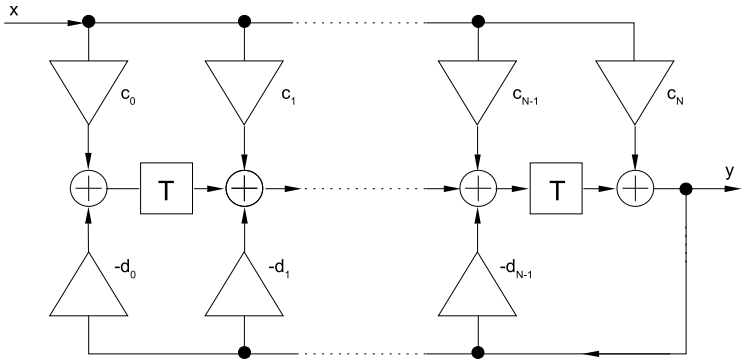


Fig. 5.33: Digital filter in the 1st canonical form (IIR filter).

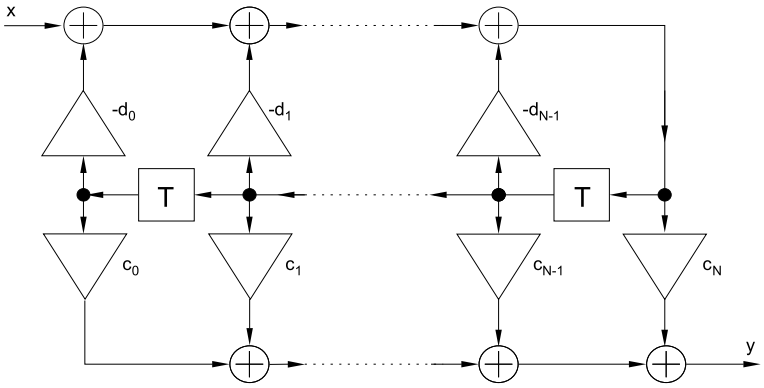


Fig. 5.34: Digital filter in 2nd canonical form (IIR filter).

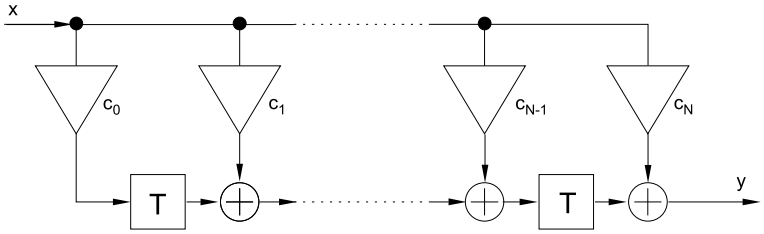


Fig. 5.35: Non-recursive digital filter in 1st, canonical direct form (FIR filter).

Equation 5.80) for the transformation of the transfer function $G_{\mathrm{hp}}(f)$ in the time domain:

$$g_{\mathrm{hp}}(nT_{\mathrm{a}}) = g_{\mathrm{tp}}(nT_{\mathrm{a}}) \cdot e^{j2\pi \frac{f_{\mathrm{a}}}{2} nT_{\mathrm{a}}} = g_{\mathrm{tp}}(nT_{\mathrm{a}}) \cdot e^{j\pi n} = g_{\mathrm{tp}}(nT_{\mathrm{a}}) \cdot (-1)^n \,. \tag{5.81}$$

Surprisingly, one now finds that the impulse response of the high-pass in this case can be calculated quite simply from the impulse response of the low-pass, if at every

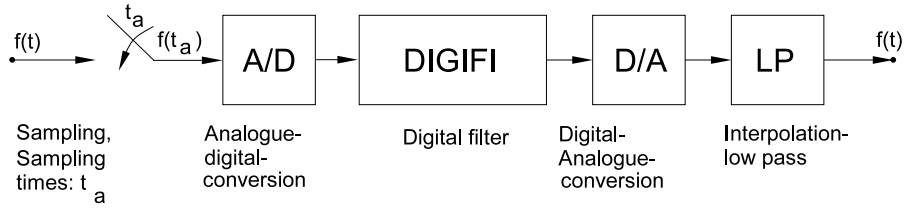


Fig. 5.36: Transmission chain when an analogue signal is processed by a digital filter.

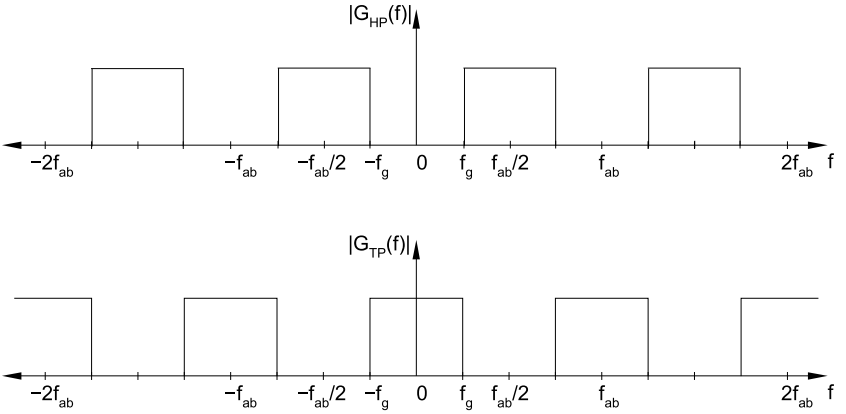


Fig. 5.37: Periodic magnitude spectra of an ideal discrete-time high-pass (top) and low-pass (bottom).

second value of the discrete impulse response of the low-pass its sign is changed. This is a very practical property when the cut-off frequency of the low-pass is $f_a/4$, i.e. in half of the total usable frequency range up to $f_a/2$. This fact is used, for example, in the half-band decomposition of the discrete wavelet-transformation.

However, it is often desirable not only to realise a high pass from a low pass, but also to design other selective filters if a suitable low pass is already available. This can be achieved by a frequency transformation. Unlike analogue filters, where the imaginary axis of the complex frequency plane $p = \sigma + j\omega$ is mapped onto itself in a suitable way by a reactance function, now in the discrete-time domain the corresponding frequency line, which now describes a circle around the origin of the coordinate system because of $z = e^{j\omega T_a}$, must be mapped onto itself again as a circle, ie, a circle is mapped onto a circle, but in such a way that the scaling changes and, for example, a low-pass becomes a high-pass or band-pass. The magnitude of z always remains unchanged at one with this transformation. Transfer functions that leave the magnitude of the input spectrum unchanged are allpasses. These only change the phase (or group delay) and can thus influence the circular frequency line in the z-range. One therefore obtains for

the most common selective filters (low pass, high pass, band pass and band stop) the following transfer functions for a frequency transformation for the case that the origin low pass has a normalised cut-off frequency of $f_a/4$ (see e.g. [9, 41]):

1. normalised low-pass-low-pass and low-pass-high-pass-transformation:

$$z_{ntp} = \pm\frac{z + a_0}{a_0 z + 1}, \quad \text{with} \quad |a_0| < 1$$
$$a_0 = \tan\left(\pi f_g T_a - \pi/4\right). \tag{5.82}$$

   z_{ntp} is the system variable in the normalised low-pass range. The positive sign applies to the low-pass-low-pass transformation, the negative sign to the "low-pass-high-pass-transformation. As can be seen from the equation, the coefficient a_0 is only dependent on the new cut-off frequency f_g, i.e. with a_0 the cut-off frequency can be varied.

2. normalised lowpass-bandpass and lowpass bandstop-transformation:

$$z_{ntp} = \mp\frac{z^2 + a_1 z + a_0}{a_0 + a_{1z+1}} \quad \text{mit} \quad |a_0| < 1 \wedge |a_1| < 1 + a_0$$
$$a_0 = \tan\left(\pi/4 - \pi[f_o - f_u]T_a\right) \tag{5.83}$$
$$a_1 = \frac{-2\sin(2\pi[f_o - f_u]T_a)}{\sin(2\pi f_u T_a) + \sin(2\pi f_o T_a) + \cos(2\pi f_u T_a) - \cos(2\pi f_o T_a)}.$$

   The negative sign applies to the lowpass-bandpass-transformation, the positive sign to the lowpass-bandpass-transformation. The coefficient a_0 depends here only on the new bandwidth $f_o - f_u$, i.e. with a_0 the bandwidth of the bandpass or the bandstop can be varied. With the coefficient a_1 the centre frequency is then influenced; because a_1 does not only depend on the difference of the upper cut-off frequency f_o and the lower f_u.

Explanatory example

A discrete-time normalised low-pass filter with sampling frequency $f_a = 200\,\text{Hz}$, equivalent to sampling period of $T_a = 1/f_a = 5\,\text{ms}$ has a cut-off frequency of $f_a/4 = 50\,\text{Hz}$. Using an all-pass transformation, a low-pass filter with a cut-off frequency of 30 Hz and 70 Hz is to be designed. Such a normalised low-pass with the transfer function $G_{ntp}(z_{ntp})$ can be realised by a simple averaging of two consecutive samples, i.e.:

$$G_{ntp}(z_{ntp}) = \frac{1}{2}\left(1 + z_{ntp}^{-1}\right). \tag{5.84}$$

After applying the low-pass-low-pass-transformation according to Equation 5.82, it then follows in general for the frequency-transformed low-pass with transfer function

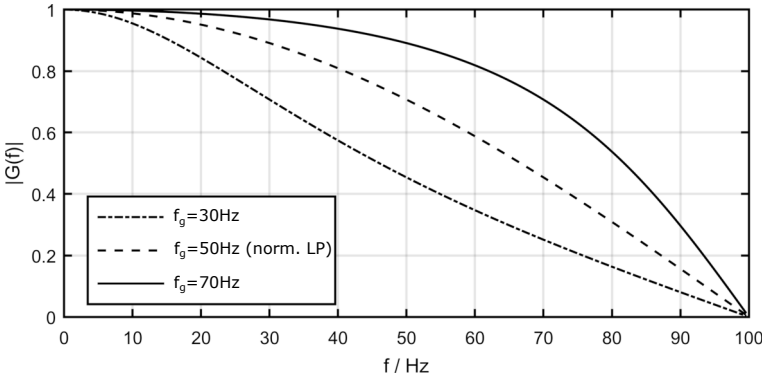


Fig. 5.38: Example of the magnitude transfer functions for low-pass-low-pass-transform using a digital all-pass according to Equation 5.82 with $f_a = 200$ Hz.

$G_{tp}(z)$:

$$G_{tp}(z) = G_{ntp}\left(z_{ntp} = \frac{z + a_0}{a_0 z + 1}\right) = \frac{1 + a_0}{2}\,\frac{z + 1}{z + a_0}\,, \qquad a_0 = \tan(\pi f_g T_a - \pi/4) \quad \text{with} \quad |a_0| < 1\,. \tag{5.85}$$

The coefficient a_0 then has the value for the transformed low-pass with the cut-off frequency of 30 Hz.

$$a_0 = \tan(\pi \cdot 30\,\text{Hz} \cdot 5\,\text{ms} - \pi/4) = -0.325 \tag{5.86}$$

and analogously for the low pass with cut-off frequency of 70 Hz

$$a_0 = \tan(\pi \cdot 70\,\text{Hz} \cdot 5\,\text{ms} - \pi/4) = 0.325\,. \tag{5.87}$$

Figure 5.38 shows the associated magnitude transfer functions $|G_{tp}(f)|$ in the context of the normalised low-pass with cut-off frequency of 50 Hz.

Another possibility to change the selectivity behaviour of a given filter with a certain property (e.g. normalised low pass) by frequency transformation is to apply the frequency transformation to an analogue filter and then to realise a digital filter with the same selectivity behaviour by transferring it to the discrete-time range. This does not apply to the frequency range from 0 to ∞ as with the analogue filter, but only from 0 to half the sampling frequency; because after that the spectrum repeats itself, since this is generally always periodic in discrete-time systems (cf. Figure 5.5). For this purpose, there is the transformation with the impulse invariance method and with the help of the bilinear transformation, which will be explained in the following.

The impulse invariance method

In this method, the digital filter is realised in such a way that the values of its impulse response $g_{di}(n)$ are the same as the impulse response of the analogue filter $g_{an}(t)$ at

the sampling times $t = t_a = nT_a$ (with $n = 1, 2, 3, \ldots$, and T_a: sampling interval), this means

$$g_{an}(t = nT_a) = g_{di}(n)\ . \tag{5.88}$$

To do this, we first assume that the transfer function $G_{an}(p)$ of the analogue filter is in a partial fraction form according to

$$G_{an}(p) = \sum_{i=1}^{N} \frac{A_i}{p - p_i}\ , \quad \text{with} \quad p_i\text{: pole } i \tag{5.89}$$

can be represented. In this case, one then obtains for the associated impulse response $g_{an}(t)$ at the sampling times $t_a = nT_a$

$$g_{an}(nT_a) = \sum_{i=1}^{N} \sigma(nT_a) A_i e^{p_i \cdot nT_a} = \sum_{i=1}^{N} A_i e^{p_i \cdot nT_a}\ , \quad \sigma(t)\text{: step function}\ . \tag{5.90}$$

Let these values be the impulse response $g_{digi}(n)$ of the digital filter to be implemented. The corresponding transfer function $G(z)$ is obtained by the discrete z- transformation:

$$G_{di}(z) = \sum_{n=0}^{\infty} g_{an}(nT_a) z^{-n} = \sum_{n=0}^{\infty} \left(\sum_{i=1}^{N} A_i e^{p_i nT_a} \right) z^{-n} = \sum_{i=1}^{N} A_i \underbrace{\sum_{n=0}^{\infty} (e^{p_i T_a} z^{-1})^n}_{\text{geometrische Reihe}}\ . \tag{5.91}$$

The sum signs may be interchanged in the last equation according to the laws of mathematics. In this case, with the sum with the upper limit ∞, one obtains a geometric series whose result can be easily calculated, and the result further follows

$$G_{di}(z) = \sum_{i=1}^{N} \frac{A_i}{1 - e^{p_i T_a} z^{-1}} = \sum_{i=1}^{N} \frac{A_i z}{z - e^{p_i T_a}} = \frac{c_0 + c_1 z + c_2 z^2 + \cdots + c_N z^N}{d_0 + d_1 z + d_2 z^2 + \cdots + z^N}\ . \tag{5.92}$$

From the pole positions p_i of the analogue filter, the coefficients c_i and d_i ($i = 0, \ldots, N$) of the digital filter can be calculated, which can then be realised by a filter in the 2nd canonical direct form according to Figure 5.34.

Note

So that for small frequencies the magnitude of the transfer function in the continuous-time domain $|G_{an}(p = j\omega)|$ is approximately equal to the amount of the transfer function $|G_{di}(z = e^{j\omega T_a})|$ for the discrete-time domain, $G_{di}(z)$ is still multiplied by the scaling factor T_a.

Explanatory Example

A simple analogue RC low-pass filter with a cut-off frequency f_g of 200 Hz with component values $R = 800\,\text{k}\Omega$ and $C = 1\,\text{nF}$ according to Figure 5.39 shall be replaced by

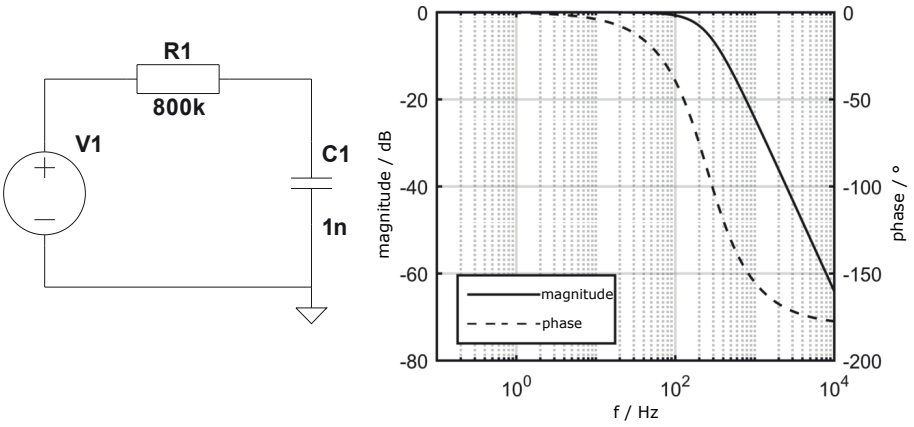


Fig. 5.39: Transfer function $G(f)_{\mathrm{RC}} = U_{\mathrm{a}}(f)/U_{\mathrm{e}}(f)$ of a 1st-order RC low-pass filter (calculation with LTspice).

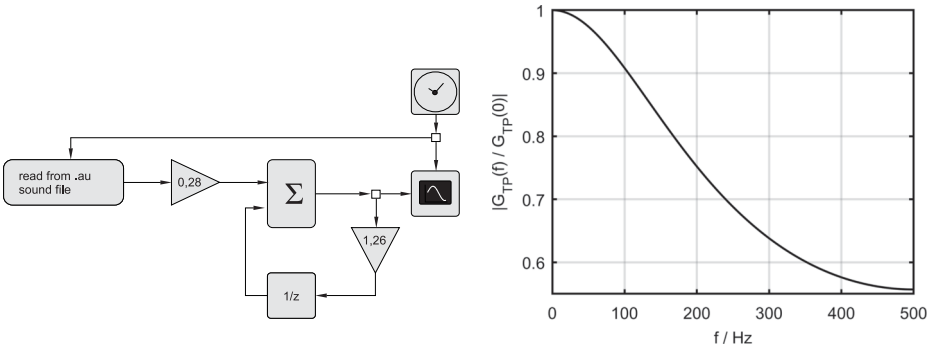


Fig. 5.40: Transfer function $G(f)_{\mathrm{RC}}$ of the 1st order digital low pass generated by the impulse invariance method from the analogue low pass according to Figure 5.39.

a digital low-pass filter with the same cut-off frequency and a sampling frequency of 1 kHz according to the pulse invariance method.

For the corresponding impulse response $g_{\mathrm{an}}(t) = g_{\mathrm{RC}}(t)$ then follows.

$$g_{\mathrm{RC}}(t) = A_1\, e^{p_1 t} \quad \text{mit } A_1 = p_1 = -\omega_{\mathrm{g}} = -1/RC$$

and thus according to Equation 5.92 for the transfer function of the digital filter to be realised:

$$G_{\mathrm{RC}}(z) = T_{\mathrm{a}} \frac{A_1}{1 - e^{p_1 T_{\mathrm{a}}} z^{-1}} = T_{\mathrm{a}} \frac{A_1 z}{z - e^{p_1 T_{\mathrm{a}}}} = \frac{1.26 \cdot z}{z + 0.28}\,.$$

The Bilinear Transformation Procedure

If one looks at the introductory example according to Figure 5.40, it is noticeable that at half the sampling frequency of 1 kHz/2 = 500 Hz the signal is only attenuated by

about half. After that, the amplification increases again, since the spectrum repeats according to the sampling theorem (cf. Figure 5.5). This effect is due to the fact that the analogue RC low-pass filter used does not have an infinitely large attenuation at the cut-off frequency f_g like an ideal low-pass filter, but only an attenuation of 3 dB, which is equivalent to a drop in the magnitude square of the transfer function by $1/\sqrt{2} \approx 0.707$. At higher frequencies, the attenuation increases but is non-zero at half the sampling frequency of $f_{ab}/2 = 500$ Hz. This means that there is overlapping of the spectrum after sampling. This cannot result in an ideal low-pass filter that has an infinitely large attenuation at its cut-off frequency. Furthermore, the property of the corresponding analogue filter according to Figure 5.39, which also has an infinitely large attenuation at a higher frequency. A remedy is achieved by not requiring the same values as those of the impulse response of the analogue filter for the sampling times in the filter design, but only striving for the closest possible approximation of the transfer function. However, because of the basically periodic course of the transfer function of the digital system, this means that the transfer function of the analogue system in the range from 0 to ∞ corresponds to that of the digital system in the range from 0 to half the sampling frequency. The frequency axis of the analogue system is virtually mapped onto the frequency axis of the digital system from 0 to half the sampling frequency. This can be done according to the figure after the bilinear transformation with the following procedure:

1. The transfer function $\mathfrak{G}(z = e^{j\omega T_a})$ shall match that of the associated analogue filter $G(p = j\omega$ up to half the sampling frequency, i.e..

$$\mathfrak{G}(z = e^{j\omega T_a}) = G(p = j\omega) \,. \tag{5.93}$$

2. substituting the inverse function of z, i.e..

$$p = \frac{\ln(z)}{T_a} \,, \tag{5.94}$$

   into the transfer function $G(p)$ of the analogue filter does *not* lead to a fractional rational function in z, and the digital filter is thereby not realisable with adders, delayers and multipliers (as, for example, in the canonical circuits).
3. The ideal transformation between p- and z-range, $z = e^{pT_a}$ or $p = \frac{\ln(z)}{T_a}$, is therefore approximated in such a way that a realisable transfer function in the z- range is obtained.
4. To do this, $\ln(z)$ is generated by a series expansion with fractional functions in z and broken off after the 1st member. This is then an approximation to be used called *bilinear transformation*; bilinear here means that there are linear functions in p or z in both the numerator and denominator:

$$\ln z = 2\left\{\frac{z-1}{z+1} + \frac{1}{3}\left(\frac{z-1}{z+1}\right)^3 + \frac{1}{5}\left(\frac{z-1}{z+1}\right)^5 + \cdots\right\} \approx 2\,\frac{z-1}{z+1} \,. \tag{5.95}$$

5. The transformation equations between the z- and p- domains are then obtained as.

$$p = \frac{1}{T_a} \ln z \approx \frac{2}{T_a} \frac{z-1}{z+1} \quad \text{und} \quad z \approx \frac{1 + p/\frac{2}{T_a}}{1 - p/\frac{2}{T_a}} , \tag{5.96}$$

resp.

$$\overline{G}(z) \approx G\left(p = \frac{2}{T_a} \frac{z-1}{z+1}\right) .$$

Explanatory example

The analogue RC low-pass already used in the above pulse invariance method according to Figure 5.39 with a cut-off frequency $f_g = 200\,\text{Hz}$ and the component values $R = 800\,\text{k}\Omega$ and $C = 1\,\text{nF}$ according to Figure 5.39 is to be replaced by a digital low-pass filter with the same cut-off frequency and a sampling frequency of 1 kHz according to the bilinear transformation method. The associated transfer function $G(p)$ of the low-pass filter with $A_1 = p_1 = -\omega_g = \frac{1}{RC}$ is given by

$$G(p) = \frac{A_1}{p - p_1} = \frac{\omega_g}{p + \omega_g} . \tag{5.97}$$

After inserting the bilinear transformation according to Equation 5.96, it follows:

$$\overline{G}(z) \approx G\left(p = \frac{2}{T_a} \frac{z-1}{z+1}\right) = \frac{\omega_B}{\frac{2}{T_a}\frac{z-1}{z+1} + \omega_B} = \frac{\frac{\omega_B}{2f_a+\omega_B}(1+z)}{\frac{\omega_B - 2f_a}{2f_a+\omega_B} + z} = \frac{c_0 + c_1 z}{d_0 + z} , \tag{5.98}$$

and further by comparing the coefficients:

$$c_0 = c_1 = \frac{\omega_B}{2f_a + \omega_B} = \frac{2\pi \cdot 200\,\text{Hz}}{2 \cdot 1\,\text{kHz} + 2\pi \cdot 200\,\text{Hz}} = \frac{\pi}{\pi + 5} = 0.385869\ldots \tag{5.99}$$

$$d_0 = \frac{\omega_B - 2f_a}{2f_a + \omega_B} = \frac{2\pi \cdot 200\,\text{Hz} - 2 \cdot 1\,\text{kHz}}{2 \cdot 1\,\text{kHz} + 2\pi \cdot 200\,\text{Hz}} = \frac{\pi - 5}{\pi + 5} = -0.22826\ldots . \tag{5.100}$$

However, digital filters can also be designed directly in the discrete-time domain without a diversion via the design of analogue filters. Of the available methods, two will be described in more detail here: i) the direct discrete-time synthesis using the *window method*, representative of methods for the design of non-recursive (FIR)-filters and ii) the *frequency sampling method* as an example for the general synthesis of recursive (IIR-) or non-recursive (FIR)-filters. Additional procedures can be taken from various publications in this field (e.g. B. [56, 62, 70]).

Direct discrete-time synthesis using window method

A non-recursive digital filter N-th order (FIR filter) according to Figure 5.35 has no feedback and according to Equation 5.79, because of which the feedback coefficients d_i are all equal to zero, has the transfer function

$$G(z) = \frac{c_0 + c_1 z + \cdots + c_{N-1} z^{N-1} + c_N z^N}{z^N} . \tag{5.101}$$

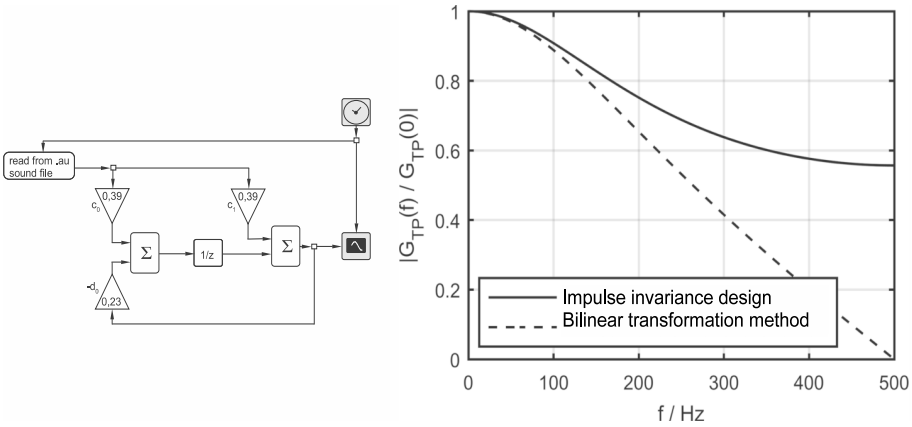


Fig. 5.41: Graph and transfer function $G(f)$ of the digital low-pass generated by the bilinear transform method from the analogue low-pass according to Figure 5.39: A comparison with the pulse invariance method shows that at $f_a/2 = 500$ Hz nothing more is transmitted.

However, in a slightly different representation, this can also be written as a function of z^{-1}, i.e..

$$G(z) = c_N + c_{N-1}z^{-1} + \cdots + c_1 z^{-(N-1)} + c_0 z^{-N} \,. \tag{5.102}$$

The element z^{-1} represents the transfer function of a delay element according to Figure 5.32. This means that a pulse applied to this element is delayed by one clock. For the element z^{-2} it is two clocks, for z^{-3} three clocks and so on. Such a digital filter thus generates several pulses weighted with c_i, $i = 0, \ldots, N$ at its output when an impulse is applied to its input. Therefore, the corresponding impulse response is:

$$g(n) = c_N\delta(n-0) + c_{N-1}\delta(n-1) + \cdots + c_1\delta(n-[N-1]) + c_0\delta(n-N) \,, \tag{5.103}$$

where $\delta(i)$ is the discrete Dirac momentum.

The filter coefficients c_i thus represent the values of the impulse response in a non-recursive filter. Conversely, however, this means that the impulse response can be specified, and one thereby also directly obtains the filter coefficients c_i. This is very practical for filter synthesis. However, it must be noted that an FIR filter always has a finite impulse response that ends after N clocks. Therefore, if a filter is to be realised that has an infinite impulse response, its impulse response can only be approximated by the impulse response of an FIR filter. In the simplest case, the given infinite impulse response is truncated after N clocks. This corresponds to multiplying the infinite impulse response $\tilde{g}(n)$ by a rectangular window function $w(t)$, which is constant for the first 0 to N beats and then drops to zero, i.e.:

$$g(n) = \tilde{g}(n) \cdot w(n) \,, \tag{5.104}$$

with $w(n) = 1$ from $n = 0$ to N, otherwise 0.

In the frequency range, however, larger deviations can occur, which are particularly noticeable with steep filter edges. If this is undesirable, the rectangular window can be replaced by another window that does not have such a steep edge, depending on the application. In this case, the original values of the given impulse response are no longer obtained in these windows, but they have a more favourable behaviour in the frequency range, e.g. no overshoots of the magnitude frequency response at a filter edge.

If the filter coefficients are symmetrical, selective filters can be realised with a non-recursive filter, which have no phase distortions or a constant group delay, which is not possible with analogue filters. Four cases can be distinguished:

1. filter order *N even*, coefficients *mirror*symmetric ($c_\mathrm{i} = c_{N-i}$).

$$G(z = \mathrm{e}^{j\omega T_\mathrm{a}}) = \left\{ c_{N/2} + 2 \sum_{i=0}^{\frac{N}{2}-1} c_\mathrm{i} \cos\left[\left(\frac{N}{2} - i \right) \omega T_\mathrm{a} \right] \right\} \mathrm{e}^{-j\omega N T_\mathrm{a}/2} , \tag{5.105}$$

2. filter order *N even*, coefficients *point*symmetric ($c_\mathrm{i} = -c_{N-i}$)

$$G(j\omega) = -j2 \left\{ \sum_{i=0}^{\frac{N}{2}-1} c_\mathrm{i} \sin\left[\left(\frac{N}{2} - i \right) \omega T_\mathrm{a} \right] \right\} \mathrm{e}^{-j\omega N T_\mathrm{a}/2} , \tag{5.106}$$

3. filter order *N odd*, coefficients *mirror*symmetric ($c_\mathrm{i} = c_{N-i}$)

$$G(j\omega) = 2 \left\{ \sum_{i=0}^{\frac{N-1}{2}} c_\mathrm{i} \cos\left[\left(\frac{N}{2} - i \right) \omega T_\mathrm{a} \right] \right\} \mathrm{e}^{-j\omega N T_\mathrm{a}/2} , \tag{5.107}$$

4. filter order *N odd*, coefficients *point*symmetric ($c_\mathrm{i} = -c_{N-i}$)

$$G(j\omega) = -j2 \left\{ \sum_{i=0}^{\frac{N-1}{2}} c_\mathrm{i} \sin\left[\left(\frac{N}{2} - i \right) \omega T_\mathrm{a} \right] \right\} \mathrm{e}^{-j\omega N T_\mathrm{a}/2} . \tag{5.108}$$

In all four cases, the phase of the filter is linearly dependent on the angular frequency ω, and it has a *constant group delay* of $t_0 = \frac{NT_\mathrm{a}}{2}$.

Explanatory Example

An ideal digital low-pass filter with a constant magnitude frequency response from 0 to the cut-off frequency $f_\mathrm{g} = 200\,\mathrm{Hz}$ and a sampling frequency of $f_\mathrm{a} = 1\,\mathrm{kHz}$ is to be realised by an 8th order FIR filter. With symmetrical filter coefficients, this filter has a group delay of $t_0 = \frac{NT}{2} = \frac{8 \cdot 1}{2}\,\mathrm{ms} = 4\,\mathrm{ms}$.

An ideal analogue low pass with a constant magnitude frequency response of 1 in the passband would have the impulse response

$$g_\mathrm{an}(t) = \frac{\sin(\omega_\mathrm{g}(t - t_0))}{\pi(t - t_0)}$$

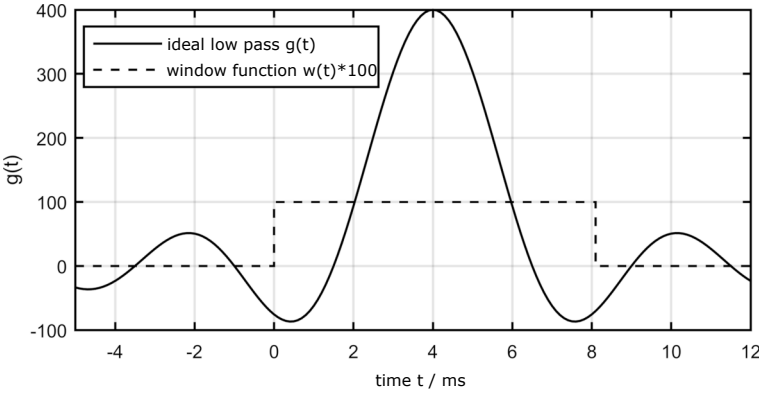


Fig. 5.42: Impulse response $g(t)$ of the ideal low-pass and rectangular window $w(t)$ enlarged by a factor of 100 for better representation.

have (cf. Figure 5.42). As with the impulse invariance method, the impulse response of the associated digital FIR filter at the sampling times nT_a ($n = 0$ to 8) should have the same values as the impulse response of the analogue filter, except for a scaling factor of the magnitude. This corresponds to a rectangular window $w(t)$ which cuts off the values of the impulse response of the ideal filter according to $N \cdot T_a = 8 \cdot 1\,\text{ms} = 8\,\text{ms}$ (see Figure 5.43). This gives the filter coefficients to

$$c_i = g(i) = T_a\, g_{an}(iT_a) = \frac{\sin(\omega_g T_a(i - N/2))}{\pi(i - N/2)}\,, \quad i = 0 \text{ to } N\,. \tag{5.109}$$

Here, the analogue impulse response $g_{an}(t)$ is additionally multiplied by the sampling interval T_a so that the impulse response of the digital filter becomes dimensionless. Furthermore, for our example it follows:

$$c_i = \frac{\sin(0.4\pi(i - 4))}{\pi(i - 4)}\,, \quad i = 0 \text{ bis } 8\,.$$

Thus, for the filter coefficients c_i:

$$\begin{aligned} &c_0 = c_8 = -0.07568\,, && c_1 = c_7 = -0.06237\,,\\ &c_2 = c_6 = 0.09355\,, && c_3 = c_5 = 0.3027\,,\\ &c_4 = 0.4\,. \end{aligned}$$

Since the filter order- N is even and the impulse response is mirror symmetric to t_o, it follows according to Equation 5.105 for the transfer function:

$$G(z = e^{j\omega T_a}) = \left\{ c_4 + 2\sum_{i=0}^{3} c_i \cos\left[\omega(4 - i)\,\text{ms}\right] \right\} e^{-j\omega 4\,\text{ms}}\,. \tag{5.110}$$

From this example it can be seen that even with a higher filtering order N not nearly the rectangular magnitude frequency response of an ideal low-pass filter is achieved.

