2010

Prognosis: A Wearable System for Health Monitoring of People at Risk

Alexandros A. Pantelopoulos
Wright State University

Follow this and additional works at: http://corescholar.libraries.wright.edu/etd_all
Part of the Computer Engineering Commons, and the Computer Sciences Commons

Repository Citation

This Dissertation is brought to you for free and open access by the Theses and Dissertations at CORE Scholar. It has been accepted for inclusion in Browse all Theses and Dissertations by an authorized administrator of CORE Scholar. For more information, please contact corescholar@www.libraries.wright.edu.
PROGNOSIS:
A WEARABLE SYSTEM FOR HEALTH MONITORING OF PEOPLE AT RISK

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

By

ALEXANDROS PANTELOPOULOS
MS., Department of Electrical and Computer Engineering, University of Patras, 2007

2010
Wright State University
COPYRIGHT BY
ALEXANDROS PANTELOPOULOS
2010
I HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER MY SUPERVISION BY Alexandros Pantelopoulos, ENTITLED Prognosis: A Wearable System for Health Monitoring of People at Risk BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Doctor of Philosophy

Nikolaos G. Bourbakis, Ph.D.
Dissertation Director

Arthur Goshtasby, Ph.D.
Director, Computer Science & Engineering
Ph.D. Program

Joseph F. Thomas Jr., Ph.D
Dean, School of Graduate Studies

Committee on Final Examination

Nikolaos G. Bourbakis, Ph.D.

Soon Chung, Ph.D.

Yong Pei, Ph.D.

Arnab K. Shaw, Ph.D.

Larry W. Lawhorne, M.D.
ABSTRACT

Pantelopoulos, Alexandros, Ph.D, Department of Computer Science & Engineering, Wright State University, 2010.
Prognosis: A Wearable System for Health Monitoring of People at Risk.

Wearable Health Monitoring Systems (WHMS) have drawn a lot of attention from the research community and the industry during the last decade. The development of such systems has been motivated mainly by increasing healthcare costs and by the fact that the world population is ageing. In addition to that, R&D in WHMS has been propelled by recent technological advances in miniature bio-sensing devices, smart textiles, microelectronics and wireless communications techniques.

These portable health systems can comprise various types of small physiological sensors, which enable continuous monitoring of a variety of human vital signs and other physiological parameters such as heart rate, respiration rate, body temperature, blood pressure, perspiration, oxygen saturation, electrocardiogram (ECG), body posture and activity etc. As a result, and also due to their embedded transmission modules and processing capabilities, wearable health monitoring systems can constitute low-cost and unobtrusive solutions for ubiquitous health, mental and activity status monitoring.

The majority of the currently developed WHMS research prototypes and products provide the basic functionality of continuously logging and transmitting physiological data. However, WHMS have the potential of achieving early detection and diagnosis of critical health changes that could enable prevention of health hazardous episodes. To do that, they should be able to learn individual user baselines and also

iv
employ advanced information processing algorithms and diagnostics in order to discover problems autonomously and detect alarming health trends, and consequently, inform medical professionals for further assistance.

In an effort to advance the capabilities of a wearable system towards these goals, we focus in this dissertation on the development of a novel WHMS, called Prognosis. The developed prototype platform includes the following innovative features, which constitute the main research contributions of this work: a) a novel and highly accurate methodology for classifying ECG recordings on a resource constrained device which is based on the Matching Pursuits algorithm and a Neural Network, b) a physiological data fusion scheme based on a fuzzy regular formal language model, whereby the current state of the corresponding fuzzy Finite State Machine signifies the current health state and context of the patient, c) the extension of the decision making methodology based on a modified Fuzzy Petri Net (FPN) model, d) the integration of a user-learning strategy based on a neural-fuzzy extension of the FPN, e) the incorporation of a system-patient dialogue interaction in order to capture non-measurable patient symptoms such as chest pain, dizziness, malaise etc and finally f) the prototyping of the system based on a smart-phone that runs multi-threaded J2ME software for handling multiple simultaneous Bluetooth connections with off-the-shelf wireless biosensors.
## TABLE OF CONTENTS

### CHAPTER 1: Introduction

1.1 Health demographics and the need for advanced IT-based health management solutions .......................................................... 1
1.2 Health monitoring devices and wearable systems ................................................. 2
1.3 The need for WHMS with embedded decision support ..................................... 6

### CHAPTER 2: A Survey on Wearable Health Monitoring Systems ................. 10

2.1 Physiological Signals and Biosensors .......................................................... 10
2.2 Wireless Communication Standards for WHMS ............................................ 11
2.3 R&D in Wearable Health Monitoring Systems .............................................. 14
2.3.1 Research Prototypes .................................................................................. 14
2.3.1.1 Systems based on a microcontroller board or on custom designed platforms .......................................................................... 14
2.3.1.2 Systems based on smart textiles .......................................................... 19
2.3.1.3 Mote-based Body Area Networks ......................................................... 23
2.3.1.4 WHMS based on commercial Bluetooth sensors and cell-phones ........ 28
2.3.1.5 Other types of WHMS .......................................................................... 31
2.3.2 Commercially available WHMS ............................................................... 33
2.4 Maturity Evaluation ....................................................................................... 35
2.5 Discussion ....................................................................................................... 40

### CHAPTER 3: ECG Signal Processing to Meet the Constraints of the WHMS .... 43

3.1 Wavelets ......................................................................................................... 44
3.1.1 The continuous wavelet transform ............................................................. 44
3.1.2 The discrete wavelet transform ................................................................ 47
3.1.3 Wavelet Packets ....................................................................................... 51
3.1.4 Matching Pursuits .................................................................................... 52
3.1.5 Denoising with Wavelets ......................................................................... 54
3.2 Dealing with Noise in the ECG ..................................................................... 56
3.2.1 The Electrocardiogram (ECG) ................................................................. 56
3.2.2 Detecting noise-corrupted segments ....................................................... 59
CHAPTER 4: The Functional Design Framework of the Prognosis Wearable Health Monitoring System ............................................. 94

4.1 Generic WHMS Architecture .................................................. 94
4.2 The Prognosis Formal Language ............................................. 101
  4.2.1 Category of value-specific physiological symptoms ............ 103
  4.2.2 Category of morphology-specific physiological symptoms – ECG 107
  4.2.3 Category of non-measurable symptoms ............................ 108
  4.2.4 The Prognosis Formal Grammar ......................................... 110
    4.2.4.1 Basic Definitions .................................................. 110
    4.2.4.2 Some notes on Medical Expert Systems and Medical
      Knowledge Representation ............................................. 113
    4.2.4.3 Production Rules and the Language’s Operation ............ 121
4.3 SPN-based Operation modeling ............................................. 131
  4.3.1 SPN Model - Level 1 .................................................... 132
  4.3.2 SPN Model - Level 2 .................................................... 134
  4.3.3 SPN Model - Level 3 .................................................... 137

CHAPTER 5: Modeling the Prognosis Language using Modified Fuzzy
Petri Nets ................................................................................. 140

5.1 Fuzzy Petri Nets .................................................................... 142
5.2 A Modified Fuzzy Petri Net (MFPN) Model .............................. 149
5.3 Knowledge-base representation as fuzzy IF-THEN rules .......... 155
5.4 Reasoning using the knowledge-base and the MFPN ............... 157

CHAPTER 6: A Patient or Population-Learning Model ...................... 161

6.1 The concept of the health-history database ............................. 162
6.2 Adapting the WHMS’ decisional capability to the user .......... 164
6.2.1 Neural-Fuzzy Systems ................................................................. 167
6.2.2 Adaptive parameter estimation for user-learning ......................... 173

CHAPTER 7: User – Device Interaction and System Simulation .......... 179

7.1 Prototyping the Human-Device Interaction .................................... 180
7.2 An example of HDI via voice ......................................................... 184
7.3 Simulation Framework .................................................................. 188

CHAPTER 8: System Implementation .................................................. 196

8.1 The WHMS’ Hardware Components ............................................. 197
8.2 Implementation Walkthrough ......................................................... 201

CHAPTER 9: Conclusions and Research Contributions ..................... 207

9.1 A wearable system prototype for health monitoring ...................... 207
9.2 Research contributions .................................................................. 209
9.3 Future Research .......................................................................... 212

REFERENCES .................................................................................. 215
LIST OF FIGURES

Fig. 1-1. The health monitoring network structure..................................................4
Fig. 2-1. Maturity scores for the systems listed in Table 2-3 (1=low, 2=medium-low,
3=good, 4=high, 5=maximum).............................................................................40
Fig. 3-1. Time-frequency boxes in case of a) a time domain signal, b) a frequency
domain signal, c) STFT and d) a wavelet basis.........................................................45
Fig. 3-2. Fast Wavelet Transform (Mallat’s algorithm) implemented with cascaded filters
and downsamplers....................................................................................................50
Fig. 3-3. Wavelet Packet Tree decomposition ................................................................51
Fig. 3-4. An “ideal” ECG signal..................................................................................57
Fig. 3-5. Example ECG recording with evident baseline drift.....................................59
Fig. 3-6. An ECG segment, part of which is severely corrupted by noise caused by
movement..................................................................................................................60
Fig. 3-7. 4 levels of approximation (using Coif5 wavelet) coefficients of the ECG shown
in Fig. 3-1..................................................................................................................61
Fig. 3-8. An ECG segment, part of which is severely corrupted by noise caused by
sensor displacement.................................................................................................63
Fig. 3-9. Detection of corrupted ECG segments with appropriate thresholding on the
energy estimates.........................................................................................................64
Fig. 3-10. Original ECG signal and estimated noise masks........................................65
Fig. 3-11. Denoising performance of different approaches on ECG segment taken from
the MIT-BIH Arrhythmia Database, Record 103 with AWGN: a) Original ECG, b) Noisy
ECG (SNR=4dB), c) Haar UWT denoising (SNR=15.08dB), d) Haar UWPT denoising
(SNR=15.11dB), e) Coif3 UWT denoising (SNR=14.60dB), f) Coif3 UWPT denoising
(SNR=14.28dB), g) Sym3 DWT denoising (SNR=9.62dB), h) Coif3 DWT denoising
(SNR=11.58dB). (Note: In the figure the UWT is denoted as SWT according to
MATLAB’s terminology).........................................................................................67
Fig. 3-12. Denoising performance of different approaches on ECG segment taken from
the MIT-BIH Arrhythmia Database, Record 208 with AWGN: a) Original ECG, b) Noisy
ECG (SNR=4dB), c) Haar UWT denoising (SNR=14.29dB), d) Haar UWPT denoising
(SNR=14.41dB), e) Coif3 UWT denoising (SNR=14.06dB), f) Coif3 UWPT denoising
(SNR=14.07dB), g) Sym3 DWT denoising (SNR=11.11dB), h) Coif3 DWT denoising
(SNR=11.92dB). (Note: In the figure the UWT is denoted as SWT according to
MATLAB’s terminology).........................................................................................68
Fig. 3-13. Denoising performance of different approaches on ECG segment taken from
the MIT-BIH Arrhythmia Database, Record 103 with additive pink noise: a) Original
ECG, b) Noisy ECG (SNR=5dB), c) Haar UWT denoising (SNR=6.73dB), d) Haar
UWPT denoising (SNR=6.79dB), e) Coif3 UWT denoising (SNR=6.48dB), f) Coif3
UWPT denoising (SNR=6.42dB), g) Sym3 DWT denoising (SNR=6.03dB), h) Coif3
DWT denoising (SNR=6.41dB). (Note: In the figure the UWT is denoted as SWT
according to MATLAB’s terminology)....................................................................70
Fig. 3-14. a) Original ECG (record 103), b) Noisy ECG with AWGN, c) Haar UWT
denoised, d) Haar UWPT denoised, e) Sym3 DWT denoised......................................71
Fig. 3-15. SNRout vs SNRin for various noise levels for the signal shown in Fig. 3-14.72
Fig. 3-16. Original ECG (record 103), b) Noisy ECG with pink noise, c) Haar UWT denoised, d) Haar UWPT denoised, e) Sym3 DWT denoised..................................................73
Fig. 3-17. SNRout vs SNRin for various noise levels for the signal shown in Fig. 3-16.73
Fig. 3-18. First three stages of the Hamilton-Tomkins algorithm.................................................75
Fig. 3-19. Last three stages of the Hamilton-Tomkins algorithm..................................................75
Fig. 3-20. Superimposed PVC beats from record 200 of the MIT-BIH Arrhythmia database. ..........................................................................................................................79
Fig. 3-21. Several beat types extracted from records 106, 208, 111, 118 and 217 of the MIT-BIH Arrhythmia database. ..........................................................82
Fig. 3-22. Normalized approximation error (e.g. normalized residual energy as defined in section 3.1.4) vs number of atoms (normal beat)......................83
Fig. 3-23. Normalized approximation error (e.g. normalized residual energy as defined in section 3.1.4) vs number of atoms (PVC beat). ....................................84
Fig. 3-24. Original signal (normal beat pattern taken from MIT-BIH Arrhythmia DB record 123) and MP approximation using 8 coiflet3 wavelet packet atoms...........85
Fig. 3-25. The first six wavelet packet atoms chosen from the MP algorithm in order to approximate the ECG beat depicted in Fig. 3-16........................................86
Fig. 4-1. Generic WHMS architecture.........................................................................................95
Fig. 4-2. Tables describing the extraction of symptoms from body signals by available biosensor technologies and human-system interaction ..................................99
Fig. 4-3. Crisp characterization of hypothetical signal values..................................................104
Fig. 4-4. Fuzzy symptoms extracted from systolic blood pressure. ...........................................105
Fig. 4-5. Fuzzy symptoms extracted from diastolic blood pressure........................................105
Fig. 4-6. Fuzzy symptoms extracted from the heart rate level................................................106
Fig. 4-7. Fuzzy symptoms extracted from the respiration rate level....................................106
Fig. 4-8. Fuzzy symptoms extracted from the body temperature level................................106
Fig. 4-9. Fuzzy symptoms extracted from the oxygen saturation level.................................106
Fig. 4-10. Correspondence between causal relevance level and weight..............................120
Fig. 4-11. Fuzzy FSM for Example 1..................................................................................124
Fig. 4-12. Fuzzy FSM for Example 2..................................................................................126
Fig. 4-13. Fuzzy FSM for Example 3..................................................................................129
Fig. 4-14. Level 1 of the SPN Model: User-Device-Medical Center-Ambulance Interaction .................................................................................................................133
Fig. 4-15. Level 2 of the SPN model: Description of the wearable system’s interaction with the peripherals/wearable sensors .................................................135
Fig. 4-16. Level 3 of the SPN model: Functional description of the central node............137
Fig. 5-1. FPN representation of a WFPR of Type 1..............................................................145
Fig. 5-2. FPN representation of a WFPR of Type 2..............................................................145
Fig. 5-3. FPN representation of a WFPR of Type 3..............................................................146
Fig. 5-4. FPN representation of a WFPR of Type 4..............................................................146
Fig. 5-5. A conjunctive WFPR with five antecedents with different attributes...............154
Fig. 5-6. MFPN for user’s health state estimation...............................................................159
Fig. 5-7. FPN representation of a sub-module of the Prognosis reasoning process.....160
Fig. 6-1. Creating higher-level health knowledge about the user........................................164
Fig. 6-2. An ANFIS type-3 neuro-fuzzy network ..............................................................169
Fig. 6-3. An adaptive neuro-fuzzy classifier (image taken from [148]). ..................170
Fig. 6-4. Bell-shaped membership functions for the fuzzy sets of systolic blood pressure. ..................................................................................................................178
Fig. 6-5. Membership functions using a difference of sigmoidal functions for the fuzzy sets of systolic blood pressure. ..................................................................................................................178
Fig. 7-1. HDI interaction initiated by the system. ...............................................................................................................................187
Fig. 7-2. WHMS Simulator start-up screen ...............................................................................................................................192
Fig. 7-3. The main screen of the WHMS simulator. ...............................................................................................................................194
Fig. 8-1. The Prognosis WHMS prototype ........................................................................................................................................197
Fig. 8-2. J2ME Software Architecture of the Prognosis prototype. ........................................................................................................................................203
Fig. 8-3. a) Real-time data display. b) real-time ECG plot. ...............................................................................................................................205
Fig. 8-4. The GUI of the remote medical server ........................................................................................................................................206
LIST OF TABLES

Table 2-1. Biosensors and Biosignals ................................................................. 11
Table 2-2. Wireless Communication Standards .................................................. 13
Table 2-3. Selected Wearable Health Monitoring Systems ................................... 36
Table 2-4. Evaluation Features ........................................................................... 37
Table 2-5. Features’ Weights ............................................................................... 39
Table 3-1. Number of beats from each class used for training and testing. .......... 82
Table 3-2. Classifier performance in various cases .............................................. 88
Table 3-3. Confusion Matrix for the case of 7 classes ......................................... 89
Table 3-4. Classification accuracy comparison for the five main beat classes
considered in our study ....................................................................................... 91
Table 4-1. Transition Information for Example 1 ............................................... 125
Table 4-2. State Transition Sequence for Example 3 .......................................... 130
Table 8-1. Hardware Specs of Blackberry Bold 9000 ......................................... 199
ACKNOWLEDGEMENT

I would like to express my gratitude to my advisor Dr. Nikolaos Bourbakis for giving me the opportunity to perform research work in a very interesting scientific area. In addition to that, his constant supervision and criticism have helped me become more efficient, responsible and creative. His help and support throughout this whole research will always be appreciated and have definitely helped me become a better professional.

I would also like to thank the members of my doctoral committee Dr. Yong Pei, Dr. Soon Chung, Dr. Arnab Shaw and Dr. Larry Lawhorne. Their suggestions have greatly contributed in this work and I am very thankful to them for doing so.

I would also like to thank my colleagues and friends Dimitris Dakopoulos, Robert Keefer, Thanasis Tsitsoulis and especially Alexandros Karargyris for our fruitful discussions, Mrs Donetta Bantle for administrative help and Dr. Danopoulos for his useful contribution. Finally, I would like to thank my co-student and friend Giovani Abuitah for his support and his technical advice.
Στους γονείς μου,
για την αγάπη τους και τη συνεχή τους υποστήριξη.

Στην Ηρώ,
για τη δύναμη που μου έδωσες.
CHAPTER 1: INTRODUCTION

1.1 Health demographics and the need for advanced IT-based health management solutions

It is a fact that the global population is both growing and ageing [1], [2]. As a consequence of this demographic change there has also been a corresponding increase of chronic age-related diseases, such as congestive heart failure, dementia, sleep apnea, cancer, diabetes and chronic obstructive pulmonary disease [3], [4]. Furthermore, the total number of people suffering from some type of disability (either life-long, or injury related or more commonly related to chronic conditions) will continue to rise [5]. In addition to that, approximately 33% of persons over the age of 65 and 50% of persons over the age of 85 experience a fall each year [6], [7]. For this population, health care costs are increasing [8], quality of life and productivity are declining and in many cases family members serve as primary care assistants.

These issues along with the challenges of effectively managing and treating postoperative rehabilitation patients, disabled people and persons with special abilities, highlight the requirement for new and innovative ways to deliver healthcare to patients. In response to that, information and communication technologies are expected to provide the means to realize personalized, low cost and citizen-centered healthcare solutions to address the pre-
viously stated challenges [8]. Moreover, as people are becoming more eager to participate actively in their own health management [10] and as healthcare providers are looking for ways to facilitate a distributed approach to delivering their services rather than through conventional large centralized institutions, a variety of health information networks is being developed. These networks seek to link hospitals, healthcare professionals, care providers, laboratories and ultimately people’s homes [11] in order to realize a closed-loop health monitoring, management and delivery system. Going even further, enabling real-time physiological, activity and mental status monitoring can in many cases lead to early detection and diagnosis of critical health changes, that are otherwise not guaranteed to be observed during routine short-time check-up visits to doctors, and could thus enable prevention of a variety of health hazards and risks [10], while also saving billions of dollars annually [12].

1.2 Health monitoring devices and wearable systems

Health monitoring devices constitute the front-end components of the health information network and are primary responsible for a) collecting physiological data from the patient/user and b) transmitting them securely and reliably to a remote monitoring location. Traditionally, “crude” and bulky monitoring devices have been employed to collect patient vital signs, for example bedside Holter ECG monitors as the ones used in Intensive Care Units (ICU).
Obviously, devices of such size and cost, that also include several wires and require the patient to be immobilized in order to acquire reliable measurements, are unsuitable when ubiquitous, unobtrusive, long-term and low-cost health monitoring is desired.

As a result, personal health monitoring devices need due to the inherent nature of their application to be of small size and weight and they have to be portable and autonomous as well, e.g. they should ensure long-term operation without supervision and without recharging. Recent advances in microelectronics, nanotechnology, miniaturized bio-sensors, wireless communication techniques, low power computing, battery technologies and effective information processing have provided the means to realize wearable non-invasive systems that can monitor the wearer’s health condition continuously and in real-time [8], [11], [13], [14].

Wearable health monitoring systems (WHMS) may comprise various types of miniature sensors, wearable or even implantable. Bio-sensors are capable of measuring significant physiological parameters like heart rate, blood pressure, body and skin temperature, oxygen saturation, respiration rate, electrocardiogram etc. The acquired measurements are in turn communicated either via a wireless or a wired link to a central node, for example a Personal Digital Assistant (PDA), a smart-phone or even to a microcontroller board, which may then in turn display the corresponding information on a user interface or
transmit the aggregated vital signs to a medical center server. Finally at the health provider’s end, the medical personnel and supervising physicians can have instant access to real-time physiological measurements and the medical history of several in and out-of-hospital monitored patients by securely connecting to the medical center’s servers. This described scenario is graphically illustrated in Fig. 1-1 for the case of one single patient using a WHMS. Fig. 1-1 also introduces the term Body Area Network (BAN), which is widely used when referring to WHMS that employ bio-sensors with wireless communication capabilities.

Fig. 1-1. The health monitoring network structure
Setting aside the required network infrastructure to realize the health service paradigm that is presented in Fig. 1-1 and focusing mainly on the WHMS or BAN side we can observe that a wearable medical system may encompass a wide variety of components: sensors, wearable materials, smart textiles, actuators, power supplies, wireless communication modules and links, control and processing units, interface for the user, software and advanced algorithms for data extracting and decision making. In addition to that, wearable systems for health monitoring need to satisfy certain strict medical criteria while operating under several ergonomic constraints and significant hardware resource limitations [14], [15], [16].

For example, wearable systems must ensure portability and unobtrusiveness to the user, which means that their form and weight factor should be small, their on-body placement must not interfere with the user’s daily activities, the bio-sensors should not irritate the patient’s skin and also there shouldn’t be any unwanted radiation or infection concerns. Technical limitations include hardware resource constraints, most importantly processor speed, memory availability, data storage space and power supply. Furthermore, WHMS are expected to perform robustly in cases of high physiological signal noise and of measurement artifacts due to patient movement and misplacement/bad contact of bio-sensors with the skin, in cases of low connectivi-
ty, low battery etc. Finally, these personal portable systems must ensure patient privacy by providing reliable and secure medical data transmission and access.

The above mentioned parameters of WHMS highlight the fact that designing such a system is a very challenging task since a lot of highly constraining and often conflicting requirements need to be taken into account by the system’s designers. Furthermore it is our point of view that there is no single ideal design for such systems, but rather a trade-off between “antagonizing” parameters that need to be balanced based on the specific area of application [14].

1.3 The need for WHMS with embedded decision support

As described so far, WHMS can serve the purpose of continuously collecting physiological measurements from the user and then storing the sensed data locally or streaming them continuously to a remote location. However this type of service on its own, while it still may provide a useful tool in the hands of physicians and care givers, is not sufficient as a standalone function to address the issues mentioned in section 1.1. To illustrate this point, we can consider the case where a supervising physician logs in at a terminal in his office to manually supervise the real-time readings and the physiological measurement history of several patients, let us suppose ten patients, that he is directly responsible for. The doctor himself will have to tediously go through a large amount of data for each patient to investigate the health status and health trend
of every individual, using up a lot of his time and subsequently costing a lot of money to the patients themselves. Furthermore, what is most important is that if a health hazard is to be identified and addressed in the moment of occurrence, the doctor or the caregiver needs to be manually supervising the collected data of each patient at any given moment. In addition to that, simple physiological readings are unlikely to provide a complete picture of the overall health status and context of any patient.

The previous discussion underlines the fact that wearable health monitoring devices need to be able to provide additional or context-aware information to physicians than just plain raw sensor readings. The measured vital parameters need to be combined with the wearer’s context and his activity level to provide a more comprehensive picture of his/her overall physiological state [8]. For example reporting sensor readings of high pulse rate does not itself provide sufficient information to the care giver regarding the user’s health status. However, high pulse rate combined with indications of high activity or low activity may indicate completely different and more specific health information and may call for different or even no actions from the care giver.

In addition to context-aware sensing capabilities, WHMS need to be able to identify alarming changes in sensor measurements that could possibly indicate a health hazard. In the simplest case, this refers to reporting values that exceed predefined safety thresholds, e.g. systolic blood pressure values
over 140 mmHg or temperature readings over 100°F. However, even in this case, different sensor readings may point to different health conditions according to the user’s context, his medical history and his individual physiology. Furthermore, as the ultimate goal is to perform early identification or even prevention of diseases and health episodes, it is made obvious that advanced logic and intelligence is required in order for the WHMS to be able to identify alarming trends in the health status of the user and also to be able to provide patient-adapted alarms or even diagnoses.

The previous discussion presents the main motivation behind the research presented in this dissertation. Specifically, our goal is to establish an operational framework for multi-sensor WHMS that can support embedded and patient-adaptive decision support, while at the same time being able to meaningfully interact with the patient by using an automated user-device dialogue system.

The organization of the rest of the chapters in this dissertation is as follows: Chapter 2 presents a survey on wearable health monitoring systems, where a wide variety of system implementations and approaches are being reviewed and discussed. In Chapter 3 we concentrate on the issue of ECG signal processing on a resource constrained device and we examine methods for detecting unusable noise-corrupted recordings and denoising the rest of the signal. We will also present a robust, highly accurate and efficient methodology
for ECG beat classification. Chapter 4 will provide a detailed insight into the organizational design and principles, the proposed physiological data fusion strategy and into the operation model of the proposed WHMS framework, which is termed “Prognosis”. In Chapter 5, we will present an alternative representation of the knowledge-base and the decision making process in the WHMS concept by employing a modified Fuzzy Petri Net model. This will be proven useful for the Patient-learning approach proposed in Chapter 6, which will be based on a neuro-fuzzy approach and will employ gradient descent methods in order to compute optimal parameters for decision making. Chapter 7 describes the interaction scheme between the patient and the device and discusses a developed simulation framework for our system. In Chapter 8, the current implementation of the proposed system based on off-the-shelf components (smart-phone and Bluetooth-enabled bio-sensors) is described. Finally, we will conclude with a discussion on the research contributions of this dissertation and also with a look into possible future research that could extend and improve the results of this study.
CHAPTER 2: A SURVEY ON WEARABLE HEALTH MONITORING SYSTEMS

In this chapter a comprehensive survey on the current state-of-the-art in research and development of wearable low-cost unobtrusive systems for health-monitoring is presented. A variety of system implementations either for general purpose or even for application specific health monitoring are reviewed in order to identify current technological achievements, potentials and shortcomings as well.

Before beginning the review on the wide variety of WHMS prototypes or commercial products that have been developed so far, a brief discussion on available biosensors for wearable technologies and on wireless communication techniques employed in wearable systems is provided.

2.1 Physiological Signals and Biosensors

In this subsection a list of several sensing technologies is provided, which can be integrated as part of a wearable health-monitoring system, along with their corresponding measured physiological signals. The measurement of these vital bio-signals and their subsequent processing for feature extraction, leads to a collection of real-time gathered physiological parameters, which can
give an overall estimation of the user’s health condition at any given time. This list can be seen in Table 2-1.

Table 2-1. Biosensors and Biosignals

<table>
<thead>
<tr>
<th>Type of Bio-signal</th>
<th>Type of Sensor</th>
<th>Description of measured data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electrocardiogram (ECG)</td>
<td>Skin/Chest electrodes</td>
<td>Electrical activity of the heart (continuous waveform showing the contraction and relaxation phases of the cardiac cycles)</td>
</tr>
<tr>
<td>Blood pressure (systolic &amp; diastolic)</td>
<td>Arm cuff-based monitor</td>
<td>Refers to the force exerted by circulating blood on the walls of blood vessels, especially the arteries</td>
</tr>
<tr>
<td>Body and/or skin temperature</td>
<td>Temperature probe or skin patch</td>
<td>A measure of the body's ability to generate and get rid of heat</td>
</tr>
<tr>
<td>Respiration rate</td>
<td>Piezoelectric/piezoresistive sensor</td>
<td>Number of movements indicative of inspiration and expiration per unit time (breathing rate)</td>
</tr>
<tr>
<td>Oxygen saturation</td>
<td>Pulse Oximeter</td>
<td>Indicates the oxygenation or the amount of oxygen that is being “carried” in a patient’s blood</td>
</tr>
<tr>
<td>Heart rate</td>
<td>Pulse Oximeter/skin electrodes</td>
<td>Frequency of the cardiac cycle</td>
</tr>
<tr>
<td>Perspiration (sweating) or skin conductivity</td>
<td>Galvanic Skin Response</td>
<td>Electrical conductance of the skin is associated with the activity of the sweat glands</td>
</tr>
<tr>
<td>Heart sounds</td>
<td>Phonocardiograph</td>
<td>A record of heart sounds, produced by a properly placed on the chest microphone (stethoscope)</td>
</tr>
<tr>
<td>Blood glucose</td>
<td>Strip-base glucose meters</td>
<td>Measurement of the amount of glucose (main type/source of sugar/energy) in blood</td>
</tr>
<tr>
<td>Electromyogram (EMG)</td>
<td>Skin electrodes</td>
<td>Electrical activity of the skeletal muscles (characterizes the neuromuscular system)</td>
</tr>
<tr>
<td>Electroencephalogram (EEG)</td>
<td>Scalp-placed electrodes</td>
<td>Measurement of electrical spontaneous brain activity and other brain potentials</td>
</tr>
<tr>
<td>Body Movements</td>
<td>Accelerometer</td>
<td>Measurement of acceleration forces in the 3D space</td>
</tr>
</tbody>
</table>

2.2 Wireless Communication Standards for WHMS

Transmission of measured data in the overall context of wearable health-monitoring systems needs to be performed for two different purposes:
a) for communicating the collected physiological signals from the bio-sensors
to the system’s central node and b) for sending the aggregated measurements from the wearable system to a remote medical station or to a physician’s cell phone.

Type a) transmission of data or else short-range transmission can be handled either by wires or by multiple wireless links. In the former case, the user’s mobility and comfortableness can be severely hindered by the use of wires and moreover there is an increased risk of system failure [17]. A more favorable approach to this matter is the use of conductive yarns to transmit the measurements collected from sensors integrated on some type of flexible smart-textile clothing [18]. Alternatively, in the latter case, autonomous sensor nodes can form a Body Area Network (BAN) or Body Sensor Network (BSN), usually in the basic configuration of a star topology, transmitting the data to the BAN’s central node, which can be a Personal Digital Assistant (PDA), a smart-phone, a pocket PC or a custom designed microcontroller-based device.

The most commonly employed wireless communication standards in BAN’s are IEEE 802.15.1 (Bluetooth) [19] and 802.15.4 (widely referred to as Zigbee [20], although Zigbee includes the specification of network, security and application layers on top of the official standard). Other technologies for short-range intra-BAN communication include Infrared (IrDA), the Medical Implant Communication Service (MICS) [21] and Ultra-Wideband (UWB) [1]. Table 2-2 summarizes some of the basic technical specifications of these technologies.
Existing standards fail to address the requirements of BANs, either due to security concerns [22] or interference problems [23] or due to form factor of hardware modules or power consumption. As a response to these issues, the 802.15.6 IEEE Task Group [24] is planning the development of a communication standard optimized for low power devices and operation on, in or around the human body. Furthermore the Bluetooth Special Interest Group (SIG) announced recently the addition of an alternative protocol stack, e.g. the Bluetooth low energy, an ultra-low power technology for devices with limited battery capacity allowing consumption of only a small fraction of the power of the original Bluetooth products, which is targeting sports & wellness, and healthcare devices.

Table 2-2. Wireless Communication Standards

<table>
<thead>
<tr>
<th>Standard</th>
<th>Range (typical)</th>
<th>Data Rate (max)</th>
<th>Power Cons.*</th>
<th>Cost per chip</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zigbee</td>
<td>10-75m</td>
<td>20kbps/40kbps/250kbps</td>
<td>30mW</td>
<td>$2</td>
<td>868MHz/ 915MHz/ 2.4GHz</td>
</tr>
<tr>
<td>Bluetooth</td>
<td>10-100m</td>
<td>1-3Mbps</td>
<td>2.5-100mW</td>
<td>$3</td>
<td>2.4GHz</td>
</tr>
<tr>
<td>IrDA</td>
<td>1m</td>
<td>16Mbps</td>
<td>-</td>
<td>$2</td>
<td>Infrared</td>
</tr>
<tr>
<td>MICS</td>
<td>2m</td>
<td>500kbps</td>
<td>25µW</td>
<td>-</td>
<td>402-405 MHz</td>
</tr>
<tr>
<td>802.11g</td>
<td>200m</td>
<td>54Mbps</td>
<td>1W</td>
<td>$9</td>
<td>2.4GHz</td>
</tr>
</tbody>
</table>

*Power consumption refers to maximum consumed power when the chip is on and it is sending/receiving.

Finally, regarding type b) data transmission or else long-range communication between the WHMS and a remote station or device, there is a wide variety of available wireless technologies that can serve that goal. Such technolo-
gies include WLAN, GSM, GPRS, UMTS and WiMAX, which can offer wide coverage and ubiquitous network access. Furthermore, future advances in 4G (fourth generation) mobile communication systems are expected to guarantee worldwide seamless access to the Internet at much higher data-rates [25] and thus to facilitate more efficiently the need for gathering real-time measurements from a wearable health-monitoring system at a remote location.

2.3 R&D in Wearable Health Monitoring Systems

In this section, several types of wearable health-monitoring systems are discussed. Since during the last ten years there have been numerous research efforts and products that can be classified as WHMS, in this review we attempt to categorize them a) based on whether they are commercial products or research prototypes and b) based on their hardware configuration, e.g. BAN-based, smart textile-based, microcontroller or custom hardware-based etc. Research efforts will be examined first, followed by a short review on commercially available systems.

2.3.1 Research Prototypes

2.3.1.1 Systems based on a microcontroller board or on custom designed platforms

Wearable systems that fall in this category are those that use some type of microcontroller board as a physiological data-collecting platform and that
are usually based on wired transmission of bio-signals from the sensors to the processing board.

The Media Laboratory of MIT, Cambridge developed LiveNet, a flexible distributed mobile platform aiming at long-term health monitoring applications with real-time data processing and streaming and context classification [26]. LiveNet used a Linux-based PDA mobile device, a modular sensor hub (SAK2) for gathering, processing and interpreting real-time contextual data and an integrated physiological board (BioSense), which incorporates a 3D accelerometer, ECG, EMG and galvanic skin conductance sensors and which allows interfacing with a wide range of commercially available sensors. Furthermore, a 3-layer software architecture has been developed, which supports communication between group-based applications, efficient distribution and processing of higher bandwidth digital signals and the implementation of real-time context classifiers for wearable applications. Finally the MIT Wearable Computing Group, in collaboration with several healthcare providers, has initiated various pilot studies using the LiveNet system, which include soldiers’ health monitoring in harsh environments, automated Parkinson symptom detection system, Epilepsy seizure detection and long-term behavioral modeling. Overall, LiveNet targets real-time feature extraction and classification of medical conditions, as well as closed-loop medical feedback systems.
AMON or the advanced care and alert portable telemedical monitor was a project financed by the EU FP5 IST program [27]. It resulted in the development of a wrist-worn device, which is capable of measuring blood pressure, skin temperature, blood oxygen saturation and a one lead ECG. In addition to that it incorporated a two-axis accelerometer for correlating user activity with the measured vital signs. The researchers designed also the GSM based secure cellular communication link, as well as the software package for the telemedicine center, where the physicians could analyze the received data from the wrist-worn device in greater detail. The AMON prototype was innovative in that it included many miniaturized sensors all integrated through nonstandard placement in a wrist-worn part and in that it also included evaluation software for real-time processing and analysis of the measured vital signs. The development of the AMON wearable health monitoring device aimed at high-risk cardiac/respiratory patients who would be confined to the hospital or their homes.

After the processing of the raw data, a classification of normal, deviant, risk, high risk or error was derived as an estimation of the patient’s health condition using specific limit values for each vital sign, taken from the World Health Organization. According to the estimated condition, appropriate actions were made from the device, like initialization of additional measurements for validation/re-estimation, risk alarms, transmission of data to the medicine cen-
Although the conducted validation study revealed problems regarding the reliability of the measured data and the wearability of the device, the general idea of the device was very well received by the users.

Lin et al. [28] describe the development of a real-time wireless physiological monitoring system (RTWPMS), which is based on a digital, low-power 2nd generation cordless phone and a custom made medical examination module measuring blood pressure, heart rate and temperature. A GPS module is also incorporated and sensor interfacing is done through serial ports. System architecture includes also a wireless base station (handles physical transmission/reception of messages/ commands), a voice/data exchange device (processes data and commands) and a network management center for overall system control. Conducted tests with system in nursing centers and hospitals validated its ability to transfer voice and data end-to-end at low error rates from multiple sources simultaneously. However the wearable examination system is too bulky for ambulatory and continuous monitoring and the utilized RF technology is out-dated as identified by the authors as well.

LifeGuard [29] is a multiparameter wearable physiological monitoring system for space and terrestrial applications, whose core element is a crew physiologic observation device (CPOD), which is capable of measuring two ECG leads, respiration rate via impedance plethysmography, heart rate, oxygen saturation, body temperature, blood pressure and body movement. Typi-
cal off-the-shelf sensors are used for measuring most of the bio-signals, which are interfaced through wired connections to CPOD data logger that can either send the data via Bluetooth to a base-station or record them for 9h continuously on a memory card. The data logger is based on a PIC μC and uses 2 AAA batteries. The authors conducted a series of verification and validation tests at extreme environments and tested the ability of satellite transmission of collected data with the obtained results indicating acceptable accuracy for the collected data and real-time transmission of measurements to remote locations.

A prototype portable system capable of measuring phonocardiography (PCG), electrocardiography and body temperature is described in [30]. The developed system consists of a capacitor-type microphone inserted on a stethoscope for PCG detection, a 3-lead ECG, a temperature sensor, a measuring board including a CPU, a Bluetooth transceiver and an A/D module and a PDA with an external memory unit. The functionality of the whole system is controlled by the PDA user by issuing commands to the measuring circuit. The PDA’s insufficient computational power is said to enforce certain function limitations to the system in terms of PCG and ECG sampling rates. In addition to that, the system consists of too many large sized external components and it is too dependable on the end user (measurements are initiated from the PDA, not automatically or event-driven), making it impractical and uncomfortable for long-term unobtrusive monitoring of a patient’s vital signs.
Finally, the development of a medical wearable device targeting brain-injured children is described in [31]. The medical device is able to measure blood oxygen saturation and heart rate through a finger-placed pulse oximeter, respiration rate by using a piezoelectric sensor placed on a belt worn on the chest and body movement by using a dual axis thermal accelerometer. The measured data are stored into a multimedia card, which are in turn transmitted at prescribed time intervals through Bluetooth to a home PC. From the home PC the data are finally transmitted through an ADSL Internet connection (or UMTS) to the Medical Service Center. The implemented system aims at detecting events such as nocturnal apnea, but has the disadvantage of low wearability due to the wired connections between the individual sensors and the wearable Bluetooth module, while no validation tests have been performed as in the case of the previously discussed system.

2.3.1.2 **Systems based on smart textiles**

Systems in this category are based on a vest or T-shirt with bio-sensors integrated on the garment.

The MyHeart project supported by the European Commission and involving 33 partners from 10 different countries, including industrial partners such as Philips, Nokia, Vodafone and Medtronic, aimed at fighting cardiovascular diseases (CVD) by prevention and early diagnosis [32]. It adopted the use of smart clothes, where the sensing modules are either garment-integrated or
simply embedded on the piece of clothing [33], [34]. The underlying concept relies on the use of tiny conductive wires knitted like normal textile yarns. In that way the wearable system is very comfortable for the user, no wireless modules are needed for the sensors and the whole system needs only one centralized on-body power supply, thus resulting in significant decrease of the overall system’s size. One main device is used to control the on-garment bus and is also responsible for the synchronization and the power supply for all the on-body components. The developed textile-sensors include an ECG and an activity sensor. An algorithm capable of classifying the activity into resting, lying, walking, running and going up/downstairs has been implemented, with very high accuracy. Finally the MyHeart project developed heart belts as well, that could be worn across the chest or attached to a standard bra or to the waistband of standard underwear [35].

The WEALTHY (Wearable Health Care System) project, part of the fifth framework program of the European Commission and completed in 2005, has developed a wearable garment, covering the whole upper body and worn under normal clothing, capable of recording biomechanical variables and physiological signals [36]. The WEALTHY system targets clinical patients during rehabilitation and other high-risk patients, such as elderly people, individuals with chronic diseases and others. The sensor elements, which have been integrated in fabric form (using conducting and piezoresistive materials) on a tex-
tile structure [37], [38], are able to monitor a 3-lead ECG, EMG placed on the arms, thoracic and abdominal respiration rate, body position and movement, skin and core temperature. On demand measurements of blood pressure and oxygen saturation can also be obtained. The wearable garment incorporates also an analog and digital signal processing module with GPRS or Bluetooth wireless transmission capabilities. Algorithms to remove artifacts introduced in the measured signals by motion have also been implemented, along with the ability to generate alert messages and synoptic patient tables.

MagIC [39], developed by researchers in Milan, Italy, is a washable sensorized vest including fully woven textile sensors for ECG and respiration rate monitoring and a portable electronic board, which evaluates the wearer’s motion level and is responsible for signal preprocessing and data transmission through Bluetooth to a local PC or PDA. The wearable system incorporates also skin temperature sensors and it mostly aims at use from elderly people or cardiac patients for home monitoring, enabling however also ambulatory daily life health monitoring. The data collected from the evaluation tests performed, showed that the system achieves very good acquired signal quality (except in the case of maximal physical activity) and that it is also able to correctly identify atrial fibrillation episodes and atrial and ventricular ectopic beats.

The MERMOTH (Medical Remote Monitoring of clothes) project, completed in 2006 and is also a European IST FP6 program and part of a wider
group of six other European Projects in Smart Fabrics and Interactive Textiles [40], has produced a low-cost, knitted, comfortable and stretchable sensing garment [41]. The developed garment incorporates conductive and electrostrictive fabrics and yarns and dry electrodes, enabling the measurement of ECG, respiratory inductance plethysmography, skin temperature and activity through accelerometers. A PDA is connected to the microcontroller that is used to interface with the sensors on the garment, providing an RF link to a local PC for display and measurement interpretation.

Finally, researchers in [42] describe Smart Vest, a wearable physiological monitoring system that consists of a vest, which uses a variety of sensors integrated on the garment’s fabric to simultaneously collect several bio-signals in a non-invasive and unobtrusive manner. The parameters measured are ECG, PPG, heart rate, blood pressure, body temperature and GSR. Furthermore it is stated that the ECG can be recorded without the use of gel and that its recording is free of baseline noise and motion artifacts due to hardware-implemented high-pass, low-pass and notch filters. Moreover, blood pressure is calculated non-invasively via PPT, where the implemented detection algorithm is individually calibrated based on the user’s ECG. Provided results from validation tests confirm to the most part the author’s claims regarding the accuracy of the measured physiological parameters. Furthermore, the sensors are
connected to a central processing unit, which is capable of correlating the acquired measurements to derive an overall picture of the wearer’s health.

2.3.1.3 Mote-based Body Area Networks

In this sub-section we take a look at WHMS that employ motes, e.g. wirelessly enabled tiny nodes, to form a Body Area Network, where every mote is responsible for collecting one or more types of physiological data and transmitting them to a central node or base station. For a more detailed review of BANs the reader should refer to [1].

In [43], the development of a prototyped wearable wireless body area network (WWBAN) is described. The proposed system consists of common off-the-self wireless sensor platforms using ZigBee (IEEE 802.15.4) compliant transceivers and ultra low-power microcontrollers. Custom sensor platforms, that perform data collection and preprocessing and that are equipped with accelerometers or bioamplifiers for ECG or EMG, have been developed and have been integrated on the commercially available Telos platforms from Moteiv, which use the event-driven TinyOS operating system. The lack of standardization in platforms, system software support and wireless communication for WBANs is underlined and discussed throughout this study. Furthermore, as a means of addressing the issue of limited power resources on the motes and extending their operational lifetime, the authors propose to perform on-board advanced signal processing of the acquired bio-signals to reduce the wireless
transmission duty cycle and thus the power consumption as well. However limited computing resources, real-time signal processing requirements and limited memory space make this task even harder.

Researchers in Harvard University have developed CodeBlue [44], a medical sensor network platform for multi-patient monitoring environments based on the ZigBee compliant MicaZ and Telos motes, including custom designed biosensor boards for pulse oximetry, 3-lead ECG, EMG and motion activity. The Codeblue project also addressed issues of reliable communication between medical sensors, multiple receivers (PDA’s carried by doctors and nurses) and various high data rates. In that direction, a software framework has been implemented, which provides protocols for device discovery, publish/subscribe multihop routing and a simple query interface allowing end users to dynamically request specific data from a specified network node. Besides that, CodeBlue uses an RF-based localization system, to track the location of patients and caregivers. The system was evaluated in a 30-node testbed in terms of network metrics such as packet loss, fairness across multiple paths etc, indicating that further work on reliable communication, bandwidth limitation issues and security is needed.

A BAN based on the IEEE 802.15.4 protocol is presented in [45]. The BAN follows a star topology and it is formed of two main types of devices: a) Sensor communication modules (SCM), which are able to interface with both
analog and digital sensors and b) a personal data processing unit (PDPU), which is in charge of coordinating the BAN, controlling the communication with all the SCMs as well as the communication with external networks (via USB, Wi-Fi or GPRS). Synchronization in the BAN is handled via the protocol-supported beacon primitives, which are also used to carry commands about sensor configuration parameters (sampling rate, gain etc), sensor activation/deactivation and data transmission. The SCMs are 802.15.4 compliant prototypes, designed as a four-layer PCB and the PDPU (powered by a Lithium-ion rechargeable battery) is also a custom design, which includes an ARM Thumb Processor, a GPRS modem, a Wi-Fi module, a Security Digital memory card, a simple two line LCD display and a five-button based joystick. The firmware of the PDPU is an embedded Linux OS. Employed biosensors include a sensor recording arterial blood pressure, micro-electrodes for ECG measurement, position sensors and respiration sensors.

Chung et al [46] present a custom developed u-healthcare system, which consists of custom 802.15.4 capable nodes that interface with ECG and blood pressure sensors as well as with a basic cell-phone device for data display and signal feature extraction. The novelty of the projects consists in transmitting only identified suspicious ECG and BP patterns to the hospital’s server. This is done by first extracting simple ECG features (such as QRS duration, RR interval, R magnitude etc) and making a decision based on simple if-
then-else rules. In the same manner, blood pressure measurements extracted from a wrist-sensor are transmitted as well in case they are found to be out of range. In [47], Chung et al. present more details regarding the hardware of the ubiquitous sensor nodes (USN), which are also extended to include accelerometers and SpO2 sensors.

Other research projects that utilize Zigbee motes for BANs include: a) the BAN described in [48], where researchers have included ECG, PPG and PCG sensors and have implemented a synchronous sampling mechanism for data synchronization, b) WiMoCA [49], developed by researchers in Italy. A star topology network is formed by sensor nodes with a palmtop PC functioning as a base station and which aims at posture recognition, gait pattern recognition and balance detection applications, c) Bi-FI [50], an embedded sensor/system architecture for wireless biosignal recording, employing ECG, EEG and SpO2 sensors, where the main goal is to perform on-board DSP to remove the bandwidth bottleneck imposed by the transceiver, d) BASUMA [51], which uses sensors equipped with the ZigBee compliant Philips AquisGrain platform. It aims at long-term monitoring of chronically ill patients and uses several custom non-invasive sensors for monitoring parameters such as ECG, air and blood content of the thorax, body temperature, breathing rate and cough control, blood pressure, pulse rate and oxygen saturation, e) a body sensor network earpiece [52], which includes a pulse oximeter and a 3D accelerometer
and which aims at monitoring postoperative recovery at home in patients undergoing abdominal surgery. The proposed device is based on the BSN node, developed from the Department of Computing in Imperial College and it includes intelligent algorithm for recognizing the type and the intensity of the user’s activity.

Besides Zigbee, other technologies have been utilized for BANs. For instance, in [53] the authors propose the use of intra-body communication for exchanging information between wearable electronic sensors within a BAN. To validate that claim, a prototype transmission system is constructed using aluminum electrodes powered by a 3V DC source and operating in the 10.7MHz frequency modulation (FM) band. The communication module was interfaced with a finger-tip pulse oximeter measuring heart rate and oxygen saturation in blood and transmitted the sensed data over a distance of 30cm. The illustrated results indicate acceptable communication quality, but probably insufficient reliability for types of data requiring high accuracy and low distortion.

Yuce et al. [54] presented a recently developed wireless body sensor network hardware, which uses the recently allocated MICS (Medical Implant Communication Service) band. The system prototype consists of a pulse rate and a temperature sensor, a Central Control Unit (CCU) and a receiver station at a medical center. The developed system has overall very small size and cost and can achieve very low power consumption. However, the latter is accom-
plished by keeping the sensor nodes in sleep mode most of the time (one measurement every 5 or 10 min) and by using a polling approach to acquire the data.