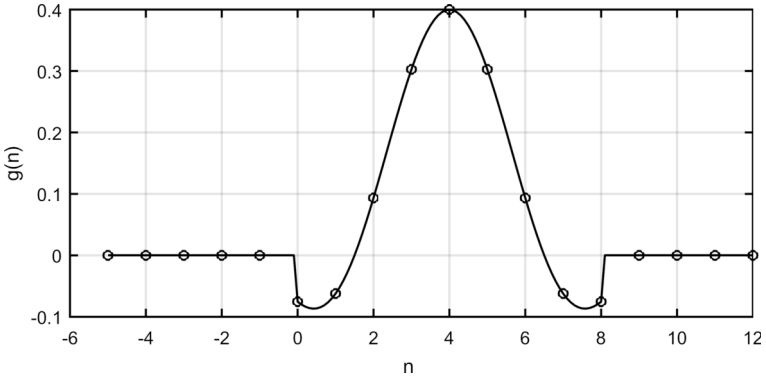


Fig. 5.43: The range of the impulse response of the ideal filter from 0 to 8 ms "cut-out" with the rectangular window $w(t)$ according to Figure 5.42 and the samples taken therein c_i with $i = 0$ to $N = 8$.

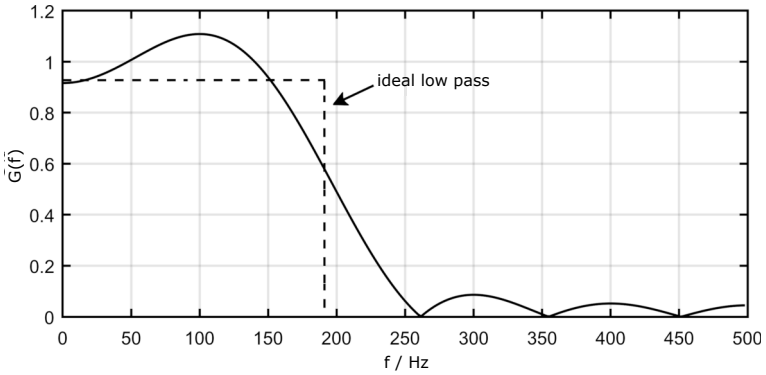


Fig. 5.44: Magnitude frequency response of the 8th order FIR filter whose filter coefficients c_i have been calculated according to Equation 5.109 after the values of the impulse response of an ideal low-pass filter in the time interval from 0 to 8 ms.

As a rule, a considerably higher filter order (e.g. $n = 30$) is necessary for this. The favourable behaviour of the FIR filter with regard to the possible constant group delay and the stability due to the finite impulse response must therefore be bought by a larger filter order.

Direct Discrete-Time Synthesis Using the Frequency Sampling Method

Instead of the values of the impulse response, the transfer function can also be chosen for approximation. In the *frequency sampling method* the desired frequency response is therefore sampled at regular intervals and can be interpreted as the spectrum of a periodic signal, which is discrete in the frequency range and only has a value in multiples of the fundamental frequency (harmonics). Furthermore, in the time domain the periodic signal is sampled with the sampling interval T_a, whereby the frequency

response is not only discrete but also periodic as in the time domain. The relationship between the time and frequency range of a periodic and discrete signal and its spectrum is determined by the discrete Fourier-transformation(*DFT*) according to

$$g(n) = \frac{1}{N}\sum_{m=0}^{N-1} G(m)e^{j2\pi\frac{nm}{N}}\,, \quad G(m) = \sum_{n=0}^{N-1} g(n)e^{-j2\pi\frac{nm}{N}} \tag{5.111}$$

where $G(m)$ is the sampled value of the desired transfer function at the mth digit and $g(n)$ is the sampled value of the impulse response at the nth digit. N corresponds to the number of samples per period in the time domain as well as the number of frequency points in the frequency domain.

In the filter synthesis, the periodic impulse response $g(n)$ is determined with the help of the DFT from the values of the transfer function $G(m)$ sampled per period and interpreted as the coefficient c_i of a non-recursive (FIR) filter (see Figure 5.35 and Equation 5.102). If, in addition, the filter is to have a constant group delay, the magnitude of the impulse response $g(n)$ must be symmetrical to $\frac{N}{2}$ if N is even, or $\frac{N-1}{2}$ (for N odd), ie:

$$|g(n)| = |g(N-n)| \quad n = 0, \ldots, \begin{cases} \frac{N}{2} & \text{for } N \text{ even} \\ \frac{N-1}{2} & \text{for } N \text{ odd} \end{cases}. \tag{5.112}$$

Here the group delay is $\frac{NT}{2}$, and the magnitude of the transfer function can be calculated according to Equation 5.105 to Equation 5.108.

If the impulse response $g(n)$ is to be real, it follows from the Fourier transform that the real part of the spectrum must be mirror symmetrical to the coordinate origin and the imaginary part must be point symmetrical to it. However, since the spectrum repeats periodically with multiples of the sampling frequency $f_a/2$, the spectral part from $-f_a/2$ to 0 is equal to the spectral part from $f_a/2$ to f_a. However, the DFT does not consider spectral components for negative frequencies but for the whole range from 0 to the sampling frequency f_a. Therefore, the range from $f_a/2$ to f_a can be obtained from the symmetry condition for the spectral range at the negative frequencies. It therefore follows:

$$G(N-m) = G^*(m) \quad m = 0 \text{ to } N\,, \tag{5.113}$$

whereby, according to Equation 5.111, the back transformation in the DFT can be simplified into the time domain. Thereby, with the help of splitting the DFT sum into two equal halves with bracketing out the function factor $e^{j2\pi\frac{N}{2}\frac{n}{N}}$ in the two partial sums, the later substitution $m' = N - m$ and from the symmetry condition according to Equa-

tion 5.113 for *even* values of N follows:

$$\begin{aligned}
g(n) &= \frac{G(0) + G\left(\frac{N}{2}\right)}{N} + \sum_{m=1}^{\frac{N}{2}-1} G(m) e^{j2\pi \frac{\left(m-\frac{N}{2}\right)n}{N}} \cdot \frac{e^{j2\pi \frac{N}{2}\frac{n}{N}}}{N} + \\
&\quad + \sum_{m'=\frac{N}{2}+1}^{N-1} G(m') e^{j2\pi \frac{\left(m'-\frac{N}{2}\right)n}{N}} \cdot \frac{e^{j2\pi \frac{N}{2}\frac{n}{N}}}{N} \\
&= \frac{2}{N}\left[G(0)\right] + \sum_{m=1}^{\frac{N}{2}-1} G(m) e^{j2\pi \frac{\left(m-\frac{N}{2}\right)n}{N}} \cdot \frac{e^{j\pi n}}{N} + \sum_{m=\frac{N}{2}-1}^{1} G(N-m) e^{j2\pi \frac{\left(\frac{N}{2}-m\right)n}{N}} \frac{e^{j\pi n}}{N} \\
&= \frac{2}{N}\Re\left[G(0)\right] + \sum_{m=1}^{\frac{N}{2}-1} \left\{ \left[G(m) e^{j2\pi \frac{\left(m-\frac{N}{2}\right)n}{N}} + G^*(m) e^{-j2\pi \frac{\left(m-\frac{N}{2}\right)n}{N}} \right] \frac{(-1)^n}{N} \right\} \\
&= \frac{2}{N}\Re\left[G(0)\right] + \sum_{m=1}^{\frac{N}{2}-1} \left\{ \left[|G(m)| e^{j\varphi} e^{j2\pi \frac{\left(m-\frac{N}{2}\right)n}{N}} + |G(m)| e^{-j\varphi} e^{-j2\pi \frac{\left(m-\frac{N}{2}\right)n}{N}} \right] \frac{(-1)^n}{N} \right\} \\
&= \frac{2}{N}\left\{ \Re\left[G(0)\right] + \sum_{m=1}^{\frac{N}{2}-1} (-1)^n |G(m)| \cos\left[\varphi + 2\pi\left(m - \frac{N}{2}\right) n/N\right] \right\}
\end{aligned} \tag{5.114}$$

and for *odd* values of N with analogous calculation

$$g(n) = \frac{1}{N}\left\{ G(0) + 2\sum_{m=1}^{\frac{N-1}{2}} (-1)^n |G(m)| \cos\left[\varphi + 2\pi\left(m - \frac{N-1}{2}\right) n/N\right] \right\}. \tag{5.115}$$

Explanatory Example

A *ideal* digital low-pass filter with a constant magnitude frequency response from 0 to its cut-off frequency of $f_g = 200\,\text{Hz}$ and a sampling frequency of $f_a = 1\,\text{kHz}$ is to be implemented by a 15th order FIR filter using the frequency sampling method. The periodic frequency response can be divided into $N - 1 = 15 - 1 = 14$ intervals and sampled with 15 values. With symmetrical filter coefficients and odd filter order, this filter has a group delay of $t_0 = T_a \frac{N-1}{2} = \frac{14 \cdot 1\,\text{ms}}{2} = 7\,\text{ms}$.

The values to be specified for the magnitude of the transfer function $|G(m)|$ with $m = 0$ to $\frac{N-1}{2} + 1 = \frac{15-1}{2} + 1 = 8$ are then to be specified at the frequency spacing of $f_a/N = \frac{1\,\text{kHz}}{15} = 66.67\,\text{Hz}$, viz. i.e. at frequency values f = 0 Hz, 66.67 Hz, 133.33 Hz, 200 Hz, 266.67 Hz, 333.33 Hz, 400 Hz and 466.67 Hz. After that, starting from half the sampling frequency, the frequency response repeats mirror-symmetrically up to the sampling frequency as shown above. If the maximum value is to be = 1, we can specify, for example, the following values:

$$\begin{aligned}
|G(0)| = |G(1)| = |G(2)| = |G(3)| = 1\,, \\
|G(4)| = |G(5)| = |G(6)| = |G(7)| = 0\,.
\end{aligned}$$

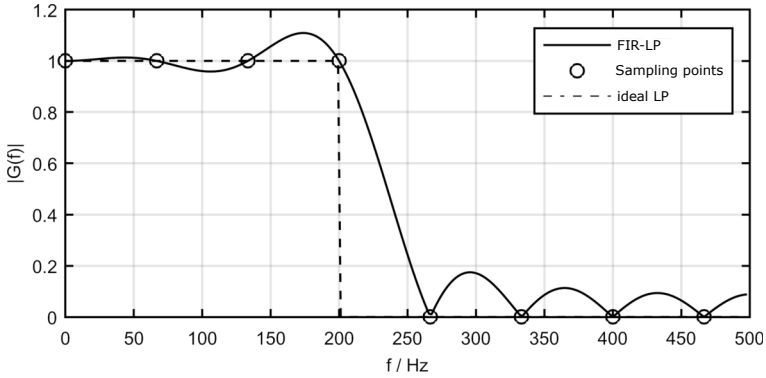


Fig. 5.45: Frequency-Samples of an ideal low-pass filter for synthesis of a 15th order FIR filter and associated transfer function

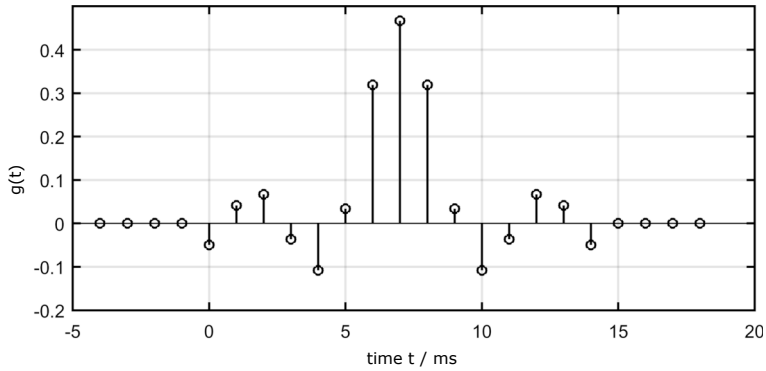


Fig. 5.46: Impulse response of an FIR low-pass 15. order according to Figure 5.45.

$|G(0)|$ here corresponds to the magnitude of the frequency response at frequency $f = 0\,\text{Hz}$, $|G(1)|$ to the magnitude of the frequency response at $66.67\,\text{Hz}$, $|G(3)|$ at $133.33\,\text{Hz}$ etc.

With these values, the impulse response g(n) of the FIR filter is now calculated according to Equation 5.115 by inverse discrete Fourier transformation, and we obtain:

$$g(0) = g(14) = -0.0498159\,; \quad g(1) = g(13) = 0.0412023\,;$$
$$g(2) = g(12) = 0.0666667\,; \quad g(3) = g(11) = -0.0364879\,;$$
$$g(4) = g(10) = -0.1078689\,; \quad g(5) = g(9) = 0.0340780\,;$$
$$g(6) = g(8) = 0.3188924\,; \quad g(7) = 0.4666667\,.$$

Finally, from the now determined impulse response of the FIR filter, the associated values of the filter coefficients can be determined, which are, after all, equal to the values of the impulse response $c_i = g(i)$ (cf. Figure 5.46).

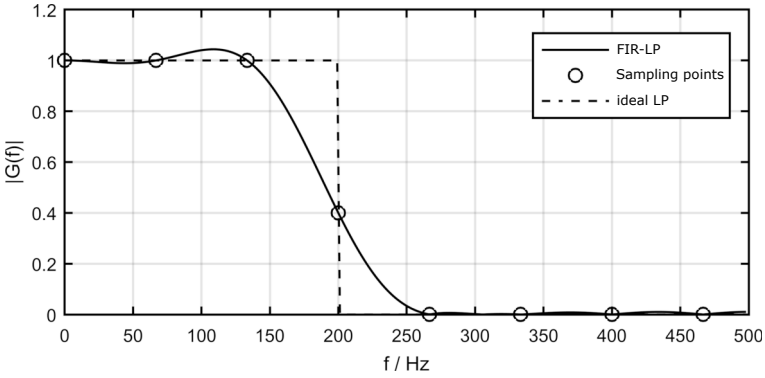


Fig. 5.47: FIR low-pass 15 order according to Figure 5.45 where the sample at the filter edge at 200 Hz has been lowered to 0.4.

In general, however, the measured biosignals are not periodic and therefore do not have a discrete frequency spectrum. If the FIR filter just designed is now applied to these signals, an error occurs. To estimate this, the impulse response is determined. In this case, the transfer function of the FIR filter is no longer discrete, but continuous in frequency. In Figure 5.45 it can be seen that between the given values of the magnitude frequency response very large deviations from the ideal lowpass can occur, which are called overshoots.

This can be remedied by no longer attempting to sample the ideal lowpass, which has a very large slope, but by replacing the steep slope with a smoother transition from the passband to the stopband. As an example of this, the sample point with the value 1 at the edge at 200 Hz is to be replaced by a value of 0.4. The result shows Figure 5.47. It can be clearly seen that the flatter edge at 200 Hz greatly reduced the overshoots in the frequency response compared to the steep edge according to Figure 5.45.

5.4 Post-Reading and Exercises

Discretisation

1. Explain the process of discretisation and quantisation in the context of analogue-digital-conversion of signals.
2. Explain the Shannon theorem.
3. What effect of analogue to digital conversion is called aliasing?
4. Explain the effect of an anti-aliasing filter.
5. What is the purpose of a sample-and-hold amplifier before an analogue-to-digital converter?

6. To what number is a voltage value of 2 V at the input of a 10-bit A/D-converter with an input span of 0 V–10 V mapped? Let the output number range start at zero and cross over only positive numbers. Note that this is a unipolar system.
7. A continuous-time signal from a sensor is to be transmitted via a digital transmission system. The spectral components of the sensor that are important for analysis are below 100 kHz. Frequency components uninteresting for the analysis extend up to 1 MHz. a) Which frequency should be sampled? Sketch the spectrum of the sampled signal! Is an "anti-aliasing" filter required? b) Assuming that each sample is encoded with 8 bits, what is the minimum data rate that the transmission system must provide (in kBit/s)?
8. An input signal with two superimposed sinusoidal signals $f1 = 3$ kHz and $f2 = 6$ kHz is sampled with a sampling frequency $f_a = 8$ kHz. Draw the spectrum of the input signal before and after sampling. What does this mean for a reconstruction of the original signal?
9. How do you recover a continuous signal from a discrete-time one?

Discrete Transformations

1. What are the special features of spectra of sampled signals?
2. What is meant by the first Nyquist range of a spectrum?
3. What is the magnitude of the highest frequency that can be represented in a spectrum, and what is the magnitude of its frequency resolution?
4. What is the leakage effect of a spectrum?
5. What is the difference between a magnitude spectrum and an amplitude spectrum? What is a phase spectrum? How are these calculated from the Fourier-coefficients?
6. What is the Fourier-transform of a square wave function?
7. What condition must a function fulfil in order to be able to develop it using a Fourier series? How can this function be described?
8. What is windowing of a signal? What is it used for?
9. Given is the sequence of a rectangular pulse $x[n] = 1, 1, 1, 1, 1, 1$. Sketch the sequence and its even and odd parts.
10. Given the two sequences $x_1[n] = 1, 2, 1$ and $x_2[n] = 1, 0, 1, 0, 1, 0, 1$. a) Draw the sequences and give their lengths. b) Fold the sequences, $x_1[n] * x_2[n] = y[n]$, and graph the result.
11. Besides the DFT, why do we need another transformation, the z-transformation?
12. What is the shift operator of the z- transform in the z- domain?
13. A sampled weight sequence of Dirac pulses (sampling sequence, discrete time series) is to be z- transformed. How can the shift operator be used for this?

Discrete Processing of Signals

1. In what other case can the alias effect occur besides A/D conversion? How can its occurrence be prevented? What losses do you have to accept?
2. Explain the amplitude spectrum of a discrete time series. Which aspects have to be considered in the interpretation? How do you calculate the frequency axis, what is the significance of the individual lines?
3. The amplitudes in the DFT spectrum from Matlab do not correspond to the amplitudes of the signal components in the time domain. How do you explain this? Which operation must always be performed for the adjustment?
4. What is meant by frequency resolution? How can the frequency resolution of a spectrum be improved?
5. What is the highest frequency that can be represented in the spectrum?
6. In the short-time Fourier transform, a signal window with a width of 100 ms is spectrally investigated. What frequency resolution can be achieved?
7. What is the advantage of the wavelet transform over the short-time Fourier transform?
8. A windowing is a multiplication in the time domain, which operation with what has to be done in the frequency domain to get the same result?
9. Explain the term impulse response. What is the Fourier-transform of the impulse response called?
10. When is it best to represent a signal in the time-frequency domain?
11. Is it advantageous to represent a sinusoidal signal with constant frequency and amplitude in the time-frequency domain?
12. What is the most important difference between the Short-Time Fourier Transform (SFTF) and the Wavelet-Transform?
13. What is the advantage of the discrete wavelet transform (DWT) over the continuous wavelet transform (CWT)?
14. Why do you use window functions for the SFTF, the DWT and the CWT?
15. Which averaging methods do you know and what can they be used for?

LTI systems and Digital Filters

1. What is a filter, what are its ideal characteristics?
2. Which filter characteristics do you know? How can they be represented in the frequency domain?
3. Why do we work mainly with digital filters? Where are analogue components indispensable?
4. What do the terms stable, causal, linear, time-invariant mean, and why are they so important when creating digital filters?
5. What is an IIR filter and what is an FIR filter? What is the difference?
6. How does a moving average filter affect signals? What extreme cases of the window can you imagine, and what would the result of the filtering be in each case?

7. Describe the operation of a simple convolution filter and the FFT method to obtain a suitable convolution kernel.
8. What is a non-recursive filter, what do the output signals consist of? How does a recursive filter differ from it?
9. What is a transfer function? What is the meaning of the real part and the imaginary part?
10. What are the poles and zeros of the transfer function?
11. Are there equivalent representations to the discrete transfer function? What are they?
12. What is the phase and amplitude response of a filter? Why do you want to create filters without phase shift?
13. What is meant by the order of a filter? Describe the differences between FIR and IIR filters.
14. What does an anti-image low-pass filter do in the context of adding zero values to subsequently increase the sampling frequency?
15. Can a time-varying system be represented in the frequency domain and if so how?
16. What is the difference between an IIR and an FIR filter?
17. What favourable property can the phase response of an FIR filter have that cannot be achieved with an analogue filter?
18. Can an FIR filter become unstable in the normal case with a finite number of filter coefficients?
19. How is the synthesis of a digital filter done after the "bilinear" transformation?

6 Applications and Methods in Biosignal Processing

In this chapter practical examples for the application of the methods of biosignal processing are presented. In principle, the methods are also applicable to other biosignals, which is why some electrical biosignals and their meaning are listed again here. Commonly used procedures are:

- Electrocardiogram (ECG): measurement of muscle excitation in the heart;
- Electroencephalogram (EEG): measurement of nerve activity in the brain;
- Electromyogram (EMG): measurement of muscle excitation in general;
- Electroretinogram (ERG): measurement of light stimulation of the eye. The essential information is in amplitudes and latencies of the detected waves;
- Electrooculogram (EOG): measurement of motor influences on the eye position.

Less frequently used methods are:

- Electroolfactogram: recording of the stimulation of the sense of smell;
- Electrogastrogram: recording of the activity of the stomach muscles;
- Elektrohysterogram: recording of the activity of the uterine musculature;
- Electrodermatogram: recording of electrical potential distribution on the skin;
- Electroneurogram: recording of intracellular electrical potentials.

6.1 Signals of the Brain

The brain is part of the central nervous system. In humans, it consists of about 10 billion individual brain cells (nervous tissue). The nervous tissue is in turn composed of neurons (nerve cells) and glial cells. Important functions of neurons of the brain are reception, processing and transmission of stimuli. Glial cells act, among other things, as supporting cells of neurons. Functionally, a brain neuron is divided into a multitude of tentacle-like dendrites for the reception of stimuli, the nucleus and the elongated axon for the transmission of stimuli. The axon branches in the terminal region to several synapses, which transmit a stimulus to another cell. The enormous performance of the brain is based on the strong interconnection of the individual neurons. In the area of the cerebral cortex (cortex), each neuron is connected with 1000 to 100,000 synapses. The transmission of stimuli from one cell to the next takes place chemically by the release of neurotransmitters from the synapse, which are taken up by dendrites of the next cell via receptors. As in all cells, the activity of neurons is accompanied by a change in transmembrane voltage, which is summarized under the term action potential. In the area of the dendrites, the neurotransmitters control the protein channels in the cell membrane and thus the action potential. There are both activating and inhibitory neurotransmitters. Since there are several synapses at the dendrites that simultaneously release neurotransmitters, it is the sum of all incoming stimuli that

https://doi.org/10.1515/9783110736298-006

determines whether an action potential is triggered in the cell. If this is the case, the action potential spreads from the dendrites along the axon and triggers the release of neurotransmitters at the end of the axon (synapse) which are deposited there in vesicles. These neurotransmitters in turn react with the receptors of the next neuron and control the cell membrane there (cf. chapter 3).

Stimulus propagation through a nerve cell occurs as the action potential advances along the axon. If depolarization is triggered at one point, Na^+ ions flow from the environment outside the cell into the interior. This results in a local decrease in the concentration of Na^+ ions in the external cellular ambience, which is compensated by diffusive ion currents from the environment. This, however, decreases the Na^+ concentration there, which briefly increases the transmembrane voltage above the threshold voltage[1]. By this the action potential is also triggered in the vicinity. However, this is only possible in the area behind the the depolarization site. The area in front of the depolarization site is still in the refractory state at this point, because the action potential has passed through there before. In the refractory phase the action potential cannot be triggered again. Therefore, the action potential, and thus the electrical stimulus, continues in only one direction of the axon fiber. The speed of this process and thus the conduction velocity is approx. 3 m/s. For neurons that excite skeletal muscles, however, this transmission speed would be too low, because due to the great length of the connection between nerve center and muscle fiber of up to one meter, the reaction e.g. to a dangerous situation would be too slow. In fact, in peripheral neurons that establish the connection to motor muscle units, a significantly higher conduction velocity of up to 120 m/s is achieved. The peripheral neurons have a different structure compared to brain neurons. The axon fiber is surrounded by Schwann cells, which prevent the triggering of the action potential. The sheath is called myelin sheath. At intervals of 0.2 to 1.5 mm, the myelin sheath is interrupted by constrictions, the Ranvier constrictions. Only there can the action potential be triggered. Since, during depolarization at a constriction, the cell-external compensatory currents extend to the next constriction, the action potential jumps from one constriction to the next. This mechanism significantly increases conduction velocity vs. brain neurons (cf. subsection 3.1.3).

The compensating currents along the axon fiber generate electric fields that propagate in the neurons of the cerebral cortex to the surface of the skull surface. The result is a potential difference between arbitrary locations on the scalp. The overall potential distribution on the scalp results from the superposition of the potentials of all active neurons and glial cells, with nearby cell regions making the strongest contribution. A spatial potential distribution is created on the scalp that varies in time with the activity of the cell regions involved. The voltage amplitude measurable on the scalp is up to 100 µV. Recording by means of electrodes is called

1 The direction of the transmembrane voltage is from the cell interior to the -exterior. Thus, a lack of positive ions in the cell exterior increases the transmembrane voltage.

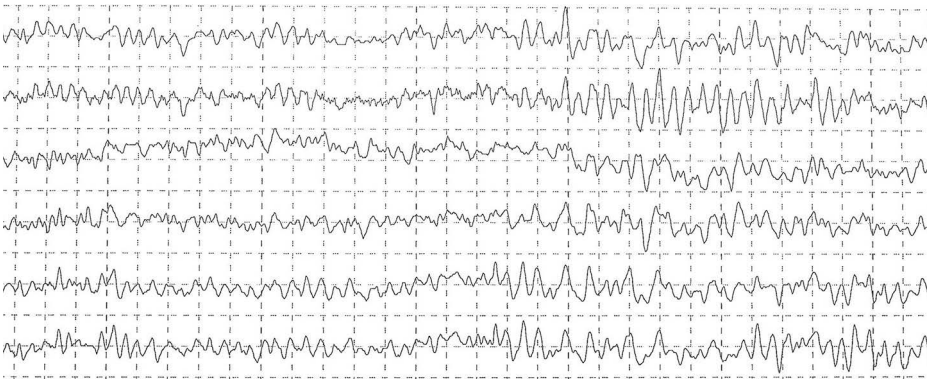


Fig. 6.1: EEG signals usually have no clear structure and thus belong to the non-deterministic signals. Nevertheless, a rhythm is recognizable in the signals.

electroencephalography (EEG). Because of the large number of cells involved and their different temporal excitation, the EEG signal usually does not have a clear structure (cf. Figure 6.1). Thus, it belongs to the non-deterministic signals. However, the EEG is underpinned by a rhythmic pattern, which depends among other things on the mental alertness of the subject. In clinical diagnostics, this rhythm is called a wave[2].

If the frequency of the wave is in the range from 0 to 4 Hz, it is called a delta wave (typical for dreamless deep sleep), from 4 to 8 Hz as theta wave (light sleep phase), from 8 to 13 Hz as alpha wave (e.g. relaxed waking state), at 13 to 30 Hz as beta wave (medium concentration, REM sleep) and above 30 Hz as gamma wave (strong concentration). In Figure 6.2, EEG waveforms at different states of wakefulness are shown.

For the determination of the frequency distribution, a Fourier transformation of the EEG time signal is performed. Figure 6.3 shows the spectrum of an EEG with a maximum at 11 Hz, suggesting alpha waves in the EEG signal. The distinction of different states in the EEG spectrum usually requires a high frequency resolution. For example, if a frequency resolution of 1 Hz is required, according to the uncertainty principle of communications engineering, an EEG sequence of at least 1 s duration is necessary[3].

The EEG electrodes have basically the same structure as ECG electrodes, but are mostly smaller in diameter. The arrangement of the electrodes on the skull follows a specific scheme. With the internationally standardized 0-20 System, 21 electrodes are placed on meridians between the root of the nose and the lower edge of the posterior skull bone at intervals of 10 % resp. 20 % of the meridian lengths. Figure 6.4 shows in side and top view the electrode arrangement of the 10-20 system. A measurement curve results from the voltage difference between a measurement electrode

2 This is not a wave in the physical sense.

3 The uncertainty principle of communications engineering states that the product of time and frequency resolution is greater than or equal to one: $\Delta t\, \Delta f \geq 1$.

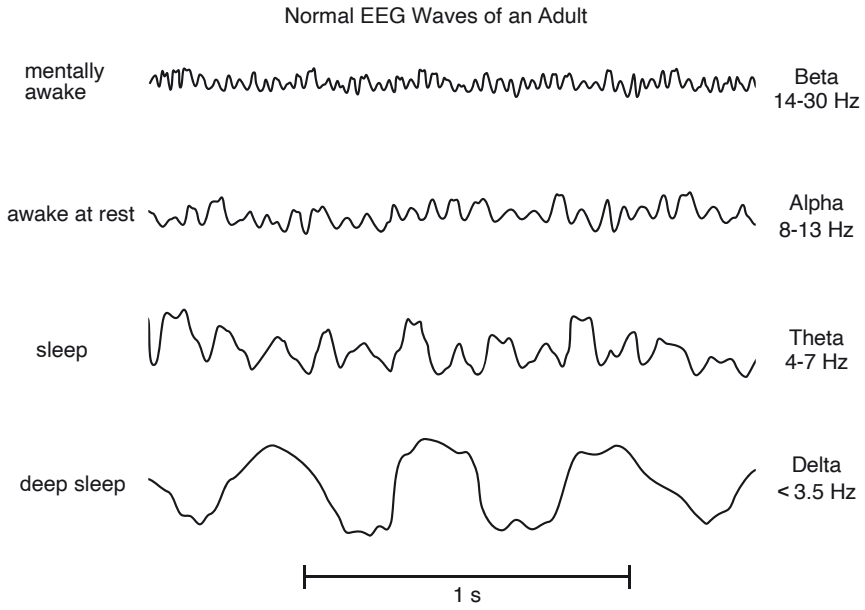


Fig. 6.2: EEG signal at different mental states. From top to bottom: mentally awake, awake at rest, sleep and deep sleep.

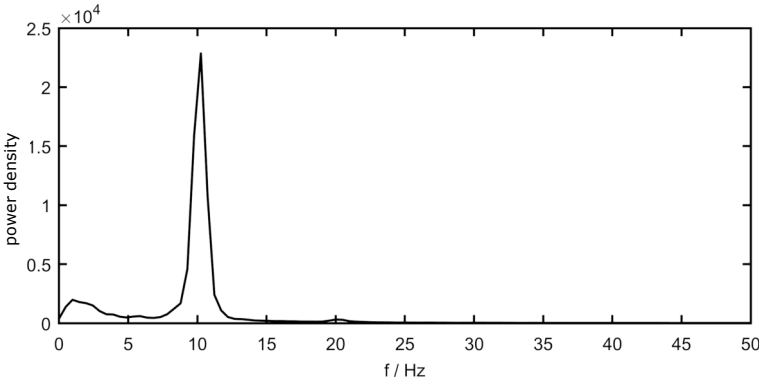


Fig. 6.3: EEG spectrum with a maximum at 11 Hz. This corresponds to alpha waves in the EEG signal that occur prominently in the relaxed awake state with eyes closed.

and a reference electrode (ear electrode) or from the voltage difference between a pair of two measuring electrodes (bipolar measurement). In an EEG measurement, several measurement curves are usually recorded and displayed simultaneously via different electrodes.

The dependence of EEG rhythmicity on wakefulness state is exploited for sleep diagnostics. A number of complaints and chronic diseases are due to sleep disorder.

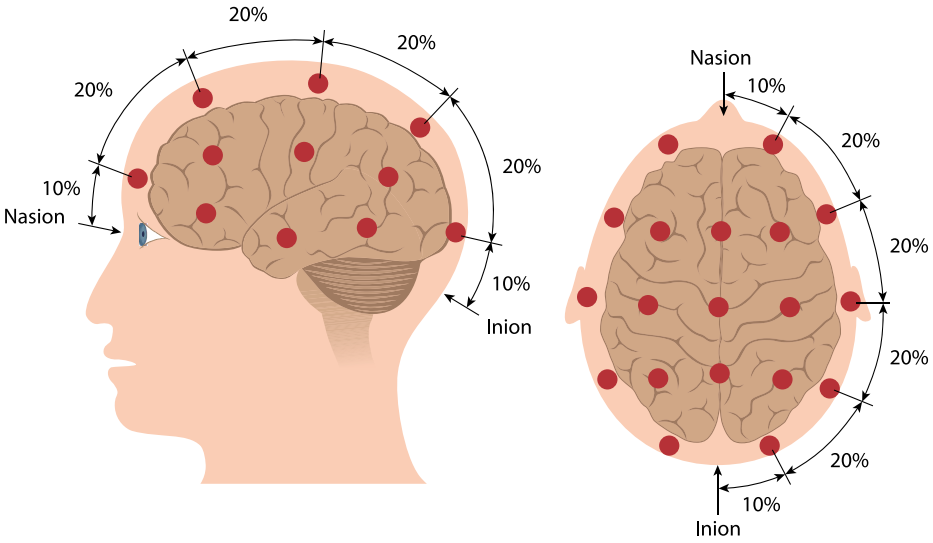


Fig. 6.4: Electrode arrangement on the skull for EEG measurement in the 10-20 system in side view (left) and top view (right).

With a continuous EEG measurement throughout the night, different sleep stages and, if necessary, sleep disorders can be identified. During falling asleep, the EEG rhythm changes from alpha to theta waves and with increasing sleep depth to delta waves. In particular, sufficiently long deep sleep phases are important for the regeneration of the body. Abrupt disturbances of the deep sleep phases for example by noise or acute oxygen deficiency due to sleep apnea lead to jumps from delta waves to alpha or theta waves in the EEG signal. In REM sleep[4], also called dream sleep, which accounts for about 20 % of the sleep time in adult humans, the brain is very active, which is expressed by a higher frequency position of the EEG signal. For sleep analysis, plotting the dominant EEG frequency as a function of time is particularly useful. The dominant EEG frequency can be determined from a maximum value analysis of the EEG spectrum.

In section 3.3 the Berger experiment was already mentioned, here now an EEG signal is to be evaluated with the help of Matlab with respect to alpha waves. The simplest way to analyze a time signal for its changes in frequency content, is the time-frequency analysis as already presented in subsection 5.3.3. The time-frequency analysis works with a time window which is slid over the signal and in which the frequency spectrum is calculated. The length of this window defines on the one hand the time resolution

4 REM: rapid eye movement; a special characteristic of this sleep phase is the rapid eye movement. Brain activity is particularly high and is perceived as a dream. The musculature, on the other hand, is largely relaxed.

and on the other hand the frequency resolution of the spectrum. For the analysis, a time signal was measured in which the patient alternates from a resting state with closed eyes (alpha state) to a resting state with open eyes (beta state). For this purpose, the EEG signal was derived at two locations at the back of the head (in the area of the optic nerves) with respect to the reference electrode at the earlobe. The evaluation of the EEG signals using the time-frequency spectrum in Matlab is given in Listing 6.1.