Finally, Human++, a research project in the Netherlands, has developed a body area network consisting of three sensor nodes and a base station [55]. Each sensor node is responsible for acquiring, amplifying, filtering, processing and wirelessly transmitting multi-channel signals of ECG, EEG and EMG, while the base station functions as a data collector in star topology, regulating the information flow. The developed system is able to run autonomously for 3 months on two AA batteries. In addition to that, Belgium research centre IMEC has developed a very promising prototype of a body-heat powered autonomous pulse oximeter [56], powered completely through a wrist-worn, watch-size, thermoelectric generator which transforms the wearer’s body heat into electrical energy and uses, instead of a battery module, a super-capacitor for short-time energy storage.

2.3.1.4 WHMS based on commercial Bluetooth sensors and cell-phones

The first example in this category is HealthGear from Microsoft [57]. The prototype released so far consists of a non-invasive blood oximeter, a sensing board providing the sampled oxygen saturation and heart rate signals, a Bluetooth module for wireless transmission of the measured signals, AAA battery power supply and a cell phone to provide the user interface. This current
exemplary application aimed at monitoring users during their sleep to detect sleep apnea events. Two methods for the automated detection of sleep apnea events are proposed. The first one works in the time domain and detects apnea events after statistically evaluating threshold values for the oxygen saturation level, while the second one works in the frequency domain trying to detect peaks in a filtered periodogram of the oximetry signal.

The system was evaluated by twenty individuals and according to the results, there were no technical problems and the system was completely successful in detecting mild and severe OSA. The users reported also that the wearability and the functionality of the whole system were good. However the measurements taken were not compared with a polysomnograph for validating the results.

HeartToGo [58] is a cell-phone-based wearable platform, capable of continuously monitoring the user’s ECG signal via a wireless ECG sensor, of analyzing the electrocardiogram in real time and possibly detecting any abnormal patterns pertaining to cardiovascular disease. The proposed system is able to adapt to the individual user’s physiological conditions through the use of artificial neural network – based machine learning schemes, which can possibly result in more accurate classification of ECG patterns.

Leijdekkers et al. [59] has developed a heart attack self-test application implemented on a Personal Health Monitor system which includes a conven-
tional mobile phone and a Bluetooth enabled ECG sensor. The mobile phone serves the function of analyzing the streaming data from the sensor in real-time and transmitting them to a heart specialist. In the illustrated application, a simple user interface on the phone is used to acquire feedback from the user about his symptoms and in case the patient’s health condition is found to be at risk, based on his answers, the emergency services will be contacted.

Another Bluetooth PAN approach is described in [60], where researchers have employed a smart-phone to collect physiological data from Bluetooth-enabled sensors (a pulse oximeter and a GPS module), to detect alarming conditions and to transmit the encrypted data via GPRS or Wi-Fi. Similarly, the Mobile Care System with Alert Mechanism presented in [61] utilizes a Bluetooth blood pressure monitor and ECG sensor and a mobile phone as the processing core for symptom recognition and alert message generation based on different urgency levels.

Finally, the work described in [62] presents a wearable device for continuous ECG monitoring, where the ECG sensor continuously transmits the measured and amplified ECG signal to the Hand Held Device (HHD), which is a common PDA. The PDA serves as the “intelligent” device of the system, in that it processes, analyzes and saves the recorded ECG measurements. An algorithm for detecting arrhythmia events, is implemented on the HHD, which has also a documented true detection rate of 99.2%. The HHD communicates
via GPRS with a remote clinical station, transmitting alarm signals along with the recorded ECG waveforms. The doctor at the station end can setup the limits for the specified alarm detections, such as bradycardia, tachycardia and arrhythmia. When an abnormal ECG activity is detected, the HHD will record 1 min of the ECG signal and transmit it to the base station and it will also calculate heart rate and some other parameters based on the measured waveform.

2.3.1.5 Other types of WHMS

In this last sub-category of research prototypes we will take a look at some interesting wearable systems that could not be directly classified to one of the previous categories.

Such a system is AUBADE [63], developed from the University of Ioannina in Greece, a wearable system that performs evaluation of the emotional state of an individual targeting environments where human subjects operate at extreme stress conditions. The developed prototype consists of a mask containing sixteen EMG textile fireproof sensors, a three-lead ECG and respiration rate sensors located on the chest of the subject and a textile sensor measuring electrodermal activity (or galvanic skin response or skin conductance activity) placed inside a glove. A 3D facial representation mechanism and an intelligent emotion recognition module have also been implemented classifying the individual’s psychological condition from a set of emotions.
Another interesting effort is described in [64], where the authors describe a system that uses a wearable ECG device to detect the motion artifacts (distortions) and classify the type of Body Movement Activity (BMA) from the ECG signal itself. This novel study targets situations where dynamic heart monitoring is required, e.g. in case of mobile patients. The pre-processed ECG beat observations are used by a supervised learning approach, based on principal component analysis (PCA), to train the BMA classifiers for sitting still, up and down, movement of left, right or both arms, walking on a level floor and climbing stairs up and down.

A wearable system for measuring emotion-related physiological parameters has been developed at Fraunhofer IGD Rostock [65]. It uses a glove as a garment hosting a sensor unit which collects data from skin conductivity and skin temperature sensors and a conventional chest belt from Polar as a heart rate sensor. The sensor unit communicates wirelessly using an ISM band transceiver with a base station, which can generate events, such as detection of certain physiological states, although this process is not described in the current study. The reader interested further in glove-based systems and their applications should refer to the comprehensive study of such systems presented in [59].
2.3.2 Commercially available WHMS

In this final category of WHMS, we will briefly discuss several commercially available wearable systems for health-monitoring application.

For example a lot of manufacturers like Nonin [66], Philips [67], Nellcor [68], Agilent [69], Redding Medical [70] and others are providing small, wearable, low-cost and lightweight finger-tip pulse oximeters, that provide real-time display of measured heart rate and blood oxygen saturation. Other examples are the heart rate monitors manufactured from Polar [71] and Omron [72], which use a chest worn belt and wristwatch for display of measurements.

Other commercial systems include Vivago WristCare [73], a wrist-worn device, which monitors skin temperature, skin conductivity and movement. A similar device is the SenseWear Armband developed from BodyMedia [74], which in addition monitors ambient temperature and heat flow. Both devices include a wireless transmitter for communicating the collected data and possible alarms to a base station for further evaluation by a professional clinician.

WelchAllyn has developed the Micropaq Monitor [75], a wearable device worn in a carrying pouch, which can perform pulse oximetry and up to 5-lead ECG monitoring. Another example of commercially available health monitoring systems is the portable polysomnography systems from CleveMed [76], which collect multiple channels of EEG, ECG, EMG, EOG, airflow, snore, thoracic and abdominal respiratory efforts, body position and pulse oximetry and
can transmit the measured data wirelessly to any location by using a simple ISM band transmitter.

VivoMetrics has developed LifeShirt [77], a washable light-weight vest which includes respiratory rate sensors, 1-lead ECG for heart rate measurement and an accelerometer for activity monitoring. Foster-Miller’s Watchdog [78] physiological monitoring tool is a garment-based system capable of monitoring heart rate, respiration rate, posture, activity, skin temperature and GPS location. Another example is the SmartShirt [79] system from Sensatex, a T-shirt-based wearable system using conductive fiber sensors to measure ECG, respiration rate and blood pressure.

CardioNet has developed a mobile cardiac outpatient telemetry system (MCOT) for ambulatory ECG monitoring, aiming at helping physicians diagnose and treat patients with arrhythmias [80]. Other products include the Bio-harness monitor from Zephyr Inc. [81], a chest belt that monitors ECG, respiration rate, skin temperature and activity and is wirelessly enabled and the SpO2, ECG and glucose wireless sensors from Alive Technologies [82]. Furthermore, Schiller [83] and Corscience [84] have also developed small portable Bluetooth ECG monitors.
2.4 Maturity Evaluation

In this section we attempt to evaluate the most representative and “prevailing” systems from the ones discussed in the previous sections. The choice of the systems to be evaluated was based upon:

- their ability to measure multiple parameters
- the amount and the detail level of their provided documentation
- the frequency of their citation by other projects
- the extent to which they utilize state-of-the-art hardware technologies
- the incorporation of intelligent algorithms for feature extraction and/or decision support

The systems that have been chosen for evaluation are listed in Table 2-3 along with a description of their hardware and communication modules, the physiological parameters that are measured and the projects’ stated application field.

Several features have been chosen for the evaluation of the various wearable systems. In Table 2-4 a list of these features is provided along with a brief description for each one.
<table>
<thead>
<tr>
<th>Project Title/Institution</th>
<th>Hardware Description</th>
<th>Communication Modules</th>
<th>Measured Signals*</th>
<th>Medical applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>A) LiveNet (MIT) [26]</td>
<td>PDA, microcontroller board</td>
<td>wires, 2.4GHz radio, GPRS</td>
<td>ECG, BP, R, T, SaO2, EMG, GSR</td>
<td>Parkinson symptom &amp; epilepsy seizures detection, behavioural modelling</td>
</tr>
<tr>
<td>B) AMON (EU IST FP5 program) [27]</td>
<td>Wrist-worn device</td>
<td>GSM link</td>
<td>ECG, BP, T SaO2, A</td>
<td>High-risk cardiac-respiratory patients</td>
</tr>
<tr>
<td>C) LifeGuard (Stanford Un. &amp; NASA) [29]</td>
<td>Custom µC-based device &amp; commercial bio-sensors</td>
<td>serial cables, Bluetooth</td>
<td>ECG, BP, R, T, SpO2, A</td>
<td>Medical monitoring in extreme environments (space &amp; terrestrial)</td>
</tr>
<tr>
<td>D) MyHeart (EU IST FP6 program) [32]-[35]</td>
<td>PDA, Textile &amp; electronic sensors on clothes + heart belt</td>
<td>conductive yarns, Bluetooth, GSM</td>
<td>ECG, R, other vital signs, A</td>
<td>Prevention and early diagnosis of CVD</td>
</tr>
<tr>
<td>E) WEALTHY (EU IST FP5 program) [29]-[31]</td>
<td>Textile &amp; electronic sensors on jacket</td>
<td>conductive yarns, Bluetooth, GPRS</td>
<td>ECG, R, T, EMG, A</td>
<td>Monitoring of rehabilitation &amp; elderly patients, chronic diseases</td>
</tr>
<tr>
<td>F) MagIC (Un. Of lan, Italy, Bioeng. Centre &amp; Cardiac Rehab.Unit.) [36]-[38]</td>
<td>Vest with textile sensors, custom electronic board, PDA</td>
<td>Bluetooth</td>
<td>ECG, R, T</td>
<td>Recording of cardiorespiratory and motion signals during spontaneous behavior in daily life and in a clinical environment</td>
</tr>
<tr>
<td>G) MERMOTH (EU IST FP6 program) [40], [41]</td>
<td>Garment with knitted dry electrodes, PDA</td>
<td>conductive yarns, RF link</td>
<td>ECG, R, T, A</td>
<td>General health monitoring</td>
</tr>
<tr>
<td>H) Smart Vest (Nat.ional Pr. On Smart Materials, India) [42]</td>
<td>Vest with woven sensors, microcontroller</td>
<td>woven wires, 2.4 GHz ISM RF</td>
<td>ECG, BP, T, PPG</td>
<td>General remote health monitoring</td>
</tr>
<tr>
<td>J) Body area network (Valencia, Spain &amp; Malta Un. &amp; Microvitaec Tech) [45]</td>
<td>Zigbee-based motes &amp; Zigbee-based custom base device</td>
<td>Zigbee, Wi-Fi, GPRS</td>
<td>ECG, BP, R</td>
<td>Detection &amp; prediction of human physiological state (wakefulness, fatigue, stress) during daily activities</td>
</tr>
<tr>
<td>K) WSN u-Healthcare system (Dongseo Un. Korea) [46], [47]</td>
<td>Custom tiny motes, cell phone &amp; commercial sensors</td>
<td>Zigbee, CDMA</td>
<td>ECG, BP, SpO2, A</td>
<td>Health monitoring, and remote identification of suspicious health patterns for further evaluation by physicians</td>
</tr>
<tr>
<td>L) Human++ (IMEC) [55], [56]</td>
<td>Miniature low-power BAN nodes, energy scavenging</td>
<td>Zigbee</td>
<td>ECG, EEG, EMG</td>
<td>Enable autonomous wearable sensor networks for general health monitoring</td>
</tr>
<tr>
<td>M) HealthGear (Microsoft) [57]</td>
<td>Custom sensing board, comm.. sensors and cell-phone</td>
<td>Bluetooth</td>
<td>HR, SpO2</td>
<td>Monitoring users during their sleep to detect sleep apnea events</td>
</tr>
<tr>
<td>N) HeartToGo (Un. of Pittsburgh) [58]</td>
<td>Cell phone &amp; comm. available BT biosensors</td>
<td>Bluetooth, GPRS</td>
<td>ECG, A</td>
<td>Individualized remote CVD detection</td>
</tr>
<tr>
<td>Q) AUDABE (Dept. of Medical Physics, Ioannina, Greece) [63]</td>
<td>Mask, glove, chest sensors</td>
<td>wires, Bluetooth, Wi-Fi</td>
<td>ECG, R, GSR, EMG</td>
<td>Evaluation of the emotional state of an individual at environments where subjects operate at extreme stress conditions</td>
</tr>
<tr>
<td>R) Lifeshirt (Vivometrics) [77]</td>
<td>Sensors embedded in vest, PDA</td>
<td>Bluetooth &amp; wires</td>
<td>ECG, R, A</td>
<td>All-day remote health monitoring</td>
</tr>
</tbody>
</table>
Table 2-4. Evaluation Features

<table>
<thead>
<tr>
<th>Feature (F1)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wearability (F1)</td>
<td>The system must have low weight and size.</td>
</tr>
<tr>
<td>Appropriate placement on the body (F2)</td>
<td>The system has to be unobtrusive and comfortable, in order not to interfere with the user’s movements and daily activity.</td>
</tr>
<tr>
<td>Aesthetic issues (F3)</td>
<td>The system should not severely affect the user’s appearance.</td>
</tr>
<tr>
<td>Data encryption and security (F4)</td>
<td>Encrypted transmission of measured signals and authentication requirement for private data access.</td>
</tr>
<tr>
<td>Operational lifetime (F5)</td>
<td>Ultra low power consumption for long-term, maintenance-free health monitoring.</td>
</tr>
<tr>
<td>Real Application (F6)</td>
<td>The developed system is applicable (and useful) to real-life scenarios/health conditions.</td>
</tr>
<tr>
<td>Real-time Application (F7)</td>
<td>The wearable system produces results, e.g. display of measurements, alerts, diagnosis etc, in (or near) real-time.</td>
</tr>
<tr>
<td>Computational &amp; Storage Requirements (F8)</td>
<td>The computational and storage resources required or utilized by the system to achieve desirable results.</td>
</tr>
<tr>
<td>Ease of use (F9)</td>
<td>The system incorporates a friendly, easy-to-use and easy-to-learn user interface.</td>
</tr>
<tr>
<td>Performance and test in real cases (F10)</td>
<td>Sufficient results and performance statistics are provided to verify the system’s functionality in real cases.</td>
</tr>
<tr>
<td>Reliability (F11)</td>
<td>The system produces reliable and accurate results.</td>
</tr>
<tr>
<td>Cost (F12)</td>
<td>The amount of money required to produce and purchase the proposed wearable system.</td>
</tr>
<tr>
<td>Interference Robustness (F13)</td>
<td>Availability and reliability of wirelessly transmitted physiological measurements.</td>
</tr>
<tr>
<td>Fault Tolerance (F14)</td>
<td>The system produces reliable results under any circumstances, such as various kinds of patient’s movements.</td>
</tr>
<tr>
<td>Scalability (F15)</td>
<td>Potentiality of upgrading, enhancing and easily incorporating additional components to the developed system.</td>
</tr>
<tr>
<td>Decision Support (F16)</td>
<td>The implemented system includes some type of diagnosis/decision mechanism or an algorithm/pattern recognition system for context aware sensing of parameters.</td>
</tr>
</tbody>
</table>

The choice of features was based upon the wide range of requirements a wearable biosensor system must meet in order for it to be used in real-life health monitoring scenarios. Since in the general case there are no ideal objective me-
trics for evaluating each feature of every system, the evaluation has been done
with relevance to the top approach in each category.

In addition to that we wanted to provide an overall evaluation of the
systems based on the perspective of the parties who are involved in the devel-
opment and the use of the systems, e.g. the manufacturers, the doctors and the
users. The reason for doing that is because specific features may have different
significance levels for each interested party, for example the on-board comput-
ing resources of the system may be a feature that is highly important to the
developer of the system, while it may not be that important to the user who
would be naturally more concerned about the ease of use of the WHMS.
Therefore, after extended discussions with colleagues, department’s students
and doctor acquaintances we decided to grade the importance of each feature
as low (1), medium (2) or high (3), which resulted in an average weight for
every feature depicting its importance in the evaluation of a WHMS. These
weights are presented in Table 2-5.

Finally an overall maturity score was produced for every system in Ta-
ble 2-3 based on the evaluation of every feature of each system under consider-
ation and by producing a weighted average score according to the weights de-
picted in Table 2-5. It should be noted that the maturity score of a system cor-
responds to the level to which the system fulfills its potentials regarding its
available resources and design approach. Therefore, a higher average score
corresponds to a system that has both reached a top level of development and
also meets the requirements listed in Table 2-4 to the fullest.

<table>
<thead>
<tr>
<th></th>
<th>Patient's perspective</th>
<th>Physician's perspective</th>
<th>Manufacturer's perspective</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>F2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>F3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>F4</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>F5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2.3</td>
</tr>
<tr>
<td>F6</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>F7</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2.3</td>
</tr>
<tr>
<td>F8</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1.7</td>
</tr>
<tr>
<td>F9</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>F10</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>F11</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>F12</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2.3</td>
</tr>
<tr>
<td>F13</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2.3</td>
</tr>
<tr>
<td>F14</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>F15</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>F16</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2.3</td>
</tr>
</tbody>
</table>

The corresponding scores are shown in Fig. 2-1. It also be noted that the
purpose of this maturity evaluation is not to criticize the developed systems,
but to identify their development level according to their potentials and to
provide possible direction for further research in the fields that these systems
show a lack of performance.
Fig. 2-1. Maturity scores for the systems listed in Table 2-3 (1=low, 2=medium-low, 3=good, 4=high, 5=maximum)

2.5 Discussion

As it can be seen from Fig. 2-1, none of the systems reached what was set as the maximum maturity level. However, from the systems we considered in the current study, there are several ones that achieved relatively high scores. These include systems based on smart textiles, which have the advantage of high wearability and comfortableness to the user and which can achieve a high degree of reliability due to guaranteeing good contact between the skin and the bio-sensors even when the subjects are in motion. In addition to those types of systems, the HeartToGo project and the Personal Health Monitor reveal also a promising alternative to the design of WHMS, as they leverage commercial cell phone devices and commercially available bio-sensors to set up a system that is either able to interact with the user to get additional feedback about his health
status or to adapt to his/her individual medical history by means of artificial neural networks.

However, there have been identified several common issues with the evaluated systems and which constitute a set of challenges that will need to be addressed by further researchers in order to improve the efficiency, the reliability and the security of WHMS. These challenges include:

**Battery technologies and energy scavenging:** Power consumption (and battery size) appears to be perhaps the biggest technical issue and performance bottleneck in current implementations. Wearable biosensor systems should be able to operate maintenance-free for long periods (e.g. years). Further research in power scavenging techniques (through body heat or motion), low-power transceivers and improvements in battery technologies promise to solve this problem.

**Security of private information:** The information collected about the user, describing his health status, has to remain secure and not be allowed to be disclosed to anyone besides to the system’s wearer and to the supervising physicians. Thus, proper encryption and authentication mechanisms are required to ensure the privacy of all the communicated data (sensor-to-sensor communication in the BAN or wearable system to base station transmissions).

**Further improvements in sensor miniaturization and efficiency:** In general a lot of the biosensors used in current wearable systems tend to have
bulky size and may require very specific on-body placement or body postures to provide reliable measurements. Further improvements in textile sensors and advanced sensor design and miniaturization are required to appropriately address these shortcomings.

**Clinical validation:** Developed systems must be exhaustively tested and validated by professional physicians.

**Standardization and cooperation at all levels:** The requirement for interoperability between different communication infrastructures and between various types of devices, sensors, actuators, health providers etc., underlines the need for standardization in communication interfaces and cooperation between researchers, medical experts, hardware and textile manufacturers, network providers and other health organizations.
CHAPTER 3: ECG SIGNAL PROCESSING TO MEET THE CONSTRAINTS OF THE WHMS

In a general remote health monitoring scenario, as discussed in the introductory section, bio-signals are recorded in an unsupervised manner. As a result two important questions that arise in this situation are: 1) under what conditions these data were recorded and 2) what do those data actually mean. Regarding the first question, the issue is whether the recorded data are actually usable and not corrupted by noise. Physiological sensor technology is currently far from being robust to measurement artifacts, which can be caused by the user’s movement, or by bad contact between the sensor and the skin or by sensor fault. As a result there is a requirement for a means to deduce whether and to what extent the recorded data should be considered reliable or not. Only in the former case should they be further analyzed by signal processing algorithms and decision making schemes to avoid producing erroneous results.

The second question raises the issue of data processing, feature extraction and interpretation. More concretely, pertaining to the ECG, the signal needs to be processed in order to 1) perform de-noising [85], 2) detect heart beats [86] and finally to classify the recorded beats and rhythms as normal or abnormal [87]. In addition to that, other signals such as the tri-axial acceleration captured from on body accelerometers can be analyzed to quantify the amount
and type of motion the user is currently performing [88]. Finally the recorded data and the corresponding extracted features can be all fused together to estimate the progression of the user’s health condition. An approach towards achieving this last goal will be discussed in the next chapter.

In this chapter we will discuss possible approaches to addressing problem 1) as stated above. We will focus on the case of the ECG signal and we will describe wavelet-based approaches to do the following: a) detect noise-corrupted unusable segments in the signal, b) de-noise the usable portions of the ECG and c) classify detected heart beats. The algorithms that will be presented constitute incremental improvements of existing methodologies that can be found in the abundant literature of wavelet-driven ECG denoising and ECG classification. Analytical and comparative results for the proposed approaches will be provided and their pros and cons will be discussed. However, before we do that it would be appropriate to provide a short but comprehensive introduction to the discrete wavelet transform.

### 3.1 Wavelets

#### 3.1.1 The continuous wavelet transform

The Wavelet Transform has emerged in the past two decades as the most powerful tool for analysis of transient non-stationary signals [89]. There have been numerous research efforts that employ the WT in signal processing
of ECG signals [90]. Its abilities to a) employ a variable window in order to realize high temporal resolution in high frequencies and high spectrum resolution in low frequencies as well as b) the large set of available wavelet functions which can utilized to detect various feature morphologies, have been proven very useful in the analysis of electrocardiograms which are non-stationary in nature. These features illustrate the superiority of the wavelet transform compared to the short time Fourier transform (STFT), which employs a single window for all frequencies and provides the same resolution at all locations in the time-frequency plane.

![Diagram showing time-frequency boxes in different domains](image)

**Fig. 3-1.** Time-frequency boxes in case of a) a time domain signal, b) a frequency domain signal, c) STFT and d) a wavelet basis.
This fact is illustrated graphically in Fig. 3-1, which shows the different time, frequency or time-frequency resolutions of a time-domain signal, of a frequency-domain signal, of a STFT and of a wavelet basis. The boxes shown in the two bottom figures are referred to as Heisenberg boxes [89] to denote the fact that one cannot have arbitrarily high resolution in both time and frequency.

The continuous wavelet transform is defined as the inner product of the function in question and of a set of dilated and translated wavelet functions:

\[
Wf(u, s) = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{s}} \psi^*(\frac{t-u}{s}) dt
\]  

Where \(s\) is the scale or dilution parameter, \(u\) is the translation or location parameter and \(\psi(t)\) is the mother wavelet. In order to characterize a function \(\psi(t)\) as a wavelet and thus to use it for wavelet analysis of signals, it must satisfy certain conditions:

1. It must be continuous and differentiable.
2. It must have compact support (or it must decrease quickly towards zero as \(t \to \infty\)).
3. It must have a band-pass-like spectrum (zero mean), e.g. \(\int \psi(t) dt = 0 \leftrightarrow \hat{\psi}(0) = 0\), where \(\hat{\psi}(f) = \int_{-\infty}^{\infty} \psi(t)e^{-j(2\pi f)t} dt\) is the Fourier transform of the wavelet function.
4. It must have finite energy, e.g. \(\int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty\).
5. Finally, it must also satisfy the admissibility criterion: \(\int_{0}^{\infty} \frac{|\hat{\psi}(f)|^2}{f} df < \infty\).
Wavelet analysis is capable of revealing the following aspects of data: trends, breakdown points, discontinuities in higher derivatives and self-similarity and as a result, it has found numerous applications in signal and image processing [89].

In addition to that the Wavelet Transform (WT) is capable of utilizing the correlation structure, which is local in both time & frequency and which might be present in neighboring points, to acquire a sparse representation of a signal. This is particularly evident with wavelets with a large number of vanishing moments $M_p = \int t^p \psi(t) dt = 0$. If a wavelet has N vanishing moments, that implies that the approximation order of the wavelet transform is also N or in other words the wavelet transform is capable of suppressing signals that are polynomials of degree less or equal to N. This fact makes the wavelet transform particularly applicable to data compression [89].

### 3.1.2 The discrete wavelet transform

The continuous wavelet transform (CWT) is easy to interpret, since its redundancy enhances its readability. However, instead of computing the inherently redundant wavelet transform for an infinite amount of scales and translations (which is both time and storage-space consuming), s and u can be chosen to be discrete. A very common choice is the dyadic grid:

(Eq. 3-2) \[ \psi_{j,k} = 2^{-j/2} \psi(2^{-j} t - k) \]
The above form expresses the discrete wavelet transform (DWT) and in this case the wavelet functions (if properly selected) constitute an orthonormal basis:

\[
\int_{-\infty}^{\infty} \psi_{j,k}(t) \cdot \psi_{j',k'}(t) dt = \begin{cases} 
1, & \text{if } j = j' \text{ and } k = k' \\
0, & \text{otherwise}
\end{cases}
\]  

(Eq. 3-3)

Using the DWT a signal can be represented by an infinite sum of translated and dilated wavelet functions multiplied by the corresponding wavelet coefficient \(d_{j,k}\):

\[
x(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} d_{j,k} \psi_{j,k}(t)
\]  

(Eq. 3-4)

In order to avoid computing an infinite number of wavelet coefficients for DWT analysis, a scaling function \(\phi(t)\) can be employed which can be expressed as an infinite weighted sum of wavelets down to a certain scale \(j\):

\[
\phi(t) = \sum_{j,k} c_{j,k} \psi_{j,k}(t)
\]  

(Eq. 3-5)

The scaling function needs to have a low pass spectrum (averaging filter), it must thus satisfy: \(\int \phi(t) dt = 1\). We can now see that a series of dilated wavelets together with a scaling function can be considered as a filter bank. An analyzed signal can now be represented as a combined series expansion using both the so called approximation coefficients (obtained by projecting the signal onto translations of the scaling function) and the detail coefficients (obtained by projecting the signal onto dilations and translations of the wavelet function):

\[
x(t) = \sum_{k=-\infty}^{\infty} a_{j_0,k} \phi_{j_0,k}(t) + \sum_{j=-\infty}^{j_0} \sum_{k=-\infty}^{\infty} d_{j,k} \psi_{j,k}(t),
\]  

(Eq. 3-6)
where $j_0$ is the coarsest scale to which the signal $x(t)$ had been analyzed.

The DWT is implemented via a pair of conjugate mirror filters, a low-pass FIR filter $h$ (associated with a scaling function $\phi$) and a high-pass FIR filter $g$ (associated with the wavelet function $\psi$), which result in what is usually referred to as subband coding. The filters $h$ and $g$ satisfy the following set of relations:

(Eq. 3-7) \[ \frac{1}{\sqrt{2}} \phi(t) = \sum_{n=0}^{N-1} h[n] \phi(t-n) \quad \text{or} \quad \hat{\phi}(2\omega) = \frac{1}{\sqrt{2}} \hat{h}(\omega) \hat{\phi}(\omega) \]

(Eq. 3-8) \[ \frac{1}{\sqrt{2}} \psi(t) = \sum_{n=0}^{N-1} g[n] \phi(t-n) \quad \text{or} \quad \hat{\psi}(2\omega) = \frac{1}{\sqrt{2}} \hat{g}(\omega) \hat{\phi}(\omega) \]

(Eq. 3-9) \[ |\hat{h}(\omega)|^2 + |\hat{h}(\omega + \pi)|^2 = 2 \]

(Eq. 3-10) \[ \hat{g}(\omega) = e^{-j\omega} \hat{h}^*(\omega + \pi), \]

where $h^*$ denotes the complex conjugate of $h$ and $\omega=2\pi f$ is the angular frequency. The first two equations are sometimes referred to as the “twin scale relations”.

In this case a signal that is decomposed up to a scale $n$ is initially filtered by $h$ and $g$ in order to produce approximation coefficients $a_i$ and detail coefficients $d_i$, which are then downsampled to remove the redundancy of the signal representation. Approximation coefficients in every scale can be further decomposed via the same scheme, in order to produce the $a_i$ and $d_i$ coefficients of the next scale. If the discrete signal is of length $N=2^M$, then this scheme allows for up to $M$ levels of decomposition (coefficients are subsampled in each
scale to avoid redundancy). This process is illustrated in Fig. 3-2. From the above discussion it is clear that the DWT acts as an octave band filter whereby by adding higher octave bands we add more detail or resolution to the signal, hence the name “multi-resolution analysis”.

Fig. 3-2. Fast Wavelet Transform (Mallat's algorithm) implemented with cascaded filters and downsamplers.

To reconstruct the original signal, the reverse procedure is employed. This fast inverse wavelet transform reconstructs progressively each $a_j$ by inserting zeros between samples of $a_{j+1}$ and $d_{j+1}$, filtering and adding the output. For the reconstruction process, the filters that are employed can be either the same that were used for the analysis or we can use a different set of filters, which must however satisfy certain (perfect reconstruction) relations with the deconstruction filters (see Chapters 7.3 and 7.4 in [89]). In the latter case, the wavelets that are employed are called biorthogonal and they no longer constitute an orthogonal basis for the respective signal space. The advantage in this case however, is that we are able to employ symmetric Finite Impulse Response (FIR) filters, which as it is known have the property of linear phase response.
If we remove the downsamplers after each filtering operation and up-sample the filter coefficients by a factor $2^{(j-1)}$ in the $j$th scale, then we acquire a translation-invariant signal decomposition, which is of course redundant. This is referred to as “algorithme à trous” or else undecimated wavelet transform (UWT).

### 3.1.3 Wavelet Packets

Wavelet packets extend the idea of the DWT by basically decomposing both approximation and detail coefficients in every scale to coarser components (the DWT decomposes only the $a_i$ in every scale). This allows the design of bases with adaptable time-frequency properties.

![Wavelet Packet Tree decomposition](image)

Fig. 3-3. Wavelet Packet Tree decomposition
Beginning with the scaling and the wavelet function and the filters $h$ and $g$ one can iteratively generate appropriate analyzing functions which can be used to produce a full binary wavelet packet decomposition tree. This is shown graphically in Fig. 3-3.

The produced binary tree is usually pruned by applying some criterion in order to select an optimal decomposition scheme. The idea behind this approach is to look at each parent node and its two children and quantify the information that is gained by splitting the parent node. If the sum of the entropies, which is the information measure that is usually adopted and is defined as:

\[(\text{Eq. 3-11}) \quad \mathcal{E}(x) = -\sum_n x^2[n] \cdot \log (x^2[n]),\]

of the two children is smaller than the entropy of the parent node, then the split is performed as there is obviously a gain in terms of compression of information.

3.1.4 Matching Pursuits

Matching Pursuit (MP) introduced by Mallat and Zhang [91] is a greedy algorithm for computing a sub-optimal decomposition of a given signal over a redundant orthogonal waveform dictionary, by selecting one by one the “optimal” vector. Let $f$ be the signal to be approximated and $D = \{\varphi_p\}_{p \in \mathbb{F}}$ be the dictionary of vectors (atoms) $\varphi_p$, which are discrete-time 1-dimensional signals
of length \( n \) and have unit norm. The MP algorithm works iteratively in order to decompose \( f \) as:

\[
\text{(Eq. 3-12)} \quad f = \sum_{m=0}^{M-1} a_m \varphi_{p_m} + R^M f
\]

where \( a_m \) is the inner product or the projection of the residual \( R^m f \) on \( \varphi_{p_m} \):

\[
\text{(Eq. 3-13)} \quad a_m = \langle R^m f, \varphi_{p_m} \rangle
\]

and

\[
\text{(Eq. 3-14)} \quad R^{m+1} f = R^m f - \langle R^m f, \varphi_{p_m} \rangle \varphi_{p_m}
\]

with \( R^0 f = f \). The question now is which atom should be chosen to project the residual \( R^m \) at each iteration. According to MP, in order to minimize the residual given by (Eq. 3-14) in every step, we need to choose the atom that maximizes the absolute value of the inner product given in (Eq. 3-13). The iteration of the algorithm can be stopped when either a satisfactory signal approximation has been achieved (measured via the normalized residual energy: \( \|R^m f\|^2 / \|f\|^2 \)) or when the maximum number of atoms for a specific task has been reached. The iterations of the MP algorithm to derive a sub-optimal approximation of a signal \( f \) using \( M \) atoms are the following:

1. \( m = 0, R^0 f = f \) and \( \hat{f}_0 = 0 \)
2. \( i = \arg \max_j \| \langle R^m f, \varphi_{p_j} \rangle \| \)
3. \( a_m = \langle R^m f, \varphi_{p_i} \rangle \)
4. \( \hat{f}_{m+1} = \hat{f}_m + a_m \varphi_{p_i} \)
5. \( R^{m+1} f = f - \hat{f}_{m+1} \)
\textbf{6. } m = m + 1

\textbf{7. } If \( m < M \) then go to 2. else stop and \( \hat{f}_m \) is the requested sub-optimal signal approximation.

A variety of waveform dictionaries can be possibly employed with MP, for example wavelet packets, cosine packets or Gabor atoms. In this work we studied the use of wavelet packet dictionaries in the approximation of ECG beats. Our implementation is done in MATLAB and it is based on the WaveLab library [92].

\textbf{3.1.5 Denoising with Wavelets}

The discrete wavelet transform can be very effectively utilized to remove undesired components (noise) from a signal. The conventional wavelet denoising scheme as proposed by Donoho [93], consists of employing a DWT decomposition up to a certain scale and then applying soft or hard-thresholding techniques to the detail coefficients at all levels in order to suppress the noise components (assuming that all wavelet scales contain noise).

Soft thresholding is defined by the following operation:

\begin{equation}
\text{Eq. 3-15} \quad d_j[k] = \begin{cases} 
\text{sign}(d_j[k]) \cdot (|d_j[k]| - T), & \text{if } |d_j[k]| > T \\ 
0, & \text{if } |d_j[k]| \leq T 
\end{cases}
\end{equation}

Similarly, hard thresholding is performed by:

\begin{equation}
\text{Eq. 3-16} \quad d_j[k] = \begin{cases} 
|d_j[k]|, & \text{if } |d_j[k]| > T \\ 
0, & \text{if } |d_j[k]| \leq T 
\end{cases}
\end{equation}
T denotes the applied threshold, which can be selected according to various estimation procedures [89], [94]:

- Stein’s unbiased risk estimate (SURE):

(Eq. 3-17) \[ T = \sqrt{2 \cdot \log_e(n \cdot \log_2(n))} \]

- Fixed threshold (FIXTHRES):

(Eq. 3-18) \[ T = \sqrt{2 \cdot \log_e(n)} \]

- HEURISTIC SURE, which replaces in very noisy conditions the SURE with the FIXTHRES estimate.

- MINIMAXI, which applies a fixed threshold given by:

(Eq. 3-19) \[ T = 0.3936 + 0.1829 \cdot \log_2(n) \]

As stated in [94], the underlying signal model assumes that the noise is additive and normally distributed with zero mean and unit variance. As a result, we have to rescale the estimated thresholds when dealing with unscaled and non-white noise. A very common rescaling approach is given by:

(Eq. 3-20) \[ \hat{T} = T \cdot \hat{\delta}, \quad \text{where:} \]

(Eq. 3-21) \[ \hat{\delta} = \frac{\text{Median}(|d_f(k)|)}{0.6745} \]

and \( \hat{\delta} \) is the estimated standard deviation of the noise [95], which can either be computed from the coefficients in the finest decomposition scale or in a scale-by-scale manner.
The de-noising algorithm is concluded by reconstructing the signal using the inverse DWT with the modified (shrunken) detail coefficients. For more details on the algorithm by Donoho the interested reader can refer to [93] and [95].

A similar approach consists of performing optimal wavelet packet decomposition and then thresholding the resulting coefficients before reversing the whole process to reconstruct the denoised signal estimate. However, looking for the optimal tree is a time-consuming process, especially in the case of processing real-time data. As a result thresholding can be applied directly to the full wavelet packet tree.

3.2 Dealing with Noise in the ECG

3.2.1 The Electrocardiogram (ECG)

The ECG is a recording of the electrical activity of the heart. It can be recorded and monitored in a non-invasive and relatively unobtrusive manner as we saw in the survey of WHMS. The standard ECG (or EKG) is measured by appropriately placing 10 electrodes on the patient’s body and by measuring 12 ECG leads as the potential difference between a pair of electrodes or between an electrode and a ground reference point. An idealized ECG can be seen in Fig. 3-4.
In this figure we can see the types of waves that compose the ECG waveform, e.g. the P wave, the QRS complex and the T wave. For a physiological interpretation of the mechanisms behind the formation of these waves, the reader can refer to [96] or to any other of the wide variety of books regarding cardiology and ECG interpretation. The R peak is usually referred to as the “fiducial” point and its detection is associated with precisely determining the occurrence of a heart beat. In Fig. 3-4 we can also see annotations regarding the most important ECG features, such as the QRS width and the ST segment, the accurate detection of which carries significant important diagnostic information [87], as depressed ST segments may be a sign of ischemic episodes and elevated ST segments may indicate myocardial infarction.
As it was stated before, the ECG signal depicted in Fig. 3-4 constitutes an idealized case, whereby noise is absent and detection of the prominent ECG waves and their corresponding boundaries is rather straightforward. However, in real-life scenarios ECG recordings will be contaminated by several types of noise: low frequency noise (causing what is referred to as baseline drift or wander) due to respiration and movement, high frequency noise due to muscle noise (EMG noise) and noise at 50 or 60Hz due to power-line interference. Furthermore, according to electrode placement and to corresponding lead measurement, some of the ECG peaks and waves depicted in Fig. 3-4 may or may not be present (or may even be inverted) [96].

An example of an ECG signal contaminated by several noise sources is given in Fig. 3-5. This signal has been taken from record 121 of the MIT-BIH Arrhythmia database [99], which has been widely adopted as the reference database when evaluating QRS detection and ECG beat classification algorithms. The ECG recording depicted in Fig. 3-5 is clearly contaminated with low-frequency noise, which causes a slowly varying drift of the signal’s baseline. Furthermore the presence of high frequency noise is evident as well as it can be observed as high frequency ripples on the segments between two R peaks. As a result, in order to detect significant diagnostic information and extract features of interest from the ECG, appropriate signal filtering and conditioning needs to be performed at a first stage.
3.2.2 Detecting noise-corrupted segments

An important issue in ambulatory ECG recordings is that of complete signal corruption of the signal due to measurement artifacts or sensor malfunction. The DWT can be employed in this case in order to analyze the energy content of the recorded signal at different scales and if values are detected over or under empirically set thresholds, then we may deduct that the signal is corrupted or that there is a low-energy unusable section (in case the wearable sensor has been removed or misplaced). A similar approach utilizing FIR and IIR filters is described in [97].

Here we want to describe the detection of noise-corrupted ECG segments using a DWT based approach. In Fig. 3-6 we can see part of an ECG recording recorded with the WHMS platform (which will be fully described in Chapter 8) and processed offline using MATLAB. An important thing to note
here is that the ECG trace provided by the Bioharness ECG sensor [81] is band-pass filtered, where the band is about 15 to 75 Hz.

![Original Signal]

**Fig. 3-6.** An ECG segment, part of which is severely corrupted by noise caused by movement.

It is evident that there is a part of the signal, starting at about the 28th second, that is severely corrupted by noise (in this case from abrupt chest movements which significantly affected the contact surface between the ECG sensors and the subject’s chest).

In Fig. 3-7 we can see four levels of approximation coefficients (obtained using the Coiflet wavelet of order N=5) of the same signal that is depicted in Fig. 3-6. We notice that the noise we are dealing with here has significant low frequency contents which of course interfere with the heart beat
waves, which at the first half of the signal, where the ECG is almost noise-free, have been smoothed down. This is particularly obvious at the coarsest decomposition scale. As a result, the approximation coefficients at level 4 offer a means by which we could discriminate between normal ECG and the noise type in question.

Fig. 3-7. 4 levels of approximation (using Coif5 wavelet) coefficients of the ECG shown in Fig. 3-6.

A way to take advantage of this feature is to compute the amount of energy that is being carried in a window of length $N$ of DWT coefficients by calculating:

$E_{a4} = \sum_{i=0}^{N-1} |a_{4,i}|^2 \quad \text{(Eq. 3-22)}$

For a more smooth power estimate over time it is more favorable to first square the approximation coefficients and then to apply a low-pass filter, as proposed in [97], to smoothly average the power estimate. This is done by em-
ploying a variable length Hamming window, whereby the length of the window is adjusted according to the instantaneous heart rate estimate. The length of the Hamming window is computed as:

\[
L = \frac{60}{HR} \cdot \frac{F}{2^n} + b
\]

(Eq. 3-23)

where \(F_s\) is the sampling frequency of the ECG (250 Hz in the case of the Bioharness [81]), \(HR\) is the instantaneous heart rate estimate (which can be taken from the estimate given by the Bioharness or by an averaged HR estimate from recent RR intervals), \(n\) is the decomposition level (level 4 in this case) and finally \(b\) is a bias to account for the heart rate variability as a result of which successive RR intervals may show significant variation (here \(b=10\), chosen empirically).

Before showing how the resulting energy estimate can be utilized to detect corrupted ECG segments, we should briefly refer to other types of signal corruption in the ambulatory ECG. Of particular interest is the case whereby the ECG becomes corrupted due to permanent displacement of the wearable sensor or even due to the user taking it off. Such a case is depicted in Fig. 3-8, where a noisy signal is produced initially due to sensor displacement and then due to the user taking off the chest belt. It would be very useful if we were able to instantly detect such a scenario as well.
Fig. 3-8. An ECG segment, part of which is severely corrupted by noise caused by sensor displacement.

By employing the energy estimation scheme described before we are indeed able to detect the described cases where severe "non-reversible" artifacts are created in the ECG. This is schematically illustrated in Fig. 3-9. To detect events such as the one depicted in Fig. 3-6 we can employ a heuristically-set threshold (depicted as $T_2$ in Fig. 3-9b), above which we can deduce that the ECG has been corrupted. Similarly for low energy portions such as in the case shown in Fig. 3-8, we can employ a lower threshold (depicted as $T_1$ in Fig. 3-9b) on the smoothed energy estimate, under which samples are discarded as well. To improve the detection of corrupted ECG segments we can utilize the high frequency detail coefficients ($d_i$) to recognize segments of abnormally high concentration of high frequency components (keeping in mind that the ECG transmitted by the Bioharness chest band has been low-pass filtered to remove
high frequencies). In this case the employed threshold is shown in Fig. 3-9c as \( T_3 \).

![Fig. 3-9. Detection of corrupted ECG segments with appropriate thresholding on the energy estimates.](image)

Putting everything together we get the power estimates shown in Fig. 3-9. By selecting heuristically appropriate thresholds we can detect the severely distorted signal portions. In the 80 sec ECG segment shown in Fig. 3-9, there are 3 types of occasions whereby noise is introduced in a nonreversible manner in the ECG. By empirically estimating two thresholds for the energy estimate shown in the second graph of Fig. 3-9 (computed from the 4th level of approximation coefficients) and another threshold for the lower graph (taken from the energy estimate of \( d_1 \)) we can recognize the parts of the signal that need to be
discarded. This thresholding operation creates three separate masks and by combining them we can identify the samples to be discarded. This is depicted graphically in Fig. 3-10, where the original ECG signal is plotted on top, while on the bottom we can see the mask that corresponds to each of the three thresholds. Combining these three masks together yields the final noise mask, which is shown as a thick black line in Fig. 3-10.

Fig. 3-10. Original ECG signal and estimated noise masks.

3.2.3 *ECG de-noising with the undecimated wavelet packet transform*

The DWT has been widely employed in the de-noising of the ECG [85], [90], [98]. Research has proved [98] that employing the undecimated (or sometimes referred to as stationary) wavelet transform, can lead to better denoising
results, due to the redundancy which is inherent in the transform. In addition to that, the effectiveness of using wavelet packet thresholding has been studied by Tikkanen [94] and has been shown to be inferior to DWT denoising. However, there has not been a study utilizing the undecimated wavelet packet transform (UWPT) of the ECG to perform coefficient thresholding in order to remove the noise of the signal. Such an approach could combine both the advantages of the wavelet packet transform (coarser analysis of the signal’s details) and of the undecimated wavelet transform (time-invariant analysis). This is the approach that has been adopted in the current study, e.g. after each decomposition step both approximation and detail coefficients are decomposed and the decimation stages are removed while the filters in each stage are upsampled.

To examine the effectiveness of employing the UWPT for ECG denoising we compare it with DWT and UWT denoising. By experimenting with various wavelet types we were surprised to find that UWPT denoising performed at best when the Haar basis (or else the Walsh wavelet packets) was used (especially in very noisy environments). Also the optimal decomposition stage for using Stein’s unbiased risk estimate (SURE) was found to be level 4. The noise level on each scale was estimated adaptively as described by (Eq. 3-20) and (Eq. 3-21). We utilized ECG recordings from the MIT-BIH Arrhythmia database [99] to compare the above approaches. In general, the UWPT proved to be slightly
more effective than UWT denoising. To illustrate the fact that the Haar wavelet yields better denoising performance than other wavelet families we can refer to Fig. 3-11, where an ECG segment taken from record 103 of the MIT-BIH Arrhythmia database has been extracted and then “contaminated” with Additive White Gaussian Noise (AWGN) resulting in a signal with signal-to-noise ratio (SNR) equal to 4dB. Then this signal is denoised using several approaches and as we can see, the approach that utilizes the undecimated Haar wavelet packet transform to perform denoising yields the higher output SNR.

![Diagram of denoising performance of different approaches on ECG segment taken from the MIT-BIH Arrhythmia Database, Record 103 with AWGN: a) Original ECG, b) Noisy ECG (SNR=4dB), c) Haar UWT denoising (SNR=15.08dB), d) Haar UWPT denoising (SNR=15.11dB), e) Coif3 UWT denoising (SNR=14.60dB), f) Coif3 UWPT denoising (SNR=14.28dB), g) Sym3 DWT denoising (SNR=9.62dB), h) Coif3 DWT denoising (SNR=11.58dB). (Note: In the figure the UWT is denoted as SWT according to MATLAB’s terminology)]

We should note here that the SNR has been computed by:
(Eq. 3-24) \[ SNR = 10 \log_{10} \left( \frac{\sum x^2[n]}{\sum (x[n] - x_e[n])^2} \right), \]

where \( x[n] \) is the original sequence and \( x_e[n] \) is the noisy or the denoised signal.

To further evaluate the performance of the denoising approaches compared in Fig. 3-11 we examined their efficiency in denoising an ECG segment which contains irregular heart beats (such as Premature Ventricular Contractions). Such a case is depicted in Fig. 3-12 where an ECG extracted from record 208 of the MIT-BIH Arrhythmia database has been used. In this case as well, the Haar UWPT denoising approach yields the best performance.

Fig. 3-12. Denoising performance of different approaches on ECG segment taken from the MIT-BIH Arrhythmia Database, Record 208 with AWGN: a) Original ECG, b) Noisy ECG (SNR=4dB), c) Haar UWT denoising (SNR=14.29dB), d) Haar UWPT denoising (SNR=14.41dB), e) Coif3 UWT denoising (SNR=14.06dB), f) Coif3 UWPT denoising (SNR=14.07dB), g) Sym3 DWT denoising (SNR=11.11dB), h) Coif3 DWT denoising (SNR=11.92dB). (Note: In the figure the UWT is denoted as SWT according to MATLAB’s terminology)
So far we have considered only additive Gaussian white noise. Although this noise model provides the most common means of estimating and comparing the performances of denoising algorithms, the type of noise that is usually introduced in the ECG does not have such ideal characteristics (if we consider the overall noise component as a superposition of different noise sources such as electrode noise, motion artifacts, electromyographic noise etc). Additionally, as Clifford et al. have stated in an excellent book about ECG signal processing and analysis [100]: “...the assumption that most noise is Gaussian in nature is approximately correct (due to the central limit theorem)”.

The authors of [100] also state that a noise model that is very representative of the type of observation noise that is introduced in the ECG signal is “pink noise”, which is characterized by a power spectrum which is proportional to $1/f$, where $f$ is the frequency.

Thus, in order to evaluate the performance of the various wavelet denoising approaches under a more realistic noise model, we compare their performances under additive pink noise as well. This is shown in Fig. 3-13 for the case of the same signal segment as the one analyzed in Fig. 3-11. Additive pink noise has been added to the ECG signal of the first plot resulting in a noisy signal with SNR 5dB. The denoised waveforms do not have a SNR as high as the ones of Fig. 3-11 (white noise), but the basic morphological characteristics of the waves are retained after the denoising process. In addition to that, the
Haar-based UWPT denoising yields the best performance in this case as well (SNR=6.79dB).