Listing: 6.1 Time-frequency analysis of an EEG signal from the Berger experiment.

```
A = importdata('eeg.txt');              % import EEG-data
EEG_raw = A.data(:,6);                  % select correct column
EEG_mean = EEG_raw-mean(EEG_raw);  % remove direct component
Ts = 0.001;                             % sample-period duration
Fs = 1/Ts;                              % sample-frequency
[N,nu] = size(EEG_mean);           % determined data length
t = (1:N)*Ts;                           % generated time vector

% Calculation of the time-frequency spectrum of the EEG
winlength = 2048;                       % window length
[S1,F,T] = spectrogram(EEG_mean,chebwin(winlength,100),...
    ceil(winlength/2),0:0.1:300, Fs);
S1 = abs(S1);                           % amount formation

% Calculation of the power density spectrum of the EEG
[P1xx, F1xx] = pwelch(EEG_mean, blackman(winlength), [],...
    winlength, Fs);

% graphical representation of the results
figure;
subplot(2,1,1)
plot(t,EEG_mean);
xlabel('t / s');
ylabel('Voltage U / mv');
axis([0 120 -600 600]);
title('Electroencephalogram for the Berger experiment');

subplot(2,1,2)
contourf(T,F,S1);
axis([0 120 0 30])
xlabel('t / s');
ylabel('f / Hz');
title('Time-frequency spectrum of the EEG');
```

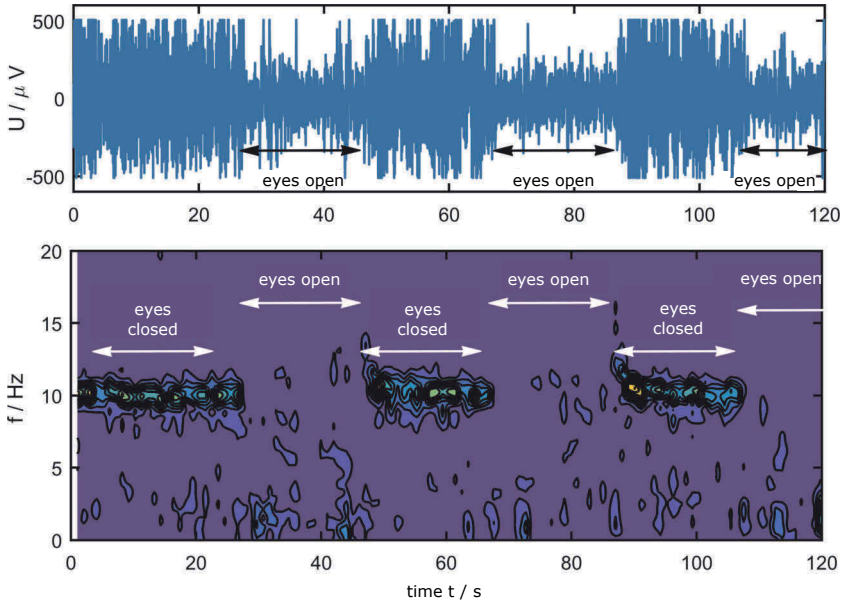


Fig. 6.5: EEG signal from the Berger experiment (top); the patient initially had his eyes closed and opened them at approximately 25, 65, and 105 s for 10 s, respectively. The time-frequency spectrum (below) shows the time course of the spectrum. In time periods with eyes closed, alpha waves occur in the frequency range of 8–13 Hz occur with varying intensity (red, yellow, green regions in the spectrum). If the eyes are closed these frequency components in the spectrum disappear again.

```
figure
plot (F1xx, (P1xx))
title('Power density spectrum of an EEG');
xlabel('f / Hz')
ylabel('Power density')
axis([0 50 0 2.5e4])
```

The results of the evaluation with Listing 6.1 are shown graphically in Figure 6.5. When the patient is in a resting state with eyes closed, dominant frequencies in the range between 8 and 13 Hz (alpha waves) occur in the time-frequency spectrum. If the eyes are opened (at 25, 65, and 105 s for the duration of 10 s, respectively), a suppression of alpha waves can be observed.

In addition to the evaluation of the EEG spectrum, analysis in the time domain is also of high diagnostic importance. In epileptic seizures, additional spikes, which can occur individually or in groups, often overlay the EEG signal. In evoked[5] potentials, an

5 Lat. *evocare* = to bring about

external stimulus is applied to a sensory organ and the EEG response is recorded. The stimulus may be, for example, visual, auditory, or olfactory[6]. Evoked potentials usually result in only a very small deflection in the EEG. The detection is done by repeating the stimulation at specific time points. The EEG sequence following each stimulus consists of a stimulus-independent part and an evoked potential sequence characteristic of the stimulus. The stimulus-independent EEG component can be stochastic signal. If many EEG sequences are added up, the stimulus-independent parts average out, and the evoked signal components of the individual EEG sequences with the same course are emphasized in the sum. The larger the number of stimuli and thus of EEG sequences, the more clearly stands out the sum of the evoked potentials from the stimulus-independent part. The evoked potentials method is used in neurology to examine the functionality of sensory organs and nerve tracts.

6.2 Signals of the Muscles and Motions

The first measurements by electromyography (EMG) date back to H. Piper in the year 1912 [59]. Since then, a large number of research groups have studied the measurement methodology and the evaluation in relation to muscle physiology and the basic processes. The fields of application of electromyography are manifold. EMG provides direct access to the action potentials of the involved muscle cells and their propagation along the fibers. Thus, myopathies such as myasthenia, amyloidosis or multiple sclerosis, but also disorders in the neuronal stimulation of muscle cells can be investigated. In sports science, EMG measurement offers the possibility to monitor the training condition, to follow the muscle build-up after injury or to analyze and optimize the course of movement. Active orthopedic prostheses are controlled by EMG signals, among other things. In ergonomics, stress conditions can be detected.

The force and speed of movement resulting from muscle contraction depend both on the number of stimulated muscle cells and on the frequency with which the action potential of the cells is triggered. It is therefore obvious to be able to draw conclusions about physiological processes via the EMG measurement. With increasing force, both the EMG amplitude and the frequency of stimulations increase. The axon of a neuron branches out to up to 500 muscle cells in the limbs. In a healthy muscle fiber, the action potential propagates from the coupling point of the electrical excitation with a velocity of about 4 m/s over the elongated muscle fiber, whereas the propagation velocity along the axon is approximately 100 m/s.

Individual motor units have dimensions on the order of 10 mm. The purpose of needle EMG is to measure the action potential of single to multiple muscle fibers in order to study their physiology or pathology. Pathologies may involve the muscle fibers

6 Lat. *olfacere* = to smell

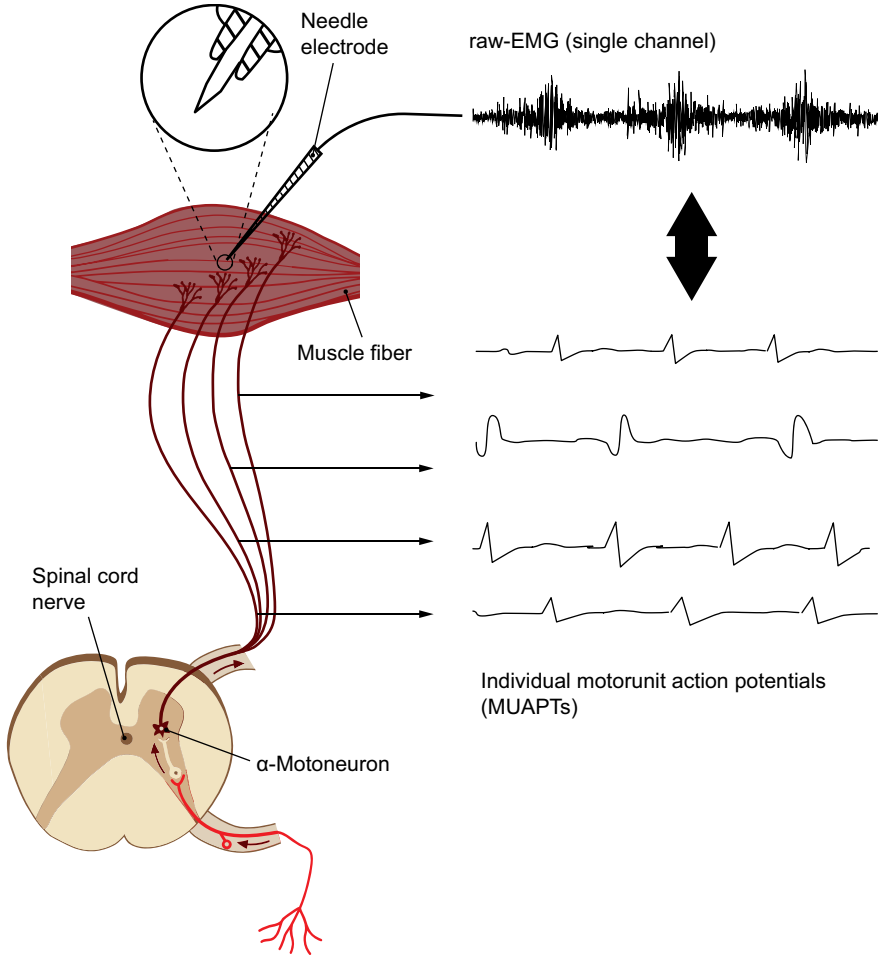


Fig. 6.6: Schematic representation of a motor unit (left) consisting of a α neuron and several muscle fibers: The cell body of the α neuron in the anterior horn of the spinal cord emits the action potential, which then innervates action potentials in the muscle fibers via the motor end plates of the branched axon at the junctions. The formation of the EMG sequence results from the sum of the individual action potentials of the involved motor units (MU)[7] (right).

themselves (myopathy) or the neurons that excite the muscle fibers (neuropathy) (cf. Figure 6.6).

Basically, an EMG can be recorded with the same technique as already presented for the measurement of an ECG. EMG needles allow the contacting of single motor units up to the detection of the action potential of single muscle fibers. This is not possible with surface electrodes due to their much poorer spatial resolution. Surface electrodes in the form of electrode arrays, on the other hand, are better suited for studies of the spatial and temporal propagation of the action potential in the individual

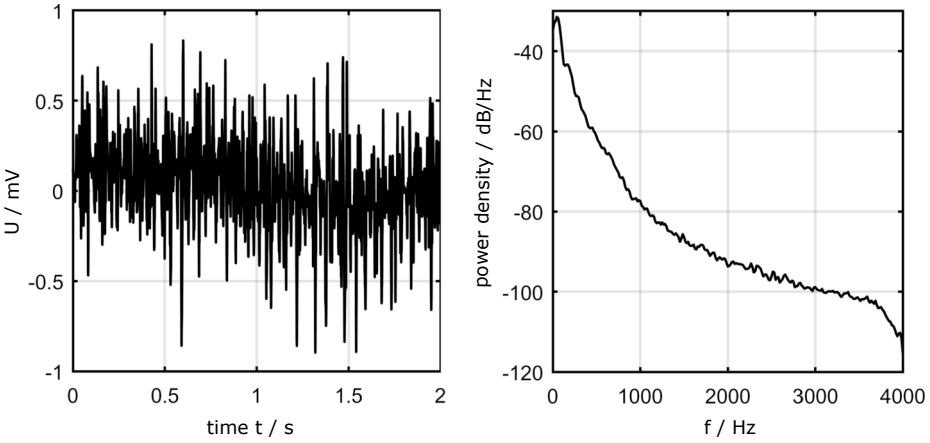


Fig. 6.7: One-channel EMG sequence from biceps over a period of 2 s (left) and power density spectrum (right).

motor units. However, even in a one-channel monitoring via two surface electrodes, information can be derived about the physiology of the underlying muscle group. A thorough summary on electromyography can be found in [50].

Figure 6.7 shows a one-channel EMG from the biceps of the upper arm measured with skin electrode. The temporal signal pattern results from stimulation of multiple muscle cells. Unlike EMG measurements with needle electrode, the bioelectrical processes of individual muscle fibers cannot be investigated here. Rather, the signal carries more global information about the stimulated muscle unit. Thus, the signal amplitude changes with the strength of the tension. Or, with sustained tension, the metabolic exhaustion of the muscle unit can be observed by a shift in the frequency position of the spectral center of gravity toward lower frequency. Overall, the EMG signal is composed of the sum of the action potentials of all muscle fibers involved and accessible via the electrodes, with muscle fibers remote from the electrodes contributing less to the overall signal strength (cf. Figure 6.6). In skeletal muscles, the action potential lasts only approx. 4 ms. To maintain a force, the action potential of the fibers is triggered repeatedly at a frequency between 8 Hz and 35 Hz.

6.2.1 Spectral Analysis of the One-Channel EMG

As already mentioned, the progressive exhaustion of a muscle unit as a result of sustained contraction is reflected in the fact, that the action potentials are triggered less frequently. Therefore, the frequency spectrum of a late EMG sequence, when meta-

7 MU: motor unit

bolic exhaustion[8] of this motor unit has already set in, is shifted to lower frequencies compared to the spectrum at the beginning of the contraction. A suitable parameter for observing the frequency shift is the frequency centroid of the power density spectrum. The power density spectrum $\Phi(\omega)$ of a function $s(t)$ is calculated from the Fourier transform of the autocorrelation of $s(t)$ according to

$$\Phi(\omega) = \int \varphi(\tau) e^{-j\omega\tau} \mathrm{d}\tau \,, \tag{6.1}$$

with the autocorrelation function

$$\varphi(\tau) = \int s(t)\, s(t+\tau) \mathrm{d}t \,. \tag{6.2}$$

The integration in Equation 6.1 and Equation 6.2 runs over all times τ and t, respectively. $\Phi(\omega)$ describes the power contained per infinitesimal frequency interval. In the case of the sequence in Figure 6.7 (left), the result for the power density spectrum is as shown in Figure 6.7 (right).

The spectral centroid f_{mean} of the power density spectrum is calculated according to

$$f_{\mathrm{mean}} = \frac{\int f\, \Phi(f) \mathrm{d}f}{\int \Phi(f) \mathrm{d}f} \,. \tag{6.3}$$

The integration limits in Equation 6.3 correspond to the first Nyquist range, i.e. from 0 to half the sample frequency. The numerator in Equation 6.3 corresponds to the first moment of the power density spectrum, the denominator describes the power contained in the signal. The implementation of the described algorithm in Matlab from Listing 6.2.1 is given below.

Listing 6.2.1: Analysis of an EMG signal in relation to the median frequency.

```
%% reads the electromyogram s1(t) from wav file, the power density
%% spectrum is formed and displayed graphically as both a
%% logarithmic as well as linear plot,
%% finally, the centre frequency of the spectrum is determined.
% Reading in and graphical representation of the signal
[s1,F1s] = audioread('emg.wav');
L1 = length(s1);        % sample number
p1 = 1:L1;
t1s = p1/F1s;
str_Fs = ['Sampling frequency F1s = ', num2str(F1s), ' 1/s.']
str_L = ['Signal length L1 = ', num2str(L1)]
```

8 Metabolic exhaustion occurs when the nutrient reserve within the cell has been within the cell has been depleted, causing the cell's action potential to be triggered less frequently.

```
str_ts = ['Time Interval t1s = [0 : ', num2str(L1/F1s), ' s].']
figure(1);
plot(t1s,s1,'b');
title('EMG-Signal');
xlabel('t [s]')
ylabel('s(t) [w.E.]')

% Calculation of power density spectrum in dB with Blackman window
% Pxx: Power density spectrum
% Fxx: Vector of frequency values
[P1xx, F1xx] = pwelch (s1, blackman(512), [], 512, F1s);
figure(3);
plot (F1xx, 10 * log10 (P1xx))
title('Power Density of an EMG Signal');
xlabel('Frequency f / Hz')
ylabel('Power density dB/Hz')

% Linear plot
figure(4);
plot (F1xx, P1xx)
hold on title('Power Density of an EMG Signal');
xlabel('Frequency f / Hz')
ylabel('Power density')

% Calculation of frequency centroid
F1xx_S=sum(P1xx.*F1xx)/sum(P1xx);
str_F1s=['Der Schwerpunkt des Signals S1 liegt bei f1 = ',...
    num2str(F1xx_S), ' Hz']
```

6.2.2 Acoustic-Kinetic Analysis of Osteoarthrosis Patients

Authors of Subsection: Jörg Subke and Benedict Schneider

The analysis of the joint angles and the ground reaction force plays an important role in the medical assessment of the human movement. If the human movement is limited by diseases - in our case osteoarthrosis [65] - the medical physician and the physiotherapist need objective criteria for their clinical diagnosis [34, 37, 64].

Osteoarthrosis is a disease of the joints where the cartilage is damaged. In the worst case the cartilage is disappeared and the bone to bone contact in the joint will produce extraordinary pain. In consequence there is the loss of mobility and the patients cannot perform e.g. their everyday sporting activities like hiking, jogging and

biking. In the worst case the patients lose their earning capacity and their independence.

However, the human joint angles deliver crucial information of the pathological mobility of the body. In this regard the ground reaction force of the feet is an important variable for the discussion about the load on the human body and the properties of the balance between the sides of the body. For the documentation of the pathological behaviour of the body, a acoustic-kinetic analysis system developed for the use in gonarthrosis [36] is used. The system consists of three measurement systems to detect:

1. the acoustic emission of the knee,
2. the ground reaction force of the feet and
3. the joint angles of hip, knee and ankle.

The three independent systems were not technical coupled and on account of this the recordings of the data were carried out in an asynchronous manner with different sampling rates. For an effective assessment the three data sets had to be synchronised. In the following discussion the approach is shown using the example of force and joint angle signals. This would be a special challenge because two different physical measured variables had to be coupled. Attributes of the signals had to be worked out to achieve the synchronisation of the measured data. The process will be exemplified by means of a clinical case.

Clinical Case

The patient of the presented clinical case is a female, 68 years old and has a body mass of 60 kg. The height of the patient is 163 cm. The patient suffers from pain in the left knee joint that occurs occasionally during routine movements and frequently under physical strain, during sports and gymnastics. Arthrosis of the left knee joint was suspected in previous investigations. In the rehabilitation clinic the physiotherapist has examined the patient to develop the therapy. To have some more objective information for the development of the therapy the acoustic-kinetic joint analysis system was applied during the examination of the patient (cf. Figure 6.8). During the examination the patient has to execute the standard movement of the acoustic-kinetic joint analysis. The standard movement are three knee bends executed one after another in the timespan of 10 seconds [78]. Prior to this the patient was invited to execute the three knee bends in her own manner [35].

To document the motion and the effect of balance of the patient the ground reaction force under the feet and the joint angles of the lower extremities were measured. The third measurement of the acoustic emission of the knee joint will not be the issue in this discussion. For further information see [37], [36] and [35].

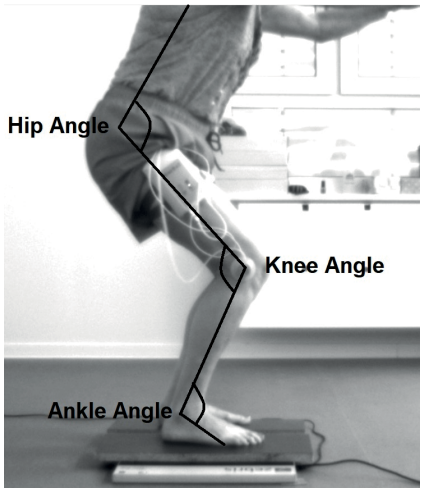


Fig. 6.8: Patient in motion instrumented with the acoustic emission system BoneDias®(near the hip) and the force plate FDM-S of Zebris®; anatomical landmarks of the patient to define the joint angles of the lower extremity (hip, knee, ankle and toe joint).

Generation of Force Data

The measurement of the ground reaction force under the feet was established by a force plate FDM-S of Zebris®. The force plate includes 2500 sensors with per square centimetre and the data is sampled at a rate of 120 Hz. The sensor area has the dimension of $54cmx33cm$. The range of force is limited by $F = 120N/cm^2$ with an accuracy of about ±5%. The Zebris® force plate thus measures the distribution of the force under the left and the right foot. The center of pressure of the left and right foot are used to conclude the distribution of the load in the left and right knee area. The total force calculated for the left and right side are one signal for the development of the synchronisation procedure (top Figure 6.9).

Generation of Kinematic Data

The measurement of the joint angles of the patient was achieve by two video cameras (Basler® Scout scA640-120gm) in the frontal and sagittal plane, respectively. The patient was not prepared with photogrammetric markers thus the kinematic analysis had to be done only with anatomical attributes like the toe, ankle, knee and hip joint as well as the shoulder joint of the trunk (cf. Figure 6.8). The analysis by means of the anatomical attributes is sufficient, a requirement of the accuracy was defined within the range of ±5 degrees. Time resolution was determined by the frame rate of 30 frames per second, this is sufficient for a photogrammetric analysis, because the motion of the patient was relatively slow during the 10 seconds of the three knee bends. As a result of the photogrammetric analysis with Tracker 5.1® there are periodic trajectories of the joint angles (bottom Figure 6.9). The trajectories exhibit the standing phase at the beginning, the turning points of the maximal flexion and the maximal extension as well as the standing phase at the end of the motion. The upright standing phases between the knee bends are recognized clearly by the maximal joint angle values.

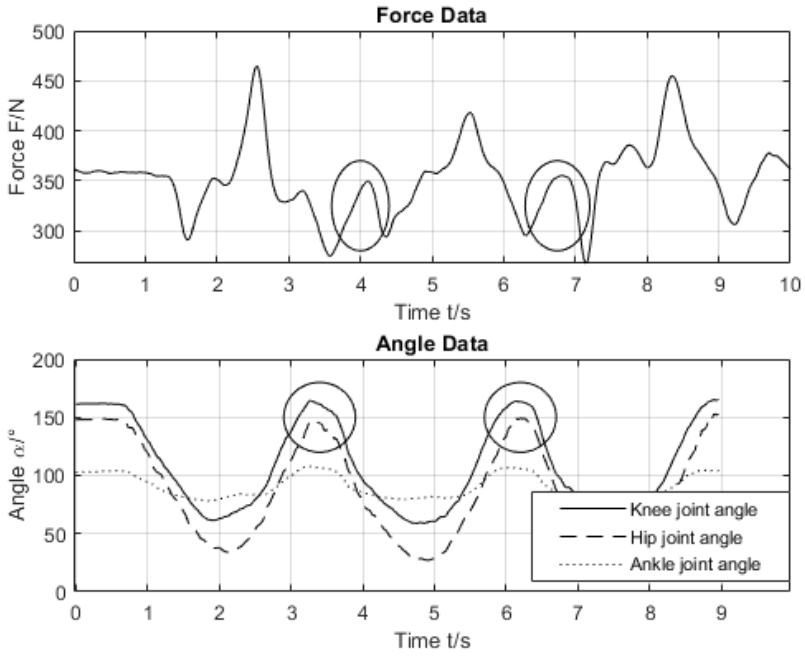


Fig. 6.9: The total force of the three knee bends of the patient (above), The result of the two-dimensional photogrammetric analysis of the three joint angles (below).

Synchronisation

The measured data sets of the technical uncoupled measurement systems had to be synchronised to get a correct temporal correlation between the two measured variables force and joint angle. The precondition for this procedure are characteristic attributes and phases of the course of the signals of the measured variables which had to be defined [36].

The periodic course of the knee angle is obvious (bottom Figure 6.9, lined), however in the course of the total force an approximate periodic course could be detected if some phases of the biomechanical behavior during the knee bends were described (cf. Figure 6.10). The periodic course in both force and joint angle data is also referred to as a characteristic pattern. By means of a detailed biomechanical comparison of the force and the knee angle curve seven phases could be defined for the patient data (cf. Figure 6.10):

1. P1: Upright stand phase
2. P2: Descending phase
3. P3: Deceleration phase - contraction of muscle force
4. P4: Deceleration - acceleration phase - center of mass slowed to halt and muscle activation for the following upward direction
5. P5: Reduction of muscle force phase

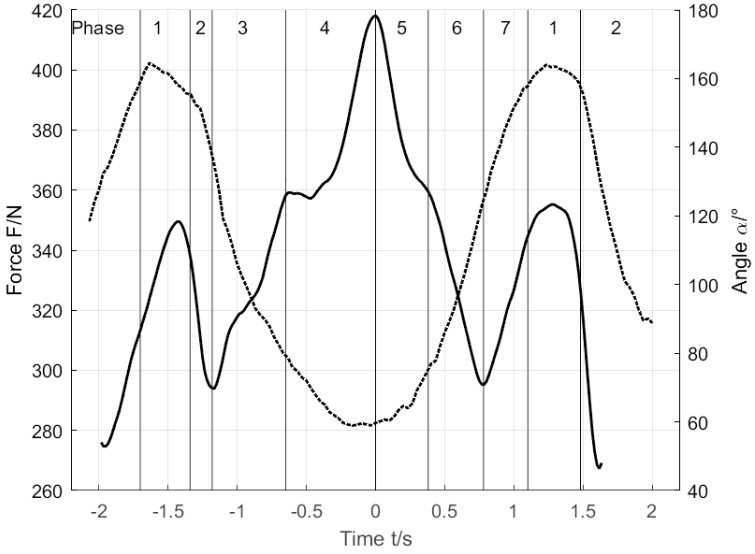


Fig. 6.10: Superimposed curves of the total force and the knee angle of the second knee bend; force (lined) and knee angle (dotted); spreading of the seven phases.

6. P6: Moment of inertia phase – further reduced support of muscle force
7. P7: Deceleration of the upward movement – activation of muscle force and center of mass slowed to halt

In preliminary assessments it appeared that the second upright stand phase between the first and the second knee bend as well as the third upright stand phase between the second and the third knee bend would be a well attribute for synchronisation [36]. In these two stand phases the patients have two physical releases between the knee bends which result in a roughly symmetric part of the force curve in these two sections. These characteristics of the force signal provide an attribute for the development of a synchronisation algorithm.

The joint angle achieves his maximal value in the upright stand phase whereas the total force achieves approximately the body weight of the patient. Sometimes there would be a local minimum in the peak of the upright stand phase [36], but in the patient data a clear maximum is apparent. These characteristic properties of the force curve in the second and third upright stand phase are suitable for the determination of two instants of time for the synchronisation (cf. Figure 6.11). There is the possibility to use these two upright stand phases as features for the superimposition of the force and the angle curve in the algorithm development.

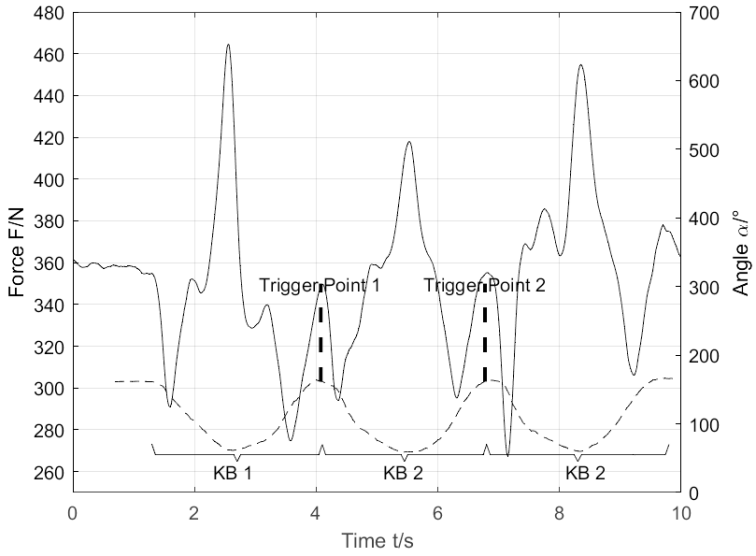


Fig. 6.11: Determination of two instants of time for the synchronisation called trigger points; KB:= knee bend, force (lined), knee angle (dotted).

Design of MATLAB algorithm

Based on manual analysis [11, 73] of the points of synchronisation an algorithm was developed which automates the first steps of synchronisation [69]. Attention should be paid on the patient data, because the results achieved, can vary individually. The selected features should minimise the inaccuracy of the synchronisation based on the measured data.

As above mentioned there are characteristic patterns in the force and joint angle data which can be used as features. Prominent features are the times of rest between the knee bends, which are further referred to as standing phases. In these phases the patient legs are mostly extended in an upright position. In Figure 6.9 the standing phases are circled black. The time difference between the standing phases in the force and joint angle curve is roughly the same thus the centres of the standing phases can be used as anchors for synchronising force and joint angle data. The centre points of the standing phases are further referred to as trigger points (TP). Regarding the joint angles we will discuss furthermore the knee joint angle because the data of the joint angle of hip, knee and ankle are in already synchronous.

Synchronisation procedure

In the following part a variety of algorithms will be presented regarding the finding of the relevant characteristic points discussed previously. In MATLAB there are sev-

eral ways to find points of interest in our data. Each way has their own advantages and disadvantages. MATLAB algorithms will be shown for the use in finding maxima, minima, and inflections in the course of the signals of the force and the knee joint angle. In the discussion of the results the focus is set on the second and third stand phases of the three knee bends which the patient had to execute. As one can see in the Figure 6.9 the signals of the ground reaction force and the knee joint angles are more or less oscillating. The second and third stand phases are represented as local maximums in the signals of the force and the knee joint angle.

First approach

The MATLAB command `findpeaks()`, is a function that allows the determination of local maxima of a vector data set. However, function options to find the correct values of the two stand phases have to be found. To find the peaks in the force and the knee joint angle data options like "MinPeakProminence" and "MinPeakHeight" are used. In the Listing 6.2.2 an example code determines peaks in both force and knee joint angle data. The results are shown in the Figure 6.12 (a and b). In the angle data the two relevant peaks of the stand phases are found but in the force data four peaks are detected., i.e. no significant prominence and other characteristic of the force signal the peaks of the stand phases could be isolated from the other peaks.

Second approach

A further approach to overcome the above problem, is to apply MATLAB functions for the derivation. As the measured data is discrete values instead of analytical functions are given Listing 6.2.2. For that reason the MATLAB the functions `diff()` and `gradient()` are used to find maxima, minima, and inflections of the force and joint angle data. This kind of approach is called numerical derivation.

Listing 6.2.2: Cut-out of the discrete measured data of the force and the knee angle; the time columns reveal the different time steps of the force and the joint angle data.

time_force	force_value	time_angle	angle_value
0	361.4984	0.00	161.1090808
0.017	361.1752	0.03	161.4235965
0.033	360.7442	0.07	161.6854311
0.05	360.3492	0.10	161.6854311
0.067	359.6668	0.13	161.6854311
0.083	359.1999	0.17	161.6854311
0.1	358.733	0.20	161.6854311
0.117	358.2661	0.23	161.6854311
0.133	357.9429	0.27	161.6854311
0.15	357.7274	0.30	161.7405815

```
0.167       357.7633    0.33        161.8509461
0.183       357.7992    0.37        161.9061601
0.2         357.8711    0.40        161.9061601
0.217       358.0147    0.43        161.5905727
0.233       358.1225    0.47        161.5905727
0.25        358.338     0.50        161.7010635
0.267       358.9126    0.53        161.3859179
0.283       359.4154    0.57        161.3859179
0.3         359.6309    0.60        161.3859179
```

To achieve improved results of the numerical derivation the measured data of the force and the joint angle are being smoothed with a moving-average filter with a window size of ten. Using the `gradient()` function the first and second derivative of both force and angle data are calculated. It must be pointed out that the underlying data is discontinuous and the numerical derivation will not deliver the exact values of zeros of the first and second derivatives. There is a need to find the zero values of the derivatives by an approximate approach. This is done by analysing the values of the derivatives where a sign change of the values appeared. If a sign change is found the indices of the points prior in the data arrays are taken and used as approximation of the zero value of the derivatives. The procedure is shown in the Listing 6.2.2 below.

In the Figure 6.12 (c) the maxima and minima of the joint angle data are found by means of the numerical derivatives. In comparison with the data of diagram of Figure 6.12 (b) the approach delivers no unique values of the two stand phases. The numerical second derivative of the joint angle is presented in the Figure 6.12 (d). The number of inflections is very high despite the smoothing of the discrete measured data. In the result of the second derivation there is no unique information about the two stand phases and a further treatment of the data seems too extensive.

In the treatment of the force data the first and the second derivation deliver many zero values in the first 1.5 seconds of the data Figure 6.12 (e and f). In comparison with the `findpeaks()` function (a) the first and second derivative of the force data deliver some more significant maxima and minima in the course of the signal. Finally, both approaches yield no unique information about the second and third stand phases of the 3 knee bends.