Finally we will compare the performance of UWT, UWPT and DWT denoising on the same signal segment for various SNRs in order to validate our previous results in cases of higher or lower signal-to-noise ratios. The noisy ECG will be decomposed using UWP and UWPT with Haar basis function and DWT with Symmlet 3 basis (which is very effective in the DWT case, although coif3 denoising tends to yield better performance for lower SNRs).
In Fig. 3-14 we can see the same ECG segment from record 103 that was used before in the first plot. The second plot shows that signal with AWGN and SNR=5dB. The SNR of the UWT denoised signal (third plot) was 15.73dB, of the UWPT (fourth plot) 16.62dB and of the DWT (fifth plot) 12.64dB. In addition to that, UWPT produced 2.17% Mean Square Error (MSE) compared to 2.67% in the UWT case, while the MSE in the QRS complexes area was 1.18% and 1.62% accordingly. The latter result along with visual supervision of the obtained waveforms proves also that UWPT denoising performs better in retaining important beat features, such as the Q, R and S waves.

Fig. 3-14. a) Original ECG (record 103), b) Noisy ECG with AWGN, c) Haar UWT denoised, d) Haar UWPT denoised, e) Sym3 DWT denoised.
Fig. 3-15 plots the SNRout against the SNRin that was obtained for the above signal for various white noise levels.

![SNRout vs SNRin for various noise levels for the signal shown in Fig. 3-14.](image)

Finally in Fig. 3-16 we can see the ECG signal of Fig. 3-14 on the top plot, while on the second plot the same signal with added pink noise is depicted (SNR=5dB). The third plot shows the Haar UWT denoised signal (7.84dB), the fourth one shows the Haar UWPT denoised ECG (7.97dB) and the last one shows the result of sym3 DWT denoising (6.72dB). Moreover, UWPT produced 15.95% Mean Square Error (MSE) compared to 16.41% in the UWT case, while the MSE in the QRS complexes area was 3.40% and 4.67% accordingly, a fact which further highlights the ability of the UWPT-based denoising in reconstructing more accurately the high frequency components of the signal. Fig. 3-17 shows the SNRout against the SNRin that was obtained for the above signal for various pink noise levels.
Fig. 3-16. Original ECG (record 103), b) Noisy ECG with pink noise, c) Haar UWT denoised, d) Haar UWPT denoised, e) Sym3 DWT denoised.

Fig. 3-17. SNRout vs SNRin for various noise levels for the signal shown in Fig. 3-16.
3.3 Classifying ECG beats

3.3.1 A brief note on the MIT-BIH Arrhythmia database

The MIT-BIH Arrhythmia database includes various recordings, of many common and life-threatening arrhythmias. It contains 48 recordings, each one containing two 30-min ECG signals, each one taken from a different lead. The data have been band-pass filtered at 0.1-100 Hz and they have been digitized using a 360 Hz sampling frequency and an 11 bit resolution.

3.3.2 Beat detection

As it was stated previously, the most important part in ECG signal processing is the detection of the fiducial points, e.g. the R peaks or what is referred to as the QRS complex. While there have been proposed several methods in order to achieve that goal [86], the method that was proposed by Pan and Tomkins [101] and which was later improved by Hamilton and Tomkins [102] has been proven to be very reliable and has been widely employed by several researchers. This method employs low-pass and high-pass filters to get rid of noise and frequency components that correspond to waves other than the QRS complex and then uses a differentiation filter, a squaring stage and a moving window integration to derive a feature signal. By applying some simple rules on the feature signal the QRS complexes can be reliably detected. This process can be seen in Fig. 3-18 and Fig. 3-19 for the case of the same ECG recording.
that was depicted in Fig. 3-5. The output of each one of the 6 stages of the Hamilton-Tomkins algorithm is shown, leading up to the detection of the ECG peaks.

![Fig. 3-18. First three stages of the Hamilton-Tomkins algorithm](image)

![Fig. 3-19. Last three stages of the Hamilton-Tomkins algorithm](image)

The displayed QRS complex detection algorithm has high accuracy and provides close to real-time performance while not being very computationally...
intensive. The next step of an ECG processing unit is the characterization of each detected ECG beat as normal or abnormal. Examples of abnormal ECG beats are: Premature Ventricular Contraction, Atrial Premature Beat, right or left Bundle Branch Blocks and escape beats. The detection of these abnormal beats allows the heart rhythm evolution to be followed and various types of arrhythmia to be recognized [103]. This is discussed in greater detail in the following section.

3.3.3 Using matching-pursuit-based features to derive an efficient ECG classification method

The electrocardiogram (ECG) is a well studied medical tool for the assessment of the heart’s function. As it provides a continuous recording of the electrical activity of the heart, it is utilized by physicians in order to diagnose heart conditions such as arrhythmias and cardiovascular disease. However, arrhythmic patterns tend to appear infrequently and thus long-term 24-hour ECG recordings are usually needed which are acquired using the so called Holter monitors [104]. Moreover, advances in wearable health monitoring technology have propelled the development of a variety of research prototypes and commercially available systems that are capable of taking ambulatory ECG recordings [14]. Automatic interpretation of electrocardiographic data is thus required to identify patterns of interest in unsupervised ECG recordings.
Automated ECG classification has been a topic of high interest in the biomedical engineering research community for more than 3 decades now [105], [106]. Several approaches have been proposed for extracting features from the ECG beats in order to aid in correct beat recognition. Examples of such approaches include the use of Discrete Wavelet based statistics [107], FFT spectral components [108], Coefficients of Hermite basis functions representation [109], ECG wave intervals, amplitudes and/or selected sample points [87], [110], [111], heuristically defined morphological features [112] and higher order statistics [113]. However, as pointed out in [113], morphological descriptors and wave attributes extracted in the time domain, as well as frequency domain features tend to vary significantly for beats belonging to the same ECG class, even for recordings extracted from the same patient. In addition to that, morphological features rely heavily on accurate detection of beat boundaries, a computationally intensive process which is sometimes defined heuristically and can be severely impaired in the presence of considerable signal noise [114].

An alternative approach to ECG feature extraction is proposed by Herreiro et al. in [115] (which is also utilized in [112] as well by the same group). Their approach consists of utilizing the Matching Pursuit algorithm [91] in order to decompose an ECG beat into a linear expansion of waveforms selected from a redundant wavelet packet dictionary. Then the expansion coefficients along with the normalized residual energy are used as inputs to an Artificial
Neural Network (ANN) to classify the beat. The main disadvantage of their approach is that for every beat that needs to be detected, a different decomposition is required and an individual neural network is employed, e.g. five in total, which is equal to the number of distinct classes. The computational overhead is further increased by the use of two ECG leads and by the computation of the normalized residual energies.

Motivated by the approach presented in [115], we propose in this section a variation of that scheme, which is considerably more computationally efficient due to the following reasons: a single ECG lead is used, less features are derived and a single ANN is required for classification and the residual energies do not need to be computed. To illustrate the effectiveness of our approach, we describe in detail the feature extraction process and we evaluate our method using the MIT-BIH Arrhythmia database [99]. We also study the effect of the following attributes on the classification performance: the number of atoms used for approximation, the type of wavelet employed for constructing the dictionary and the number of ECG samples around the R peak that are processed. Finally, we discuss the improvement in terms of the significantly reduced computational overhead, a fact which allows for implementation of the proposed scheme in a resource constrained portable device.

As it was mentioned above, beat morphologies may show significant intra-class variability. For example in Fig. 3-20 we can see all the PVC beats ex-
tracted from record 200 of the MIT-BIH Arrhythmia database superimposed on each other. The pattern corresponding to that particular beat class shows significant variability even in the one single recording taken from one patient. As a result, employing the MP algorithm on a single ECG beat template of each class (in order to decide the atoms to be used for approximating that particular class) could result in dismissing signal details of same-class beats with considerable morphological deviations from the chosen template.

As a result, we have followed a modified MP approach in our study which works as follows:

Fig. 3-20. Superimposed PVC beats from record 200 of the MIT-BIH Arrhythmia database.
1) First we select 10 representative beats for each class of interest from each MIT-BIH record that actually contains beats corresponding to that specific class (resulting in a set containing 380 normal beats, 250 PVC beats, 40 LBBB beats, 60 RBBB beats, 110 APC beats, 20 paced beats and 20 PFUS beats). For each beat we collect 256 samples around the R peak using the database’s annotations (90 samples before the R peak and 165 samples after). We also save the R-to-R interval before the R peak as well as the average RR interval before the previous beat, which is taken as the average of the five preceding intervals.

2) Then for each class we try to determine the set of most representative atoms by identifying in each step the atom which maximizes the sum of absolute values of the projections of all beats in that class on the specified atom, e.g.:

\[ p_i = \arg \max_{p_m} \sum ||< R^m f, \varphi_{p_m} > || \]  

After \( p_i \) has been determined, we subtract the projection of the ECG beats on that wavelet packet from their previous residuals (individually) and then the process repeats until we have a satisfactory number of atoms. Our approach here varies from the one from Herrero et al., who initially perform the MP algorithm for every beat in their training set individually and then choose the set of atoms that have maximum overall correlation with the “denoised” waveforms, without considering the overall residual energy in their choice of atoms.
3) For every class we collect the top-ranking atoms returned from the previous steps. Then we look at all the atoms collected for all classes and we discard the ones that are repeated. The remaining wavelet packet waveforms are kept and are utilized for projecting the ECG beats. The inner products of these projections together with the RR interval features mentioned in 1) are utilized as inputs to a single ANN to perform beat identification. Before feeding the inputs to the ANN, all data are normalized by subtracting the mean and diving by the standard deviation, while for the testing data the mean and standard deviation obtained during the training phase are employed.

It should be noted here that before performing the above procedures, we preprocessed all MIT-BIH Arrhythmia records using a) a median filter with a 300ms window to remove baseline wander and b) 4-level undecimated wavelet denoising to remove high frequency noise. Also, only a single-lead of ECG data was used, namely the modified limb lead II (MLII), which has been obtained via chest electrodes, producing a trace that is morphologically similar from that recorded with a wearable chest band. Furthermore, the beat types that were considered in our study are: Normal (N), Premature Ventricular Contraction (PVC), Left Bundle Branch Block (LBBB), Right Bundle Branch Block (RBBB) and Paced. Examples of these beat patterns are shown in Fig. 3-21. We also considered classifiers that identified Atrial Premature Contractions (APC) and Fusion of Normal and Ventricular Beat (PFUS).
Fig. 3-21. Several beat types extracted from records 106, 208, 111, 118 and 217 of the MIT-BIH Arrhythmia database.

All beats pertaining to above classes were extracted from all MIT-BIH Arrhythmia records (except for records 102 and 104 that do not include a MLII lead trace) using the first half of each record as part of the training set and the second half in our testing set. The precise number of beats that were utilized from each class for training and testing are shown in Table 3-1.

Table 3-1. Number of beats from each class used for training and testing.

<table>
<thead>
<tr>
<th>Number of Beats</th>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>37373</td>
<td>37352</td>
</tr>
<tr>
<td>PVC</td>
<td>3542</td>
<td>3531</td>
</tr>
<tr>
<td>LBBB</td>
<td>4036</td>
<td>4033</td>
</tr>
<tr>
<td>RBBB</td>
<td>3626</td>
<td>3623</td>
</tr>
<tr>
<td>Paced</td>
<td>1808</td>
<td>1808</td>
</tr>
<tr>
<td>PFUS</td>
<td>367</td>
<td>367</td>
</tr>
<tr>
<td>APC</td>
<td>1237</td>
<td>1233</td>
</tr>
<tr>
<td>Total</td>
<td>51989</td>
<td>51947</td>
</tr>
</tbody>
</table>
3.3.4 Results

In order to determine the most suitable wavelet type to construct the wavelet packet dictionary, we experimented with Daubechies (d), Coiflet (c) and Symmlet (s) wavelets. Fig. 3-22 and Fig. 3-23 depict the approximation error versus the number of atoms used for approximating a normal beat and a PVC beat. The most efficient wavelet from each family (s8, d8 and c3) was determined empirically after several tests, whereby the ability of each wavelet to approximate selected patterns of interest was evaluated. These figures show also the approximation error in the case that 64 samples around the R peak are used for each beat.

Fig. 3-22. Normalized approximation error (e.g. normalized residual energy as defined in section 3.1.4) vs number of atoms (normal beat).
It can be noticed that in this case, by using just 8 atoms of either family, the approximation error for both types of beats is less than 1%.

In addition to that, when performing steps 1-to-3 as described in the previous section, we found that in general the Symmlet 8 family achieved the best performance in terms of average and maximum residual error energy for most of the beat classes. However, as we can see for example in Fig. 3-23 for the case of the coif3 wavelet family and the PVC beat, other types of wavelets might be particularly efficient in representing a specific beat-type.

---

![Fig. 3-23](image.png)

**Fig. 3-23.** Normalized approximation error (e.g. normalized residual energy as defined in section 3.1.4) vs number of atoms (PVC beat).
To illustrate how ECG beat patterns get approximated by a sub-optimal linear combination of wavelet packet atoms we show in Fig. 3-24 a normal ECG beat (taken from record 123 of the MIT-BIH Arrhythmia database) and its approximation using 8 coiflet3 wavelet packet atoms. The approximation error in this case is 2.53%. Also in Fig. 3-25 we can see the top 6 wavelet packet atoms that were selected by the Matching Pursuits algorithm. The 6 waveforms (together with the next 2 wavelet packet atoms) are weighted according to the projection coefficients and then added together to produce the approximation of the ECG beat that we saw in Fig. 3-24.

Fig. 3-24. Original signal (normal beat pattern taken from MIT-BIH Arrhythmia DB record 123) and MP approximation using 8 coiflet3 wavelet packet atoms.
Considering the fact that this level of approximation is obtained by a simple linear transform (a multiplication of a 256-point vector with a 256×8 matrix) and that we end up with an 8-dimensional pattern word in contrast to the original 256 element vector, we automatically notice the advantages of employing the MP algorithm for feature extraction and compact waveform representation. However, we still need to verify that the extracted MP coefficients (e.g. inner products between ECG beats and wavelet atoms) constitute good features for beat discrimination and classification.

To do that, as a first test we examined the discrimination of N and PVC beats using 10 atoms of a Symmlet 8 wavelet packet dictionary. It turns out
there are no common atoms selected for the two classes and as a result we end up representing each beat with 22 features (10 atoms for each type of beat plus the 2 RR interval features). Using a single 2-layer ANN with 22 hidden neurons and 2 output neurons, all employing hyperbolic tangent transfer functions, the sensitivity of the classifier for normal beats was 99.6% and for PVC beats 97.8%, which are almost identical with the accuracy of the corresponding classifier proposed in [115], which employs 48 features (2 ECG leads, MP residuals and some Independent Component Analysis derived features) and a 3 layer ANN.

Further tests included the classification of the 5 beat types mentioned in the previous section. In this case we experimented with the number of samples around the R peak that were utilized (256, 128 or 64), with the type of wavelet employed and with the number of atoms used for approximation. Sensitivity and Specificity results for the five beat classes in various cases are shown in Table 3-2. In all cases a 2-layer MLP ANN was employed, with the number of hidden neurons equal to the number of input features and 5 output neurons, one corresponding to each class. It was found that employing a 3-layer MLP did not result in greater classifier accuracy, while it significantly increased the training time (plus there is added computational load due to the extra layer).

We notice from Table 3-2 that using more samples around the R peak (to possibly include parts of the P and T waves) does not necessarily increase
the classification accuracy, to the contrary performance tends to be slightly degraded when using more samples.

Table 3-2. Classifier performance in various cases

<table>
<thead>
<tr>
<th>Class</th>
<th># Beats</th>
<th>PVC</th>
<th>LBBB</th>
<th>RBBB</th>
<th>Paced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sens. (1)</td>
<td>37352</td>
<td>3531</td>
<td>4033</td>
<td>3623</td>
<td>1808</td>
</tr>
<tr>
<td>Spec. (1)</td>
<td>99.4%</td>
<td>95.3%</td>
<td>94.7%</td>
<td>77.2%</td>
<td>99.9%</td>
</tr>
<tr>
<td>Sens. (2)</td>
<td>96.5%</td>
<td>99.1%</td>
<td>99.5%</td>
<td>99.9%</td>
<td>99.4%</td>
</tr>
<tr>
<td>Spec. (2)</td>
<td>99.2%</td>
<td>97.9%</td>
<td>92.6%</td>
<td>99.3%</td>
<td>99.9%</td>
</tr>
<tr>
<td>Sens. (3)</td>
<td>99.3%</td>
<td>97.7%</td>
<td>94.8%</td>
<td>98.2%</td>
<td>99.9%</td>
</tr>
<tr>
<td>Spec. (3)</td>
<td>99.9%</td>
<td>99.0%</td>
<td>99.8%</td>
<td>99.9%</td>
<td>99.9%</td>
</tr>
<tr>
<td>Sens. (4)</td>
<td>99.4%</td>
<td>97.3%</td>
<td>92.3%</td>
<td>99.1%</td>
<td>99.9%</td>
</tr>
<tr>
<td>Spec. (4)</td>
<td>97.5%</td>
<td>99.6%</td>
<td>99.8%</td>
<td>99.9%</td>
<td>99.9%</td>
</tr>
<tr>
<td>Sens. (5)</td>
<td>99.1%</td>
<td>96.4%</td>
<td>93.8%</td>
<td>97.9%</td>
<td>99.9%</td>
</tr>
<tr>
<td>Spec. (5)</td>
<td>98.6%</td>
<td>99.3%</td>
<td>99.6%</td>
<td>99.9%</td>
<td>99.9%</td>
</tr>
<tr>
<td>Sens. (6)</td>
<td>99.1%</td>
<td>97.3%</td>
<td>95.3%</td>
<td>98.2%</td>
<td>99.8%</td>
</tr>
<tr>
<td>Spec. (6)</td>
<td>98.5%</td>
<td>99.4%</td>
<td>99.6%</td>
<td>99.9%</td>
<td>99.9%</td>
</tr>
<tr>
<td>Sens. (7)</td>
<td>99.0%</td>
<td>97.3%</td>
<td>93.3%</td>
<td>98.3%</td>
<td>99.9%</td>
</tr>
<tr>
<td>Spec. (7)</td>
<td>98.9%</td>
<td>99.0%</td>
<td>99.6%</td>
<td>99.9%</td>
<td>99.9%</td>
</tr>
<tr>
<td>Sens. (8)</td>
<td>99.3%</td>
<td>97.3%</td>
<td>93.6%</td>
<td>99.3%</td>
<td>99.9%</td>
</tr>
<tr>
<td>Spec. (8)</td>
<td>99.2%</td>
<td>99.5%</td>
<td>99.8%</td>
<td>99.9%</td>
<td>99.7%</td>
</tr>
<tr>
<td>Sens. (9)</td>
<td>99.0%</td>
<td>96.7%</td>
<td>95.2%</td>
<td>98.0%</td>
<td>99.9%</td>
</tr>
<tr>
<td>Spec. (9)</td>
<td>98.6%</td>
<td>99.3%</td>
<td>99.7%</td>
<td>99.9%</td>
<td>99.9%</td>
</tr>
</tbody>
</table>

Case (1): 6 atoms/class; 22 features; s8 wavelet; 64 samples
Case (2): 8 atoms/class; 31 features; s8 wavelet; 64 samples
Case (3): 10 atoms/class; 39 features; s8 wavelet; 64 samples
Case (4): 8 atoms/class; 29 features; s8 wavelet; 128 samples
Case (5): 10 atoms/class; 32 features; s8 wavelet; 256 samples
Case (6): 10 atoms/class; 32 features; d8 wavelet; 64 samples
Case (7): 10 atoms/class; 32 features; c3 wavelet; 64 samples
Case (8): 15 atoms/class; 52 features; s8 wavelet; 64 samples
Case (9): same as case (9) only PCA was employed to reduce the number of features to
As a result, extracting just 64 samples around the R peak, which corresponds to about a 178ms window for 360Hz sampling frequency, was proven sufficient for achieving satisfactory performance. In addition to that, less samples translates to less computations as we will see in the next section.

We can also see that utilizing 10 atoms to approximate the beats in each class leads to the best overall performance, while using significantly more atoms, as in the last 2 cases, did not lead to better results (even when we employed Principal Component Analysis to reduce the number of features). Additionally, it is also verified that employing Symmlet wavelet packets yields the best performance compared to other wavelet types, at the expense of employing more wavelet packets than in the other 2 cases, e.g. 39 versus 32.

Table 3-3. Confusion Matrix for the case of 7 classes

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>PVC</th>
<th>LBBB</th>
<th>RBBB</th>
<th>PACED</th>
<th>PFUS</th>
<th>APC</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>36644</td>
<td>62</td>
<td>54</td>
<td>59</td>
<td>1</td>
<td>35</td>
<td>170</td>
</tr>
<tr>
<td>V</td>
<td>236</td>
<td>3420</td>
<td>21</td>
<td>4</td>
<td>0</td>
<td>70</td>
<td>77</td>
</tr>
<tr>
<td>L</td>
<td>247</td>
<td>25</td>
<td>3958</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>R</td>
<td>36</td>
<td>4</td>
<td>0</td>
<td>3557</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>P</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1805</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>106</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>262</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>83</td>
<td>12</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>981</td>
</tr>
<tr>
<td>SE%</td>
<td>98.1%</td>
<td>96.9%</td>
<td>98.1%</td>
<td>98.2%</td>
<td>99.8%</td>
<td>71.4%</td>
<td>79.6%</td>
</tr>
<tr>
<td>SP%</td>
<td>97.4%</td>
<td>99.2%</td>
<td>99.4%</td>
<td>99.9%</td>
<td>100%</td>
<td>99.8%</td>
<td>99.8%</td>
</tr>
</tbody>
</table>
Finally an additional test was performed attempting to consider also the case of APC and PFUS beats by using 64 samples/beat and 8 atoms/class selected from the s8 wavelet family (requiring 34 features). Table 3-3 presents the corresponding confusion matrix, yielding degraded sensitivity for the two additional beat types, as PFUS and APC tend to be misclassified as either normal or PVC.

To further evaluate the performance of the proposed ECG beat classification scheme, we will now consider some of the most popular approaches found in the literature for ECG classification in order to compare the performance of the described scheme with other state-of-the-art ECG classifiers. However, one thing that needs to be mentioned at this point is that although the MIT-BIH Arrhythmia database has been adopted by the majority of the researchers working on ECG classification, the comparison metrics, the number of beats used and the number and types of arrhythmic beats that are classified tend to vary significantly. For example, about 70% of the ECG beat annotations are Normal Beats [87]. As a result, some researchers have chosen to utilize more balanced datasets in their work, by including almost same number of beats that they are interested in classifying [116]. Consequently, such an approach cannot yield directly comparable results since they are based on significantly different training and testing subsets of the database.

Table 3-4 compares the classification accuracy (for the five main beat classes considered in our work e.g. normal, premature ventricular contraction,
left bundle branch block, right bundle branch block and paced beat) of our work with other popular approaches. We are aware that there are other advanced methods that can be employed for ECG classification [100], but the purpose of our comparison here is not to strictly prove which classifier derives the most accurate performance, but to indicate that after extensive training and tests, the proposed approach, as simple as it is, it derives classification results compared to high-accuracy algorithms while at the same time being computationally inexpensive so that it may be easily implementable and executable in real-time on a resource constrained device such as a smart-phone.

Table 3-4. Classification accuracy comparison for the five main beat classes considered in our study.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>PVC</th>
<th>LBBB</th>
<th>RBBB</th>
<th>Paced</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Our Method</strong></td>
<td>99.1%</td>
<td>97.3%</td>
<td>95.3%</td>
<td>98.2%</td>
<td>99.8%</td>
</tr>
<tr>
<td>Christov et al. [112]</td>
<td>98.9%</td>
<td>94.8%</td>
<td>98.9%</td>
<td>98.7%</td>
<td>99.9%</td>
</tr>
<tr>
<td>Herrero et al. [115]</td>
<td>99.7%</td>
<td>95%</td>
<td>99%</td>
<td>99.6%</td>
<td>99.6%</td>
</tr>
<tr>
<td>Oresko et al. [111]</td>
<td>99%</td>
<td>92.6%</td>
<td>-</td>
<td>98%</td>
<td>96%</td>
</tr>
<tr>
<td>Suarez et al. [117]</td>
<td>99.5%</td>
<td>81.6%</td>
<td>99.4%</td>
<td>98.9%</td>
<td>99.5%</td>
</tr>
<tr>
<td>Yu et al. [118]</td>
<td>99.7%</td>
<td>98.5%</td>
<td>96.3%</td>
<td>99.2%</td>
<td>99.4%</td>
</tr>
<tr>
<td>Rodriguez et al. [110]</td>
<td>97.9%</td>
<td>92.8%</td>
<td>98.9%</td>
<td>96.7%</td>
<td>98%</td>
</tr>
<tr>
<td>de Chazal et al. [87]</td>
<td>99.2%</td>
<td>81.2%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Osowski et al. [113]</td>
<td>98.1%</td>
<td>96.6%</td>
<td>97%</td>
<td>94%</td>
<td>-</td>
</tr>
</tbody>
</table>
3.3.5 Conclusions

To conclude our contribution in this section, we have studied the effectiveness of using the projections of ECG beat samples on wavelet packet atoms obtained from the Matching Pursuit algorithm, as features for ECG classification. We examined the effects of the number of atoms, number of data samples, of the wavelet type and of the number of classes in the performance of an ANN-based classifier. Our goal was to derive a compact set of features, which can be easily computed from a single-lead ECG, to be used as inputs of a single MLP network.

Compared to the approach of Herrero et al. [115] our approach achieves comparable classification accuracy, with considerably less features and less computational effort, for all beat types, except for the LBBB class where the sensitivity is at best 95%. This highlights the fact that to better identify the beats of this class two leads might be more useful.

Calculation of the projection coefficients is achieved by simple matrix multiplication and is proportional to the number of features used (which depends on the number of classes, the type of wavelet and the number of atoms) and to the number of data samples extracted around the R peak. For example, in the case (3) of Table I, to project each beat we need to multiply a 1×64 data vector with a 64×39 atom matrix, which requires 39×64=2496 multiplications and 39×63=2457 additions. Then to classify the beat we need to feed it to a 39-5
MLP, which requires $39 \times 39 = 1521$ multiplications and $39 \times 39 = 1521$ additions in the first layer and $5 \times 39 = 195$ multiplications and as many additions in the second layer (plus applying the hyb. tangent 44 times). In total we have: 4212 multiplications and 4173 additions.

In the case of [115], 11 features are extracted with 256 samples per beat and per lead, leading to 110 features per beat, while 5 different ANNs are required for classification. In total, if we follow all the calculations, we see that there are totally 147345 multiplications and 149785 additions required (plus applying the hyb. tangent 830 times), which are about 35 times than in our case, a fact which highlights the effectiveness of our proposed approach.
CHAPTER 4: THE FUNCTIONAL DESIGN FRAMEWORK OF
THE PROGNOSIS WEARABLE HEALTH MONITORING SYSTEM

In this chapter we introduce the concept of the proposed Prognosis wearable health monitoring system by providing an in-depth analysis of the underlying functional design framework and reasoning mechanism. First a generic modular architectural model of a multi-sensor wearable health monitoring system will be presented and discussed in order to identify the key hardware and software components of a WHMS. Then, a novel physiological data fusion strategy based on a context-free formal language will be presented. Finally, we will provide a detailed functional description of the Prognosis WHMS in terms of a Stochastic Petri Net (SPN) model, which is able to capture and illustrate the concurrent and synergistic tasks that are present in the proposed system.

4.1 Generic WHMS Architecture

In this subsection we will discuss a generic architecture for wearable health-monitoring systems in order to provide more insight into the specific hardware and software components, of which an intelligent multi-sensor personal health system is composed. We will be using this architectural model
throughout this chapter as the starting point or reference, on which the novel components that are presented in this dissertation will be added.

The generic architectural model is depicted in Fig. 4-1. Physiological biosensors constitute the front-end components of the system and they can be employed to measure a variety of bio-signals produced by the human body, including vital signs such as blood pressure, pulse rate, body temperature and respiration rate as well as other physiological signals like electrocardiogram, electromyogram and oxygen saturation. A comprehensive list of currently available biosensors along with the type of physiological data they are used to measure was given in Table 2-1.

![Generic WHMS architecture](image-url)

Fig. 4-1. Generic WHMS architecture
However it is not within the scope of this research to investigate new sensing technologies or to analyze current ones. In contrast to that, the work presented and proposed here is based on current sensor availability, however it is scalable and easily expandable and could thus incorporate new sensing technologies in the future.

As it was seen in the various systems that were reviewed in the second chapter, the wearable physiological sensors can be either embedded in clothing as smart textiles or they can be integrated on other types of wearable devices, such as wrist devices, ear-lobe sensors, finger sensors, arm bands, chest belts, waist belts etc. In the latter case, the distributed bio-sensors are capable of wirelessly communicating their measurements and thus constitute a Body Area Network (Fig. 1-1), which can be either formed through Bluetooth enabled devices or through Zigbee motes. Basic signal conditioning operations such as filtering, amplifying and analog-to-digital conversion or even basic feature extraction is usually performed by dedicated hardware, which is either embedded on the sensor as a single integrated chip (IC) or on the central node.

The central node of the WHMS consists of some type of portable platform, such as a personal digital assistant (PDA), smart-phone, pocket PC or even a custom designed microcontroller board. In either case, the WHMS central node is responsible for several tasks:
1. Handling the communication with the on-body distributed biosensors, which involves: collecting physiological measurements, communication synchronization, sending control signals for adjusting sensors’ parameters, e.g. sample rate, accuracy etc and finally also receiving sensor status data.

2. Performing additional digital signal processing on the acquired signals for feature extraction.

3. Verifying the received data, e.g. checking the validity of the received data via an advanced algorithm and discarding those that are found to be erroneous.

4. Comparing the extracted features or values from each signal with the thresholds, limits or patterns located in the local signal database, which may contain patient-specific information about abnormal states, in order to possibly detect any health risks. This task is what we will be referring to as embedded decision support.

5. Generating alarm signals for the user.

6. Displaying the collected measurements on the Graphical User Interface (GUI) in real-time.

7. Transmitting the extracted medical information about the user to a remote medical station, e.g. to a medical center or to a physician’s cell phone, either in real-time or in terms of report forms upon request or upon detection of events.
The functional description of the WHMS given above pertains to the majority of the up-to-now developed wearable system prototypes or commercially available products in the field of remote health monitoring. However, features 3 and 4 of the central node, e.g. validity check and **embedded decision support**, as seen in the previous chapter, are either barely or partially supported in most systems, for example by employing simple if-then rules to provide notifications when certain measurements exceed certain predefined thresholds.

In addition to the previously mentioned WHMS components, we propose to include an additional feature in the system design in order for the system to be able to get additional feedback from the user about non-measurable symptoms, namely to add a **voice-driven interaction** mechanism between user and WHMS. This is shown graphically in Fig. 4-1 where the system is capable of verbally interacting with the user by means of a speech recognition/synthesizer component and an automated dialogue system. The motivation behind this approach is illustrated in Fig. 4-2.
Namely, the fact is that for an accurate estimation of one’s health condition and the diagnosis of many, if not the most, diseases several other symptoms than just the ones detected from biosensor measurements, need to be taken into consideration [119], [120], [121]. These symptoms, like cough, malaise or chest discomfort for example, are either not measurable at all or they cannot be estimated without using invasive methods, e.g. as in the case of determining electrolyte levels in the body.

Coming back to Fig. 4-2, the green frame includes a comprehensive list of most of the physiological parameters, which can be measured non-invasively and which need to be taken into consideration and properly evaluated in order to diagnose a patient. This list is not exhaustive and it does not
include findings, which can only be obtained from thorough clinical examinations and tests like MRI, CT scan, chest radiology and other medical and laboratory examinations typically performed in a hospital. By using bio-sensor technologies such as those listed in Table 2-1, the symptoms listed in the light blue frame in Fig. 4-2 can be detected (symptoms 1-11).

Entries 12 and 13 in the green frame and all the entries in the red frame in Fig. 4-2 constitute additional symptoms and physical signs, which are associated with a great variety of diseases. These symptoms, once detected or quantified, provide important information, which together with the measured vital signs provide a more comprehensive description of what is referred to as the clinical presentation, which under proper interpretation may lead to a specific diagnosis. However, in order to get feedback from the patient about the possible existence of these symptoms either the patient himself has to describe them or in case of some of these physical signs (12 and 13), they can only be measured in an invasive way using current sensor technologies.

As a result, by incorporating a speech recognition module in the system design along with an automated speech dialogue system between the device and the user, additional non-measurable symptoms related to the physical condition of the patient could be captured by the WHMS.

Finally, returning to Fig. 4-1, alarm signals and measured physiological data along with the feedback from the patient can be transmitted through the
cellular network or the Internet to the medical center and possibly also to a
dispatched ambulance in case of an emergency, as it was also described in Fig.
1-1. As the healthcare center keeps a database with the long-term detailed med-
ical history of the patient, the received data and the detected patient symptoms
can be put into a wider context and can also be further evaluated by a supervis-
ing physician to derive a more accurate estimation of the patient’s health.

4.2 The Prognosis Formal Language

The Prognosis language is a theoretical model, around which the weara-
ble monitoring and early prognosis system has been designed. The basic hypo-
thesis of this model is that the various body or physiological signals produced
by the human body are composed of “symptoms of health” whose occurrence
under certain conditions may indicate the presence of a specific health risk. The
aim of the Prognosis formal language is to provide an efficient and compact re-
presentation of the multiple combinations of extracted physiological measure-
ments in order to aid in the association of “pathological” symptoms and pat-
terns with the detection or estimation of a corresponding health condition.

The proposed formal language model is coupled with the generic
WHMS architecture described in the previous section. Specifically, the sensors
that are included in the WHMS provide real-time measurements of physiologi-
cal data, from which corresponding symptoms of health are extracted. These
symptoms may be considered as normal (benign) or alarming (e.g. dangerous or hazardous). However the degree of dangerousness (or severity) and the degree of actual occurrence of a specific symptom are fuzzy in nature [122]. As the philosopher of medicine Kazem Sadegh-Zadeh has stated in [123]: “health is a matter of degree, illness is a matter of degree and disease is a matter of degree”. Furthermore, the founder of fuzzy logic, Lofti A. Zadeh was the first one to suggest that a fuzzy set could be used to denote the degree of relation of a disease to various symptoms, which are also fuzzy in nature as well [124].

As a result, the fuzzy symptoms that can be extracted from the physiological sensors generate the set of terminal symbols of the Prognosis formal language. The three basic types of symptoms that can be extracted from the acquired physiological signs correspond to the three basic types of signals or information the system is able to collect about the patient, namely: a) signals that are “value-specific”, e.g. their instantaneous value carries the actual diagnostic content, b) signals that are “morphology-specific”, e.g. their structural morphology and timing are the elements that carry important diagnostic information and c) voice recordings that may reveal the presence of a specific health symptom as described by the user.

In the following subsections we will first describe how fuzzy symptoms are extracted from each family of collected data and how these symptoms are finally described in terms of terminal (fuzzy) symbols in the formal language.
This section will finally be concluded by giving a formal definition of the Prognosis formal grammar and by providing simple illustrative examples to show the language’s operation.

4.2.1  *Category of value-specific physiological symptoms*

The most typical and common physiological data, which can indicate the presence of symptoms that fall in this category, are systolic and diastolic blood pressure, respiration rate, body temperature, glucose level, heart rate and oxygen saturation. These types of physiological signals are characterized as value-specific because according to their current value we may extract a certain symptom (or decide that the signal level is not alarming). For example, a body temperature measured over 100°F is normally considered as fever (or light fever) [120]. Furthermore, a heart rate under 50 BPM (beats per minute) is considered as abnormally low.

The value range of this category of physiological data can be divided in several levels of importance. For example, blood pressure measurements are commonly classified as indicating *hypotension* (low blood pressure), a normal blood pressure level, *prehypertension* (slightly high blood pressure), *stage 1 hypertension* (high blood pressure) and *stage 2 hypertension* (very high blood pressure) [120]. This categorization of signal values into levels of importance serves the purpose of indicating the severity level of a physiological measure-
ment and also the purpose of formally describing the progression of a possible abnormal health incident. A rather simple and straightforward manner according to which signal levels could be characterized is depicted in Fig. 4-3, e.g. a value can be characterized as being either normal, low or high or even very low or very high according to crisp thresholds.

![Fig. 4-3. Crisp characterization of hypothetical signal values](image)

However, the characterization of signal values in a crisp manner as shown in Fig. 4-3 is rather subjective and strict and also it does not capture the level of progression of a specific symptom. A better alternative is to employ linguistic variables, e.g. fuzzy sets [125], to describe the degree of occurrence of a certain symptom. This approach has also been adopted successfully in several other works concerned with *computer aided medical diagnosis*, such as in the research efforts presented in [126], [127]. We should also stress the fact here that it is generally accepted that there are no universal and fixed thresholds for characterizing signals levels as being normal or abnormal or very risky etc [120]. On the other hand, some rules of thumb or generally accepted signal limits do
exist. However, in this research we want to consider the fact that different individuals due to their individual physiology and medical history, nutrition, genes etc may have different normal signal values. As a result, the thresholds for characterizing signal levels will be adaptable in the system and not hard-coded. This issue will be thoroughly discussed and addressed in Chapter 6.

Finally, coming back to the fuzzy characterization of health symptoms extracted from raw physiological measurements, we can make use of the approach presented in Fig. 4-3 and then by fuzzifying the limits between signal levels we can get the corresponding fuzzy symptoms. Fig. 4-4 through Fig. 4-9 depict how fuzzy symptoms are extracted from the systolic and diastolic blood pressure and from the heart rate, respiration rate, body temperature and oxygen saturation levels.

Fig. 4-4. Fuzzy symptoms extracted from systolic blood pressure.

Fig. 4-5. Fuzzy symptoms extracted from diastolic blood pressure.
Fig. 4-6. Fuzzy symptoms extracted from the heart rate level.

Fig. 4-7. Fuzzy symptoms extracted from the respiration rate level.

Fig. 4-8. Fuzzy symptoms extracted from the body temperature level.

Fig. 4-9. Fuzzy symptoms extracted from the oxygen saturation level.

The reason for deciding to define symptoms as fuzzy sets on the range of measurements of each bio-signal will be made obvious later on when we de-
fine the *Prognosis* formal language, where the fuzziness of the symptoms will help us determine a confidence level (or “degree of support” as stated in [126]) for the currently estimated health state of the patient.

We should also note here that the membership functions chosen in the fuzzy sets shown in Fig. 4-4 through Fig. 4-9 are trapezoidal. However this is just an assumption that we adopt in our model since there are numerous types of different memberships functions that can be employed for describing a fuzzy set [125], for example triangular, sigmoidal or Gaussian. For now we can assume that these functions are trapezoidal, however there are techniques that can be used to derive an optimal estimation of the shape of the membership functions for a given problem as we will explain in chapter 6.

### 4.2.2 Category of morphology-specific physiological symptoms – ECG

This category includes body signals, which describe the electrical activity of various body parts, amongst which the most common ones are the electrocardiogram (ECG), the electroencephalogram (EEG) and the electromyogram (EMG). Detecting healthy and pathological symptoms in such kind of signals is a complicated process and it requires careful conditioning of the signal (e.g. filtering, amplifying etc) and intelligent signal processing for feature extraction [96].
In this work we will only consider electrocardiographic signals. ECG’s are a well-studied medical tool, which has also been widely employed in WHMS (as we saw in Chapter 2). Furthermore the automatic detection of heart beats, the extraction of ECG related parameters and the classification of beats and rhythms as normal or arrhythmic has been widely researched by the engineering and medical community (as discussed in Chapter 3). Furthermore in Chapter 3 we presented a methodology for classifying ECG beats on a resource constrained device, while achieving close to real-time performance and classification accuracy within medically acceptable margins.

In any case, the goal of the ECG processing stage is to derive a confidence grade for the detected type of heart beat, which will be utilized by the Prognosis formal language in the estimation of the user’s health condition.

4.2.3 Category of non-measurable symptoms

This category of symptoms includes conditions and information which is not measurable and not-quantifiable by conventional bio-sensors of any type. As mentioned in the beginning of this chapter, patient symptoms such as malaise, headache, dizziness, chest-pain, sputum, disorientation, back pain, running nose etc are conditions that cannot be measured by sensory devices. As a result, only the patient himself can provide information of this sort to the diagnostician or to the wearable system via a question-answer process.
For example, a patient may signify during an examination that he has been experiencing chest discomfort. The physician will then try to acquire detailed information regarding that symptom e.g. does the symptom show instantaneous or periodic occurrence, is it mild or severe, is the symptom aggravated by movement, are there any other relevant symptoms etc. This extra information can aid the physician in diagnosing the cause of the corresponding condition of the patient.

As it will be discussed in chapter 7, we propose the integration of a user-device dialogue scheme in order for the WHMS to be able to capture and also to take into consideration such non-measurable symptoms. For example the user may indicate the presence of a condition, such as headache for example and then the system could proceed to ask the user to quantify that symptom, e.g. as being benign, mild or weak, moderate, strong or severe. These linguistic variables describing the severity level of the corresponding symptom may be translated to a score in the [0, 1] scale in order to provide a quantitative estimation of the symptom’s strength. This quantification of the symptom’s strength can be utilized from the Prognosis language in the process of estimating the health status of the user with a specific degree of confidence, a process that will be explained thoroughly in the subsequent sections.
4.2.4  The Prognosis Formal Grammar

4.2.4.1  Basic Definitions

As we know a formal language [128] is produced by a grammar which is defined as:

Definition 1: A formal grammar $G$ is defined as: $G = (V, \Sigma, P, S)$, whereby: $V$ is the set of non-terminals or variables, $\Sigma$ is the alphabet, $P$ are the production rules and $S$ is the start symbol.

In addition to that the definition of a regular language is:

Definition 2: A regular language is a formal language produced by a grammar $G$, whose production rules $P$ have the form: $A \rightarrow \alpha B | \beta | \lambda$, where: $\alpha, \beta$ are terminal symbols, $A$ and $B$ are non-terminal symbols and $\lambda$ is the empty symbol.

We also know that a regular language is equivalent to a Finite State Machine:

Definition 3: A Finite State Machine (FSM) is denoted $M = (Q, \Sigma, \delta, q_0, F)$, where: $Q$ is a finite set of states, $\Sigma$ is a finite alphabet, $\delta$ is a transition function from $Q \times \Sigma$ to $Q$, $q_0$ is the initial state and $F$ is the set of final/accepting states.

Extending the above definitions to fuzzy regular languages and fuzzy automata as described in [129], we can define a fuzzy finite automaton as:

Definition 4: A fuzzy automaton is a 7-tuple $A = (Q, \Sigma, \delta, q_0, F, \varphi, \oplus)$, where: $(Q, \Sigma, \delta, q_0, F)$ is a crisp automaton, $\varphi: (Q \times \Sigma \times Q) \rightarrow [0,1]$ associates a weight to every transition and $\oplus$ is a t-norm.

In the case of fuzzy finite automata as defined above, we can also see the equivalence between a FSM and a regular language. Namely, the above fuzzy automaton is equivalent to a fuzzy regular language, which has a weight
associated with every transition. At the end of a string production, the derived word has a membership degree associated with it, which is obtained by applying the t-norm on the sequence of derivations.

Adopting the fuzzy automaton model for our case and defining Prognosis as a fuzzy regular language, we can give the following definition:

**Definition 5:** The set of states $Q$ or equivalently the set of non-terminals $V$ denotes the set of all possible health states of the patient/user. These states are defined as the states resulting from all possible combinations of health symptoms which are extracted from the measured physiological data, from the user’s context and from the user’s non-measurable symptoms. There are no explicitly defined accepting/final states as any state included in the FSM signifies a possible health state of the user and as such we will always consider a continuous transition (trajectory) between states from the moment the system is turned on until it is turned off.

**Definition 6:** The alphabet $\Sigma$ consists of the set of all observable symptoms (and contexts) in the system. Examples of these symptoms include: tachycardia, hypertension, fever, high respiration rate, low oxygen saturation, ectopic heart beat, abnormal heart rhythm, cough, chest-pain, lying on back, running etc. These symptoms are defined as fuzzy (linguistic) variables and each one has a degree of membership (DOM) $0 \leq \mu(i,j,x) \leq 1$ associated with it, where:

1. $\{\text{set of all physiological signals measured by the WHMS}\}$
2. $\{\text{set of fuzzy symptoms that can be extracted from the } i^{th} \text{ sensor}\}$

and $x$ is the actual measured value. The DOM denotes the certainty or strength of the corresponding symptom.

**Definition 7:** The start symbol $S$ or the initial state $q_0$ signifies the initial (normal) health state of the user.
Finally, before moving on to formally defining the set of production rules of Prognosis (or equivalently the transition function in the FSM representation) we need to consider how diseases or disorders are usually specified in medical terms and how their correlation to a variety of symptoms is analyzed by diagnosticians. As it is described in [122], a general outline of a systematic diagnostic approach is:

1. Obtain case facts (physiological measurements, symptoms, medical history etc).
2. Evaluate the importance of different signs and symptoms.
3. Make a differential diagnosis, e.g. list all possible diseases which the specific case can reasonably resemble.
4. Try to exclude diseases by further evaluation of additional signs and symptoms.

Following that way of thinking we may be tempted towards trying to imitate that process in deriving a specific diagnosis. However this is not the goal of the proposed model. What we are actually aiming at here is a way of following the progression of health symptoms so as to be able to derive at any given time an estimation of the user’s health condition in order to possibly detect health risks by identifying dangerous health trends.

To be able to do that we need to embed knowledge into our model regarding how the occurrence of several symptoms is related to a variety of
disorders and to what degree the presence of a specific symptom under a certain context points towards a specific medical disorder or health state. This is what is commonly regarded as medical knowledge and it is the element that will help us determine the confidence factors relating the occurrence of a symptom to the detection of a certain disorder. In other words, medical knowledge will aid us in defining the approximate values of the connection weights or equivalently the degrees of confidence associated with each production rule in the language or with each transition in the FSM representation of the model.

4.2.4.2 Some notes on Medical Expert Systems and Medical Knowledge Representation

As we explained, gathering medical knowledge in order to establish a well-defined, easy-to-access and standardized knowledge representation scheme is of crucial importance to the current research goals of this dissertation. Before elaborating on this part, we need to briefly mention some research efforts and corresponding decision support systems (DSS) which constitute pioneer works in the field of medical expert systems and which have greatly inspired the author in his current research.

One of the earliest and most influential medical expert systems was MYCIN [130], [131] and particularly its essential version called EMYCIN [132], [133]. MYCIN was a rule-based systems using weighted rules and combining functions to calculate the global degree or certainty of suggested diagnoses. It
was used to diagnose several types of infections and to suggest treatment options. Two basic measures of uncertainty were used in MYCIN, the *measure of belief* \( MB[h,e] \) which expresses the degree to which an observed piece of evidence \( e \) increases the belief in a hypothesis \( h \) and the *measure of disbelief* \( MD[h,e] \) that expresses the degree to which an observed piece of evidence \( e \) decreases the belief in a hypothesis \( h \):

(Eq. 4-1) \[
MB[h,e] = \begin{cases} 
1, & \text{if } P(h) = 1 \\
\frac{\max[P(h|e)P(h)] - P(h)}{\max[1,0] - P(h)}, & \text{otherwise}
\end{cases}
\]

(Eq. 4-2) \[
MD[h,e] = \begin{cases} 
1, & \text{if } P(h) = 0 \\
\frac{\min[P(h|e)P(h)] - P(h)}{\min[1,0] - P(h)}, & \text{otherwise}
\end{cases}
\]

Where \( P(h|e) \) and \( P(h) \) represent conditional and a priori probabilities and the final certainty factor of the hypothesis \( h \) given the evidence \( e \) is given by:

(Eq. 4-3) \[
\]

CADIAG-II (Computer Assisted DIAGnosis) [133] - [135] was an expert system for internal medicine, aiding physicians in diagnoses and utilizing fuzzy set theory and the compositional rule of inference by Zadeh [124]. Adlassnig [134], [135] defined two fuzzy relations to describe the “medical knowledge” connecting symptoms, findings, test results and disorders: occurrence \( \mu_{SD}^0 \) (of frequency of occurrence) and confirmability \( \mu_{SD}^c \) (how strongly does a symptom confirm a disorder). These relations or metrics can be quantified by
the following formulas which are based on specifying relative frequencies ($F(\cdot)$ denotes frequency of an event):

(Eq. 4-4) \[ \mu_{SD}^a = F(S|D) = \frac{F(S \cap D)}{F(D)} \]

(Eq. 4-5) \[ \mu_{SD}^b = F(D|S) = \frac{F(S \cap D)}{F(S)} \]

These functions could be determined by (a) linguistic documentation by medical experts and (b) medical database evaluation by statistical means (as described by the two formulas above), or a combination of both. Adlassnig used max-min inference rules to quantify the degree of membership of a given patient $P$ to a specific disorder $D$. By doing that one could distinguish between three resulting qualities of examined diagnoses: confirmed diagnoses, excluded diagnoses and generated diagnostic hypotheses.

As it is shown by the example of the two medical expert systems discussed above, fuzzy quantities utilized in fuzzy reasoning can either be specified using statistical means or by expert-knowledge acquisition. In many cases, such as the ones addressed by the EMERGE expert system for evaluating diagnoses of patients suffering from chest pain and the Hypernet expert system for diagnosing heart disease (reviewed by Hudson et al. in [136]) there might be labeled experimental data available to construct the rule base. Large evidence bases are required especially when parameters such as the ones described by (Eq. 4-1) - (Eq. 4-5) are to be determined, since these metrics are statistical in
nature. However, as it is elegantly emphasized by Lovell et al. [137], “the largest hurdle facing the design and validation of the DSS for monitoring chronic disease patients is a lack of an evidence-base for training such a system”. To put it in different words, there is a serious lack of an evidence-base with labeled physiological training data in order to derive accurate estimates of the quantities that were defined above for the case of unsupervised physiological monitoring.