Third approach

In a third approach a threshold method is used to determine the points of interest. The application of thresholding to determine trigger points in force and knee joint angle data has already been proven successfully [69]. The method is explained by means of the force data.To get the threshold value an average over the force values is calculated and then multiplied with a factor of 95%. All force values above the threshold are set to NULL. With the `find()` function the indices of all elements unequal to zero are

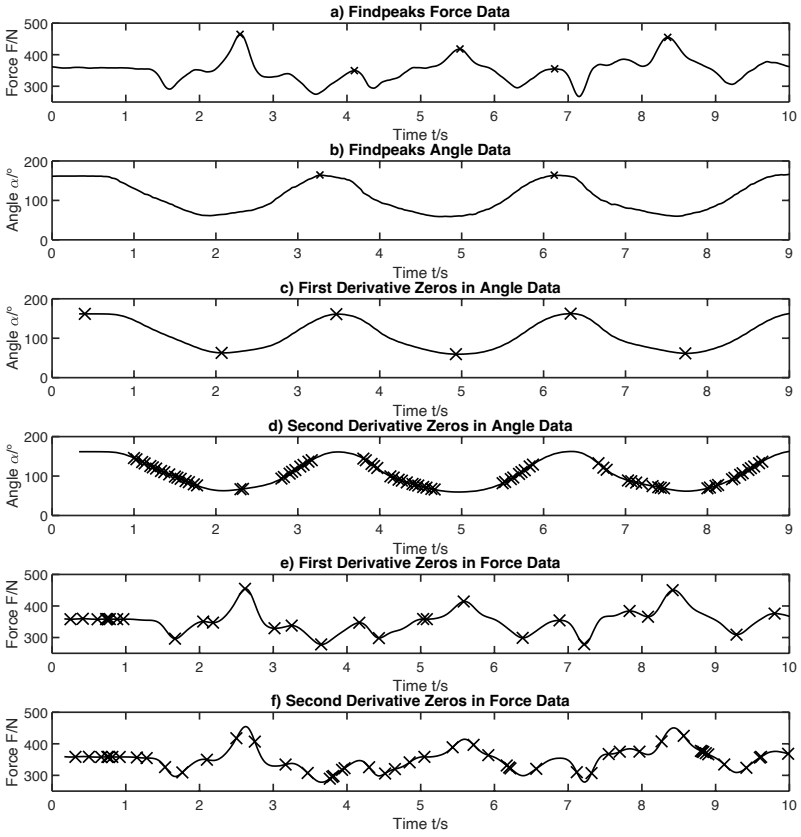


Fig. 6.12: Results of the analysis approaches by means of the MATLAB functions `findpeak()` and `gradient()`.

added to an array. These force values are further referred as local minima. The start and end points of the local minima are used to calculate the trigger points for the synchronisation. According to additional conditions of the characteristic pattern of the 3 knee bends [11], [73], the number of the local minima in the force signal must be six (cf. Figure 6.13, a). If the amount of the local minima is above six, which is very likely starting with a threshold factor of 95%, the threshold factor has to decrease by 1% and the determination of local minima is performed again until the number of six local minima is reached. The determination of the local minima of the knee joint angle data follows the same principle. But due to the different characteristic pattern of the knee joint angle signal the desired number of local minima is three (cf. Figure 6.13, b). The result of the threshold manipulation and the stored values of the local minima of the force and the knee joint angle can be seen in Figure 6.13 (c and d), where the start and end points of the extracted local minima are marked. After determining the

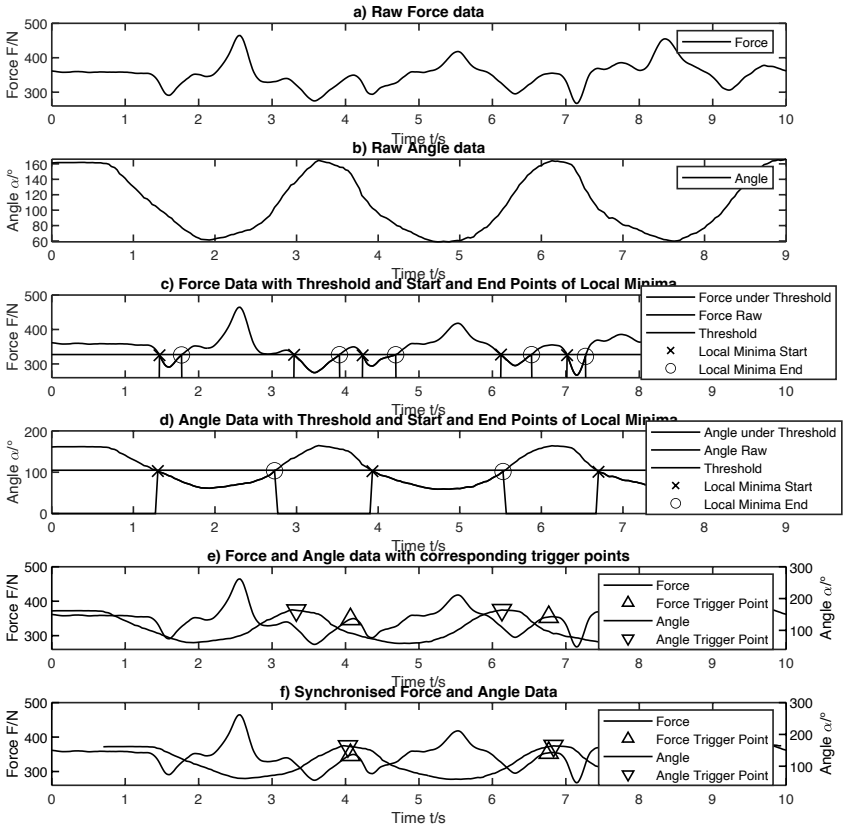


Fig. 6.13: Application of a threshold and average procedure with biomechanical conditions (the number of local minima as characteristic pattern in the force and joint angle data of the three knee bends); the second and third stand phase as synchronisation features.

local minima for knee joint angle and the force data the points which are required for the synchronisation of both data sets, the so called trigger points, are calculated. Due to the characteristic pattern of the force curve the two trigger points in the standing phases are between the second and the third local minimum and the fourth and fifth local minimum (cf. Figure 6.11).

To determine the first trigger point (TP1) the average of the time of the first force value (x Local Minima Start) in the third local minimum (cf. Figure 6.13, c) and the time of the last force value (o Local Minima End) in the second local minimum is calculated. To determine the second trigger point (TP2) the average of the time of the first force value (x Local Minima Start) in the fifth local minimum (cf. Figure 6.13, c) and the time of the last force value (o Local Minima End) in the fourth local minimum is calculated. The resulting trigger points are shown as upward pointing triangles in Figure 6.13, (e).

Because the knee joint angle curve follows a characteristic pattern too the first trigger point (TP1) is between the first and second local minimum (cf. Figure 6.13, d). The second trigger point (TP2) lies between the second and third local minimum. The average TP1 of the time of the first element (x Local Minima Start) of the second local minimum (cf. Figure 6.13, d) and the last element (o Local Minima End) of the first local minima is calculated. The average TP2 of the time of the first element (x Local Minima Start) of the third local minimum (cf. Figure 6.13, d) and the last element (o Local Minima End) of the second local minima is calculated. The resulting trigger points are also shown in Figure 6.13 (e) as downward pointing triangles.

By means of the four individual trigger points (force: TP1, TP2; joint angle: TP1, TP2) a single synchronisation time has to be generated. To synchronise the force and the joint angle data one data set has to be shifted by means of a time interval which has to be calculated. The time difference between TP1 force and TP1 joint angle should be roughly the same as the time difference between TP2 force and TP2 joint angle (cf. Figure 6.13, e). To consider the uncertainty of the measurement an average of the two time differences is calculated (see MATLAB code) and called synchronisation time. As last step the synchronisation time is added as time shift to the time of the angle data. The result can be seen in Figure 6.13 (f). Both data sets are now synchronised to a level of uncertainty of 5% [69] and can be further analysed for correlation of force and angle data.

Listing 6.2.2: Matlab example for the synchronisation and analysis of force and kinematic data of osteoarthrosis patients.

```
% Reading data
data = readmatrix('force&angle.txt', ...
    'HeaderLines', 1, 'ExpectedNumVariables', 9);
time_force = data(:,1);
value_force = data(:,2);
time_angle = data(:,3);
value_angle = data(:,4);

%show variety of algorithms to find points
f1 = figure('Name', 'Point Algorithms');
value_angle_t = value_angle(¬isnan(value_angle));
time_angle_t = time_angle(¬isnan(time_angle));
windowWidth = 10;
kernel = ones(windowWidth,1) / windowWidth;
filt_force = filter(kernel, 1, value_force);
filt_force = filt_force(11:end);
filt_time_force = time_force(11:end);
filt_angle = filter(kernel, 1, value_angle_t);
```

```
filt_angle = filt_angle(11:end);
filt_time_angle = time_angle_t(11:end);

%findpeaks
subplot(6,1,1)
[value, location] = findpeaks(value_force, "MinPeakProminence",50,...
    "MinPeakHeight",340);
plot(time_force, value_force,'k-')
hold on
scatter(time_force(location), value, 'kx')
xlabel('Time t/s')
ylabel('Force F/N')
axis([0 10 250 500])
title('Findpeaks Force Data')
subplot(6,1,2)
[value, location] = findpeaks(value_angle_t, "MinPeakProminence", ...
    100, "MinPeakHeight",150);
plot(time_angle_t, value_angle_t, 'k-')
hold on
scatter(time_angle_t(location), value, 'kx')
xlabel('Time t/s')
ylabel('Angle \alpha/deg')
axis([0 9 0 200])
title('Findpeaks Angle Data')

%Derivatives
subplot(6,1,3)
%calc first derivative
dydx = gradient(filt_angle) ./ gradient(filt_time_angle);
AA = find(dydx(1:end-1)>0 & dydx(2:end) < 0); %find maxima
AB = find(dydx(1:end-1) <0 & dydx(2:end) > 0); %find minima
plot(filt_time_angle, filt_angle,'k-')
hold on
%plot(filt_time_angle,dydx, 'k:')
scatter(filt_time_angle(AA), filt_angle(AA), 100, 'kx') %display ...
    maxima
scatter(filt_time_angle(AB), filt_angle(AB), 100, 'kx') %display ...
    minima
xlabel('Time t/s')
ylabel('Angle \alpha/deg')
axis([0 9 0 200])
title('First Derivative Zeros in Angle Data')
```

```
subplot(6,1,4)
%calc second derivative
ddyddx = gradient(dydx) ./ gradient(filt_time_angle);
BA = find(ddyddx(1:end-1)>0 & ddyddx(2:end) < 0); %find maxima
BB = find(ddyddx(1:end-1) <0 & ddyddx(2:end) > 0); %find minima
plot((filt_time_angle), filt_angle,'k-')
hold on
%plot(time_angle(1:length(ddyddx)),ddyddx, 'k:')
scatter(filt_time_angle(BA), filt_angle(BA), 100, 'kx') %display ...
    maxima
scatter(filt_time_angle(BB), filt_angle(BB), 100, 'kx') %display ...
    minima
xlabel('Time t/s')
ylabel('Angle \alpha/deg')
axis([0 9 0 200])
title('Second Derivative Zeros in Angle Data')

subplot(6,1,5)
%calc first derivative
dval = gradient(filt_force) ./ gradient(filt_time_force);
CA = find(dval(1:end-1)>0 & dval(2:end) < 0); %find maxima
CB = find(dval(1:end-1) <0 & dval(2:end) > 0); %find minima
plot(filt_time_force, filt_force, 'k-')
hold on
%plot(time_force(1:length(dval(11:end))), dval(11:end), 'k:')
scatter(filt_time_force(CA), filt_force(CA), 100, 'kx') %display ...
    maxima
scatter(filt_time_force(CB), filt_force(CB), 100, 'kx') %display ...
    minima
xlabel('Time t/s')
ylabel('Force F/N')
axis([0 10 250 500])
title('First Derivative Zeros in Force Data')

subplot(6,1,6)
%calc second derivative
ddval = gradient(dval)./ gradient(filt_time_force);
DA = find(ddval(1:end-1)>0 & ddval(2:end) < 0); %find maxima
DB = find(ddval(1:end-1) <0 & ddval(2:end) > 0); %find minima
plot(filt_time_force, filt_force, 'k-')
hold on
```

```
%plot(time_force(1:length(ddval(11:end))), ddval(11:end), 'k:')
scatter(filt_time_force(DA), filt_force(DA), 100, 'kx') %display ...
    maxima
scatter(filt_time_force(DB), filt_force(DB), 100, 'kx') %display ...
    minima
xlabel('Time t/s')
ylabel('Force F/N')
axis([0 10 250 500])
title('Second Derivative Zeros in Force Data')

%thresholding algorithm
%Calculating force local minima
threshold_value = 0.95; %set beginning threshold value
while true
    %threshold array calculation
    threshold_force = threshold_value*abs(mean(value_force));
    local_minima_force = value_force';
    local_minima_force(value_force > threshold_force) = 0;
    %segment seperated by 0 calculation
    %find nonzero elements
    force_ne0 = find(local_minima_force);
    %find local minima start indices
    force_ixstart = force_ne0(diff([0 force_ne0])>1);
    %find local minima end indices
    force_ixend = force_ne0([find(diff(force_ne0)>1) ...
    length(force_ne0)]);
    if length(force_ixstart) == 6
        break
    else
        %lower threshold value by 1%
        threshold_value = threshold_value - 0.01;
    end
end

%Calculating angle local minima
%data preparation
time_angle = time_angle';
time_angle = time_angle(¬isnan(value_angle))';
value_angle = value_angle';
value_angle = value_angle(¬isnan(value_angle))';
%threshold array calculation
threshold_angle = 0.95*abs(mean(value_angle));
```

```
local_minima_angle = value_angle';
local_minima_angle(value_angle > threshold_angle) = 0;
%segment seperated by 0 calculation
%find nonzero elements
angle_ne0 = find(local_minima_angle);
 %find local minima start indices
angle_ixstart = angle_ne0(diff([0 angle_ne0])>1);

angle_ixend = angle_ne0([find(diff(angle_ne0)>1) length(angle_ne0)]);

%Plotting raw force and angle data with thresholds and
%start and end points of local minima
f2 = figure('Name','Synchronization');
subplot(6,1,1)
plot(time_force(1:601), value_force(1:601), 'k')
xlabel('Time t/s')
axis([0 10 260 500])
ylabel('Force F/N')
title('Raw Force data')
legend ('Force')
subplot(6,1,2)
plot(time_angle(1:271),value_angle(1:271), 'k')
xlabel('Time t/s')
ylabel('Angle \alpha/deg')
title('Raw Angle data')
legend ('Angle')
subplot(6,1,3)
plot(time_force(1:601), local_minima_force(1:601), ...
    'k-')
hold on
plot(time_force(1:601), value_force(1:601), 'k:')
t_f(1:609,1) = threshold_force;
plot(time_force(1:601), t_f(1:601),'k--')
scatter(time_force(force_ixstart), ...
    local_minima_force(force_ixstart), 100, ...
    'kx')
scatter(time_force(force_ixend), local_minima_force(force_ixend), ...
    100, ...
    'ko')
xlabel('Time t/s')
axis([0 10 260 500])
ylabel('Force F/N')
```

```
title('Force Data with Threshold and Start/End Points of Local ...
    Minima')
legend ('Force under Threshold','Force Raw', 'Threshold', ...
    'Local Minima Start', 'Local Minima End')
subplot(6,1,4)
plot(time_angle(1:271), local_minima_angle(1:271), 'k-')
hold on
plot(time_angle(1:271), value_angle(1:271), 'k:')
t_a(1:278,1) = threshold_angle;
plot(time_angle(1:271),t_a(1:271),'k--')
scatter(time_angle(angle_ixstart), ...
    local_minima_angle(angle_ixstart), 100,....
    'kx')
scatter(time_angle(angle_ixend), local_minima_angle(angle_ixend), ...
    100, ...
    'ko')
xlabel('Time t/s')
ylabel('Angle \alpha/deg')
title('Angle Data with Threshold and Start/End Points of Local ...
    Minima')
legend ('Angle under Threshold', 'Angle Raw', 'Threshold', ...
    'Local Minima Start', 'Local Minima End')

%force trigger point calculation
%get times for last and first element of local minimum
%section 2&3 and 4&5 surround the two trigger points
%extract last element of 2&4 and first 3&5
idx_force_last = time_force(force_ixend)';
idx_force_first = time_force(force_ixstart)';
%calculate trigger point 1
tp_force_1 = (idx_force_first(1,3)+idx_force_last(1,2))/2;
%calculate trigger point 2
tp_force_2 = (idx_force_first(1,5)+idx_force_last(1,4))/2;

%angle trigger point calculation
%get times for last and first element of local minimum
%section 2&3 and 3&4 surround the two trigger points
%extract last element of 1&2 and first 2&3
idx_angle_last = time_angle(angle_ixend)';
idx_angle_first = time_angle(angle_ixstart)';
%calculate trigger point 1
tp_angle_1 = (idx_angle_first(1,2)+idx_angle_last(1,1))/2;
```

```
%calculate trigger point 2
tp_angle_2 = (idx_angle_first(1,3)+idx_angle_last(1,2))/2;

%calculate time adjustment for synchronization
st_1 = tp_force_1 - tp_angle_1;
st_2 = tp_force_2 - tp_angle_2;
synchronization_time = (st_1+st_2)/2;
adj_time = time_angle + synchronization_time;

%finding index of time in force and angle closest to trigger points
[¬,idx_f1]=min(abs(time_force-tp_force_1));
[¬,idx_f2]=min(abs(time_force-tp_force_2));
[¬,idx_a1]=min(abs(time_angle-tp_angle_1));
[¬,idx_a2]=min(abs(time_angle-tp_angle_2));

%plot trigger points
subplot(6,1,5)
plot(time_force, value_force,'k-')
hold on
yyaxis right
plot(time_angle, value_angle,'k:')
ylabel('Angle \alpha/deg', 'Color','k')
ax = gca;
ax.YColor = 'k';
axis([0 10 40 300])
scatter(time_angle([idx_a1 idx_a2]), value_angle([idx_a1 idx_a2]),...
    100, 'kv')
yyaxis left
axis([0 10 260 500])
xlabel('Time t/s')
ylabel('Force F/N')
scatter(time_force([idx_f1 idx_f2]), value_force([idx_f1 idx_f2]),...
    100, 'k^')
title('Force and Angle data with corresponding trigger points')
legend('Force', 'Force Trigger Point', 'Angle', 'Angle Trigger Point')

%plot synchronized force and angle data
subplot(6,1,6)
plot(time_force(1:601), value_force(1:601), 'k-');
hold on
yyaxis right
plot(adj_time, value_angle, 'k:');
```

```
scatter(adj_time([idx_a1 idx_a2]), value_angle([idx_a1 idx_a2]),...
    100, 'kv')
ylabel('Angle \alpha/deg', 'Color','k')
ax = gca;
ax.YColor = 'k';
axis([0 10 40 300])
yyaxis left
scatter(time_force([idx_f1 idx_f2]), value_force([idx_f1 idx_f2]),...
    100, 'k^')
xlabel('Time t/s')
ylabel('Force F/N')
axis([0 10 260 500])
title('Synchronised Force and Angle Data')
legend('Force', 'Force Trigger Point', 'Angle', 'Angle Trigger Point')
```

6.3 Signals of the Cardiovascular System

The analysis of signals of the cardiovascular system includes, in addition to the already discussed electrocardiogram, the evaluation of the phonocardiogram, i.e. the heart sounds and the evaluation of the photoplethysmogram to determine the oxygen saturation of the blood. As already indicated in section 2.2, these signals contain different information of the cardiovascular system and accordingly also lead to different results in the evaluation. The ECG allows conclusions about the conduction system of the heart, the phonocardiogram about the valve activity and the blood flow in the heart, and the photoplethysmogram contains information about the propagation of the pulse waves and the oxygen saturation of the blood. On the other hand, all the above signals contain, for example, information on the heart repetition rate and can be analysed in different ways in this respect. However, their determination from the ECG sequence is much more accurate than their determination from the other two signals due to the distinctive QRS complex. For diagnostic purposes, the robustness of an analysis method with regard to the interpersonal variability of signals is also of great importance, since a reliable statement is to be made for all patients.

6.3.1 Electrocardiogram

The section 3.2 on electrophysiology of the heart describes the origin and propagation of electromagnetic fields, which can also be measured outside the body. Since the measurement outside the body is more patient-friendly and less time-consuming, the standard examination consists of the measurement on the surface of the body, the time

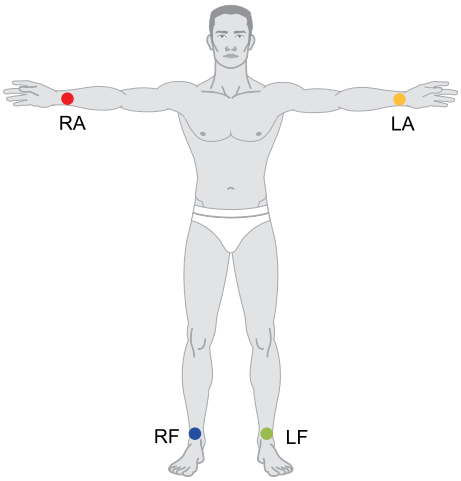


Fig. 6.14: Influence of cardiac excitation on the body surface by a central cardiac Vector P_Q in 3D representation during a vector ECG. Here, the change in the heart vector is described by a brown or light brown locus curve, which indicates the position of the vector peak as a function of time.

course of which is called electrocardiogram (ECG). The aim is to be able to examine the excitation of the heart muscles as precisely as possible so that the physician can diagnose any heart disease that may be present. If, as already shown in section 3.2, the excitation of the heart is represented by a single electrical heart vector P_Q at a central location Q in the heart, this changes its strength and direction during cardiac activity as a function of time (see Figure 6.14).

The electromagnetic field generated by the cardiac vector influences the potential at the body surface. The heart vector is a current vector that summarises all the currents of the heart cells generated by individual current dipoles (cf. subsection 3.2.2). If this relationship is approximately linear and the superposition principle applies, the potential Φ_P at any point P on the body surface can be described as a scalar product of the heart vector with a so-called *lead vector* C_Q, which has its origin at the same place as the heart-current-dipole [47], i.e. h.

$$\Phi_P = C_Q \cdot P_Q \,. \tag{6.4}$$

The difference between two potentials $\Phi_{Pi} = C_{Qi} \cdot P_{Qi}$ and $\Phi_{Pj} = C_{Qj} \cdot P_{Qj}$ on the body surface is the voltage V_{Pij} measured between the recording points i and j, which can also be expressed by a lead vector:

$$V_{Pij} = \underbrace{(C_{Qi} - C_{Qj})}_{C_{Qij}} \cdot P_Q = C_{Qij} \cdot P_Q \,.$$

The *new* Vector C_{Qij} now point not from the origin of the heart vector to the body surface, but from one recording point on the body surface to another.

In order to assess the excitation of the heart muscle well, the potentials are measured at different locations on the surface of the body. Depending on the location, number and presentation of the recording points, a distinction is made between different leads. The standard "12-channel" lead includes:

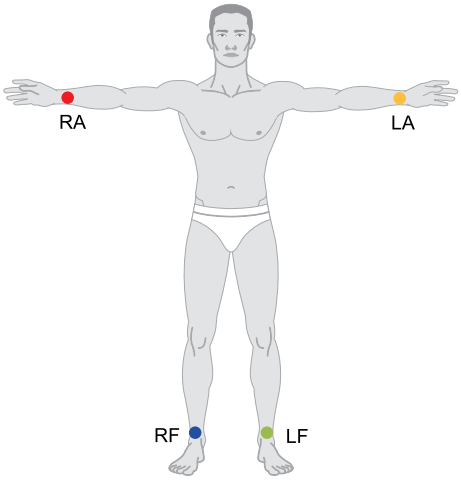


Fig. 6.15: ECG-Leads according to Einthoven and Goldberger with 3 recording Points in the Frontal Plane. The potentials are measured at the Left arm (LA), Right arm (RA) and Left Foot (LF). Grounding is applied to the right Foot (RF), which should also help to Reduce Interference such as the 50 Hz mains hum.

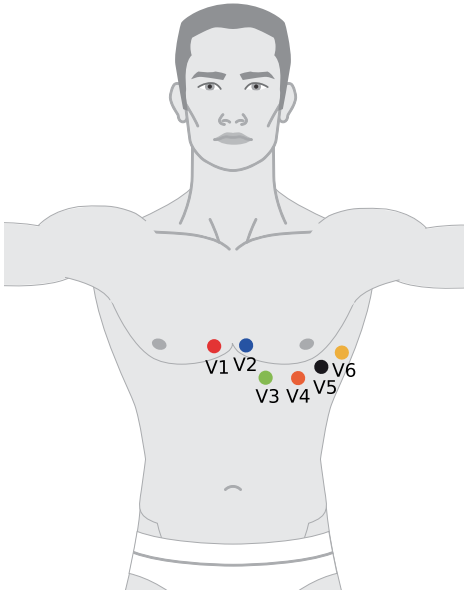


Fig. 6.16: ECG-Leads according to Wilson with 6 recording points *V*1 to *V*6 in the horizontal plane: the measuring points are approximately parallel to the 5th rib from the top.

- three potentials each according to Einthoven (I, II, III) and Goldberger (aV_R, aV_L, aV_F) in the *frontal plane*, cf. Figure 6.15,
- six potentials according to Wilson (V_1 to V_6) in the *horizontal* plane, cf. Figure 6.16.

Einthoven-Leads

As already mentioned in subsection 3.2.2, measurements are taken between two recording points each, the locations being on the left and right arms and on the left foot. The foot serves as a reference point and the measured voltages V_I, V_{II} and V_{III}

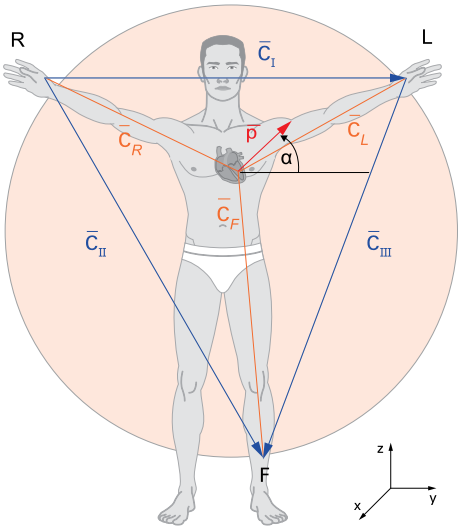


Fig. 6.17: Einthoven-Triangle with heart vector P and lead vectors C_{I}, C_{II}, C_{III} und C_{L}, C_{R}, C_{F}.

are referred to it (see Figure 6.17):

$$\begin{aligned} V_{\mathrm{I}} &= \Phi_{\mathrm{L}} - \Phi_{\mathrm{R}} = C_{\mathrm{I}} \cdot P \\ V_{\mathrm{II}} &= \Phi_{\mathrm{F}} - \Phi_{\mathrm{R}} = C_{\mathrm{II}} \cdot P \\ V_{\mathrm{III}} &= \Phi_{\mathrm{F}} - \Phi_{\mathrm{L}} = C_{\mathrm{III}} \cdot P \,. \end{aligned} \tag{6.5}$$

In order to be able to make statements about the individual unipolar voltages at the recording points in the case of voltage differences according to Einthoven, Wilson and colleagues [83, 85] suggested creating a central recording point to which all voltages can refer, and used the mean value of all recorded voltages as the central recording point, cf. Figure 6.18.

The sum of all the currents must vanish at this point because the measuring lead is highly resistive and the current can thus be neglected. The following therefore applies:

$$I_{\mathrm{L}} + I_{\mathrm{R}} + I_{\mathrm{F}} = 0 \,. \tag{6.6}$$

The voltages across the recording resistors of $R = 5\,\mathrm{k\Omega}$ then form the new unipolar voltages of the Einthoven-triangle.

$$V_{\mathrm{L}} = R \cdot I_{\mathrm{L}} \,, \quad V_{\mathrm{R}} = R \cdot I_{\mathrm{R}} \quad \text{and} \quad V_{\mathrm{F}} = R \cdot I_{\mathrm{F}} \,. \tag{6.7}$$

Since the values of the recording resistors R are all equal, it follows:

$$V_{\mathrm{L}} + V_{\mathrm{R}} + V_{\mathrm{F}} = 0 \,. \tag{6.8}$$

These new voltages can be obtained from the bipolar voltages V_{I}, V_{II} and V_{III} by applying Kirchhoff's mesh rule. One obtains e.g. according to Figure 6.18, subframe b

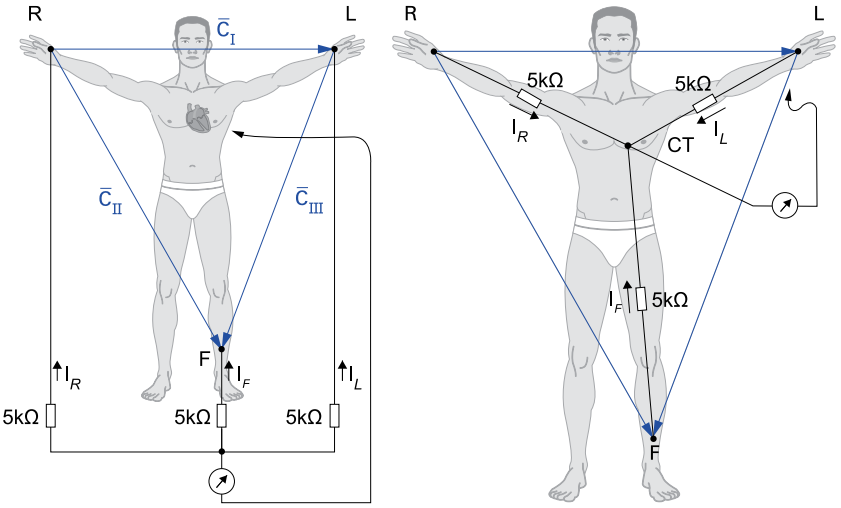


Fig. 6.18: Central reference point according to Wilson – here the mean value of all measured voltages is formed. The respective measurement is carried out with high impedance via the 5-kΩ-resistors [83, 85]. The left figure shows the actual measurement and the right figure the equivalent location in the body.

and Equation 6.8:

$$\begin{aligned} V_{\mathrm{I}} &= R \cdot (I_{\mathrm{R}} - I_{\mathrm{L}}) = V_{\mathrm{R}} - V_{\mathrm{L}} \\ V_{\mathrm{II}} &= R \cdot (I_{\mathrm{R}} - I_{\mathrm{F}}) = V_{\mathrm{R}} - V_{\mathrm{F}} = 2V_{\mathrm{R}} + V_{\mathrm{L}} \\ V_{\mathrm{III}} &= R \cdot (I_{\mathrm{L}} - I_{\mathrm{F}}) = V_{\mathrm{L}} - V_{\mathrm{F}} = 2V_{\mathrm{L}} + V_{\mathrm{R}} \end{aligned} \tag{6.9}$$

and further:

$$\begin{aligned} V_{\mathrm{R}} &= (V_{\mathrm{I}} + V_{\mathrm{II}})/3 \\ V_{\mathrm{L}} &= (V_{\mathrm{III}} - V_{\mathrm{I}})/3 \\ V_{\mathrm{F}} &= -V_{\mathrm{L}} - V_{\mathrm{R}} = -(V_{\mathrm{II}} + V_{\mathrm{III}})/3\ . \end{aligned} \tag{6.10}$$

Goldberger Leads

The unipolar voltages V_{L}, V_{R} and V_{F} with the central recording point according to Einthoven are smaller than the voltages V_{I}, V_{II} and V_{III} of the bipolar derivatives according to Einthoven. Therefore Goldberger [21, 22] proposed another reference point by which, using the same measuring points , the larger measuring voltages aV_{L}, aV_{R} and aV_{F} arise. [9]. Here the measured voltages aV_{L}, aV_{R} and aV_{F} are not formed on the central potential according to Wilson by averaging, but are related to the potential at a

9 The subscript *a* before the designations of the stresses stands for *augmented*, which means augmentation.

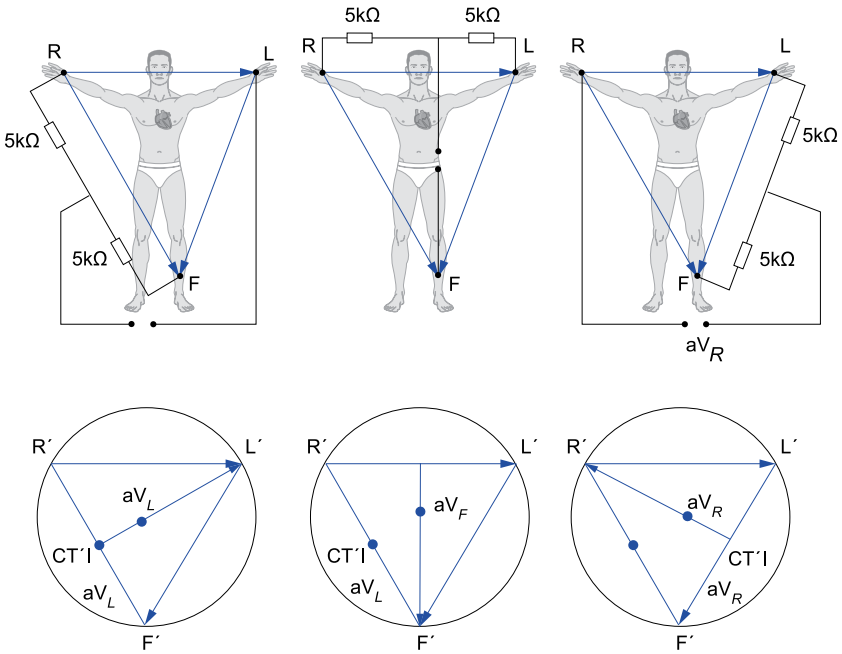


Fig. 6.19: Magnification of the unipolar measurement voltages in the ECG lead according to Goldberger [21, 22]: The figure above shows the recording arrangement, the figure below the associated recording of the voltage vectors.

virtual mass point, which is composed of the average of the other potentials as follows:

$$aV_{\mathrm{L}} = V_{\mathrm{II}}/2 - V_{\mathrm{III}}$$
$$aV_{\mathrm{R}} = V_{\mathrm{III}}/2 - V_{\mathrm{II}}$$
$$aV_{\mathrm{F}} = V_{\mathrm{I}}/2 - V_{\mathrm{II}}\,.$$

Figure 6.19 shows how the associated voltage vector increases by approximately 30 %.

Wilson Leads

In the Wilson ECG leads, the heart vector is measured in the horizontal plane. In this cross-section, it is sometimes easier to detect various heart diseases (e.g. special myocardial infarction) than in the frontal plane, because the electrodes are closer to the heart than in the Einthoven or Goldberger leads. As already shown in Figure 6.16, six electrodes ($V1$ to $V6$) are placed in the same upper body plane, at the level of the 4th and 5th rib from the top (see Figure 6.20).

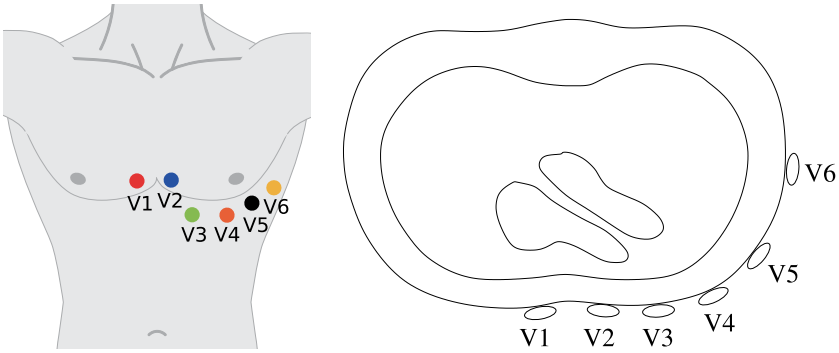


Fig. 6.20: Electrode arrangement in leads according to Wilson [84]: The illustration on the left shows the arrangement in the vertical plane, on the right in the horizontal plane with the projections of the heart vector.

6.3.1.1 Analysis of the ECG

For the physician, it is not enough just to record an ECG. He must be able to recognise from its course whether a disease is present. To be able to do this, various characteristics must be determined from the ECG curve. These include not only the heart rate, but also the regularity and the shape of the ECG curve. From this, for example, a dysrhythmia or a heart attack can be recognised.

The ECG measured in real life contains various disturbances and distortions. These include: 50 Hz mains hum, electromagnetic influences such as those caused by neon lights, radios, TVs or PCs, as well as short-term influences when the patient moves during the measurement and the contact resistance between the skin and the electrode is changed by it. Within the framework of this book, not all procedures for recognizing the features of the ECG curve can be dealt with, but two important determinations will be presented in more detail as examples, namely i) the recognition of the QRS complex in a noisy ECG according to Pan-Tompkins and ii) the determination of the heart rate and its variability.