In order to construct such a knowledge-base, one would need to monitor several patients over an extended period of time while working with an expert/physician who would be responsible for labeling the acquired multidimensional data and quantifying the patients’ health condition. Then by having such data at our disposal, we could not only apply approaches such as the ones adopted in MYCIN and CADIAG-2, but one could also train statistical pattern recognition schemes [138] (or well-studied probabilistic knowledge-representation models such as Bayesian Networks and Markov Models) in order to enable the system to accurately recognize certain deterioration events as they are occurring. Again as Lovell et al. denote [137]: “Given the obstacles involved in designing a DSS for telehealth applications, a practical approach will likely involve a combination of a heuristic system design, inspired by existing guidelines and clinical expertise, which facilitates incremental improvement through user feedback and ongoing assessment of classification performance.”
As a result, the approach adopted in this research is to embed medical knowledge in the WHMS in terms of documented medical experience and guidelines in the basic form of IF-THEN rules. However in the next two chapters we will elaborate more on how initial choices can be made for several rules, the parameters of which can be then fine tuned during a patient-learning process. For now, we will focus on describing how medical knowledge is usually found in common medical textbooks (or even how it is captured by professional physicians), how we can capture this information and represent it in a standardized manner and finally how information relating symptoms and disorders can be integrated in our model. To illustrate the above point we will use the example of a relatively common disorder, e.g. pneumonia. The medical knowledge that will be referenced in this part is taken from [120]. Summarizing the knowledge and the sign and symptom description regarding pneumonia we can define the following knowledge structure:

**Disorder type:** Pneumonia

**General Information**

- Frequency of occurrence: common
- Risk level: very high

**Risk factors**

- Alcoholism
- Respiratory rate ≥ 30 breaths/min
- Hypotension

**Typical setting**

- History of predisposition to aspiration <sometimes>

**Clinical Presentation and findings (Symptoms)**

- Acute or subacute onset of fever <usually>
- Cough <usually>
- Purulent sputum <usually>
Dyspnea <usually>
Rales <very common>
Sweat <common>
Chills <common>
Chest Discomfort <common>
Hemoptysis <common>
Fatigue <common>
Anorexia <common>
Headache <common>
Abdominal pain <common>
Hypothermia <common>
Tachypnea <common>
Tachycardia <sometimes>
Mild oxygen desaturation <sometimes>
Malaise <sometimes>
Weight loss <sometimes>

Differential Diagnosis
Upper respiratory tract infections
Reactive airway diseases
Congestive heart failure
Pulmonary hemorrhage
...

It is evident from the structure of the above data that such type of medical knowledge can be captured and stored using an object markup language such as XML. An example representation of the above medical knowledge is given below in XML representation (where for simplicity we have not defined a schema and the corresponding required tags):

<Disorder type="Pneumonia">
  <info>
    <frequency>common</frequency>
    <risk_level>high</risk_level>
  </info>
  <risk_factors>
    alcoholism##high respiration rate##hypotension
  </risk_factors>
  <setting frequency="usually">
    predisposition to aspiration@sometimes
  </setting>
</Disorder>
<symptoms frequency="usually">
  fever##cough##purulent sputum##dyspnea##
</symptoms>

<symptoms frequency="very common">
  rales
</symptoms>

<symptoms frequency="common">
  sweat##chills##hemoptyis##fatigue##
  anorexia##headache##abdominal pain##hypothermia##
  tachypnea##chest discomfort
</symptoms>

<symptoms frequency="sometimes">
  tachycardia##mild oxygen desaturation##
  malaise##weight loss
</symptoms>

<diff_diagnosis>
  upper resp tract infection##reactive airway disease##
  congestive heart failure##...
</diff_diagnosis>
</Disorder>

By repeating the above process for every disorder of interest we can create a standardized knowledge base using the same template, which allows for quick parsing and searching. The way this knowledge base can be utilized to determine the dependence or relation between a certain disorder and a set of manifestations (symptoms) is by using linguistic variables. For example, in the above case of pneumonia there are certain symptoms that give a stronger indication of pneumonia and others that might or might not be present at all, thus in a sense they are “contributing less” to the occurrence or the confirmation of that specific disorder. More specific, if the following symptoms are observed: headache, anorexia and tachycardia, which are labeled as being either common or less common (e.g. “sometimes”) symptoms in the case of pneumonia, we
will only have a very weak indication that the patient is suffering from pneumonia. However in the presence of fever, cough and dyspnea there will be a very strong indication towards that specific disorder.

The previous example serves the purpose of illustrating the process of determining causal associations between symptoms and disorders and assigning corresponding weights to these relationships. We can consider the following causal relevance levels: never, sometimes (occasionally), often, usually and always. This set of levels is practically a set of linguistic variables and can be assigned a corresponding causal relevance grade. An example of how this could be done is given in Fig. 4-10 (a similar approach to what is proposed in [126]).

Another example can be given in the case of hypothermia. Clinically it is described as the reduction of the body temperature below 35°C (95°F) and it includes several stages. The low body temperature value is a symptom that
needs to be present in order to be able to diagnose hypothermia. Other clinical findings that further suggest the presence of hypothermia are: bradycardia, bradypnea, hypotension and the onset of atrial (or even ventricular) fibrillation. However in the case that some of these “side-symptoms” have occurred, but decreased body temperature has not been observed, we may not come to the conclusion that the patient is experiencing hypothermia. Of course there might still be some other hazardous health condition(s) that is present and that could be the cause of these observed symptoms.

4.2.4.3 Production Rules and the Language’s Operation

In the previous section we described intuitively how causal weights between symptoms and disorders are utilized together with the level of fuzziness of the corresponding symptoms (as described in the previous subsection) in order to derive a quantitative estimation about our confidence regarding the presence of a specific health state. These weights constitute the function \( \varphi: (Q \times \Sigma \times Q) \rightarrow [0,1] \) that was first introduced in Definition 4: and which assigns a weight to every possible transition in the FSM. More formally we can give the following definition:

**Definition 8:** The weighting function \( \varphi: (Q \times \Sigma \times Q) \rightarrow [0,1] \) associates a weight to every transition or production rule in the language and represents the causal associations between symptoms and disorders/health states.
Finally, coming back to the definition of the Prognosis fuzzy regular language (or the equivalent fuzzy FSM), we still need to define two more elements in order to provide a complete definition of our proposed model: a) what are the production rules \( P \) (or the transition function \( \delta \)) and b) how the causal weights and the fuzziness of the detected symptoms are fused together to derive a confidence level (or degree of membership) at a given derivation level of the language.

**Definition 9:** The production rules \( P \) (and equivalently the transition function \( \delta \)) of the Prognosis language are of the form \( A \rightarrow \alpha B \), where \( A \) signifies the current health state of the user, \( B \) is the new estimated health state (\( B \) can be equal to \( A \)) and \( \alpha \) is a new observed symptom that is being processed (consumed) by the language.

From the previous definition we can understand that the number of possible transitions can be quite large, since by taking into account ten different bio-signals for example we can have a very large set of combinations (which depends upon how many symptoms or fuzzy sets we may define for each bio-signal). To understand the magnitude of this problem we may consider that even in the case that each bio-signal derives a binary result, we will still have \( 2^{10} = 1024 \) different combinations!

In order to clarify how transitions are made and how the corresponding confidence score of the current state of the model is evaluated we will now
consider a somewhat trivial example, which we will then gradually make more complex to illustrate the operation of the Prognosis regular language:

**Example 1:** Let us consider the trivial case whereby the WHMS measures only 1 parameter, e.g. heart rate, which can only give rise to three individual fuzzy symptoms: *bradycardia*, *normal heart rate* and *tachycardia*. We will denote these states as *B*, *N* and *T* respectively and consider *B* and *T* as pathologic and *N* as healthy. In addition to that let us assume that the symptoms that can be extracted from the heart rate measurements are three as well, e.g. “low HR” (denoted ‘b’), “normal HR” (denoted ‘n’) and “high HR” (denoted ‘h’). In this simplistic example we consider all the weights corresponding to transitions (or production rules) to be equal to one, e.g. \( w_{N \rightarrow N} = w_{N \rightarrow T} = w_{N \rightarrow B} = w_{T \rightarrow T} = w_{T \rightarrow N} = w_{B \rightarrow B} = w_{B \rightarrow N} = 1. \)

The FSM that corresponds to Example 1 can be seen in Fig. 4-11 (*N* is the starting state). Now assume that the following sequence of symptom-confidence pairs is extracted: \( n (1.0), n (0.8), h (0.7), h (0.8), h (0.9), h (1.0), h (1.0), h (0.8), h (0.6), n (0.6), b (0.6) \), where in this case we have only considered the symptom with the highest confidence in each time instant.

We now want to go through the sequence of extracted symptoms and derive the estimated state of the fuzzy FSM along with the corresponding confidence level. We assume that we begin at state *N* with confidence 1.0.
At every derivation step, we will apply the compositional rule of inference as defined by Zadeh [122]:

\[
\mu(D) = \max_{s \in S} \left\{ \min(\mu_{S_i}(s), w_R(s)) \right\}
\]

In the above formula, \(\mu_{S_i}(s)\) signifies the degree of membership or the confidence level of the \(S_i\) symptom and \(w_R(s)\) denotes the connection weight between the current state and the state we are transitioning to. What the above formula means, is that when a new symptom is acquired, we will look for the most plausible transition (or production rule), by trying to find the one that maximizes the confidence level. The confidence level of a transition is defined as the “fuzzy AND” e.g. the minimum value of the symptom fuzziness level and the weight of the transition. Finally after the value \(\mu(D)\) has been computed, we will evaluate the confidence of the new state as the average of the previous state and the new computed confidence \(\mu(D)\) (the reason for this is more thoroughly explained in the beginning of the next chapter but simply put
it is done to prevent confidence levels from being driven towards zero and to account for the “momentum” of the state model).

Illustrating the above scheme with the provided example we have the sequence of transitions presented in Table 4-1:

<table>
<thead>
<tr>
<th>State</th>
<th>N</th>
<th>N</th>
<th>N</th>
<th>T</th>
<th>T</th>
<th>T</th>
<th>T</th>
<th>T</th>
<th>T</th>
<th>N</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>St. Conf.</td>
<td>1</td>
<td>1</td>
<td>0.9</td>
<td>0.4</td>
<td>0.55</td>
<td>0.73</td>
<td>0.87</td>
<td>0.94</td>
<td>0.87</td>
<td>0.74</td>
<td>0.43</td>
</tr>
<tr>
<td>Transition</td>
<td>n</td>
<td>n</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>h</td>
<td>n</td>
<td>b</td>
<td>-</td>
</tr>
<tr>
<td>Tr. Conf.</td>
<td>1</td>
<td>0.8</td>
<td>0.7</td>
<td>0.7</td>
<td>0.9</td>
<td>1</td>
<td>1</td>
<td>0.8</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
</tr>
</tbody>
</table>

**Example 2:** We now consider a more complex example which is however based on the previous paradigm. We assume that the WHMS can measure the heart rate and the body temperature as well and that the corresponding states that interest us are the same states as before for the heart rate and the following three states for the temperature: *fever* (F), *normal temperature* (N) and *hypothermia* (C). As a result the total number of states in the current FSM will be $3 \times 3 = 9$ states: NN, NF, NC, TN, TF, TC, BN, BF and BC, whereby the first letter signifies the state of the heart rate value and the second letter the one of body temperature.

Furthermore, in our current example we account for the case where the symptoms extracted from the heart rate measurement include additional and finer fuzzy sets, e.g. “very low hr” (vlhr), “low hr” (lhr), “normal” (nhr), “high hr” (hhr) and “very high hr” (vhhr). Correspondingly for the temperature val-
ues we will have “very low temp” (vlt), “low temp” (lt), “normal” (nt), “high temp” (ht) and “very high temp” (vht). Consequently, in the current case the weights on the transitions will have a bigger impact as there is a different contribution of a “high hr” value to the state tachycardia than the contribution from the symptom “very high hr”. The corresponding FSM of Example 2 is depicted in Fig. 4-12.

**Fig. 4-12. Fuzzy FSM for Example 2**

In this case, where we have used a “fuzzier” approach since the domain of symptom values is divided into more fuzzy sets, the transition weights play a more significant role. For example consider the case of the transition from the initial NN state to the TN state. This can be done by either detecting a “high heart rate value” or a “very high heart rate value”. However the corresponding weight for the latter transition (or production) will be stronger than the one for
the former transition, as a very high heart rate will contribute more towards an actual tachycardic patient state.

In Example 2, we can consider the Prognosis model as simulating two fuzzy FSMs in parallel. For every type of biosignal, corresponding fuzzy symptoms are continuously extracted and contribute to state transitions in the manner that was described in Example 1. By combining the confidence level of parallel states, we can deduct an overall confidence for the current state in Fig. 4-12. Using this approach, the aforementioned model can be expanded to consider all types of symptoms extracted from the physiological measurements described in the previous subsection. However, in the case of detectable conditions, symptom contributions to the corresponding state should be weighted appropriately, according to the causal association relevance. This is illustrated in the following example.

Example 3: Finally, we will consider a more concrete example, which will illustrate the application of the proposed formal language approach in estimating the user’s health, and consequently, accessing the risk level of the user’s health status. In this example, we will assume that the system is capable of continuously monitoring the following physiological parameters: ECG, HR, blood pressure, oxygen saturation, respiration rate, and that, it is also able to capture verbal feedback from the user. Furthermore, let the set of extractable symp-
toms be of the form $\text{lspo2, vhhr, "low blood pressure" (lbp)}$, etc. Additionally, consider the following symbolic representations for user health states.

1) $S_1$: Hypoxemia.

2) $S_2$: Coughing.

3) $S_3$: Hypotension.

4) $S_4$: Tachycardia.

5) $S_5$: Dizziness, weakness, or nausea.

In Fig. 4-13, a part of the Prognosis fuzzy FSM that corresponds to the aforementioned system is shown. In Fig. 4-13, several selected user health states are depicted along with a small subset of symptoms that cause transitions between these states (not all possible transitions are being shown to avoid confusion). The combination of $S_3$, $S_4$, and $S_5$ have been known to be an indication (are always present) of acute cardiogenic shock [120], [121] and possible additional symptoms, such as tachypnea, arrhythmias, and sweating can further enhance that indication and also increase the severity level of the condition. In addition to that, the state $S_1S_2$ corresponds to a strong indication of anoxic syncope, a hypothesis that is enhanced by the presence of dizziness or nausea.
We can now consider the following scenario: The user is initially in state $N$ and then gradually his blood pressure levels start dropping and his pulse rate starts increasing. This transition is depicted in Table 4-2, where a series of hypothetical fuzzy symptoms (along with their confidence level) are extracted and the corresponding estimated health state is shown (of course this could happen in a more gradual or complex manner, but we assume relatively rapid transitions here for the sake of discussion and without loss of generality). As the state $S_3S_4$ is close to the “high risk” state $S_3S_4S_5$ the system decides to inquire the user regarding the presence of additional symptoms. Given the case
that the user indicates the presence of dizziness, the system may deduct a transition to the neighboring state $S_5$ which requires immediate attention and thus the system will generate an alarm and notify the health care provider or a supervising physician.

### Table 4.2. State Transition Sequence for Example 3

<table>
<thead>
<tr>
<th>Current State</th>
<th>HR</th>
<th>SpO2</th>
<th>BP</th>
<th>Resp Rate</th>
<th>VF</th>
<th>ECG</th>
</tr>
</thead>
<tbody>
<tr>
<td>N 1.0</td>
<td>n 0.9</td>
<td>n 1</td>
<td>n 0.7</td>
<td>n 1</td>
<td>-</td>
<td>N 0.9</td>
</tr>
<tr>
<td>N 0.95</td>
<td>n 0.8</td>
<td>n 1</td>
<td>lbp 0.6</td>
<td>n 1</td>
<td>-</td>
<td>N 0.9</td>
</tr>
<tr>
<td>$S_3$ 0.64</td>
<td>n 0.7</td>
<td>n 0.9</td>
<td>lbp 1</td>
<td>n 1</td>
<td>-</td>
<td>N 0.8</td>
</tr>
<tr>
<td>$S_3$ 0.78</td>
<td>n 0.6</td>
<td>n 0.9</td>
<td>vlbp 0.6</td>
<td>n 0.9</td>
<td>-</td>
<td>N 0.9</td>
</tr>
<tr>
<td>$S_3$ 0.79</td>
<td>hhr 0.6</td>
<td>n 0.9</td>
<td>vlbp 0.8</td>
<td>n 0.8</td>
<td>-</td>
<td>N 0.8</td>
</tr>
<tr>
<td>$S_3S_4$ 0.61</td>
<td>hhr 0.9</td>
<td>n 0.8</td>
<td>vlbp 1</td>
<td>n 0.7</td>
<td>-</td>
<td>N 0.8</td>
</tr>
<tr>
<td>$S_3S_4$ 0.73</td>
<td>vhhr 0.7</td>
<td>n 0.8</td>
<td>vlbp 1</td>
<td>n 0.6</td>
<td>-</td>
<td>N 0.7</td>
</tr>
<tr>
<td>$S_3S_4$ 0.76</td>
<td>vhhr 0.9</td>
<td>n 0.8</td>
<td>vlbp 1</td>
<td>hrr 0.6</td>
<td>dizziness</td>
<td>PVC 0.6</td>
</tr>
<tr>
<td>$S_3S_4S_5$ 0.57</td>
<td>vhhr 1</td>
<td>n 0.8</td>
<td>vlbp 1</td>
<td>hrr 0.7</td>
<td>-</td>
<td>PVC 0.6</td>
</tr>
</tbody>
</table>

To evaluate the confidence or degree of support of a user health state, after every “collection cycle” of physiological measurements (Eq. 4-7) is applied, where $N$ is the number of bio-signals that did not change state, $M$ is the number of bio-signals that did change state, $\mu_{Si}$ is the confidence for the state of the $i^{th}$ bio-signal and $\mu(n)$ is the overall confidence of the health state estimation at discrete time $n$.

(Eq. 4-7) \[ \mu(n) = \frac{\frac{1}{N+1}(\sum_{i=1}^{N} \mu_{Si}(n-1)) + \mu(n-1)) + \frac{1}{M+1} \sum_{i=1}^{M} \mu_{Si}(n-1)}{2} \]
In estimating the confidence of the state $S_3S_4S_5$ a weighted average of the strength of the symptoms that are causally associated with the specific condition will be calculated. In addition to that a bias $b$ will be used, representing the confidence level that the user is indeed in the previous stage $S_3S_4$:

(Eq. 4-8) \[ \mu_D = \frac{b + \sum_{i=1}^{N} w_{R}(s_i) \mu(s_i)}{1 + \sum_{i=1}^{N} w_{R}(s_i)} \]

### 4.3 SPN-based Operation modeling

Using Stochastic Petri Nets (SPN) [139] as a graphical modeling tool, we are able to provide a detailed and at the same time easy-to-comprehend functional description of the wearable health-monitoring system. Furthermore SPNs enable hierarchical top-down modeling of the system while at the same time being able to capture the effects of concurrency and synchronization of events that take place in the system.

It should be noted that in the SPN model to be presented the following convention has been adopted: While tokens are removed from input places and deposited in output places in a deterministic manner (e.g. a transition fires either immediately or after a random time delay if and only if there is a token in each one of its normal input places and also there are no tokens in any of its inhibitor input places and it deposits a token in each of its output places), the firing of transitions that have common normal input places and/or common
inhibitor input places will happen in an non-deterministic manner. We will refer to this property of the SPN as the random transition property.

Specifically, consider a set of transitions $T_1, T_2, ..., T_N$ that all have at least one common input or one common inhibitor place and that happen to be all enabled at a given time instant. We assume that these co-enabled transitions are each assigned a probability $p_i$ where: $\sum_{i=1}^{N} p_i = 1$, such that when they are all enabled, only one of the enabled transitions will fire with probability $p_i$. In addition to that we assume that the described firing probabilities are marking-dependent, e.g. a specific model state (or marking) may “favor” the firing of a certain transition over the rest of the co-enabled transitions.

4.3.1 SPN Model - Level 1

Fig. 4-14 presents the first level of the hierarchical SPN model of the Prognosis WHMS, which simulates the high-level interaction between the basic entities in the system, e.g. the human user, the device and the medical center.

As we can see from the diagram, there is always a token at the place representing the user/patient (as due to the immediate feedback transition there will be one token generated when the current token is consumed) denoting the fact that the user is constantly “ON”, meaning that the human body provides the WHMS with body signals in a continuous fashion. Furthermore
the user is able to provide voice feedback to the system to record non-
measurable health symptoms.

The WHMS device is capable of directly communicating with the user
in terms of an automated HMI dialog system, whenever the system detects a
health status of high risk or the user notifies the system about a symptom like
for example chest pain that may require “further investigation”.

![Diagram](image)

**Fig. 4-14. Level 1 of the SPN Model: User-Device-Medical Center-Ambulance Interaction**

Furthermore, the WHMS provides either regular updates of the user’s
health condition or upon request from the medical center (which is also a place
in the SPN that constantly has a token). Finally the system may “decide” that
the user requires immediate medical attention based on the aggregated physio-
logical symptoms and thus may send an alert to the medical center or even re-
quest for an ambulance to be dispatched in case the medical center approves that action as well. This is denoted in Fig. 4-14 by the use of places \( P4 \) and \( P5 \), where a token might be generated in one of the cases described above.

Regarding the property described earlier about the SPN model and which we have adopted here, e.g. that enabled transitions with at least one common input place may be chosen to fire in a stochastic manner, we can see this in the following example:

Transitions \( T1 \) (system “talks” to the user), \( T2 \) (inform the medical center about an emergency), \( T3 \) (send a report or regular update to the medical center) and \( T10 \) (consume or process a series of measurements) all share place \( P1 \) (the WHMS) as input. When there is one (or more) token in \( P1 \), all the above mentioned transitions will be enabled. However, in such a case, due to the adopted property, only one of these transitions will fire, which will result in one token from \( P1 \) to be consumed and in one token being deposited in each output place of the firing transition.

4.3.2 **SPN Model - Level 2**

In the following we describe the second level of the hierarchical SPN model of the WMHS, which is depicted in Fig. 4-15. In this level, the functionality of the wearable device is simulated in a scenario in which the following sensors are assumed to be available: ECG, \( \text{SpO}_2 \), thermometer, respiration sensor,
blood pressure monitor, accelerometer, GPS and a voice recognition/synthesizer system to capture direct patient feedback.

The central node of the WHMS (microcontroller board, PDA, smartphone etc) continuously gathers physiological parameters from the wearable sensors in a round-robin fashion to create input symbols for the Prognosis language. For example the pulse oximeter sends an estimation of the oxygen saturation and the heart rate of the user every 1 second. Similarly, the central node polls the thermometer and the respiration rate sensor every second to acquire a measurement.

However, since available blood pressure monitors are cuff-based, the central node needs to request from the BP device to initiate a measurement and it should also be able to collect it when it is made available. Regarding the ECG
collection, because the electrocardiogram is sampled at a high rate, e.g. 250 Hz or even more, it is more reasonable to transmit packetized samples of the ECG every second (as denoted by the weight 250 over the arc leaving place $P2$). That approach allows also for easier synchronization with the collection of the rest of the bio-signals.

In this SPN level, we can see again how the random transition property affects the sequence of events. For instance, place $P0$, e.g. the central node, is a common inhibitor place for all sensors (or sensor buffers). This way we are trying to simulate the fact that whenever there is a token at $P0$, the central node is busy collecting data from one of the sensors. On the other hand, no token at $P0$ means that the central node is ready to request or collect data from a sensor. However at any given time only one sensor-place will remove a token and then deposit that token at $P0$.

Finally the voice recognizer module has the ability to deliver information regarding non-measurable symptoms to the central node, in case the user has recorded such a phenomenon. As in the case of the ECG sensor, to favor the round-robin-based synchronized collection of signals, the central node polls the voice recognizer for any possible detected symptoms. This polling scheme can be thought of either as sequential polling of digital and analog ports on a microcontroller board or as the beaconing mechanism available in Zigbee which allocates specific time intervals to distributed sensor nodes or finally as a
Bluetooth piconet, where master and slaves take turns communicating in an inherent round-robin scheme.

### 4.3.3 SPN Model - Level 3

In the final level of the top-down hierarchical SPN model depicted in Fig. 4-16 we describe the way the central node operates. We provide an insight to how the collected physiological (and voice) data are used to create input vectors, e.g. to form words/strings of symptoms of the Prognosis language, which in turn parses the derived string to make a decision/estimation about the user’s health condition.