Both procedures actually belong together. If the QRS complex is known, it is possible to determine the heart rate and from its temporal course –the heart never beats at exactly the same frequency, because this is influenced by the nervous system depending on the load –the frequency spectrum. It is not the person whose pulse is almost always uniform who is particularly healthy. On the contrary, several studies have shown that a healthy heart has variable pulse rates (within certain limits).

6.3.1.2 Determination of the QRS Complex by the Pan-Tompkins Method

J. Pan and W. J. Tompkins, in their 1985 publication [58], introduced a method for detecting the QRS complex in an ECG that is very robust and insensitive to interference. The advantage of this method is also that it can be applied in real time and used, for ex-

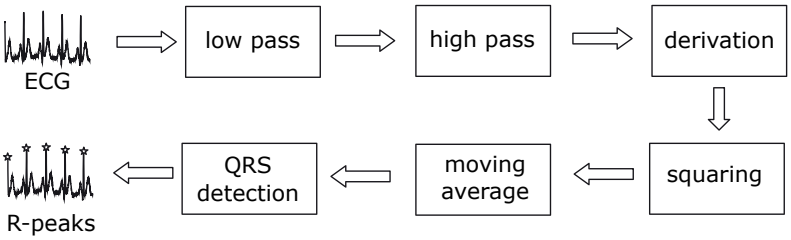


Fig. 6.21: Basic Principle of Pan-Tompkins-algorithm.

ample, in an implantable pacemaker. It was adapted and modified several times after publication. The basic principle, however, has remained the same, cf. Figure 6.21.

Pan and Tompkins were using a simple Z80- microprocessor at the time and had to be careful to use simple algorithms that did not overtax this processor, mainly because of real-time requirements. Therefore, they could not use common standard filters with higher filter degrees as well as set coefficients that had to be accurate, for example, up to the fourth digit after the decimal point. Their goal was to be able to implement the algorithms using simple values that can be expressed as powers of 2, such as $8 = 2^3$; because such coefficients can be represented by simple shift-operations of the binary memory contents. In addition, the analogue biosignals must also be converted into digital values, i.e. the analogue values are limited in their spectral range beforehand so that the sampling theorem is fulfilled. They then have to be sampled and digitised. Fortunately, the biological signals of the heart are not so high-frequent and do not place excessive demands on the sampling rate. Here, a sampling frequency of 200 Hz is sufficient, which could be realised well with A/D- converters even back then. However, the A/D converters at that time only had a low resolution, and so the developers had to use 8 bit, i.e. 256 values for their filter coefficients.

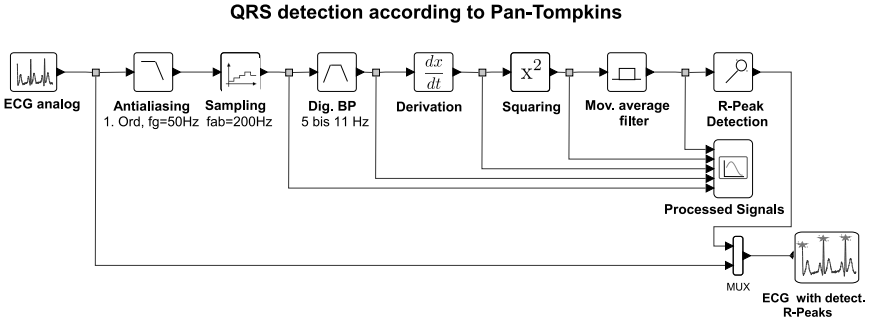


Fig. 6.22: Simulation model of Pan-Tompkins-Algorithm for Matlab/Simulink or Scilab/Xcos.

Nowadays, of course, more powerful hardware can be used. The development engineer can design the algorithm on the PC and also test it in hardware by coupling it to the PC in real time. Hardware today does not always have to be built in the lab to be tested. An intensive simulation with PC hardware-coupling shortens the development enormously. The exact structure for simulating the Pan-Tompkins-algorithm, as it can be realised e.g. in Matlab/Simulink or Scilab/Xcos, is shown Figure 6.22. Of course, it is also possible to write programs in Matlab or Scilab, C or Fortran, but the graphical representation is often much easier to understand than a complicated program where after years even the developer has difficulties to understand his program.

Pre-filtering

An analogue low-pass filter first limits the magnitude frequency spectrum to 50 Hz. This serves to comply with the sampling theorem when subsequently sampling at 200 Hz.[10]

Bandpass Filtering

In the next step, the ECG, which is available in digitised form, is additionally bandpass filtered in the discrete-time domain. Since most of the energy of the QRS complex lies between about 5 to 15 Hz [77], a bandpass in the range of 5 to 11 Hz is chosen. In this range, however, a bandpass is difficult to realise with various methods. Therefore, it is designed as a series circuit of a low-pass and a high-pass. The low pass has easily realisable coefficients,

$$A_{\mathrm{TP}}(z) = \frac{\left(1 - z^{-6}\right)^2}{\left(1 - z^{-1}\right)^2}\,, \quad \text{mit} \quad |A_{\mathrm{TP}}(\omega T_a)| = \frac{\sin^2(3\,\omega T_a)}{\sin^2(\omega T_a/2)}\,, \tag{6.11}$$

a cut-off frequency of 11 Hz and a gain of 36 (cf. Figure 6.23).

Its group delay is constant 5 samples. This corresponds to $5 \cdot (1/200\,\mathrm{Hz}) = 25$ ms for a sampling frequency of 200 Hz. Because of

$$A_{\mathrm{TP}}(z) = \frac{Y_{\mathrm{TP}}(z)}{X_{\mathrm{TP}}(z)}\,, \quad X_{\mathrm{TP}}, Y_{\mathrm{TP}}\text{: input-, output spectrum} \tag{6.12}$$

is further obtained with Equation 6.11:

$$Y_{\mathrm{TP}}(z) \cdot (1 - 2z^{-1} + z^{-2}) = X_{\mathrm{TP}}(z) \cdot (1 - 2z^{-6} + z^{-12}) \tag{6.13}$$

and after transforming back to the time domain, the corresponding algorithm for the output signal $y_{\mathrm{TP}}(n)$ as a function of the input signal $x_{\mathrm{TP}}(n)$:

$$y_{\mathrm{TP}}(n) = 2y_{\mathrm{TP}}(n-1) - y_{\mathrm{TP}}(n-2) + x_{\mathrm{TP}}(n) - 2x_{\mathrm{TP}}(n-6) + x_{\mathrm{TP}}(n-12)\,. \tag{6.14}$$

10 The upper cut-off frequency must be less than half the sampling frequency.

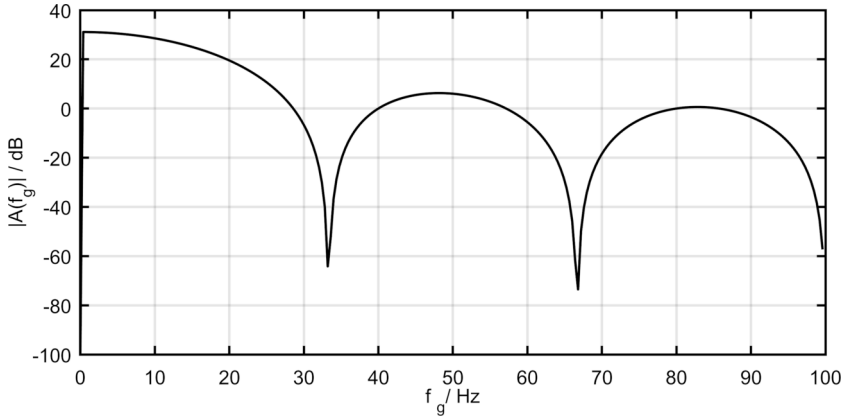


Fig. 6.23: Amount-frequency-response $|A_{TP}(f)|$ for the low-pass part of the bandpass according to Figure 6.22.

The high pass is realised by subtracting the output of a first order low pass from the output of an all pass, which delays the signal by an additional 16 samples. This corresponds to 80 ms at a sampling frequency of 200 Hz. It has no gain (or a gain of 1) and a cut-off frequency of approximately 5 Hz, see Figure 6.24. Thereby one obtains for the transfer function $A_{\mathrm{HP}}(z)$

$$A_{\mathrm{HP}}(z) = \frac{-1/32 + z^{-16} - z^{-17} + z^{-32}/32}{1 - z^{-1}} \tag{6.15}$$

and the corresponding magnitude frequency response

$$|A_{HP}(\omega T_a)| = \sqrt{1 + \left(\frac{\sin(16\omega T_a)}{32\,\sin(\omega T_a/2)}\right)^2 - \frac{\sin(16\omega T_a)}{16\,\tan(\omega T_a/2)}}\,. \tag{6.16}$$

Analogous to the procedure for the low-pass filter, the following is obtained for the associated algorithm for the output signal $y_{\mathrm{HP}}(n)$ as a function of the input signal $x_{\mathrm{HP}}(n)$:

$$y_{\mathrm{HP}}(n) = y_{\mathrm{HP}}(n-1) - x_{\mathrm{HP}}(n)/32 + x_{\mathrm{HP}}(n-16) - x_{\mathrm{HP}}(n-17) + x_{\mathrm{HP}}(n-32)/32\,. \tag{6.17}$$

Differentiate

The QRS complex in the ECG has steep edges, especially at the R wave. In order to emphasise these, the output signal of the bandpass filter is now digitally filtered using a fifth-degree non-recursive filter according to the numerical approximation

$$\widetilde{y}_{\mathrm{Dif}}(n) = \big(2x_{\mathrm{Dif}}(n+2) + x_{\mathrm{Dif}}(n+1) - x_{\mathrm{Dif}}(n-1) - 2x_{\mathrm{Dif}}(n-2)\big)/8 \tag{6.18}$$

and differentiated. However, since the digital filter must be causal in a real-time application, a delay of 2 clock instants (equivalent to $2 \cdot 5$ ms $= 10$ ms) must be applied

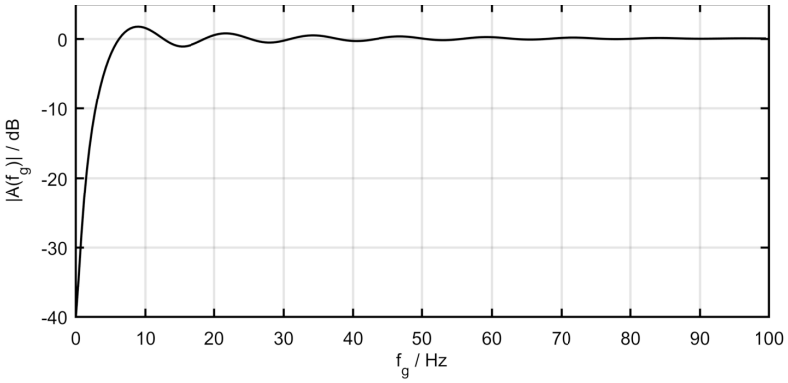


Fig. 6.24: Amount frequency response $|A_{HP}(f)|$ for the high-pass component of the bandpass according to Figure 6.22.

so that only samples that lie in the past can be taken. The associated filter algorithm will therefore produce a delayed differentiation:

$$y_{\mathrm{Dif}}(n) = \tilde{y}_{\mathrm{Dif}}(n-2) = (2x_{\mathrm{Dif}}(n) + x_{\mathrm{Dif}}(n-1) - x_{\mathrm{Dif}}(n-3) - 2x_{\mathrm{Dif}}(n-4))/8\,. \quad (6.19)$$

The transfer function $A_{\mathrm{Dif}}(z)$ is then obtained according to

$$A_{\mathrm{Dif}}(z) = \frac{Y_{\mathrm{Dif}}(z)}{X_{\mathrm{Dif}}(z)} = \frac{1}{8}(2 + z^{-1} - z^{-3} - 2z^{-4}) \quad (6.20)$$

with the associated magnitude frequency response represented in Figure 6.25 in double logarithmic measure.

$$|A_{\mathrm{Dif}}(\omega T_a)| = \frac{1}{4}[2\sin(2\,\omega T_a) + \sin(\omega T_a)]\,. \quad (6.21)$$

From the double logarithmic plot, it can be seen that the differentiator is up to approximately 30 Hz the magnitude frequency response rises linearly and then drops steeply. Up to approximately 30 Hz, however, most spectral ranges are contained. The drop thereafter causes an additional attenuation of higher-frequency interference components (e.g. 50-Hz-mains hum).

Square

This non-linear limb causes additional amplification of the QRS complex highlighted by the differentiator according to the simple instruction

$$y_{\mathrm{quad}}(n) = x^2_{\mathrm{quad}}(n)\,, \quad x_{\mathrm{quad}}, y_{\mathrm{quad}}\text{: input and output of the squarer} \quad (6.22)$$

Window Integration (MA – Moving Average)

In the subsequent moving-window-integration, 30 samples (corresponding to $30 \cdot 5\,\mathrm{ms} = 150\,\mathrm{ms}$) are averaged in a time interval and continuously output depending

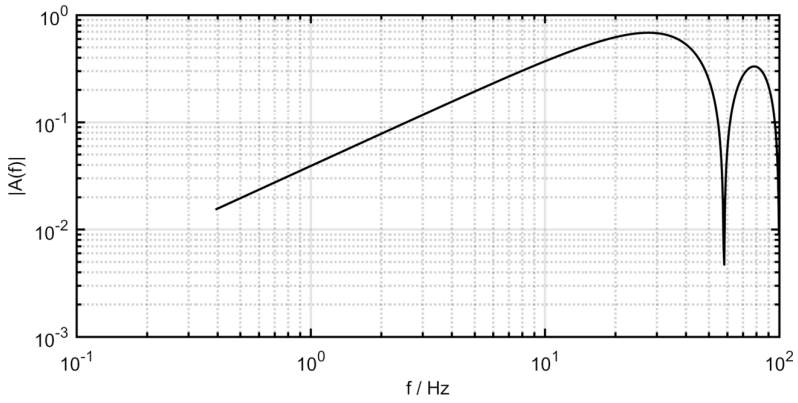


Fig. 6.25: Amount frequency response of the differentiator $|A_{\mathrm{Dif}}(\omega T)|$ according to Figure 6.22.

on the window position of the n-th sample in which the $N = 30$ samples are located. The algorithm is

$$y_{\mathrm{MA}}(n) = [x_{\mathrm{MA}}(n) + x_{\mathrm{MA}}(n-1) + \cdots + x_{\mathrm{MA}}(n-(N-1))]/N\,.$$

In the time domain, we then obtain for the transfer function $A_{\mathrm{MA}}(z)$:

$$A_{\mathrm{MA}}(z) = \frac{Y_{\mathrm{MA}}(z)}{X_{\mathrm{MA}}(z)} = \frac{1}{N}\sum_{i=0}^{N-1} z^{-i} = \frac{1}{N}\frac{1-z^{-N}}{1-z^{-1}}\,, \quad \text{geometcic series} \tag{6.23}$$

and thus for the magnitude frequency response (cf. Figure 6.26):

$$|A_{\mathrm{MA}}(\omega T)| = \left|\frac{1}{N} z^{-\frac{N-1}{2}} \cdot \frac{\sin(\frac{N}{2}\omega T)}{\sin(\omega T/2)}\right| = \frac{1}{N}\left|\frac{\sin(\frac{N}{2}\omega T)}{\sin(\omega T/2)}\right|. \tag{6.24}$$

From the transfer function in Equation 6.24, we see that the window integrator has a signal delay of $(N-1)/2$ samples. With a window width of 30 samples this corresponds to $(30-1)/2 \cdot 5\,\mathrm{ms} = 72.5\,\mathrm{ms}$.

In the subsequent investigation of whether a QRS complex is present, it must be noted that the ECG signal at the output of the sliding window integrator is delayed compared to the original ECG (see Figure 6.28). The delay results from the 25 ms for the low-pass, the 80 ms for the high-pass, the 10 ms for the differentiator and the 72.5 ms for the window integrator at a sampling frequency of 200 Hz. The total delay time is therefore 25 ms + 80 ms + 10 ms + 72.5 ms ≈ 190 ms and must be considered when determining the location of the QRS complex.

This can also be seen in the example shown in Figure 6.27, where for an ECG disturbed by noise, the signals at the output of each processing block are shown.

Searching the QRS Complex

After integrating 30 samples through the sliding window, the search for the QRS complex begins both in the output signal after window integration and in the ECG signal

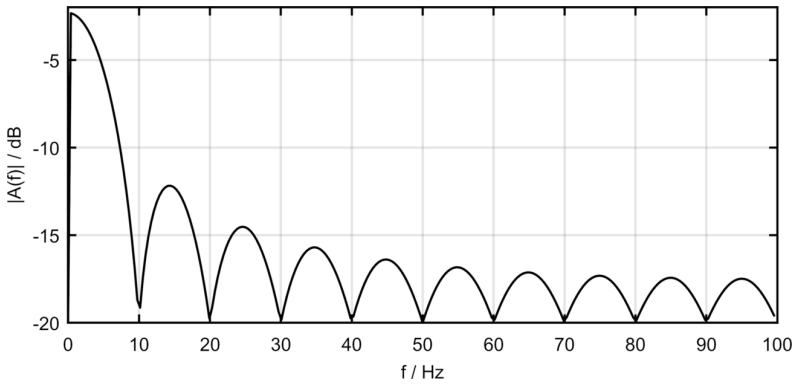


Fig. 6.26: Amount frequency response of the moving window integrator (MA) $|A_{\mathrm{MA}}(\omega T)|$ according to Figure 6.22.

after bandpass filtering. This search is quite complicated, as the QRS complex should be found even in a very disturbed ECG signal. Many subsequent publications on this problem have attempted to simplify this procedure, especially when it is not to be used in real time. The original procedure is divided into several sections:

1. **Measurement preparations**: Since the rising edge of the output signal after window integration is related to the location of the QRS complex (cf. Figure 6.28), this allows the associated locations to be approximated. This preliminary investigation is carried out exclusively on the output signal of the window integrator (MA). In Matlab, however, there is still the possibility to determine local maxima in the output signal *PEAKI* of the window integrator with the command `findpeaks()`, which makes the programming very easy. A minimum distance between the local maxima can also be specified for this command; because it is not possible, due to the required recovery time of the heart cells, for two QRS complexes to have a shorter distance than approximately 200 ms from each other. Two thresholds are provided: *THRESHOLDI*1 to further investigate if it is a QRS complex, and *THRESHOLDI*2 if the first threshold is not exceeded and decided whether to initiate a reverse search for a QRS complex not found. Then the local maximum must be greater than the second threshold. In [72] another "training phase" of 2 seconds is provided to initially determine the threshold values.
2. **Adaptive threshold adjustment:** After measurement preparation, further maxima are searched for in the output signal *PEAKI* of the window integrator. The threshold values are adjusted adaptively. If the current local maximum exceeds the threshold *THRESHOLD*_I1, this maximum is interpreted as belonging to a QRS complex (*signal peak*), if it lies between the thresholds *THRESHOLD*_I1 and *THRESHOLD*_I2, it is classified as a maximum in the noise signal (*noise peak*), which initiates a later reverse search. Here, not the local maxima in the signal *PEAKI* are chosen as a measure for the size of the QRS complex, but estimated

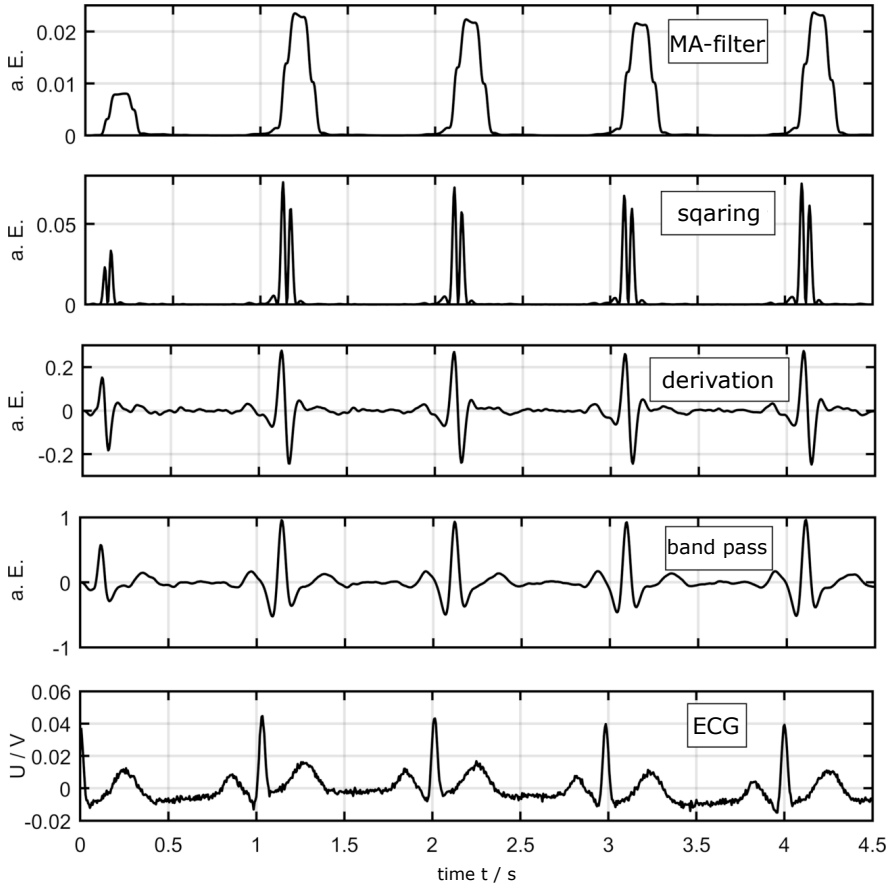


Fig. 6.27: Output signals of the individual blocks in Figure 6.22 according to the Pan-Tompkins-algorithm.

values *SPKI* and *NPKI*, which are calculated according to

$$SPKI = 0.125 \cdot PEAKI + 0.875 \cdot SPKI \quad \text{at signal peak}$$
$$NPKI = 0.125 \cdot PEAKI + 0.875 \cdot NPKI \quad \text{at Noise Peak}$$

adaptively. The thresholds are modified as follows:

$$THRESHOLD\_I1 = NPKI + 0.25 \cdot (SPKI - NPKI)$$
$$THRESHOLD\_I2 = 0.5 \cdot THRESHOLD\_I1 \,.$$

If another QRS complex was actually found during the backward search initiated by exceeding the second threshold, the estimated value of the QRS complex *SPKI* is further corrected:

$$SPKI = 0.25 \cdot PEAKI + 0.75 \cdot SPKI \,.$$

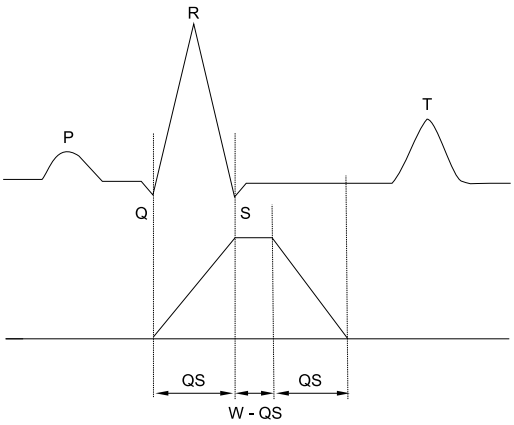


Fig. 6.28: Delay of the sliding window integrator (MA) output signal by 190 ms in the lower part of the graph compared to the schematised original ECG in the upper part; the time interval of the rising edge of the window integrator corresponds to the width QS of the QRS complex.

In addition to finding the QRS complex in the output signal of the window integrator, they must also be present in the output signal of the original ECG after bandpass filtering, and they must be able to be found there. There, ECG QRS signal and noise values are estimated and threshold values are determined in an analogous way:[11]

$$\begin{aligned} SPKF &= 0.125 \cdot PEAKF + 0.875 \cdot SPKF \quad \text{at signal peak} \\ NPKF &= 0.125 \cdot PEAKF + 0.875 \cdot NPKF \quad \text{at Noise Peak} \\ THRESHOLD\_F1 &= NPKF + 0.25 \cdot (SPKF - NPKF)\,, \\ THRESHOLD\_F2 &= 0.5 \cdot THRESHOLD\_F1\,, \end{aligned}$$

and if the reverse search was successful:

$$SPKF = 0.25 \cdot PEAKF + 0.75 \cdot SPKF\,.$$

In the case of irregular heartbeats, the first threshold value for both signals is reduced:

$$\begin{aligned} THRESHOLD\,I1 &\leftarrow 0.5 \cdot THRESHOLD\,I1 \\ THRESHOLD\,F1 &\leftarrow 0.5 \cdot THRESHOLD\,F1\,. \end{aligned}$$

3. **Adjustment of intervals** between QRS complexes. (*RR intervals*) and the pulse rate limits: The algorithm determines two average values of the intervals between the QRS complexes (*RR intervals*).
   (a) The first average value *RR_AVERAGE*1 is the average of the last 8 heartbeats.
   (b) The second average *RR_AVERAGE*2 is the average of the last 8 heartbeats that have a heart rate between 92 % and 116 % of the previous average rate

11 The variables differ only by a single letter at the end. *I* indicates the reference to the integrated output signal and *F* the reference to the output signal filtered by the bandpass.

*RR_AVERAGE*2.

$$RR\_LOW\_LIMIT = 0.92 \cdot RR\_AVERAGE2$$
$$RR\_HIGH\_LIMIT = 1.16 \cdot RR\_AVERAGE2$$
$$RR\_MISSED\_LIMIT = 1.66 \cdot RR\_AVERAGE2\ .$$

If no QRS complex within 166 % of the previous. Heart rate *RR_AVERAGE*2 the largest local maximum between the previously thresholds set so far is chosen as the QRS complex.

If all heart rates of the last RR intervals between the thresholds *RR_LOW_LIMIT* and *RR_HIGH_LIMIT*, the heart rate is considered to be regular, and the average value *AVERAGE_*1 is equal to the average value *AVERAGE_*2. In this case, there is a normal sinus rhythm.

4. **Detection of T-wave:** If the time between the R peaks should be less than 360 ms the time between the "R" spikes in an ECG must be greater than the 200 ms recovery time of the cardiac cells which corresponds to a heart rate greater than 167/min, it is investigated whether the maximum slope of the QRS complex found is is smaller than half of the average slopes of the QRS complexes found so far. If not, the "new" QRS complex found will be indentified as false and is interpreted as a T-wave. (cf. Figure 3.16).

There are several implementations of this algorithm in Matlab, but they have often been modified and simplified. A complete program also with backward search according to the method published by Pan and Tompkins [58] was written in Matlab by Hooman Sedghamiz, for example [72].[12] An application example of this programme for determining QRS complexes is shown Figure 6.29. There, both the output signal of the window integrator and the ECG signal after bandpass filtering are shown. The adaptive thresholding and signal and noise level estimation procedures are also evident.

6.3.1.3 Determination of heart rate and its variability

Heart rate variabilityheart rate variability (HRV) describes the variation of the time interval between successive heartbeats, i.e. the heart rate HF_n with the definition:

$$\begin{aligned} HF_n &= \frac{1}{RR_n} \qquad \text{with} \\ RR_n &= R_n - R_{n-1} \\ R_n &: \text{ location of } n\text{-th R-pinch}\ . \end{aligned} \tag{6.25}$$

12 It can be found in ResearchGate at https://www.researchgate.net/publication/313673153/_Matlab_Implementation_of_Pan_Tomp-kins_ECG_QRS_detector

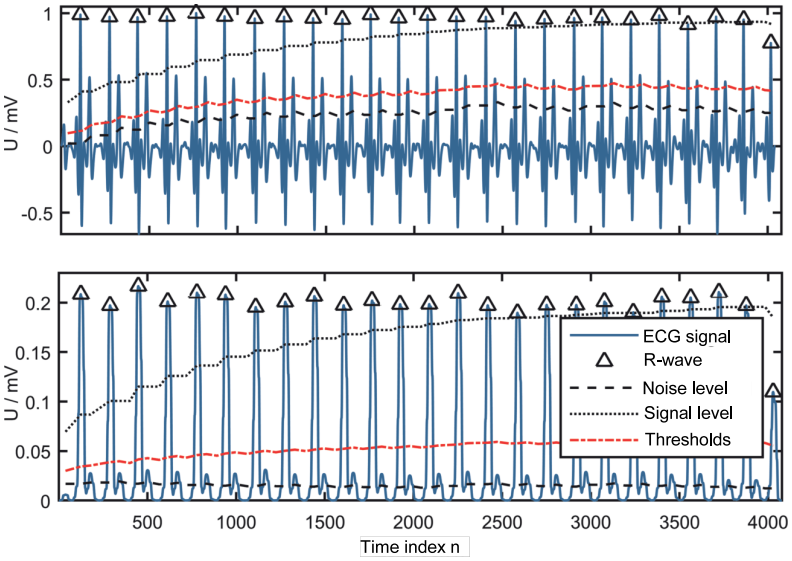


Fig. 6.29: Example of QRS search in an ECG signal using Sedgehamiz's Matlab-program with Pan-Tompkins adaptive threshold adjustment procedure [72]: In this illustration, the adaptive threshold adjustment can be seen, shown as dashed horizontal bars. The top figure shows QRS search in the filtered ECG at the output of the bandpass, and the bottom figure shows QRS search after window integration (MA filtering).

Both the sympathetic and parasympathetic parts of the autonomic nervous system influence heart rate. The sympathetic part can increase the heart rate, while the parasympathetic can decrease it. The control of autonomic regulation involves several interconnected areas of the autonomic central nervous system. Furthermore, additional external nervous system areas such as arterial regulation of blood pressure through the baroreceptor reflex and regulation of respiration can also cause rapid changes in heart rate through the sympathetic and parasympathetic nervous systems. The baroreflex is based on baroreceptors, which are located on the walls of some large blood vessels and register the stretching of the vessel walls when blood pressure increases. The spectral range of heart rate variation can therefore be divided into different sections (cf. Figure 6.30):

– Typically, the most prominent oscillatory component of heart rate variability is represented by the *Respiratory Sinus Arrhythmia* (RSA), whereby stimulation by the vagus nerve is suppressed during inhalation and the heart rate increases as a result. On exhalation, it can take effect again, so that the heart rate rises again. This high frequency component (HF: *high frequency*) of the heart rate is thus mainly generated by the respiratory rate in the range of 0.15 to 0.4 Hz.

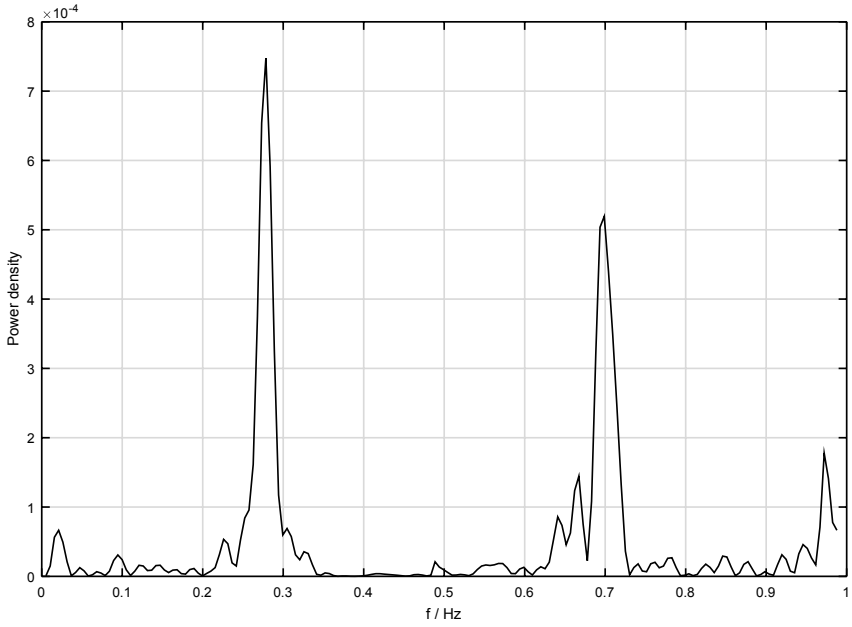


Fig. 6.30: Example of a spectrum of heart rate with a pronounced peak of respiration at approximately 0.28 Hz.

- Another conspicuous component of HRV is the low frequency component (LF: *low frequency*) in the range of 0.04 to 0.15 Hz. This component is almost exclusively caused by the activity of the parasympathetic nervous system.
- The fluctuations below 0.04 Hz have been studied less intensively than at higher frequencies. These frequency ranges are often divided into very low (VLF: *very low frequency*) bands from 0.003 to 0.04 Hz and extremely low (ULF: *ultra low frequency*) bands from 0 to 0.003 Hz. When HRV is measured for a short time, the ULF band is usually also omitted. The low-frequency rhythms are attributed, for example, to hormonal factors, thermoregulatory processes and the renin-angiotensin-system, which also regulates blood pressure.

Determination of heart rate variability is a commonly used method when assessing the functioning of the heart and autonomic regulation. It has been used in numerous studies related to cardiovascular research. One of the main clinical areas in which HRV has been found to be valuable is in the risk assessment of sudden cardiac death following acute myocardial infarction. In addition, decreased HRV is generally considered an early warning sign of diabetic cardiovascular autonomic neuropathy, with the most significant decrease in HRV associated with diabetes found within the first 5-10 years of diabetes.

In addition to these two main clinical applications, HRV has also been studied in relation to various cardiovascular conditions such as renal failure, exercise, occupational and psychosocial stress, gender, age, drugs, alcohol, smoking and sleep. The term HRV analysis generally refers to changes in the heartbeat interval to investigate precursors of disease.

The study of cardiac rhythm variability can be performed in both the time and frequency domains, as well as in the time-frequency domain.

Time domain

Statistical methods are used for the investigation in the time domain, including i) the determination of *statistical parameters*, such as the mean duration of all RR intervals and their standard deviation, ii) the performance of *analysis of variance* , such as the frequency of occurrence of different heart rates in the time domain. e.g. the frequency of occurrence of different heart rates in a histogram, and iii) the use of *correlation methods* such as the changes in heart rate HF_n between successive RR intervals according to Poincaré (see Figure 6.31).

Frequency Domain

When investigating in the frequency domain, for example, long and short time-*Fourier analysis* as well as a *wavelet-transformation* can be applied to the course of the heart rate in the time domain. The wavelet transformation has the advantage that the time interval in which the transformation is to be carried out is adapted to the frequency range, i.e. a large time interval is selected for the analysis of low frequencies and a small time interval for the analysis of high frequencies (cf. section 2.4). For the short-time-Fourier transform and the wavelet-transform, different "windows" can be used to select the time range, which must be examined for optimal analysis behaviour depending on the signal type.

For Analysis in the Time Domain

In addition to the basic parameters such as mean and standard deviation

$$\begin{aligned} E[HF] &= \frac{1}{N}\sum_{n=1}^{N} HF_n\,, \quad \text{mean}\,, \quad N\text{: number of frequency values} \\ \sigma^2[HF] &= \frac{1}{N}\sum_{n=1}^{N} (HF_n - E[HF])^2\,, \quad \text{standard deviation} \end{aligned} \tag{6.26}$$

the frequency of occurrence of different heart rate values is often also presented in the form of a histogram (cf. Figure 6.32), the width of which is additionally a measure of variability.

As an alternative to the histogram, the variability of the heart rate can also be seen very well, as already mentioned, in a Poincaré-diagram, in which the heart rates are

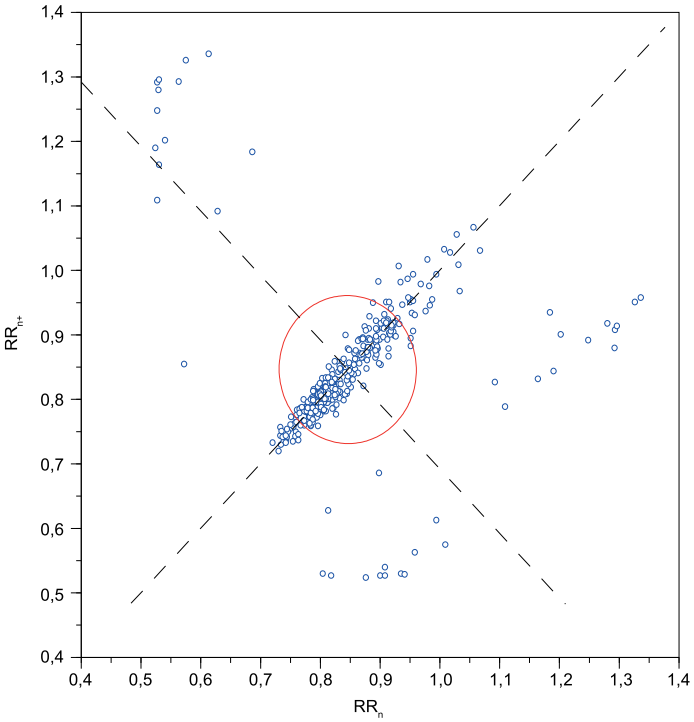


Fig. 6.31: Example of a Poincaré-representation: for this, the time between the n+1-th and the n+2-th R-prong (RR_{n+1}) is plotted as a function of the time between the n-th and the n+1-th R-prong (RR_n) of the QRS complex.

plotted slightly shifted: HF_N on the X-axis, opposite HF_{n+1} on the Y-axis. However, here the "cloud", i.e. the width of the variability, is not necessarily a measure of the patient's health. For some diseases, it can also be the other way round. Figure 6.33 shows examples of this. In (a) the "heart rate cloud" is shown during an attack of atrial fibrillation, and in (b) when the atrial fibrillation is over. It can be seen that the heart rate variation is significantly greater during atrial fibrillation than afterwards. This is because the atrium can no longer be regularly excited by the sinus node, but beats at a very high rate of about 200 to 300 per minute due to additional "interfering excitations" (mostly from the mouths of the pulmonary veins). However, since the atrioventricular node (or AV node for short, cf. section 3.2) acts like a filter in such a case and only transmits every second to third excitation to the main chamber, the heart does not beat as frequently (at a rate of about 100). This excitation, however, is not as regular as with excitation by the sinus node and is very clearly noticeable in the Poincaré-diagram. During atrial fibrillation, a very broad cloud is therefore seen, and in the phase thereafter a much smaller cloud.

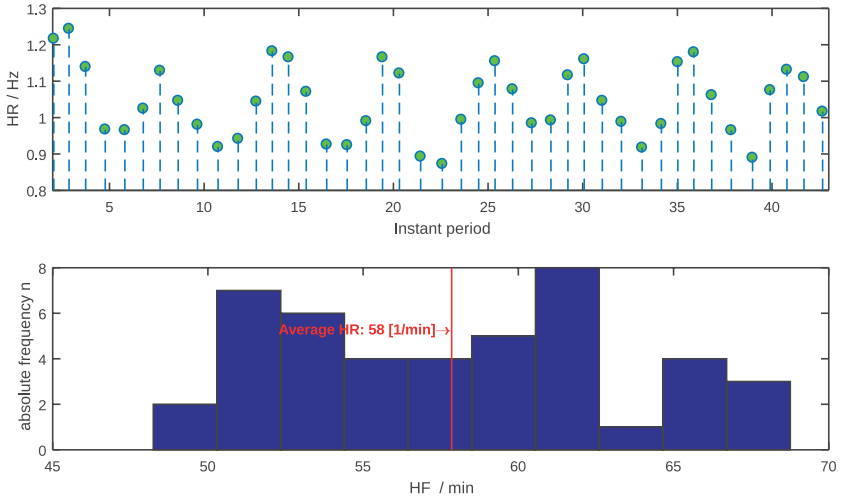


Fig. 6.32: Instantaneous heart rate (top) with associated histogram (bottom).

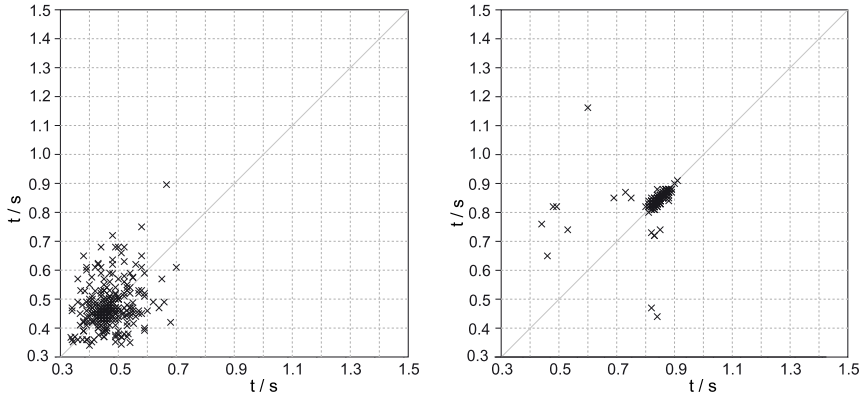


Fig. 6.33: Poincaré-diagram in atrial fibrillation by seizure (left) and after seizure (right) [own measurement of atrial fibrillation by co-author K.-H. Witte].

The correlation methods examine the similarity between successive values of two functions. In the case of the heart rate, it is the similarity between different sections of the same function (autocorrelation). For this, it is usually assumed that this is not time-dependent, which is often not true in practice. The calculation of the autocorrelation for heart rate is done according to

$$R_{HF\,HF}(n, m) = E\left[HF_n \cdot HF_{n+m}\right] = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} HF_n \cdot HF_{n+m}\ . \tag{6.27}$$

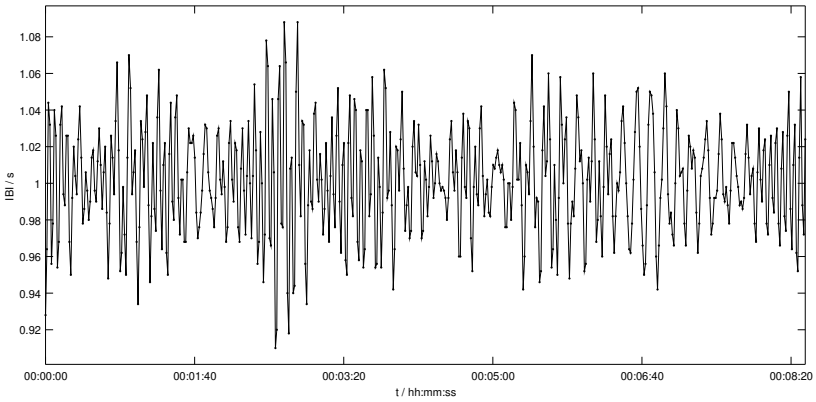


Fig. 6.34: Example of the time intervals between the R-waves of an ECG (IBI: Interbeat intervals) with an average heart rate of 60 beats per minute, generated with the software tool ECGSYN, see [48].

Analysis in the Frequency Domain

When analysing in the frequency domain, it must be noted that the discrete values for the heart rate in Equation 6.25 do not have equal intervals, but act like an irregular sampling of a continuous heart rate course due to the different lengths of the RR intervals, i.e. a Fourier transformation cannot be applied directly, as this presupposes equal time intervals between the samples. In order to create this prerequisite, the heart rate can first be converted into a time-continuous form by interpolation, which is then digitised again by uniform sampling in order to be able to transform it into the frequency domain (cf. Figure 6.34).

The simplest form of interpolation is to interpolate the irregular values of the measured heart rate linearly or with spline functions. However, this increases the ratio of the spectral components of the lower to those in the higher frequency range [7]; because this interpolation can act like a low-pass filter. With linear interpolation, this still does not receive a smooth (i.e. continuously differentiable) course, but many corners, which lead to additional frequency components in the spectrum, which are not present in the measured heart rate. With the interpolation by spline-functions, this problem is reduced. A further reduction can be achieved by dividing the recorded interval into several smaller intervals which overlap and are supplemented with a weighting function (e.g. Hamming window), see [81]. The basis for these methods is the Fourier transformation. The prerequisites for their application or models with recursive feedbacks, in which the coefficients of the model are estimated (Autoregressive (AR) models) presuppose that the signal is stationary and sampled uniformly. However, this is rarely the case in biological systems, so other methods, such as the wavelet transform or spectral estimation using delay and least square error minimisation, can

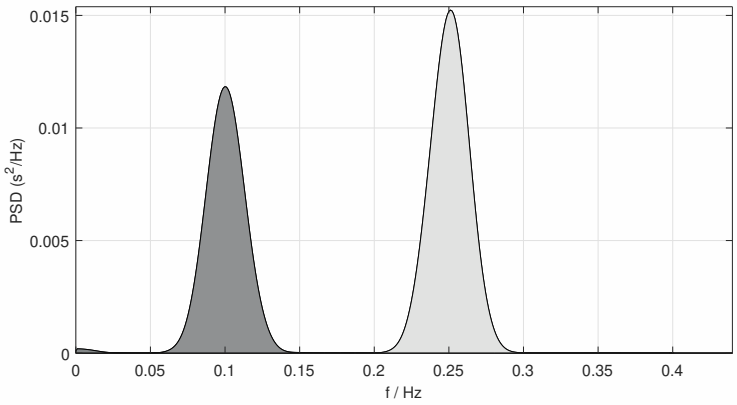


Fig. 6.35: With HRVAS created associated power spectral density according to Welch, see [81] .

give better results. With the wavelet transform the signal need not be stationary and with Lomb's method [44] need not be uniformly sampled.

In the method of Lomb [44] the time intervals between the samples (interbeat interval IBI) are determined by a sine function with delay according to

$$IBI(t_n) + \varepsilon_n = a\ \cos(2\pi f[t_n - \tau]) + b\ \sin(2\pi f[t_n - \tau])\ . \tag{6.28}$$

for all N values of the signal $IBI(t_n),\ n = 1, \ldots, N$ at a given frequency f is approximated by varying the values of a,b and τ such that the error ε is smallest. A previous interpolation and sampling in a uniform manner is not necessary here. As can be seen by comparing the power spectral density according to the Welch and Lomb method, the spectrum according to Welch is more bell-shaped than that according to Lamb, which speaks for a more detailed resolution of the Lomb method. Both spectral power densities were created with the software tool HRVAS by Ramshur [66], which can not only perform the spectral power densities but also analyses in the time and time-frequency composite range. In addition, various preprocessings are possible, such as the suppression of very low-frequency components (detrending) or artefacts caused, for example, by patient movement with electrode displacements. The tool can be downloaded free of charge from the internet at https://github.com/jramshur/HRVAS and can be installed either as an extension of Matlab and Scilab or without them as a standalone package.

The autocorrelation serves for further spectral investigation. It can also be understood as a convolution with its own mirrored course. By Fourier transforming the autocorrelation of the heart rate according to Equation 6.27, its spectral power density can be calculated and compared with the quadratic amount of the Fourier transform in order to better estimate the actual spectrum. In practical recordings, only a section of the actual ECG is available. Ideally, a long-term ECG of up to seven days would be used,

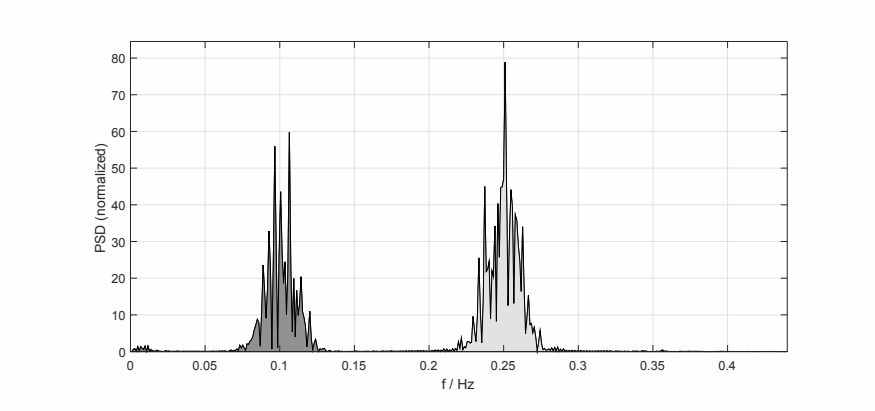


Fig. 6.36: Associated power spectral density according to Lomb created with HRVAS, see [44].

but often only a three-day ECG is obtained. In order to still obtain accurate measurements, the windows can be adjusted (e.g. according to Gauss, Hanning, Han, Kaiser, etc.) so that no major errors occur in the frequency range, e.g. due to overshoots.

The discrete-time short-time-Fourier transformation is defined by the relation

$$X(m,\omega) = \sum_{b=-\infty}^{\infty} x(n) \cdot w(n-m) e^{-j\omega n} \tag{6.29}$$

Here $x(n)$ is the sampled signal (here the ECG), and $X(m,\omega)$ is the sliding window defined by the window function $w(n-m)$ whose position is described by m. This spectrum is *not* discrete, as in the case of the discrete-time Fourier transform. Only when the input signal is periodic does the discrete-time Fourier transform change to the discrete one. Since the heart beats rhythmically, the periodicity is virtually assumed in the recording. However, since the heart does not beat purely periodically, but varies, errors occur in the determination of the spectrum.

In order to improve the spectral resolution, wavelet transformations can be used, which have a shorter window at higher frequencies, i.e. at lower frequencies one has a poor spatial resolution with good frequency resolution, while in the higher frequency range it is exactly the opposite (cf. section 2.4).

6.3.2 Phonocardiogram

In section 4.3 the technique of registration of heart tones was presented. Figure 6.37 (top) shows a phonocardiogram of 20 s duration. In it, the two heart sounds are rhythmically repeated.

The first heart sound is produced by the contraction of the ventricles and is also called the force sound. Under physical stress, the amplitude of the first heart sound increases, because the increased oxygen demand of the organism causes a greater

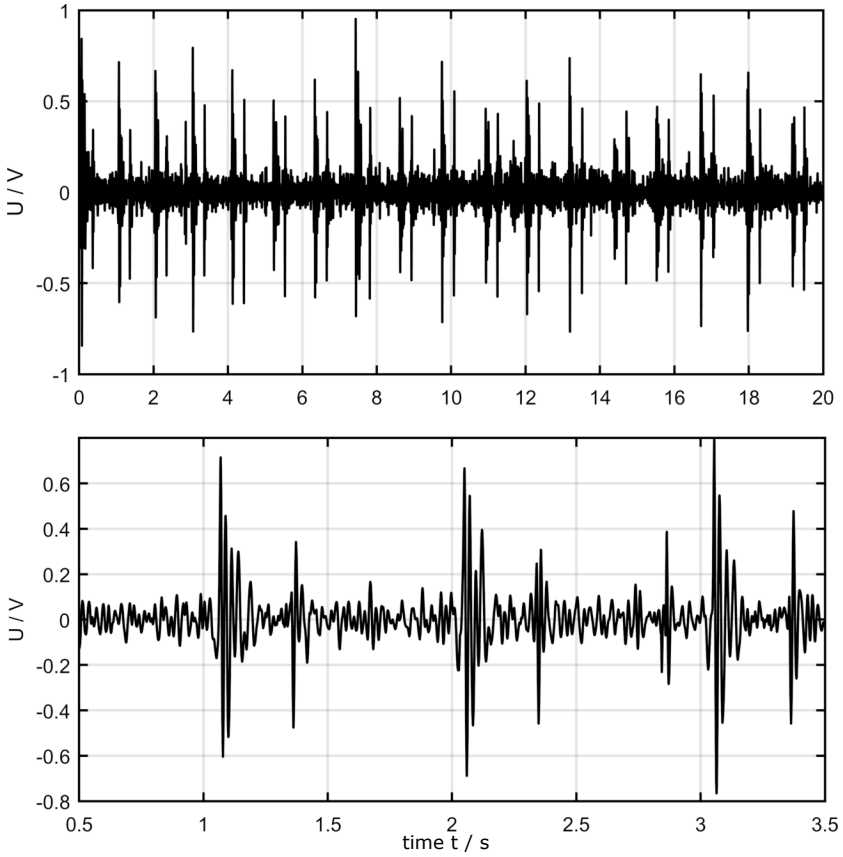


Fig. 6.37: Phonocardiogram on a healthy heart with 20 s duration (top): First and second heartbeats alternate in a regular sequence, with the interval between first and second heartbeats being shorter than between second and first. Section of phonocardiogram showing three heartbeats (bottom).

pumping capacity of the heart, which in a healthy heart is accompanied by an increase in the strength and frequency of contractions. The second heart sound is produced by the synchronous closing of the aortic and pulmonary valves and marks the end of blood ejection. The amplitude of the two heart sounds varies with the respective measurement position, which is indicated by the number of the intercostal space and the lateral distance to the sternum (breastbone) according to convention. Thus, measurement position 2R2 denotes a point in the second intercostal space, right side, 2 cm from the sternum. At this position, the aortic valve can be easily heard. 4L4 means accordingly: fourth intercostal space, left side, 4 cm from the sternum. This corresponds to a point centrally located over the heart. Other common measurement positions are 2L2 and apex (apex of the heart, in the region of the left costal arch). Both the time course and the spectral composition of the heart sounds contain dia-

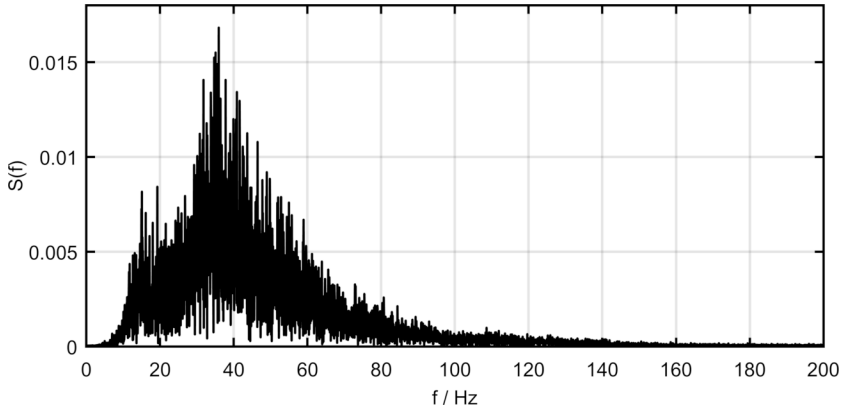


Fig. 6.38: Spectrum of the phonocardiogram from Figure 6.37.

gnostic information. Examples of heart sounds conspicuous in the time domain are the systolic or a third heart sound in the diastolic phase. A systolic murmur is a noise between the first and second heart sounds and may be a indication of aortic or pulmonary valve regurgitation. A third heart sound may be associated with ventricular dilatation due to chronic heart failure. A complete review of heart murmurs and their origin can be found in [31]. The spectral composition of heart murmurs (Figure 6.38) provides additional information. For example, if the systolic murmur has strong high-frequency components above 400 Hz, this can be attributed to increased transvalvular flow velocity due to a heart valve stenosis.

As mentioned in the section 4.3, phonocardiography has been largely displaced in the cardiological functional diagnostics superseded by ultrasound imaging sonography. However, the particular simplicity of use in conjunction with modern methods of remote data transmission opens up new fields of application for phonocardiography. For example, heart patients can use an electronic stethoscope to record heart sounds themselves and send them to experts for further evaluation. This enables close monitoring in the home environment, which is particularly advantageous for monitoring disease progression.

Furthermore, a method for the automatic determination of the left ventricular ejection time (LVET) will be presented. LVET is an important cardiological parameter and is used as such or normalized to the heart rate to assess cardiac output. At normal cardiac pumping rate and a heart rate of approximately 70 beats per minute, LVET is 260 to 320 ms. If the contractility of the heart muscle and thus the pumping capacity is reduced due to disease, the organism reacts by prolonging the pumping process, which increases the LVET. In patients with chronic heart failure and poor pumping capacity (ejection fraction < 35 %), LVET values up to more than 450 ms can be found. With massively impaired pump function, LVET may then shorten again because the pumping process stops prematurely. Thus, LVET comes into consideration for the

follow-up of chronic heart failure. The relationship of this parameter to the phonocardiogram becomes clear when the causes of the two heart sounds are recalled. The first heart sound is produced by the rapid contraction movement of the ventricles, which is associated with the onset of the expulsion phase. At the end of the expulsion phase, the aortic and pulmonary valves close, which in turn causes the second heart sound. Thus, LVET corresponds quite closely to the time between the first and second heart sounds.

When designing a suitable algorithm for the determination of LVET, it must be taken into account that phonocardiograms are in practice overlaid by artifact noise and that the amplitude of heart sounds in cardiac patients is often weak. A simple algorithm for determining maximum values and their location would quickly lead to incorrect results. Instead, an algorithm based on the autocorrelation function (AKF) (cf. subsubsection 5.3.1.2) is presented here. The AKF offers the advantage of highlighting signal components that occur systematically. First, the phonocardiogram is rectified and subjected to low-pass filtering using a moving-average filter. The Figure 6.39 shows in each case the processing result for the sequence from Figure 6.37 (top).

Like the Fourier or Laplace transform, the AKF belongs to the class of integral transforms. In the continuous-time domain the transformation rule is

$$\rho_{xx}(\tau) = \int x(t)x(t+\tau)\mathrm{d}t \,, \tag{6.30}$$

where $x(t)$ is the function to be transformed, here the function from Figure 6.39 (below), and τ is a shift parameter in the function $x(t)$, which becomes the variable of the AKF. Equation 6.30 can be translated as the function $x(t)$ is shifted against itself by τ, the shifted function and the original function are multiplied together, and then the area of the multiplication result is calculated by integration. This is repeated for all shift values τ, resulting in the AKF $\rho(\tau)$. In the discrete-time domain, the integral becomes a sum with the discrete-time variable m and the displacement value n:

$$\rho_{xx}(n) = \sum(x(m)x(m+n)) \,. \tag{6.31}$$

Applying the AKF to the processed phonocardiogram from Figure 6.39 (below), the AKF value will be becomes large whenever the function $x(t)$ or $x(m)$ is shifted exactly one heart cycle. Then the first heart sound falls back to the first and the second heart sound to the second. Product and integral in Equation 6.30 take large values in that case. Smaller maxima occur when the first heart sound is switched to the second or the second is shifted onto the first. The first of these two cases is particularly interesting, because the value for the shift of the first to the second heart sound just corresponds to the LVET we are looking for. Figure 6.40 (above) shows the result of AKF formation from the signal in Figure 6.39 (bottom) and a magnification of the AKF around the maximum value (bottom). The LVET we are looking for corresponds to the distance in τ of a main maximum to the next right side maximum. The LVET, i.e. the distance between the main maximum and the first right minor maximum, can be calculated using the algorithm given in Listing 6.3.2.

Listing 6.3.2: Matlab example for the analysis of a phonocardiogram in relation to the LVET.

```
%% reads the phonocardiogram x(t) from wav file,
%% it is rectified x1 = abs(x),
%% it is gated with a moving-average filter,
%% from which the auto-correlation is calculated and displayed.
%% Subsequently, the LVET is calculated and output.

% Read signal and plot
[x,Fs] = audioread('pkg.wav');
Ns = 1:length(x);                    % signal length
ts = Ns/Fs;                          % sample time
x = x(1:160000);                     % selection of first 20 sec.
ts = ts(1:160000);
```

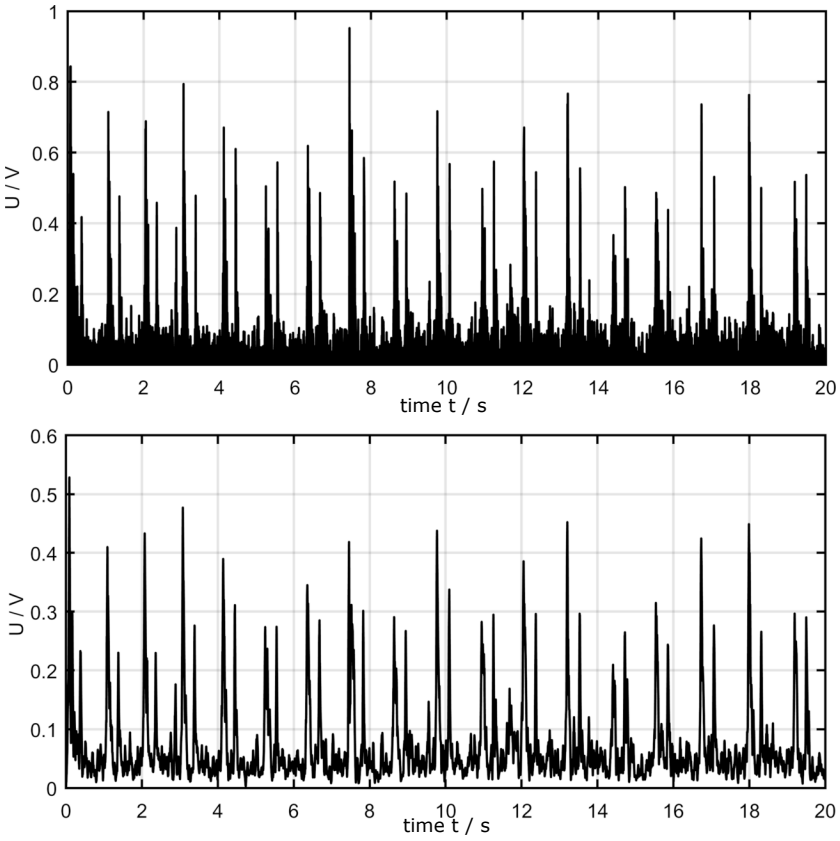


Fig. 6.39: Phonocardiogram from Figure 6.37 after rectification (top): Phonocardiogram after rectification and low-pass filtering using moving-average filter (window width: 200 samples) (bottom).

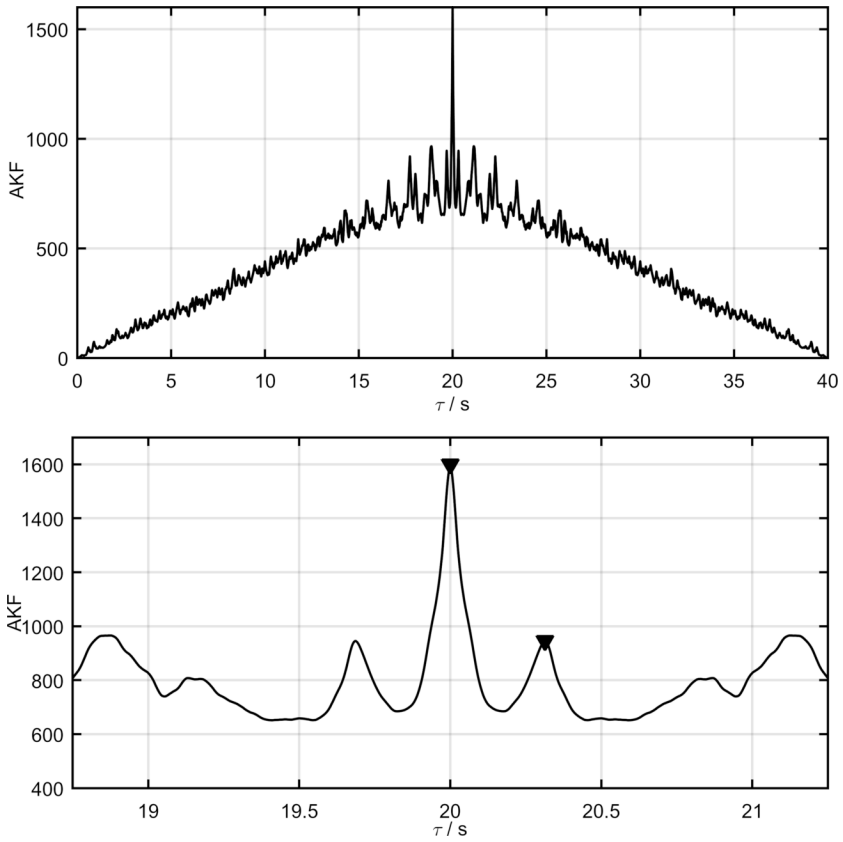


Fig. 6.40: Top: Result of the autocorrelation of the signal from Figure 6.39 (bottom). The main maxima correspond to a shift of one heart period, the first minor maximum to the right of a major maximum corresponds to a shift by LVET. On the abscissa the displacement is given in samples. The sample frequency here is 8 kHz. Bottom: Magnification of the AKF around the main maximum. The distance of the main maximum to the next right side maximum corresponds to the searched LVET.

```
% Plot of raw signal
figure1 = figure;
plot(ts, x);
xlabel('t / s');
ylabel('s(t)');
title('Phonokardiogram');

% Plot of absolute values
x1 = abs(x);
figure2 = figure;
```

```
plot(ts, x1);
xlabel('t / s');
ylabel('s(t)');
title('Rectified Phonokardiogram');

% Moving Average Filter
windowSize = 200;           % window sizeFensterbreite
b = (1/windowSize)*ones(1,windowSize);  % FIR ilter coeffizient
a = 1;                                  % IIR ilter coeffizient
x_av = filter(b,a,x1);                  % apply filter

% Plot of absolute values after filtering
figure3 = figure;
plot(ts, x_av);
xlabel('t / s');
ylabel('s(t)');
title('Rectified Phonokardiogram after Filtering');

% Autocorrelation function AKF
x_c = xcorr(x_av,x_av);
tau = 1:length(x_c);
tau_1 = tau/8000;

% Plot of autocorrelation function
figure4 = figure;
plot(tau_1, x_c);
xlabel('\tau / s');
ylabel('AKF');
title('Autokorrelation');

% Calculation of peaks in the autocorrelation function
[s_pk, tau_pk] = findpeaks(x_c,tau_1);
[s_p1, pos] = max(s_pk)
tau_p1 = tau_pk(pos)
s_p2 = s_pk(pos+1)
tau_p2 = tau_pk(pos+1)

% Plot of peaks
figure5 = figure;
plot(tau_1, x_c);
hold on;
plot(tau_p1,s_p1,'rv','MarkerFaceColor','r');
```

```
plot(tau_p2,s_p2,'rv','MarkerFaceColor','r');
axis([150/8, 170/8, 400, 1700])
xlabel('\tau / s');
ylabel('AKF');
title('Autocorrelation');
hold off

%  Calculation of LVET
L_T = (tau_p2-tau_p1)/8
figure(6); hold on;
% Plot of LVET result
strb = ['The medium LVEST is: ', num2str(L_T), ' ms.']
annotation('textbox',[0.1 0.4 0.8 0.4],'String',strb,'FontSize',11,...
    'LineStyle','none','BackgroundColor', [1 1 1]);
set (gca, 'visible', 'off')
```

For the presented phonocardiogram, this results in an LVET of 314 ms.

6.3.2.1 Phonocardiogram of Mechanical Prosthetic Heart Valves

A completely different form of phonocardiography application is the control of mechanical prosthetic heart valves. These usually consist of a ring-shaped housing in which two valve leaflets are suspended. The valve leaflets can be opened in one direction so that they are perpendicular to the blood flow. In the other flow direction, the flap sails close with the housing and thus block the backflow. In this way, the desired valve function is achieved. A textile ring is attached to the outside of the housing, with which the prosthesis is sewn into the heart. The material of the housing and valve leaflet is usually pyrolytically deposited carbon, which has a very smooth surface, to prevent the adhesion of thrombi. Mechanical prosthetic heart valves are the treatment option along with biological prostheses, when a native heart valve needs to be replaced due to a defect. Despite the very smooth surfaces of mechanical prosthetic heart valves, there is an increased risk of thrombus adhesion, which is why anticoagulants must be used to reduce blood clotting in patients with mechanical heart valve replacement. If anticoagulation therapy is inadequate, thrombi can adhere to the valve body and leaflets and impair valve function (see Figure 6.41) – sometimes with fatal consequences. Mechanical prosthetic heart valves generate high-frequency sounds when the valve leaflets open and close, which are closely related to the mechanical function of the prosthesis. Figure 6.42 shows the sound pattern of a two-leaflet prosthesis in the aortic position over one cardiac cycle. An initial fainter murmur occurs when the valve leaflets open. This begins the ventricular expulsion phase. Once the ventricle is emptied at the end of the expulsion phase, reflux begins, causing the valve leaflets to close. This is accompanied by a hard impact of the valve leaflets, which produces a loud im-

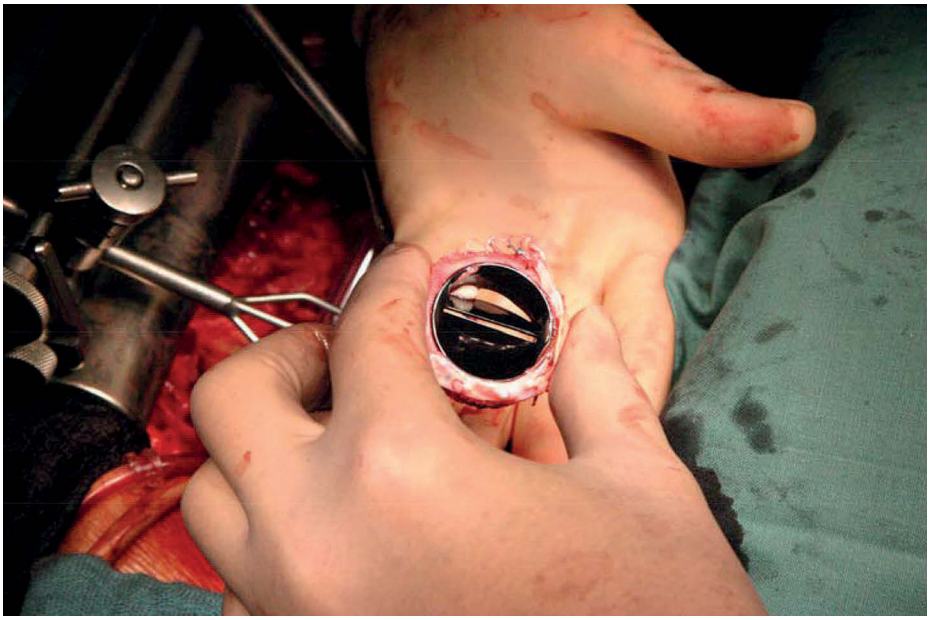


Fig. 6.41: Explanted mechanical prosthetic heart valve with white thrombus between the housing ring and valve leaflet.

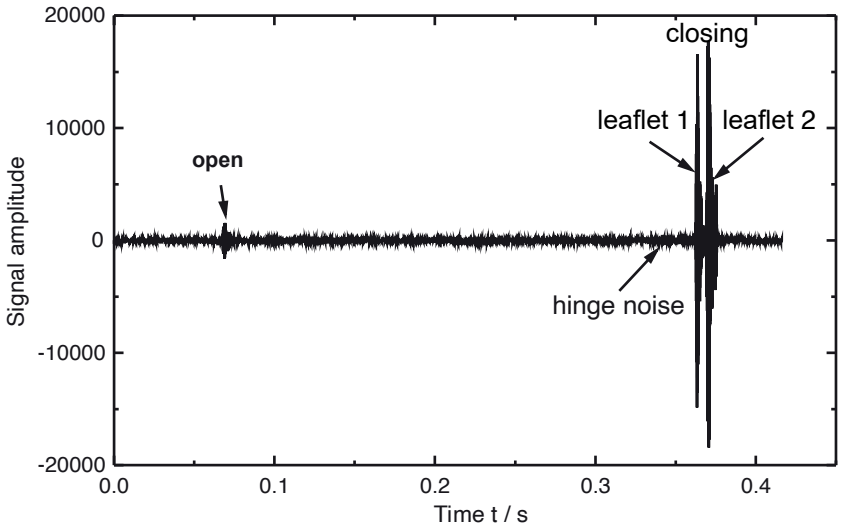


Fig. 6.42: Sounds of a mechanical heart valve prosthesis (two-leaflet prosthesis) over one cardiac cycle after high-pass filtering with 8 kHz.

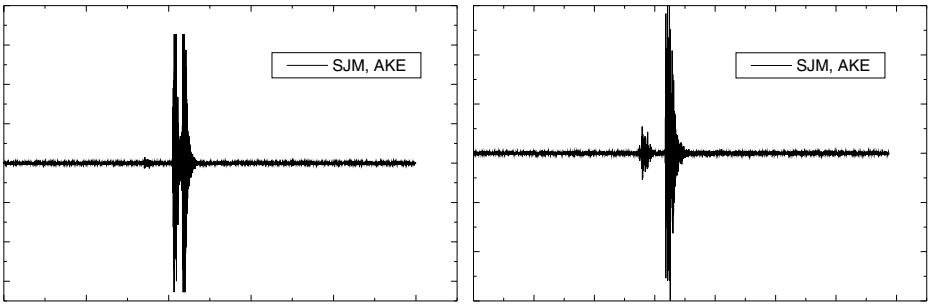


Fig. 6.43: Closing sound of a normal two-leaflet prosthesis (left) and with thrombus adhesion (right): The impact noise of one of the two leaflets is clearly attenuated by the thrombus in the right image.

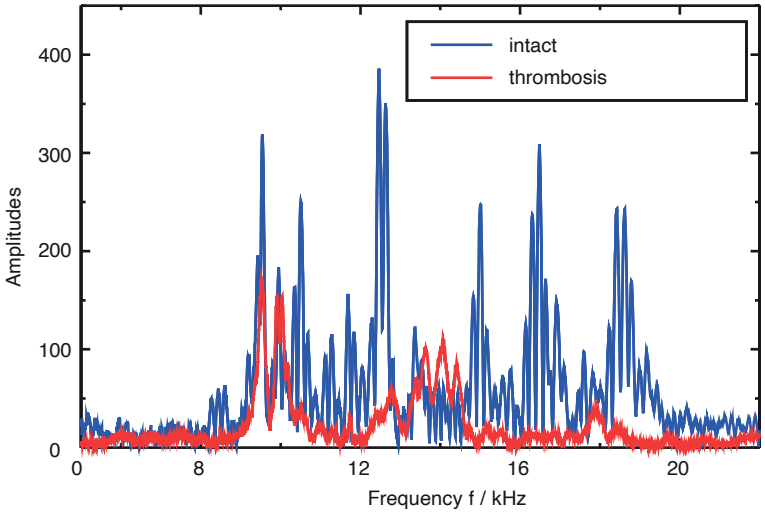


Fig. 6.44: Spectrum of a normal impact sound of a mechanical heart valve (blue) and in the presence of thrombus (red).

pact sound. Since the two valve leaflets are never exactly perpendicular in the blood stream, the two valve leaflets close with a short time lag of up to 12 ms, which is why two impact noises can be found in the sound image. Shortly before the impact, there is a weak hinge noise, which is caused by the rotation of the flap leaflets in their housing suspension.

If a thrombus obstructs the hard impact of a valve leaflet, as in Figure 6.41, the amplitude of the closing sound is strongly damped (Figure 6.43). A thrombus also alters the spectrum of the valve closure sound (cf. Figure 6.44). In particular, the high frequency components above about 11 kHz largely disappear. In the spectrum, the essential difference to conventional phonocardiography becomes clear, because there the bandwidth reaches only up to a maximum of 1.2 kHz. Noise analysis in the time

and frequency domain is a very sensitive method for assessing the integrity of mechanical Prosthetic heart valves [17, 91]. Differentiation of the respective impact sounds of both valve leaflets requires a time resolution of less than 1 ms, implying a frequency resolution of 1 kHz. This relatively poor frequency resolution is, however, still sufficient to be able to detect conspicuous changes in the spectrum.

6.3.3 Determination of Oxygen Saturation and Photoplethysmography

As already announced in section 4.3, the oxygen saturation of the blood can be determined by measuring the light absorption at two different wavelengths. In this section, the relationship between the measured quantity and the oxygen saturation will be derived mathematically. The measured quantity is the light intensity I, which is measured at the transmission on the opposite side of the irradiation point by the photodiode (cf. Figure 6.45). In addition to absorption in the tissue, the light is also attenuated by scattering from internal finger structures. The scattering effect is not taken into account in the SpO_2 calculation.

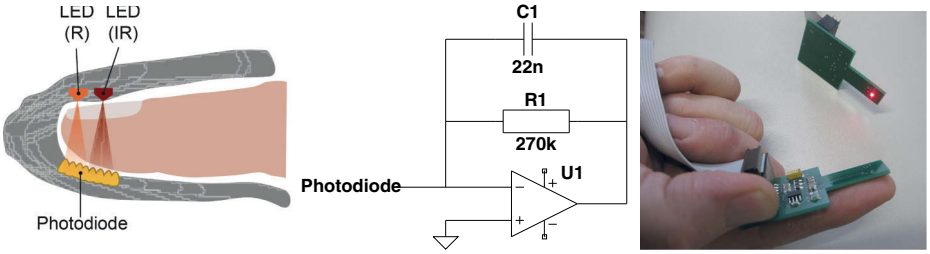


Fig. 6.45: Arrangement of LED and photodiode for transmission measurement at the fingertip (left): A transimpedance amplifier with low-pass filter is used to amplify the signals of the photodiode. (see electronic circuit diagram in the middle). A measurement of the pulse curve can also be performed with a reflection sensor on any part of the body (here on the finger) (right).

For a simple homogeneous layer, the intensity according to the Lambert-Beer law depends exponentially on the substance-specific and wavelength-dependent absorption coefficient α and the layer thickness d:

$$I = I_0 e^{-\alpha d}\,, \tag{6.32}$$

where I_0 is the intensity at the irradiation point. If the light path is composed of several layers i with different absorption coefficients α_i and layer thicknesses d_i, then the intensity after the passage of all partial paths is

$$I = I_0 e^{-\sum \alpha_i d_i}\,. \tag{6.33}$$

If one wanted to calculate the intensity for the fluoroscopy of the fingertip or the earlobe, according to Equation 6.33 the layer structure of the tissue with the corresponding absorption coefficients would have to be known exactly. This is practically impossible. Since, as will be shown below, the temporal change in intensity is used to calculate oxygen saturation, it is sufficient to use a simplified two-layer model. The first layer represents all absorption processes in venous vessels, fat and connective tissue, skin, etc., which do not depend on time. The time dependence of intensity is due to the change of thickness of arterial vessels as a result of rhythmic pulse waves. This is represented by the second layer in terms of a time-varying thickness $d_2(t)$. With this model, Equation 6.33 simplifies to.

$$I(t) = I_0 e^{-(\alpha_1 d_1 + \alpha_2 d_2(t))} = I_0 e^{-\alpha_1 d_1}\, e^{-\alpha_2 d_2(t)} = C(\lambda)\, e^{-\alpha_2 d(t)}\,. \tag{6.34}$$

In Equation 6.34 all time-independent factors are combined to the constant C, which, however, depends on the wavelength because of α_1. α_2 is the absorption coefficient of blood. The quantity $d(t)$ describes the time-dependent vessel thickness, but it is not directly accessible and therefore must be eliminated. For this purpose, the auxiliary quantity Γ is introduced as follows:

$$\Gamma = \frac{\ln \frac{I(\lambda_1 t_1)}{I(\lambda_1 t_2)}}{\ln \frac{I(\lambda_2 t_1)}{I(\lambda_2 t_2)}}\,. \tag{6.35}$$

For the auxiliary quantity Γ, therefore, a total of four measurements are required, two measurements each at the wavelengths λ_1 and λ_2 at two different times t_1 and t_2. Substituting Equation 6.34 into Equation 6.35 leaves,

$$\Gamma = \frac{\alpha_2(\lambda_1)}{\alpha_2(\lambda_2)} \tag{6.36}$$

so $d(t)$ is dropped.

To determine oxygen saturation, a correlation must now be established with the concentration of oxygen-saturated hemoglobin ($C_{\mathrm{HbO_2}}$) and unsaturated hemoglobin (C_{Hb}). For this purpose, the blood absorption coefficient is considered in more detail. In general, for media consisting of several substances, the absorption coefficient is the sum of the molar extinction coefficients ε of the individual substances multiplied by the corresponding concentration. Transferred to blood this means

$$\alpha_2(\lambda) = \varepsilon_{\mathrm{HbO_2}}(\lambda) C_{\mathrm{HbO_2}} + \varepsilon_{\mathrm{Hb}}(\lambda) C_{\mathrm{Hb}}\,. \tag{6.37}$$

Of course, blood does not consist only of hemoglobin, so in Equation 6.37 also the other components would have to be included accordingly in the sum. However, these are insignificant for further consideration and can be mentally added to the quantity $C(\lambda)$. The extinction coefficients $\varepsilon_{\mathrm{HbO_2}}(\lambda)$ and $\varepsilon_{\mathrm{Hb}}(\lambda)$ are well known from the literature [39, 40, 79].

In the next step, for Equation 6.37, the sought oxygen saturation

$$\mathrm{SpO_2} = \frac{C_{\mathrm{HbO_2}}}{C_{\mathrm{HbO_2}} + C_{\mathrm{Hb}}} = \frac{C_{\mathrm{HbO_2}}}{C_{\mathrm{ges}}} \tag{6.38}$$

is introduced. If Equation 6.38 is converted to $C_{\mathrm{HbO_2}}$ or C_{Hb} and inserted into Equation 6.37, we obtain

$$\alpha_2(\lambda) = C_{\mathrm{ges}} \left[(\varepsilon_{\mathrm{HbO_2}}(\lambda) - \varepsilon_{\mathrm{Hb}}(\lambda)) \, \mathrm{SpO_2} + \varepsilon_{\mathrm{Hb}}(\lambda) \right] . \tag{6.39}$$

Substituting Equation 6.39 into Equation 6.36, the unknown quantity C_{ges} truncates out, leaving an equation with the sought oxygen saturation $\mathrm{SpO_2}$, the four measured values $I(\lambda_{1,2}, t_{1,2})$ contained in Γ and the known molar extinction coefficients $\varepsilon_{\mathrm{HbO_2}}(\lambda)$ and $\varepsilon_{\mathrm{Hb}}(\lambda)$:

$$\Gamma = \frac{(\varepsilon_{\mathrm{HbO_2}}(\lambda_1) - \varepsilon_{\mathrm{Hb}}(\lambda_1)) \, \mathrm{SpO_2} + \varepsilon_{\mathrm{Hb}}(\lambda_1)}{(\varepsilon_{\mathrm{HbO_2}}(\lambda_2) - \varepsilon_{\mathrm{Hb}}(\lambda_2)) \, \mathrm{SpO_2} + \varepsilon_{\mathrm{Hb}}(\lambda_2)} . \tag{6.40}$$

Converted to $\mathrm{SpO_2}$ finally results in

$$\mathrm{SpO_2} = \frac{\varepsilon_{\mathrm{Hb}}(\lambda_1) - \Gamma \varepsilon_{\mathrm{Hb}}(\lambda_2)}{\varepsilon_{\mathrm{Hb}}(\lambda_1) - \varepsilon_{\mathrm{HbO_2}}(\lambda_1) + \Gamma(\varepsilon_{\mathrm{HbO_2}}(\lambda_2) - \varepsilon_{\mathrm{Hb}}(\lambda_2))} . \tag{6.41}$$

From Equation 6.41 and Equation 6.35 it is immediately clear why the spectral sensitivity of the photodiode is not included in the $\mathrm{SpO_2}$ determination. Only intensity ratios measured at a specific wavelength appear in Γ. The spectral sensitivity, which enters as a factor in the intensity, thus truncates in Γ. However, there are other causes that affect the accuracy of the $\mathrm{SpO_2}$ determination. This becomes clear by looking at the individual quantities in Equation 6.41. On the one hand, the molar extinction coefficients of hemoglobin and oxyhemoglobin are subject to slight individual variations [79]. However, a larger error may be due to the inaccuracy in the LED wavelength that can occur between different LEDs of the same type or within an LED due to temperature variations. For commercially available LEDs the variation is in the range of 1 %, relative to the centroid of the wavelength distribution. The molar extinction coefficients, which enter into Equation 6.41, depend on the wavelength (cf. Figure 6.46). Therefore, the uncertainty in the wavelength implies an uncertainty in the molar extinction coefficients and, according to the rules of the error propagation, also an inaccuracy in the $\mathrm{SpO_2}$ value. In particular, below 800 nm, the absorption spectrum of Hemoglobin exhibits a high gradient. In this range, the wavelength uncertainty results in a large uncertainty of the molar extinction coefficients. Therefore, in this wavelength range, special care should be taken to use LEDs with high wavelength stability. Another source of error is external stray light that hits the photodiode in addition to the LED light and increases the measured intensity. High-quality sensors therefore have housings that shade the measuring range well from external light.

At this point, the limitations of spectrometric $\mathrm{SpO_2}$ measurement with respect to clinical use should be noted. The output value is expressed as a percentage and is a measure of the saturation of the hemoglobin present with oxygen. This says nothing

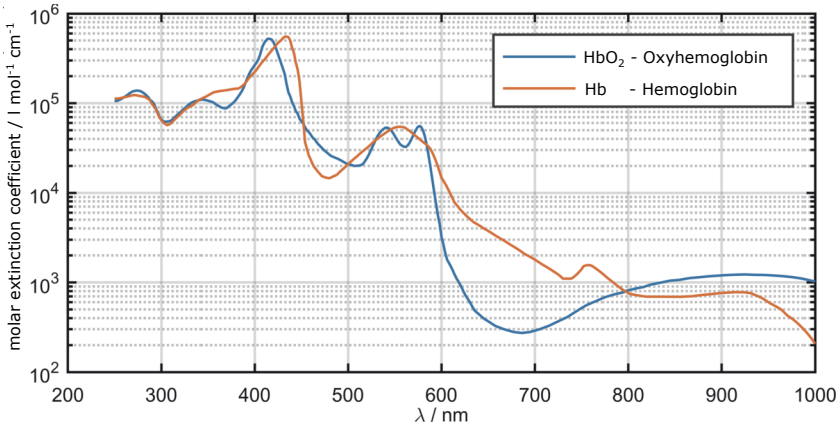


Fig. 6.46: Molar extinction coefficient of hemoglobin and oxyhemoglobin as a function of wavelength [40].

about the absolute oxygen supply of the organism. In the case of blood anemia, the hemoglobin concentration is reduced with the consequence of an oxygen deficiency. However, the SpO_2 value could be inconspicuous at the same time because the decreased Hb fraction is well oxygenated. Furthermore, there is a risk of misinterpretation in carbon monoxide poisoning. Then CO binds to Hb instead of O_2. Since the molar extinction coefficient ε_{HbCO} differs from ε_{Hb} and is similar at some wavelengths ε_{HbO_2}, a statement about the remaining oxygen saturation with this method is no longer possible in the situation.

6.3.4 Signal Classification of Multichannel Photoplethysmography

Author of Subsection: Urs Hackstein

In the previous subsection 6.3.3 photoplethysmography (PPG) signals were used to quantifiy the oxygen saturation of blood. Due to their simple acquisition, PPG signals are used in an increasing number of areas for diagnostic purposes, e.g. for the determination of vascular age via the arterial stiffness or the blood pressure by the pulse transit time. Furthermore, PPG signals are used for the diagnosis of diseases [43], as an example the case of aneurysms and stenoses detection using machine learning techniques and classification is considered.

Cardiovascular diseases are the leading cause of death worldwide. Particularly aortic aneurysms, i.e. abnormal aortic bulges, are problematic and underdiagnosed: According to [10], they affect 12-14% of the population. An aortic aneurysm can be dangerous if it is not spotted early on. It can get bigger over time and could burst, causing life-threatening bleeding. However, the asymptomatic course of the diseases in most patients and the misinterpretation of the initial symptoms in about one third of the

symptomatic patients lead to a categorical underdiagnosed clinical situation. Existing diagnostic techniques like transesophageal echocardiography, doppler sonography or computed tomography either lack sensitivity or require specialised expertise, are invasive and costly which hampers effective screening. Therefore it is of great interest to establish a new diagnostic method that is suitable for effective screening on the family physician level.

One idea for such a new diagnostic method is based on the analysis of biosignals. As described in previous sections, the heart initiates pulse waves of blood pressure and flow that propagate from the heart through the aorta and the other arteries to the periphery of the body. It is observed that the shape and properties of the wave change if it meets irregularities in the structure of the arterial system such as aneuryms, stenoses (constrictions) or bifurcations (branching). Thus the forms of the pressure and flow pulse wave depend on the structure of the cardiovascular system and can be modelled if the structure is known.

The attempt for an early diagnostic method of aortic aneurysms is to solve the inverse problem. Let us assume that the pulse pressure or flow waves are given at some peripheral measuring locations. Is it possible to deduce information on the structure of the cardiovascular system from these measurements? This question is called the inverse hemodynamic problem. The general version has infinitely many solutions [63], but the solution to the constraint case, that we will consider here, remains unclear. All methods and results reported here are originally published in [24] and [25].

In [24], the authors simulated pressure waveforms at arteria brachialis and arteria femoralis respectively and could show that one can deduce information on the structure of the cardiovascular system from these two simultaneous measurements. Therefore, coefficients of the pressure-pressure transfer functions between these two measuring locations were evaluated. The use of transfer function parameters bears the hope to eliminate the strong influence of different boundary conditions between patients (heart, peripheral resistance due to auto-regulation) and isolate the arterial channel properties.

These results based on in-silico data and the functionality of the proposed diagnostic method was checked in [25] on in-vivo data from a clinical study performed at the University Hospital Tübingen, Germany. This study consists of 55 patients thereof 28 with aneurysms at different locations (thoracial, abdominal and thoracic-abdominal aorta) and 27 vascular healthy patients. The results were compared with the diagnoses of the gold standard computed tomography angiography (CTA). As it is much easier to collect PPG-signals than pressure signals, PPG-signals at six different locations were measured simultaneously.

The 55 patients were divided into two study groups based on the available CTA data sets, with each group containing 27 (Group a)) and 28 (Group b)) respectively patients:

(a) Control group (CG): Vascularly healthy subjects. No stenoses in the aorta or the pelvic arteries, diameter of the aorta in the entire thoracic and abdominal course $< 40mm$ (except aortic sinus).
(b) Aneurysm (A): detection of a thoracic or abdominal aortic aneurysm with a diameter of $\geq 45mm$.

Six photoplethysmography probes of the measuring device were attached to the patient who is in supine position, two on the temples, two on the thumbs, and two on the toes (see Figure 6.47).

The position of the sensors was optimized using the single signal of the respective channel. ECG electrodes were placed. Measurement was started as soon as clear signals were visible in all channels. The raw signals were recorded with a sampling rate of 2048 Hz. The patient was measured in quiescent condition (no physical strain 15 min before the measurement, no meal or smoking 1h before the measurement) and the room temperature was fixed to 23 °C to ensure that there is normal blood flow in the peripheral parts of the body. The room was darkened to avoid perturbation by other light sources.

A cut-out of the raw signals are shown in Figure 6.47, the raw signals undergo the following pre-procssing:

1. the spikes at the start and the end of the measurement due to the removal of the sensors have been removed by cutting the signal between 5 s and 55 s, leaving us with 50 s of data
2. the baseline was corrected by subtracting a signal filtered by a sharp lowpass filter at 0.5 Hz
3. the signal was filtered with a sharp filter at 30 Hz (digital filter that does not shift phase). This was done due to the removal of low and high frequency noise.
4. the signal was scaled such that the amplitude of the first harmonic is 1 because the absolute amplitude is assumed not to be relevant due to large differences in positioning the sensor and skin thickness etc.

The resulting band limited signal of frequency range 0.5 Hz - 30 Hz and a signal length of 50 s serves as basis for further computation. Based on a these pre-processed PPG signals from different locations, the above diagnostic problem of aneurysms is described as a classification problem. For all feature extraction methods described below, we split the signal in five 10 s intervals and perform an average on the resulting features. This is beneficial if e.g. a single extrasystole would be present, which could throw off the transfer function computation. Subsequently, the characteristic Fourier-coefficients were calculated from these pre-processed signals and specific harmonics of the transfer-functions between two measuring locations were used as features for a subsequent classification of the aneurysm (A) and control group (CG).

These classification problems are a special case of machine learning which is in turn a part of artificial intelligence. With the help of machine learning, IT systems are

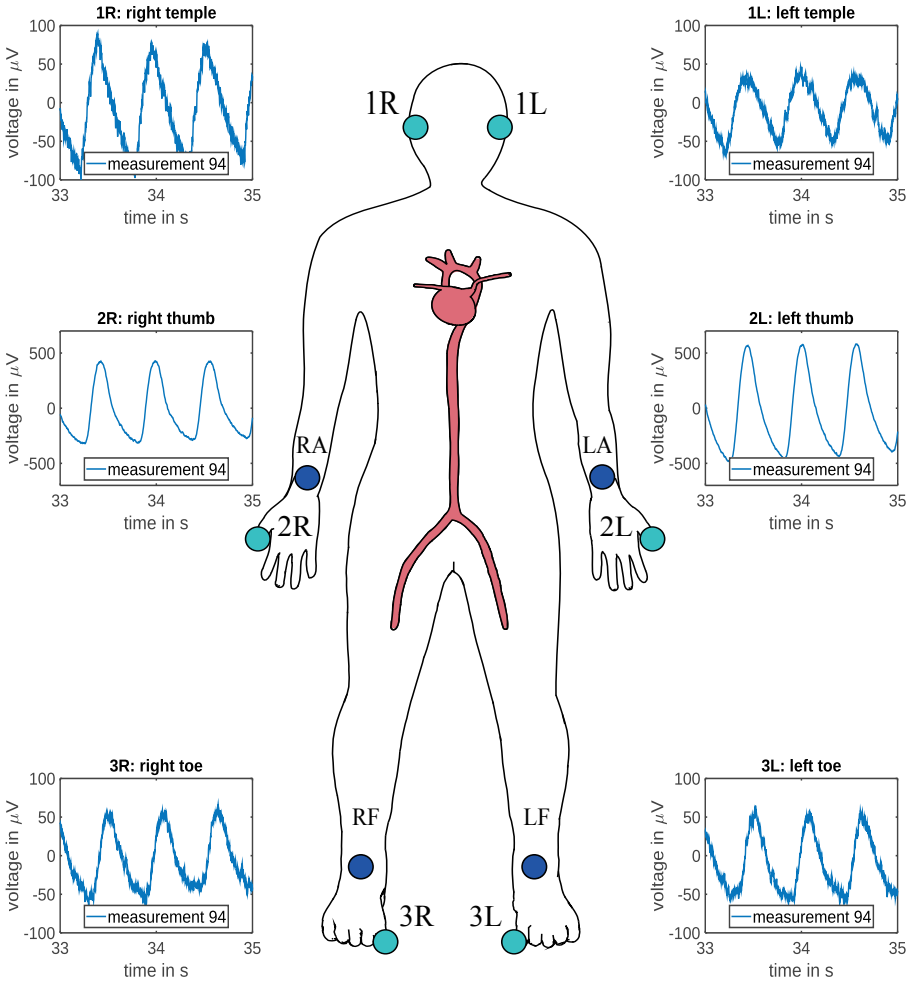


Fig. 6.47: Measuring locations, the six photoplethysmographic probes (light blue) were measured on left and right temples (1L and 1R), left and right thumbs (2L and 2R) and left and right toes (3L and 3R). The ECG-electrodes (dark blue) were placed on left arm (LA), right arm (RA), left foot (LF) and right foot (RF). Quoted from [25, Figure 1].

enabled to recognise patterns and regularities on the basis of existing data and algorithms and to provide intelligent solutions. Artificial knowledge is generated from experience, so to speak. Classification is a specific process of categorising a given set of data into classes, i.e. into sets of objects with the same or similar properties. It can be performed on both structured or unstructured data. The process starts with predicting the class of given data set. The classes are often referred to as target, label or categories.

In many applications meaningful features are extracted from the raw data to form feature vectors. Mathematically a feature is an individual measurable property or characteristic of a phenomenon and a feature vector is a n-dimensional vector of numerical features that represent some object [3]. The goal of the classification problems is to deduce the correct class of an object from its features using appropriate classification algorithms or short classifiers. The result of a classification is therefore the assignment of a class to each object. As it can happen that a classifier assigns not the correct class to an object, there are several measures to control the quality of a classifier. The most important measures are sensitivity, specificity and accuracy. For simplicity, let us assume that there are two different classes, diseased (A) and healthy (CG). Then the sensitivity indicates the percentage of a test that detects the disease in people who are actually ill, whereas specificity indicates the percentage of a test that classifies healthy people correctly under all healthy people. The accuracy finally is the overall percentage of the correct assigned classes. The greater the sensitivity, specificity and accuracy, the better the classifier.

In most cases, the data set of objects is divided in a training and a test set when a classification is performed. The training data serve to deduce rules for the classification of the objects. These rules are then applied to the objects in the test set. As the results depend on the choice of the split in training and test set, the so-called k-fold-cross-validation is applied: The set is divided in k-sets of equal sizes. In a first step, the first set is taken as test set, the other sets as training sets. In a second step, the second set is taken as test set, the others as training sets, and so on. After training the classifier is tested and the overall quality measures are calculated from those of the individual steps.

In the above clinical case of the given photoplethysmographic data, several sorts of coefficients are evaluated as feature vector, however the best results were received for the coefficients from the frequency response approach:

Let $F(k_i)$ be the element of the fast Fourier-transform corresponding to the discrete frequency k_i where only complex features for the harmonic frequencies are computed. This is motivated by the observation that there is a clearly visible periodicity in the spectrum of the PPG-signals and the hypothesis that the effect of an aneurysm manifest in the periodical properties of the signal and not necessarily in the aperiodical. Therefore the first five harmonic frequencies are extracted by Matlab's findpeaks from the absolute values of the corresponding spectrum. These frequencies align in almost all cases perfectly for input and output signal, if there is a slight deviation ($k_{i,in} \neq k_{i,out}$) the features for the i-th peak are still divided to the resulting coefficients

$$H_i = \frac{F_{in}(k_{i,in})}{F_{out}(k_{i,out})}.$$

This is done for five intervals of ten seconds and the resulting values (real and complex parts) are averaged. For the sake of simplicity, we will only consider the case that the input signal is that at the right thumb and the output signal is that at the right toe.

These coefficients H_i were taken as features to evaluate the performance of several classification algorithms, however the best results were achieved using a simple K-nearest neighbour algorithm (KNN) with K=10. The KNN algorithm is a non-parametric supervised learning method; it was first defined by Evelyn Fix and Joseph Hodges in 1951 [16] and later expanded to regression methods by [1]. An object is classified by a plurality vote of its neighbours in the feature space, with the object being assigned to the class most common among its K nearest neighbours (K is a positive integer, typically small). If K = 1, then the object is simply assigned to the class of that single nearest neighbour.

As explained in [25], the euclidean distance is used as metric between the data points. It is well known that the performance of the classifier depends on the correct choice of K. If K is small it may happen that noise in the training sets decreases the accuracy. If K is chosen too big, it is possible that distant points influence the classification decision. In our case, we choose $K = 10$ and use the classifier as implemented in Matlab (see Listing 6.3.4).

Obviously, the amount of data in the clinical study is small. To deal with that problem, a procedure described in Figure 6.48 is employed and histograms of the resulting accuracies are created.

Step 1 (correct labels):
Do N-times
– Create a new random M-fold cross validation
– Calculate the accuracy
Step 2 (comparison with incorrect labels):
Do N_R-times
– Randomly permute the labels
– **Do N-times**
– Create a new random M-fold cross validation
– Calculate the accuracy

Fig. 6.48: Procedure to test whether the performance of a classifier trained and tested on correct labels exceeds the performance of a classifier trained and tested on incorrect labels. For the results in this paper, we always choose $N = 100$, $N_R = 100$ and $M = 10$. Algorithm is quoted from [25, Figure 3].

We quote from [25] that this was motivated by large deviations in the achievable accuracies depending on the split in training and test set. Further, the comparison with the achievable accuracies on classifiers trained with randomly permuted labels gives us an indicator if the used method is prone to overfitting due to hyperparameter tuning. Whilst it is evident that the cross-validation could have been left out in this case, by instead creating only one training and test set in the loop, this procedure gives a

reliable way to implement it on top of an existing classification method. For the cross-validation itself, it is always ensured that two measurements of the same patient are not split between test and training set.

Due to the existing measurement devices and other limitations, the mean of the accuracies for the classifications on correct labels is around 60%, whereas it is 50% with randomly permuted labels, see Figure 6.49.

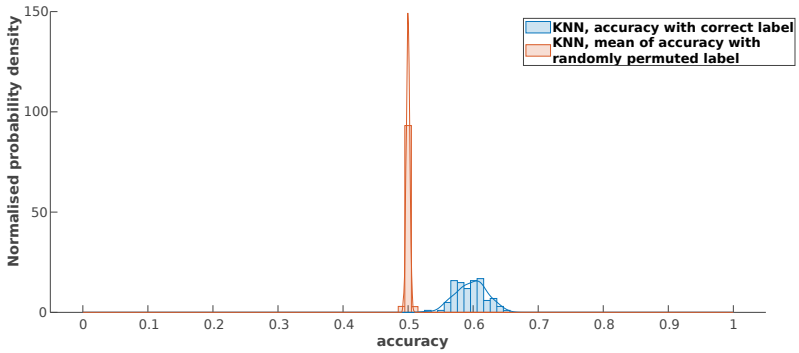


Fig. 6.49: Resulting accuracies with frequency response coefficients as features: The blue histogram shows the distribution of accuracies achieved with 100 different partitions on correctly labeled data as described in Figure 6.48. In comparison, the red histogram shows the distribution of the accuracy of a classifier trained and evaluated on randomly permuted labels 100 times, where each time the accuracy for 100 different partitions was calculated and averaged. (This figure is quoted from [25, Figure 7].)

This proves the influence of aneurysms on the collected data, but the resulting accuracies are still not suitable for technical purposes. The method can therefore detect an aneurysm with 60% accuracy, actual results using four instead of two PPG signals lead to 64%; this gives a first indication that can be further controlled by image processing methods.

Listing 6.3.4: Matlab function for K-next neighbour classification.

```
function KNN(obj,previousMethod,data,labels,previousLog,previousModel)
% Tests whether the right parameters are given
...

% Use data all together
X = reshape([data.data],length(data(1).data),[])';
Y = [labels.data];
```

```
% Fit one model on the complete data set
modelAll = fitcknn(X,Y,'NumNeighbors',neighbours,...
    'DistanceWeight','equal');
[¬,score] = resubPredict(modelAll);
predictedClass = predict(modelAll,X);
confusionMatrix = confusionmat(Y,predictedClass,...
    'Order',["K" "A"]);

% write these predictions to the results for now
for i = 1:length(data)
    obj.results(i).data = score(i,:);
end

% global Matrix to the log
logindex = 1;
obj.log(logindex).description = "confusionMatrixAll";
obj.log(logindex).data = reshape(confusionMatrix,1,[]);
logindex = logindex + 1;

obj.log(logindex).description = "accuracyAll";
obj.log(logindex).data = sum(diag(confusionMatrix))/...
    sum(confusionMatrix,'all');
logindex = logindex + 1;

%cross validation, output of confusion Matrix for each fold
if mode == "standard"
    cp = cvpartition(Y,'KFold',KFold);
elseif mode == "skip"
    cp = cvpartition(Y(2:2:end),'KFold',KFold);
else
    error("Mode not implemented");
end

for i = 1:KFold
    % our own cross-validation
    if mode == "standard"
        trainingIndizes = training(cp,i);
        testIndizes = test(cp,i);
    elseif mode == "skip"
        trainingIndizes = reshape(repmat(training(cp,i),1,2)',...
            [],1);
        testIndizes = reshape(repmat(test(cp,i),1,2)',[],1);
```

```
    else
        error("Mode not implemented");
    end

    modelTrain = fitcknn(X(trainingIndizes,:),...
        Y(trainingIndizes),'NumNeighbors',neighbours,...
        'DistanceWeight','equal');

    predictedClass = predict(modelTrain,X(testIndizes,:));
    confusionMatrix = confusionmat(Y(testIndizes),...
        predictedClass,'Order',["K" "A"]);

    obj.log(logindex).description = "confusionMatrixFold" +...
        string(i);
    obj.log(logindex).data = reshape(confusionMatrix,1,[]);
    logindex = logindex + 1;

    obj.log(logindex).description = "accuracyFold" +...
        string(i);
    obj.log(logindex).data = sum(diag(confusionMatrix))/...
        sum(confusionMatrix,'all');
    logindex = logindex + 1;
end
```

6.4 Post-Reading and Exercises

Signals of the Brain

1. What is the basis for measuring an EEG? What information do you hope to gain from it?
2. Describe and sketch the different brain regions. Name their function and assign the corresponding sensory organs.
3. Describe a sensory organ in some detail; how is the stimulus transformed and how does it reach the brain.
4. What are evoked potentials? How can they be evoked in experiments?
5. What is meant by reference points? Which ones are indicated in EEG?
6. What is event correlation?
7. How can the influences of events be filtered out of the noise?
8. What diseases can be diagnosed with the EEG?
9. What is a frequency band and which ones do you know in the EEG? What do they stand for? Make a sketch of the signals.

10. How can the activities of the bands be determined?
11. Describe the use of EEG in sleep research.
12. What is meant by short-term FFT? What is it used for in EEG?
13. What is understood by topography, what by tomography?

Signals of the Muscles and Motions

1. What is EMG used for? What results does it give?
2. How is an EMG signal generated? Describe the physiological structure and the origin of the electrical signal at the electrodes.
3. How can the fatigue of a muscle be described quantitatively on the EMG?
4. What influence do electrodes (shape, size, material) have on the measurement?
5. What characteristics should an EMG amplifier have?
6. Do you really measure the exact activity of the muscle at the surface?
7. Name the magnitude of the conduction velocity in the nerve and in the muscle.
8. In what form is this expressed in the spectrum of the EMG in muscle fatigue?
9. What information does a power density spectrum contain?
10. By which parameter can muscular fatigue be described?
11. Do the knee bends by your own and document the movement by means of a mobile camera.
12. Describe the kinetic behaviour of your knee bends.
13. Think about a biomechanical model of your body.
14. Develop a biomechanical model on the base of a multi body system.
15. Think about the moment of inertia of the different masses.
16. Think about the ground reaction force during the knee bend. Use a mechanical weighing machine and observe the scale.
17. What is in principle the difference between the parameters force and joint angle?
18. What kind of measurement system are used to document the external force and the joint angles in human movement analysis?
19. What kind of models are used to analyse the biomechanics of the human movement?
20. What is the difference between static and dynamic in biomechanics?
21. What is osteoarthrosis?
22. In which manner are the patient burdened during the examination in gonarthrosis?
23. What kind of phases are defined for the joint angle in the acoustic-kinetic analysis?
24. Why you cannot transfer the definition of the phases of the joint angle immediately to the force?
25. What kind of biomechanical defined phases are used for the force?
26. In which case you have to synchronise the signals of the force and the joint angles?
27. What kind of attributes are used to synchronise the force and the joint angle?

28. What is the strategy to synchronise automatically the signals by means of a program?
29. What are the problems if you want to assess the movement of a patient?

Signals of the Cardiovascular System

1. What diseases can be read from a patient's ECG?
2. Why does the inference from the signal to the disease of the heart work?
3. In what order of magnitude are the measured potentials of the ECG?
4. What are the physiological causes of the first and second heart sounds?
5. At what times in relation to the ECG signal do the first and second heart sounds occur?
6. How does the phonocardiogram change under severe physical stress compared to that at rest?
7. Why is the phonocardiogram broken down into different frequency bands? What diagnostic information can be obtained from it?
8. What is meant by the abbreviation 2L2 in relation to a phonocardiogram?
9. At what times in relation to the ECG signal do the first and second heart sounds occur?
10. What is the relationship between the "LVET" parameter and the phonocardiogram?
11. Why is autocorrelation used for the automatic determination of LVET?
12. How can the LVET be read from the result of the autocorrelation?
13. What is the frequency relationship of a phonocardiogram and the sounds of mechanical prosthetic heart valves?
14. Name two objectives of a photoplethysmogram. What parameters can be obtained?
15. On which physical properties of blood is the SpO_2 measurement based? How can these be measured?
16. Describe the respiration process down to the cells. What effects play a role there?
17. What is meant by an absorption spectrum? What quantities are needed to describe absorption?
18. What does the Lambert-Beer law say?
19. Why do you need two wavelengths for the determination of SpO_2?
20. Why is one time point not sufficient to determine the SpO_2 value?
21. Describe the technical construction of a pulse oximeter.
22. Describe the signal-processing structure of a pulse oximeter based on its components.
23. Which component is used to use a common photodetector for both LEDs? What other signal processing steps does this require?
24. Why do you pulse the two LEDs?
25. In which range do you use pulse oximeters?

26. Why do the histograms in Figure 6.49 prove the influence of an aneurysm in the data?
27. What other possible coefficients could you imagine to take as features for a classification „aneurysm“ vs. „control group“?
28. Why is it necessary to ensure that two measurements of the same patient are not split between test and training set?
29. Why is it important to measure the PPG-data always in quiescent condition and supine position?
30. What is the motivation to use transfer-function parameters as features for the classification?

7 Appendix: Quantity- / Unit Symbols and Important Constants

Tab. 7.1: Important natural constants, quantity and unit symbols.

Name of Quantity	Symbol	Domain / Value	Unit	Unitsymbol
chapter 2				
Information	I	$\mathbb{R}$	Shannon	sh
Probability	p	$0, \ldots, 1$		
Data volume	D	$\mathbb{R}$	Bit or Byte	b or B
Word width	W	$\mathbb{N}$		
Data rate	C	$\mathbb{R}$	$\frac{\text{Bit}}{\text{second}}$	bit/s
Symbol duration	T_s	$\mathbb{R}$	second	s
Symbol rate	f_s	$\mathbb{R}$	Baud	Bd
Power	P	$\mathbb{R}$	Watt	W
Potential	u, U	$\mathbb{R}$	Volt	V
Current	i, I	$\mathbb{R}$	Ampere	A
Resistance	R	$\mathbb{R}$	Ohm	Ω
Time	t, t_0, T	$\mathbb{R}$	second	s
Signal to noise ratio	SNR/dB	$\mathbb{R}$	Dezibel	dB
Bandwidth	B	$\mathbb{R}$	Hertz	Hz
real Fourier coefficients	a_k, b_k	$\mathbb{R}$		
Period time	T	$\mathbb{R}$	second	s
Angular frequency	ω	$\mathbb{R}$	$\frac{\text{radiant}}{\text{second}}$	rad/s
Normal distribution	$\mathbb{N}$	$\mathbb{R}$		
Variance	$\text{Var}(X)$	$\mathbb{R}$		
Standard deviation	$\sigma(X)$	$\mathbb{R}$		
Mean / expectation value	$\mu(X)$, $E(X)$	$\mathbb{R}$		
Energy	E	$\mathbb{R}$		
Integal kernel	K			
Variable in complex domain	ξ	$\mathbb{R}$		
Laplace variable	p	$\mathbb{C}$		
Real part of Laplace variable	σ	$\mathbb{R}$		
chapter 3				
Avogadro number	N_A	$6.022 \cdot 10^{23}$	Mol^{-1}	mol^{-1}
Elementary charge	e	$1.602 \cdot 10^{-19}$	Coulomb	C

https://doi.org/10.1515/9783110736298-007

Name of Quantity	Symbol	Domain / Value	Unit	Unitsymbol
Electrical charge	q, Q	$\mathbb{R}$	Coulomb	C
Boltzmann constant	k_b	$1.381 \cdot 10^{-23}$	Joule/Kelvin	J/K
Universal gas constant	R	8.314	$\frac{\text{Joule}}{\text{Kelvin} \cdot \text{Mol}}$	$\text{J}/(\text{K} \cdot \text{mol})$
Faraday constant	F	$9.64 \cdot 10^4$	$\frac{\text{Coulomb}}{\text{Mol}}$	C/mol
Electrical resistance	R	$\mathbb{R}$	Ohm	Ω
Specific resistance	r	$\mathbb{R}$	Ohm · meter	Ω · m
Electrical conductivity	σ	$\mathbb{R}$	Siemens	S
Electrical capacitance	C	$\mathbb{R}$	Farad	F
Electrical current	i, I	$\mathbb{R}$	Ampere	A
Electrical potential	u, U	$\mathbb{R}$	Volt	V
Charge number	z	$\mathbb{N}$		
Membrane potential	U_{rev}	$\mathbb{R}$	Volt	V
Time constant	τ	$\mathbb{R}$	second	s
Time	t	$\mathbb{R}$	second	s
Nabla operator	Δ	$\mathbb{R}$	1/meter	m^{-1}
Electrical field intensity vector	E	$\mathbb{R}$	$\frac{\text{Volt}}{\text{meter}}$	V/m
Electrical permeability	ϵ	$\mathbb{R}$	$\frac{\text{Ampere} \cdot \text{second}}{\text{Volt} \cdot \text{m}}$	As/Vm
Vacuum electrical permeability	ϵ_0	$8.853 \cdot 10^{-12}$	$\frac{\text{Ampere} \cdot \text{second}}{\text{Volt} \cdot \text{m}}$	As/Vm
Magnetic field intensity vector	H	$\mathbb{R}$	$\frac{\text{Ampere}}{\text{meter}}$	A/m
Magnetic permeability	μ	$\mathbb{R}$	$\frac{\text{Volt} \cdot \text{second}}{\text{Ampere} \cdot m}$	Vs/Am
Vacuum magnetic permeability	μ_0	$4\pi \cdot 10^{-7}$	$\frac{\text{Volt} \cdot \text{second}}{\text{Ampere} \cdot \text{m}}$	Vs/Am
Current density vector	J	$\mathbb{R}$	$\frac{\text{Ampere}}{\text{meter}^2}$	A/m^2
Current density vector of electrical field	J_E	$\mathbb{R}$	$\frac{\text{Ampere}}{\text{meter}^2}$	A/m^2
Current density vector of ion membrane currents	J_i	$\mathbb{R}$	$\frac{\text{Ampere}}{\text{meter}^2}$	A/m^2
Arbitrary vector field	A	$\mathbb{R}$	undefined	undefined
Arbitrary scalare function	ϕ	$\mathbb{R}$	undefined	undefined
Volume charge density	ρ_v	$\mathbb{R}$	$\frac{\text{Ampere}}{\text{meter}^3}$	As/m^3
Specific electrical conductivity	κ	$\mathbb{R}$	$\frac{\text{Siemens}}{\text{meter}}$	S/m
Angular frequency	ω	$\mathbb{R}$	Hertz	Hz
Distance between current source and measuring location	r	$\mathbb{R}$	meter	m

Name of Quantity	Symbol	Domain / Value	Unit	Unitsymbol
Volume	V	$\mathbb{R}$	cubicmeter	m^3
Wave vector	k	$\mathbb{R}$	1/meter	1/m
Potentials on the left and right arm and foot	Φ_L, Φ_R, Φ_F	$\mathbb{R}$	Volt	V
Lead voltages according to Einthoven	V_I, V_{II}, V_{III}	$\mathbb{R}$	Volt	V
Lead voltages according to Goldberger	aV_L, aV_R, aV_F	$\mathbb{R}$	Volt	V
Lead voltages according to Wilson	$V1$ bis $V6$	$\mathbb{R}$	Volt	V
Potential matrix on the heart surface	$\mathbf{S}$	$\mathbb{R}$	Volt	V
Potential matrix on the body surface	Φ	$\mathbb{R}$	Volt	V
Transfer matrix	$\mathbf{A}$	$\mathbb{R}$	dimensionless	none
chapter 4				
Sampling time	T_a	$\mathbb{R}$	second	s
Sampling frequency	f_a	$\mathbb{R}$	Hertz	Hz
Bandwidth	f_B	$\mathbb{R}$	Hertz	Hz
Time constant	τ	$\mathbb{R}$	second	s
Input resistance	R_E	$\mathbb{R}$	Ohm	Ω
Cut-off frequency	f_g	$\mathbb{R}$	Hertz	Hz
Group delay	T_g	$\mathbb{R}$	second	s
Magnetic flux	Φ	$\mathbb{R}$	Tesla $\cdot$ m^2	T $\cdot$ m^2
Magnetic induction	B	$\mathbb{R}$	Tesla	T
Temperature	T	$\mathbb{R}$	Kelvin	K
Magnetic permeability in vacuum	μ_0	$4\pi \cdot 10^{-7}$	$\frac{\text{Volt} \cdot \text{second}}{\text{Ampere} \cdot \text{m}}$	Vs/Am
Oxygen saturation	SpO_2	$\mathbb{R}$	percent	%
chapter 5				
Sampling instance	t_a	$\mathbb{R}$	second	s
Sampling interval	T_a	$\mathbb{R}$	second	s
Sampling frequency	f_a	$\mathbb{R}$	Hertz	Hz
Amplitude of a signal	A	$\mathbb{R}$		
Width of a sampling interval	ΔT	$\mathbb{R}$	second	s
Frequency resolution	Δf	$\mathbb{R}$	Hertz	Hz
Continuous signal	$f(t)$	$\mathbb{R}$		
Fourier-transform of $f(t)$	$F(f)$	$\mathbb{C}$		
Discrete signal	$f(n)$	$\mathbb{R}$		

Name of Quantity	Symbol	Domain / Value	Unit	Unitsymbol
z- transformed of $f(n)$	$F(z)$	$\mathbb{R}$		
Sampled signal with Dirac pulse	$f_T(t)$	$\mathbb{R}$		
Fourier-transformed of $f_T(t)$	$F_T(t)$	$\mathbb{R}$		
Sampled signal with rectangular-pulse	$f_{\Delta T}(t)$	$\mathbb{R}$		
Fourier-transform of $f_{\Delta T}(t)$	$F_{\Delta T}(t)$	$\mathbb{R}$		
Short notation for sampled signal at the times $f_T(j \cdot T)$	$f_T(j)$	$\mathbb{R}$		
Discrete signal after undersampling	$f_{T\mathrm{u}}(n)$	$\mathbb{R}$		
Discrete signal after oversampling	$f_{T\ddot{\mathrm{u}}}(n)$	$\mathbb{R}$		
Fourier matrix	$\mathbf{W}$	$\mathbb{R}$		
Circular cut-off frequency	$\omega_{\mathrm{g}}(t)$	$\mathbb{R}$	$\frac{\text{Radiant}}{\text{second}}$	rad/s
Cut-off frequency	$f_{\mathrm{g}}(t)$	$\mathbb{R}$	Hertz	Hz
Averaged discrete signal	$f_{\mathrm{M}}(k)$	$\mathbb{R}$		
Maximum signal duration	t_g	$\mathbb{R}$	second	s
Complex frequency	p	$\mathbb{C}$		
Periodic time	t_{p}	$\mathbb{R}$	second	s
Periodic signal	$f_{\mathrm{p}}(t)$	$\mathbb{R}$		
Periodic Sequence of Dirac pulses in the frequency domain	$P(f)$			
Rectangular pulse	$g_F(t)$	$\mathbb{R}$		
Random variable x	X	$\mathbb{R}$		
Expectation value X	$E(X)$	$\mathbb{R}$		
Correlation function of X_1 and X_2	$R_{X_1X_2}$	$\mathbb{R}$		
Auto-correlation function of x at time t_i with a shift of τ_j	$R_{XX}(t_i, \tau_j)$	$\mathbb{R}$		
Auto-correlation function of $R_{XX}(t_i, \tau_j)$ averaged over t_i	$\overline{R}_{XX}(j)$	$\mathbb{R}$		
Discrete Fourier-transform of $\overline{R}_{XX}(j)$ discrete power spectral density	$\overline{S}_{XX}(j)$	$\mathbb{R}$		

Name of Quantity	Symbol	Domain / Value	Unit	Unitsymbol
Auto-covariance for $R_{XX}(t_i, \tau_j)$ averaged over t_i	$\overline{C}_{XX}(j)$	$\mathbb{R}$		
Discrete-time function	$f(n)$	$\mathbb{R}$		
Discrete-time Fourier-transform of $f(n)$	$F_D(m)$	$\mathbb{C}$		
Mathematical operator applied to $x(n)$	$T\{x(n)\}$	$\mathbb{R}$		
Discrete Dirac pulse	$\delta(n)$	$\mathbb{R}$		
Discrete pulse response	$g(n)$	$\mathbb{R}$		
Fourier-transform of $g(n)$	$G(m)$	$\mathbb{R}$		
Discrete transfer matrix	$\mathbf{G}$	$\mathbb{R}$		
Circular matrix with $g_p(n)$	$\mathbf{Zykl}\{g_\mathrm{p}\}$	$\mathbb{R}$		
Triangular matrix with $g(n)$	$\mathbf{Dr}\{g\}$	$\mathbb{R}$		
Discrete Dirac-pulse sequence	$\delta_\mathrm{p}(n)$	$\mathbb{R}$		
Discrete pulse response	$g_\mathrm{p}(n)$	$\mathbb{R}$		
Number of values per period of a periodic sequence	N_p	$\mathbb{R}$		
Discrete sequence	$x(n), y(n)$	$\mathbb{R}$		
Number sequence at node i of a digital filter	w_i	$\mathbb{R}$		
Filter coefficients of a digital filter	c_i, d_i	$\mathbb{R}$		
Multiplier coefficient of a digital filter	K	$\mathbb{R}$		
Discrete-time Fourier-transform of $x(n), y(n)$	$X(n), Y(n)$	$\mathbb{R}$		
Time-dependent window function	$w(t)$	$\mathbb{R}$		
Fourier-transformed of the window function $w(t)$	$W(f)$	$\mathbb{R}$		
Inverse window function of $w(t)$	$h(t)$	$\mathbb{R}$		
Fourier-transformed of the inverse window function $h(t)$	$H(f)$	$\mathbb{R}$		
Signal faded out with the window function $w(t)$	$f_w(t)$	$\mathbb{R}$		

Name of Quantity	Symbol	Domain / Value	Unit	Unitsymbol
Fourier-transformed of the signal $f_w(t)$	$F_w(f)$	$\mathbb{R}$		
Pulse width in the time domain	T	$\mathbb{R}$	second	s
Pulse width in the frequency domain	B	$\mathbb{R}$	Hertz	Hz
Measure for the width of a Gaussian window	σ	$\mathbb{R}$		
Impulse function in the time domain	$i(t)$	$\mathbb{R}$		
Impulse function in the frequency domain	$I(f)$	$\mathbb{R}$		
Low-frequency signal of the DWT of the i-th level	$X_{\mathrm{A}i}$	$\mathbb{R}$		
High-frequency signal of the DWT of the i-th level	$X_{\mathrm{D}i}$	$\mathbb{R}$		
Impulse responses of the analogue and digital signal in the impulse invariance method	$g_{\mathrm{an}}(t)$, $g_{\mathrm{di}}(n)$	$\mathbb{R}$		
z- transformed of $g_{\mathrm{an}}(t)$, $g_{\mathrm{di}}(n)$	$G_{\mathrm{an}}(t)$, $G_{\mathrm{di}}(n)$	$\mathbb{R}$		
Impulse responses of a RC-circuit	$g_{\mathrm{RC}}(t)$	$\mathbb{R}$		
chapter 6				
Power density spectrum	Φ			
Auto-correlation function	ϕ			
Median frequency	f_{mean}	$\mathbb{R}$	Hertz	Hz
Intensity	I	$\mathbb{R}$	Watt/ meter^2	W/m^2
Absorption coefficient	α	$\mathbb{R}$	1/meter	1/m
Wavelength	λ	$\mathbb{R}$	1/meter	1/m
Molar extinction coefficient	ϵ	$\mathbb{R}$	$\frac{\text{liter}}{\text{Mol} \cdot \text{centimeter}}$	l·Mol^{-1}·cm^{-1}

Bibliography

[1] Naomi S. Altman. An introduction to kernel and nearest neighbor noparametric regression. *The American Statistician.*, 46(3):175–185, June 1991.

[2] G. W. Beeler and H. J. Reuter. Reconstruction of the action potential of ventricular myocardial fibres. *Journal of Physiology 307*, 1977.

[3] Christopher Bishop. *Pattern recognition and machine learning*. Springer, Berlin, 2006.

[4] Armin Bolz and Wilhelm Urbaszek. *Technik in der Kardiologie*. Springer, 2002.

[5] M. Bresadola. Medicine and science in the life of luigi galvani. *Brain Research Bulletin*, 46(5):367–380, 1998.

[6] Eugene N. Bruce. *Biomedical Signal Processing and Signal Modeling*. John Wiley & Sons, 2001.

[7] G.D. Clifford and L. Tarassenko. Quantifying errors in spectral estimates of hrv due to beat replacement and resampling. *IEEE transactions on bio-medical engineering*, 2005.

[8] P. Colli-Franzone, L. Guerri, et al. A mathematical procedure for solving the inverse problem of electrocardiography. analysis of the time-space accuracy from in vitro experimental data. *Math. Bioscience*, 1985.

[9] A. G. Constantinides. Spectral transformations for digital filters. *POC. IEE, Vol. 117, No. 8*, 1970.

[10] M. H. Criqui, R. D. Langer, A. Fronek, H. S. Feigelson, M. R. Klauber, T. J. McCann, et al. Mortality over a period of 10 years in patients with peripheral arterial disease. *The New England Journal of Medicine*, 326(6):381–386, February 1992.

[11] E. Demirel. *Analyse von Kraftmessdaten der Standardbewegung in der Schallemissionsanalyse von Arthrosepatienten im Rahmen einer klinischen Studie.* Technische Hochschule Mittelhessen, Life Science Engineering, Bachelorarbeit, 2020.

[12] P. D. deSolla. *On the Brink of Tomorrow: Frontiers of Science.* National Geographic Society, 1982.

[13] J. Enderle. *Introduction to Biomedical Engineering*. Boston: Academic Press, 2005.

[14] Wolfgang Engelmann and K-Heinz Witte. Wie man eine biologische uhr stoppen kann: Singularitaetspunkt. *TOBIAS-lib - Hochschulschriftenserver der Universitaet Tuebingen*, 2016.

[15] R. FitzHugh. Impulses and physiological states in theoretical models of nerve membrane. *Biophysical J.*, 1961.

[16] Evelyn Fix and Joseph Hodges. Discriminatory analysis. nonparametric discrimination: Consitency properties. *USAF School of aviation medicine, Randolph Field, Texas*, February 1951.

[17] Dirk Fritzsche, Thomas Eitz, et al. Early detection of mechanical dysfunction using a new home monitoring device. *The Annals of Thoracic Surgery*, 83:542, 2007.

[18] David B. Geselowitz. On the theory of the electrocardiogram. *Proceedings of the IEEE Vol. 77, No. 6*, 1989.

[19] Werner Gitt. *Am Anfang war die Information*. Hänssler-Verlag, 2002.

[20] S. Gois and M. A. Savi. An analysis of heart rhythm dynamics using a three-coupled oscillator model. *Chaos, Solitons and Fractals*, 2009.

[21] E Goldberger. *The aVL, aVR, and aVF leads; A simplification of standard lead electrocardiography*. Am. Heart, 1942a.

[22] E Goldberger. *A simple indifferent electrocardiographic electrode of zero potential and a technique of obtaining augmented, unipolar extremity leads*. Am. Heart, 1942b.

[23] Krzysztof Grudzinski and Jan J. Zebrowski. Modeling cardiac ppacemaker with relaxion oszillators. *Physica A 336*, 2004.

[24] Urs Hackstein, Stefan Krickl, and Stefan Bernhard. Estimation of ARMA-model parameters to describe pathological conditions in cardiovascular system models. *Informatics in Medicine - Unlocked*, 18:100310, January 2020.

https://doi.org/10.1515/9783110736298-008

[25] Urs Hackstein, Tobias Krüger, Alexander Mair, Charlotte Degünther, Stefan Krickl, Christian Schlensak, and Stefan Bernhard. Early diagnosis of aortic aneurysms based on the classification of transfer function parameters estimated from two photoplethysmographic signals. *Informatics in Medicine - Unlocked*, 25:100652, January 2021.
[26] Thomas Harriehausen and Dieter Schwarzenau. *Moeller Grundlagen der Elektrotechnik*. Springer Vieweg, 2013.
[27] Jeremy Hill, Disha Gupta, et al. Recording human electrocorticographic (ECoG) signals for neuroscientific research and real-time functional cortical mapping. *Journal of Visualized Experiments*, 64:3993, 2012.
[28] A. L. Hodgkin and A. F. Huxley. A quantitative description of membran current and its application to conduchtion and exitation in nerve. *J. Physiol.*, 1952.
[29] A. L. Hodgkin and A.F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology.*, 117:500–544, 1952.
[30] Josef Hoffmann and Franz Quint. *Einführung in Signale und Systeme*. Oldenbourg Verlag, 2013.
[31] Klaus Holldack and Dieter Wolf. *Atlas und kurzgefasstes Lehrbuch der Phonokardiographie*. Georg Thieme Verlag, 1974.
[32] https://de.wikipedia.org/wiki/Fensterfunktion. Fensterfunktion. Internet, 02 2018.
[33] Peter Husar. *Biosignalverarbeitung*. Springer, 2010.
[34] Kiselev J., Wolf U., Ziegler B., Schwalbe H.-J., and Franke R.-P. Detection of early phases of osteartritis using acoustic emission analysis. *Med. Eng. Phys.*, 65:57–60, 2019.
[35] Subke J., Schneider B., Hanitz F., Krüger S., Junker H.-O., Schwalbe H.-J., and Wolf U. Clinical case study in acoustic-kinetic joint analysis: Synchronization and evaluation of kinetic measurement data in aea (acoustic emission analysis) based diagnosis of arthritic knee joint defects. *Current Directions in Biomedical Engineering*, 2021.
[36] Subke J., Schneider B., Hanitz F., and Wolf U. Acoustic-kinetic joint analysis: Synchronization and evaluation of kinetic measurement data in aea (acoustic emission analysis) based diagnosis of arthritic knee joint defects. *Current Directions in Biomedical Engineering*, 2021.
[37] Subke J., Krueger S., Junker H.-O., Schwalbe H.-J., Franke R.-P., and Wolf U. An introduction to acoustic emission analysis (aea) based medical diagnosis techniques: Screening and monitoring of cartilage defects in knee osteoarthrosis. *Current Directions in Biomedical Engineering*, 5(1):1–3, 2019.
[38] Ulrich Karrenberg. *Signale, Prozesse, Systeme*. Springer, 2017.
[39] J. G. Kim and H. Liu. Variation of haemoglobin extinction coefficients can cause errors in the determination of haemoglobin concentration measured by near-infrared spectroscopy. *Phys. Med. Biol.*, 52:6295–6322, 2007.
[40] Jae G. Kim, Mengna Xia, and Hanli Liu. Extinction coefficients of hemoglobin for near-infrared spectroscopy of tissue. *IEEE Engineering in Medicine and Biology*, 2005.
[41] Arild Lacroix and K-Heinz Witte. *Zeitdiskrete normierte Tiefpaesse*. Huethig, 1980.
[42] Claire M. Lochner, Yasser Khan, et al. All-organic optoelectronic sensor for pulse oximetry. *Nature Communications*, 2014.
[43] Hui Wen Loh, Shuting Xu, Oliver Faust, Chui Ping Ooi, Prabal Datta Barua, Subrata Chakraborty, Ru-San Tan, Filippo Molinari, and U Rajendra Acharya. Application of photoplethysmography signals for healthcare systems: An in-depth review. *Computer Methods and Programs in Biomedicine*, 216:106677, April 2022.
[44] N.R. Lomb. Least-squares frequency analysis of unequally spaces data. *Astrophysics ans Space Science*, 1976.
[45] H-Dieter Lueke. *Signalübertragung*. Springer, 1999.
[46] Stephane Mallat. *A wavelet tour of signal processing: the sparse way*. Academic Press, 2009.
[47] Jaako Malmivuo and Robert Plonsey. *Bioelectromagnetism*. Oxford University Press, 1995.

[48] PE McSharry, GD Clifford, et al. A dynamical model for generating synthetic electrocardiogram signals. *IEEE Transactions on Biomedical Engineering*, 2003.
[49] Nicole Menche (Mit Beiträgen von Stefanie Engelhardt and Bernd Guzek und anderen), editor. *Biologie Anatomie Physiologie*. Elsevier Urban & Fischer, 2012.
[50] Roberto Merletti and Philip Parker, editors. *Electromyography*. John Wiley & Sons, 2004.
[51] Otto Mildenberger. *Entwurf analoger und digitaler Filter*. Vieweg, 1992.
[52] Otto Mildenberger. *Uebertragungstechnik, Grundlagen analog und digital*. Vieweg, 1997.
[53] J. Nagumo, S. Arimoto, et al. An active pulse transmission line simulating nerve axon. *Proc IRE*, 1962.
[54] D. Noble. A modification of the hodgkin-huxley equations applicablt to purkinje fiber action and pacemaker potential. *Journal of Physiology. 160*, 1962.
[55] A.van Oosterom and T. Oostendorp. ECGSIM; an interactive tool for studying the genesis of qrst waveforms. *Heart*, 2004.
[56] Alan V. Oppenheim and Rolnalf Schafer. *Digital Signal Processing*. Prentice-Hall, 1975.
[57] H. S. Oster, B. Taccardi, et al. Noninvasive electrocardiographic imaging: Single and multiple electrocardiographic events. *Circulation*, 1997.
[58] Jiapu Pan and Willis J. Tompkins. A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering*, 1985.
[59] H. Piper. *Elektrophysiologie Menschlicher Muskeln*. Springer Verlag, 1912.
[60] R. Plonsey and R. Collin. *Principles and Applications of Electromagnetic Fields*. McGraw-Hill, 1961.
[61] Robert Plonsey. *The theoretical basis of electrocardiology*. Clarendon Press Oxford, 1976.
[62] J. G. Proakis and D. G. Manolakis. *Digital Signal Processing, Principles, Algorithms, and Application*. Prentice-Hall, 1996.
[63] C. M. Quick, W. L. Young, and A. Noordergraaf. Infinite number of solutions to the hemodynamic inverse problem. *American Journal of Physiology. Heart and Circulatory Physiology*, 280(4):H1472–1479, April 2001.
[64] Franke R.-P., Schwalbe H.-J., Kiselev J., Wolf U., Subke J., and Ziegler B. Schallemissionsanalyse zum nachweis von gelenkdefekten in der medizinischen diagnostik. *Deutsche Gesellschaft für zerstörungsfreie Prüfung 18. Kolloquium Schallemission, Berichtsband*, 2011.
[65] M. Rabenberg. *Arthrose. Gesundheitsberichterstattung des Bundes.*, volume 54. Robert Koch-Institut, Berlin, 2013.
[66] John T. Ramshur. Design , evaluation, and application of heart rate variability analysis software (hrvas). Master's thesis, The University of Memphis, 2010.
[67] Jürgen Rettinger, Silvia Schwarz, and Wolfgang Schwarz. *Electrophysiology*. Springer International Publishing, Cham, 2016.
[68] Werner Rupprecht. *Netzwerksynthese*. Spriner Berlin. Heidelberg, NewYork, 1972.
[69] B. Schneider. *Datenanalyse und –synchronisation im Rahmen einer klinischen Studie zur Gonarthrose.* Technische Hochschule Mittelhessen, Life Science Engineering, Bachelorarbeit, 2021.
[70] H. W. Schuessler. *Digital Systeme zur Signalverarbeitung*. Springer, 1973.
[71] W. Schwarz and J. Rettinger. *Elektrophysiologie*. Shaker Verlag, 2004.
[72] Hooman Sedghamiz. Matlab implementation of pan tompkins ECG QRS detector. Technical report, Rochester General Hospital, 2014.
[73] N. Shresta. *Bewegungsanalyse im Rahmen einer klinischen Studie zur Gonarthrose. Bewegung der Kniebeuge anhand der Koordinaten der Bodenreaktionskräfte.* Technische Hochschule Mittelhessen, Life Science Engineering, Bachelorarbeit, 2021.
[74] I. K. Skwirzynski. *Design Theory and Data for Electrical Filters*. D. van Nostrand Company ltd., 1965.

[75] Till Tantau. *The TikZ and PGF Packages*, 2013.
[76] TexasInstruments. Analysis of the sallen-key architecture, 2003.
[77] N. V. Thakor, J. G. Webster, et al. Optimal QRS detector. *Med. Biol. Eng. Comput.*, 1983.
[78] M. Titze. *Schallemissionsanalyse als Nachweis des Reibverhaltens im menschlichen Kniegelenk in verschiedenen Belastungssituationen.* Diplomarbeit FB Maschinenbau und Feinwerktechnik, FH-Gießen-Friedberg, 1996.
[79] O. W. van Assendelft and W.G. Zijlstra. Extinction coefficients for use in equations for the spectrophotometric analysis of haemoglobin mixtures. *Analytical Biochemestry*, 69:43–48, 1975.
[80] J. Walsh. Of the electric property of the torpedo. *Philosophical Transactions*, 63:461–480, 1773.
[81] P.D. Welch. The use of fast fourier transform for the estimation of power spectra: A methode based on time averaging over shord modified periodograms. *IEEE Trans. Audio and Electroacoustics*, 1967.
[82] Martin Werner. *Signale und Systeme*. Vieweg + Teubner, 2008.
[83] F. N. Wilson, F. D. Johnston, et al. Electrocardiograms that represent the potential variations of a single electrode. *Am. Heart*, 1934.
[84] F. N. Wilson, F. D. Johnston, et al. The precordial electrocardiogram. *Am. Heart*, 1944.
[85] F. N. Wilson, A. G. Macleod, et al. Potential variations produced by the heart beat at the apices of einthoven's triangle. *Am. Heart*, 1931.
[86] C. H. Wu. Electric fish and the discovery of animal electricity. *American Scientist*, 72:598–607, 1984.
[87] K. Yanagihara, A. Noma, et al. Reconstruction of sino-atrial node pacemaker potential based on voltage clamp experiments. *Japanese Journal of Physiology 30*, 1980.
[88] Oksana Zayachkivska and A. Coenen. Mecks-memorial an essey in honour of his 150th birthday. 01 2016.
[89] J. Zebrowski and K. Grudzinski. Nonlinear oscillator model reproducing various phenomena in the dynamics of the conduction system of the heart. *Chaos*, 2007.
[90] Eberhard Zeidler, editor. *Springer-Taschenbuch der Mathematik*. Springer Fachmedien Wiesbaden, Wiesbaden, 2013.
[91] S. Ben Zekry, A. Sagie, et al. Initial clinical experience with an hand-held device (thrombocheck) for the detection of bileaflet prosthetic valve malfunction. *The Journal of Heart Valve Disease*, 14:476, 2005.

Index

https://doi.org/10.1515/9783110736298-009