![Diagram](image)

**Fig. 4-16.** Level 3 of the SPN model: Functional description of the central node.

There are three basic types/categories of data that the WHMS can collect (as previously described in this chapter): a) “scalar” data, like the temperature value, the systolic or diastolic blood pressure etc, which according to their val-
ue give rise to specific symptoms or not, b) “morphology” specific data, which in the current scenario are comprised of the ECG signal, which need to be analyzed to possibly detect patterns of high risk and c) voice data which may correspond to non-measurable symptoms that the user has chosen to communicate to the device.

When either a new type a) or type b) measurement is collected, it is buffered in the system in a dedicated ring buffer for each bio-signal. Then that data are checked for validity, e.g. a decision on whether the data are erroneous or valid needs to be made based in terms of statistically examining the signal’s trend to possibly eliminate outliers or by using any provided sensor status info. If the data are found to be valid, they are used to extract the corresponding fuzzy symptom. At the last stage the extracted symptoms or language symbols are passed on to the Prognosis language which updates or creates the string, which gives a thorough estimation of the user’s health.

Regarding voice recordings, which can also be void in terms of null feedback, they are processed and converted to text in order for them to be “understood” by the system’s language. After cross-checking the detected word(s) with a database of non-measurable symptoms, an indication of some incident like for example “back pain” may be identified and then passed onto Prognosis as further contribution to the input vector of the string formation mechanism.
Finally, the *Prognosis* scheme produces an output in terms of a word or string, which may give rise to either a normal/healthy indication (null string) of the user’s status or to an unknown/erroneous status or more importantly to an alarming state. In such a case an alarm message will be generated and sent to the medical station and moreover the system may “decide” to take further action in terms of initiating a dialogue with the user or suggesting him to take several actions which may help prevent an escalation of the occurred or even future event.
CHAPTER 5: MODELING THE PROGNOSIS LANGUAGE USING MODIFIED FUZZY PETRI NETS

In the previous chapter we described a fuzzy formal language model for capturing the transition of the WHMS’ user between different health states and for estimating his/her health at any given time. As it was mentioned in section 4.2.4.3, by using the FFSM approach, where in practice we are simulating several fuzzy FSMs in parallel, the following problem arises: the amount of possible states in the system can become very large as the number of different physiological parameters (and their respective fuzzy sets) that are measured by the system increases. To illustrate this more concretely we can consider the case where the WHMS records the following features/signals, where we also denote the number of fuzzy terms by which each variable is represented:

- **ECG** (about 6 different beat types and 8 different rhythms resulting in 15 distinct combinations)
- **Heart rate** (6 different states)
- **Respiration rate** (4 different states)
- **Skin/body temperature** (5 different states)
- **Systolic & diastolic blood pressure** (5 different states)
- **Oxygen saturation** (4 different states)
• **Symptoms described by the user** (for 10 different symptoms we can have \(2^{10}=1024\) combinations, without considering level of severity of symptoms)

Calculating all possible combinations using the above attributes leads us to the following result: \(15 \times 6 \times 4 \times 5 \times 5 \times 4 \times 1024 = 36,824,000\) distinct states! This number can keep growing if we also take into account factors such as the user’s age, weight, height, gender, medical history, current medication, the activity that is being performed etc. It is thus obvious that the FFSM approach can become computationally very expensive when so many cases are considered, especially since our aim is to derive an embedded decision support scheme that should at least be partially implementable on a portable resource-constrained device.

Furthermore, our FFSM model does not consider certain important aspects in a diagnosis making process which will be discussed in this chapter. For example the effect of negative evidence on a decision or the exclusion of a disorder due to certain findings has not been considered. In addition to that, the FFSM model does not provide a means for adaptive parameter estimation and assumes fixed well-estimated parameters for the fuzzy IF-THEN rules that are included in its knowledge base.

In an attempt to overcome these limitations, we present in this chapter an advanced decision-making model which is based on a modification of the concept of Fuzzy Petri Nets (FPN), to which we will simply refer as “Modified
Fuzzy Petri Net” (MFPN). In what follows, we will introduce the concept of FPNs and of the MFPN and then we will explain analytically the process of decision making using the proposed reasoning mechanism.

5.1 Fuzzy Petri Nets

Fuzzy Petri Nets (FPN) are an extension of the classical Petri Net modeling tool and were introduced in [140] and [141] as a means of supporting approximate reasoning, fuzzy knowledge representation and decision making. In contrast to the conventional Petri Net model, where tokens are assigned to places and generated from transitions in a crisp manner, in the FPN case, each token is accompanied with a fuzzy value which indicates the degree of certainty of a given event or proposition. There have been several formal definitions of FPN models [140]-[148] in an attempt to capture concisely the effect of multiple weighted fuzzy production rules in rule-based decision making.

A Fuzzy Petri Net for representing Weighted Fuzzy Production Rules (WFPR) is defined as:

(Eq. 5-1) \[ FPN = (P, T, D, A, I, O, \alpha, \beta, \gamma, W), \]

where:

- \( P = \{p_1, p_2, \ldots, p_n\} \) is a finite set of places,
- \( T = \{t_1, t_2, \ldots, t_m\} \) is a finite set of transitions,
- \( D = \{d_1, d_2, \ldots, d_n\} \) is a finite set of propositions,
\[ \Lambda = \{ \lambda_1, \lambda_2, \ldots, \lambda_m \} \] is a finite set of thresholds and \( 0 \leq \lambda_i \leq 1 \) for all \( \lambda_i \in \Lambda \),

- \( I: T \rightarrow P^\infty \) is an input function, a mapping from transitions to places,
- \( O: T \rightarrow P^\infty \) is an output function, a mapping from transitions to places,
- \( \alpha: P \rightarrow [0,1] \) is an association function (fuzzy term), a mapping from places to real values between zero and one (which indicates the fuzzy truth values of propositions or the fuzzy marking of places),
- \( \beta: P \rightarrow D \) is an association function, a bijective mapping from places to propositions,
- \( \gamma: T \rightarrow \Lambda \) is an association function, a bijective mapping from transitions to thresholds,
- \( W = W_I \cup W_O \), where:
  - \( W_I: I \rightarrow [0,1] \) is an association function, a mapping from the input function to real values between zero and one (which assigns a weight to every input arc of a transition),
  - \( W_O: O \rightarrow [0,1] \) is an association function, a mapping from the output function to real values between zero and one (which assigns a weight to every output arc of a transition or in other words it represents the confidence level in a given transition for a certain consequent proposition),
- \( P \cap T \cap D \cap \Lambda = \emptyset, |P| = |D|, |T| = |\Lambda| \), where \( | | \) denotes cardinality.

Using the above definition of a Fuzzy Petri Net we can represent the following 4 types of weighted fuzzy production rules:
For each WFPR we may define weights for every antecedent (denoted as "\( w \)" and taken from the set \( W_i \)) and every consequent (denoted as "\( \mu \)" and taken from the set \( W_o \)) element and also a threshold value "\( \lambda \" which controls the firing of a rule. Let "\( \alpha(d_i) \" signify the fuzzy truth value of proposition \( i \), "\( \land \" signify the fuzzy AND operation (which can be the \( max \) function or some other t-norm) and "\( \lor \" signify the fuzzy OR operation (the \( min \) function or some other s-norm). Then the fuzzy truth values of the consequent portions of the above 4 rules are computed as follows:

\[
\begin{align*}
\text{(Eq. 5-6)} & \quad \textbf{Type 1}: \quad \text{IF } \alpha(d_1) \cdot w_1 \land \alpha(d_2) \cdot w_2 > \lambda \quad \text{THEN } \alpha(d_3) = (\alpha(d_1) \cdot w_1 \land \alpha(d_2) \cdot w_2) \cdot \mu \\
\text{(Eq. 5-7)} & \quad \textbf{Type 2}: \quad \text{IF } (\alpha(d_1) \cdot w_1) > \lambda \quad \text{THEN } \alpha(d_2) = \alpha(d_1) \cdot w_1 \cdot \mu_2 \\
& \quad \quad \text{AND } \alpha(d_3) = \alpha(d_1) \cdot w_1 \cdot \mu_3 \\
\text{(Eq. 5-8)} & \quad \textbf{Type 3}: \quad \text{IF } \alpha(d_1) \cdot w_1 \lor \alpha(d_2) \cdot w_2 > \lambda \quad \text{THEN } \alpha(d_3) = \alpha(d_1) \cdot w_1 \cdot \mu_1 \lor \alpha(d_2) \cdot w_2 \cdot \mu_2 \\
\text{(Eq. 5-9)} & \quad \textbf{Type 4}: \quad \text{IF } (\alpha(d_1) \cdot w_1) > \lambda_1 \text{ AND } (\alpha(d_1) \cdot w_2) > \lambda_2 \quad \text{THEN EITHER } \alpha(d_2) = \alpha(d_1) \cdot w_1 \cdot \mu_2 \\
& \quad \quad \text{OR } \alpha(d_3) = \alpha(d_1) \cdot w_1 \cdot \mu_3
\end{align*}
\]

The aforementioned WFPRs can be mapped directly to their FPN representations as it can be seen in Fig. 5-1, Fig. 5-2, Fig. 5-3 and Fig. 5-4.
Fig. 5-1. FPN representation of a WFPR of Type 1.

Fig. 5-2. FPN representation of a WFPR of Type 2.
As stated before, in the FPN model, each token is associated with a fuzzy truth value and since each place is mapped to a proposition $d_i$, a place containing a token with value $\alpha(d_i)$ means that the degree of truth of the corresponding logical proposition is $\alpha(d_i)$. In addition to that, transitions are the means by which knowledge and/or belief is propagated in the network. Fur-
thermore, an input weight $w$ from a place to transition denotes how much a proposition influences a given WFPR, while an output weight $\mu$ represents the belief in that specific rule with regards to a specific consequent proposition. It is worth noting that the effect of weights on a WFPR is of importance mainly in the case of type 1 rules, e.g. in the case of composite conjunctive reasoning, whereby all antecedents need to be accounted for and weighted in order for the rule to fire. In the rest of the rules, weights can be usually assumed to be 1.

One thing we should note here and which has been a common misconception in publications regarding FPNs is that the FPN of Fig. 5-4 does not represent a rule like:

If $d_1$ THEN $d_2$ OR $d_3$

Instead it represents the XOR rule that we showed before. This is in agreement with the way transitions take place in a regular Petri Net, whereby a transition that is enabled and fires will remove a token from all of its input places and place a token in all of its output places. As a result, in Fig. 5-4, if place $p_1$ has a token that results in a fuzzy value greater than both thresholds $\lambda_1$ and $\lambda_2$, then both transitions will be enabled but only one (and exactly one) can fire and not both. In the case of FPNs where tokens are not removed when a transitions takes place but rather copies of tokens are passed on [141], [143], this type of FPN will mean that both transitions will become enabled and thus there is no difference in this case from rule of type 2.
Another important issue, that we need to consider when reasoning with FPNs, is the "diminishing token syndrome" as it has been noted in [149]. Specifically, as a token progresses through the Fuzzy Petri Net and as rules are applied along the reasoning path, the token value is bound to be a decreasing function with time, as it is constantly being multiplied with weights that are non-greater than zero and is also ANDed with other tokens. This issue is also made obvious in the equivalent representation of a Fuzzy Finite State Machine (FFSM) as it is described in [150]. In this case, the constant application of max-min rules results in a monotonic decrease of the fuzzy membership values of the states in the FFSM.

We have addressed this problem in the FFSM case in our work (as we saw in section 4.2.4), where the new computed confidence level for a given state is combined with the previous one to derive the actual final degree of membership in that state. In that sense, the current momentum of the FFSM is also taken into consideration. In addition to that, the scheme we proposed does not have the form of a feed-forward network, but rather in includes loops and also several continuously repeating fuzzy events (fuzzy symptoms extracted from physiological measurements) which constantly change the state vector of the FFSM. In this way fuzzy membership values are always “refreshed” and this prevents them from continuously diminishing. However this model suffers
from the issues and limitations that were mentioned at the beginning of the current chapter.

In the next section we will introduce a model of a Modified Fuzzy Petri Net (MFPN), which overcomes the limitations of the conventional FPN models that have been proposed in the corresponding literature. It should be noted however that the MFPN will be defined in a way that is considered optimal for the problem in question (e.g. estimation of user’s health state using basically a fuzzy IF-THEN rule-base) and to address some of the issues not addressed by the FFISM and the FPN (e.g. inhibition/exclusion, obligatory/optional findings, momentum and most importantly rule adaptation/learning). As a result the presented modified FPN should not be considered as a general improvement of the FPN, but rather as an extension of it in order to better address a particular problem.

5.2 A Modified Fuzzy Petri Net (MFPN) Model

Fuzzy Petri Nets are used for knowledge representation and fuzzy reasoning rather than for system operation modeling compared to conventional Petri Nets (deterministic or stochastic). As a result, in the case of FPNs it makes intuitively more sense to not consider token consumption in a place when an output transition fires. After all, a place in a FPN represents a logical proposition and the token value stands for the degree of truth of that specific predi-
cate. If that proposition constitutes one of the antecedents of a production rule and that rule fires, the proposition does not cease to be true (to the same fuzzy level as before). In contrast to this example, tokens in conventional Petri Nets tend to represent some quantity or some resource of the system that is being modeled. For example, the number of tokens in a place could stand for the number of customers waiting in a queue or a single token could signify a semaphore in a concurrent program.

Thus, in our case we will not consider consumption of tokens, but we will rather treat token values as permanent once they get evaluated, until of course they get updated. Besides the intuitive reason behind this approach that was mentioned above, we wish to have a “momentum property” in the system. This means that once a transition fires, the token value (or fuzzy truth) of each of the output places will be evaluated based both on the level of firing of the transition as well as on the previous token value(s) of the output place. However in order to prevent enabled transitions from firing continuously, we will also introduce another property describing places in FPNs that will denote if the token value has been updated since the rule fired the last time. If none of the token values in the input places of a transition have been updated, then the transition will not be enabled. As soon as one of the input places gets updated the rule will be evaluated, as it will then be possible for the rule to produce a different and updated output.
In addition to these properties we also wish to integrate the notion of obligatory and optional predicates for a WFPR and also the concept of proving and not proving evidence [135] in the reasoning model. Specifically, a symptom $S_i$ will be considered obligatory for a disease $D_j$ if the absence of $S_i$ implies directly the ruling-out of $D_j$. In addition to that, a symptom $S_i$ will be considered as proving with respect to disorder $D_j$ if the presence of $S_i$ constitutes sufficient evidence to conclude that $D_j$ is present.

Given the previous discussion, we now provide the definition for the proposed modified FPN:

A Modified Fuzzy Petri Net (MFPN) for representing Weighted Fuzzy Production Rules (WFPR) is defined as:

(Eq. 5-10) \[ FPN = (P, T, D, A, I, O, \alpha, \beta, \gamma, W, o, r, u, m), \]

where $P, T, D, A, I, O, \alpha, \beta, \gamma$ and $W$ are defined as in (Eq. 5-1) and:

- $o: I \rightarrow \{0,1\}$ is an association function, a mapping from the input function to either the value 0 or 1 (used to check substantiation of the obligatory antecedents in a WFPR).
- $r: I \rightarrow \{0,1\}$ is an association function, a mapping from the input function to either the value 0 or 1 (denoting whether the corresponding input attribute is considered as proving for the corresponding rule).
• \( u: P \rightarrow \{0,1\} \) is an association function, a mapping from places to either the value 0 or 1 (denoting whether the corresponding proposition, e.g. its fuzzy truth, has been updated or not).

• \( m: P \rightarrow \{0,1,2,3,\ldots\} \) is an association function, a mapping from places to positive integers (denoting the number of momentum terms used in a place \( p_i \)).

Finally, we will also consider that the mapping \( \alpha \) from places to the interval \([0,1]\) can also be a mapping from places to a fuzzy set (this will be explained more when we describe the reasoning scheme based on the MFPN in the following section).

To illustrate the usefulness of the newly defined attributes, let us consider a hypothetical case where the following rule \( R_i \) needs to be evaluated:

\[
\text{IF } d_1 \text{ (obligatory and proving for } R_i) \text{ AND } d_2 \text{ (obligatory and not proving for } R_i) \text{ AND } d_3 \text{ (optional and proving for } R_i) \text{ AND } d_4 \text{ (optional and not proving for } R_i) \text{ AND NOT } d_5 \text{ (presence of } d_5 \text{ excludes } R_i) \text{ THEN } d_6
\]

This rule is shown schematically in Fig. 5-5. Considering the MFPN definition given in (Eq. 5-10), the output of the rule will be computed now as:

\[
\alpha(d_6)_n = m(p_6)_0 \cdot u_{R_1} \cdot o_{R_1} \cdot \max\{f(R_1), r_{R_1}\} + m(p_6)_1 \cdot \alpha(d_6)_{n-1} + m(p_6)_2 \cdot \alpha(d_6)_{n-2} + \cdots + m(p_6)_{M_6} \cdot \alpha(d_6)_{n-M_6'}
\]

where:
• $u_{R_1} = \max_i(u_{t_1})$, where $u_{t_1}$ signifies whether antecedent $i$ of rule $R_1$ has been updated since the rule’s last evaluation (independently of whether the rule fired or not).

• $o_{R_1} = \min_i(o_{t_1})$, where $o_{t_1}$ signifies whether antecedent $i$ of rule $R_1$ is considered obligatory ($o_{t_1} = 0$ if $a(d_i) = 0$ and $o_{t_1} = 1$ otherwise) or optional ($o_{t_1} = 1$ always, since optional evidence of a rule should not prevent a transition from firing). In the case of the inhibitory input place $p_5$ we will compute $o_{5_1}$ as: $o_{t_1} = 1 - a(d_5)$.

• $r_{R_1} = r_{1_1} \oplus r_{2_1} \oplus r_{3_1} \oplus r_{4_1} \oplus r_{5_1}$, where $\oplus$ is the probabilistic t-conorm, e.g. $x \oplus y = x + y - x \cdot y$ and $r_{t_1} = a(d_i)$ for proving propositions and $r_{t_1} = 0$ for not-proving propositions. The reason that the probabilistic sum is employed to evaluate $r_{R_1}$ is because intuitively it makes more sense to increase the confidence in the rule’s substantiation whenever more than 1 proving propositions are true (or true to a certain degree).

• $f(R_1) = \frac{\sum_{t=1}^{6} w_{t_1} a(d_i) + w_{5_1} (1 - a(d_5))}{\sum_{t=1}^{6} w_{t_1}}$, is the weighted level of firing of rule $R_1$.

These weights represent what we referred to in section 4.2.4.2 as “causal relevance”, e.g. a fuzzy value indicating the likelihood or possibility of displaying symptom $S_i$ when disorder $D_i$ is present.

• $m(p_6)_i$ (with $0 \leq m(p_6)_i \leq 1$ and $\sum_i m(p_6)_i = 1$), stands for a momentum term indicating the degree to which the current firing strength of the rule ($m(p_6)_0$) and the previous fuzzy truths of the proposition
\( \{m(p_6)_1, m(p_6)_2, \textit{etc}\} \) contribute to the computation of the new fuzzy truth of proposition \( d_6 \). \( M_6 \) denotes the amount of “memory” used in place \( p_6 \).

![Diagram of a WFPR with five antecedents with different attributes](image)

Fig. 5-5. A conjunctive WFPR with five antecedents with different attributes

In the above WFPR we have assumed that the threshold for firing is \( \lambda = 0 \) and that the output weight (e.g. the belief or trust in the rule) is \( \mu = 1 \). Also we can notice that we have used a \textit{white bar} to represent this particular rule (a similar distinction is made in Stochastic Petri Nets, where this bar denotes probabilistic transitions). The distinction is made here to denote the fact that the particular IF-THEN rule is a rule representing medical knowledge and it needs to be handled differently. However, whenever we will employ the regular \textit{black bar}, we will evaluate the rule as it was defined by (Eq. 5-6)-(Eq. 5-9), while still considering the \textit{updating property} of rules and the momentum terms.
Before moving on to describing the reasoning process, which uses the modified FPN model described in this section, we need to define the knowledge-base that is employed in order to derive estimation of the user’s health state. This will be done in the following section by listing several fuzzy IF-THEN rules that have been integrated in the Prognosis prototype. After we do that, we will be able to describe in detail how a decision is reached and how the system generates hypotheses about the user’s condition.

5.3 Knowledge-base representation as fuzzy IF-THEN rules

In this section we will provide a list of some of the IF-THEN rules that constitute the initial knowledge base for the Prognosis WHMS. Propositions (in the antecedent part of the rules) will be described linguistically and they will have the general form:

1. `<physiological parameter> is <fuzzy set>`, for value-specific physiological parameters (section 4.2.1)
2. `<ECG> shows <beat/arrhythmia type>`, for ECG data (chapter 3)
3. `patient has <symptom type>`, for non-measurable symptoms indicated from the patient (section 4.2.3)

Given the above definitions, we provide a list of rules that show the medical knowledge that has been incorporated in the current prototype (where
the parameters of the rules that are required for calculating their substantiation according to (Eq. 5-11) have been omitted here for presentation purposes):

- **IF** heart rate is very high **AND**
  blood pressure is very low **AND**
  respiration rate is (very low OR very high) **AND**
  temperature is low **AND**
  oxygen saturation is low **AND**
  ECG shows ventricular arrhythmia
  THEN acute cardiogenic shock

- **IF** temperature is very high **AND**
  respiration rate is very high **AND**
  patient has cough
  THEN pneumonia

- **IF** blood pressure is very low **AND**
  respiration rate is very low **AND**
  temperature is very low **AND**
  ECG shows ventricular arrhythmia **AND**
  patient has (dizziness OR weakness)
  THEN hypothermia

- **IF** blood pressure is (very high OR very low) **AND**
  respiration rate is very high **AND**
  heart rate is (very high OR very low) **AND**
  ECG shows ventricular arrhythmia **AND**
  patient has (dyspnea AND cough)
  THEN acute left ventricular failure

- **IF** blood pressure is high **AND**
  patient is experiencing chest pain OR shortness of breath **AND**
  ECG is abnormal
  THEN angina

- **IF** blood pressure is very high OR
  blood pressure is very low OR
  heart rate is very low OR
  respiration rate is very high OR
  respiration rate is very low OR
  temperature is very high OR
  temperature is very low OR
  oxygen saturation is very low OR
ECG is abnormal
THEN alarm

• IF heart rate is very high AND
  activity level is very low
THEN alarm

5.4 **Reasoning using the knowledge-base and the MFPN**

As the previous section shows, there might be a considerable number of production rules that will need to be evaluated whenever new sensor readings are received by the WHMS. This can be graphically depicted in a parallel structure as the one shown in Fig. 5-6, which displays schematically part of the reasoning process using the MFPN. The functionality of each of the 4 layers is explained below:

- **Layer 1**: This layer holds the actual crisp inputs of the sensors. The places in this layer deviate from their conventional definition in Petri Net or an FPN since they can hold real values.

- **Layer 2**: As the real values from Layer 1 get propagated to Layer 2 they pass through a set of transitions that have a single input place and multiple output places. Each of the output arcs has a different effect on the propagated values, as each arc is associated with a fuzzy term. As a result, what this process represents is the fuzzification of the input values, where the membership of the crisp parameters in the fuzzy sets (fuzzy symptoms) defined in 4.2.1 is computed. The places in this layer do not
have momentum terms and they hold the fuzzy membership values of the most recent physiological measurements.

- **Layer 3**: The purpose of this layer is to compute an averaged/smoothed confidence in the symptoms extracted in the previous layer. As a result, the places found here utilize momentum terms and permit the confidence in a specific symptom to rise as more measurements indicating the same state are collected. In addition to that, and depending on the number of momentum terms used, outlier values can be suppressed.

- **Layer 4**: This layer holds the degree of support of the consequent part of all the IF-THEN rules in our knowledge-base. The places in the previous layer represent the antecedent parts and their fuzzy token values represent the strength of confirmation of the corresponding symptoms. The transitions (e.g. the WFPRs) between layer 3 and layer 4 are computed according to the MFPN model.

- **Layer 5**: Finally, this last layer holds a single place which we define as the overall “unhealthiness of the user”. The fuzzy truths of the disorders computed in the previous layer are combined together using logical disjunction (probabilistic t-conorm) and their aggregation produces a single value, which provides an indication of the patient’s health instability level (“0”→100% healthy and “1”→100% unhealthy or unstable). The intuition behind the computation of this metric is that as the num-
ber of disorders, which are found to be possible (non-zero degree of support), increases then the more likely it is for the patient to reside in or to be entering an unstable or alarming health state. The threshold for which the condition of the user should be considered critical must be determined through a learning process and an initial approach towards achieving that will be presented in the next chapter.

Fig. 5-6. MFPN for user’s health state estimation

To fully understand the above MFPN architecture, we can look at a sub-module of the MFPN shown in Fig. 5-6. This sub-module, which is shown as a
dashed rectangle in Fig. 5-7, displays how the confidence about the user’s heart rate level/status is continuously estimated and how confidence levels of other symptoms might be combined to fire one of the transitions between the 3rd and 4th layers in Fig. 5-6. Analogously we can consider a similar sub-module for estimating the certainty factor for the fuzzy sets that divide the universe of discourse of the rest of the recorder physiological parameters.

Fig. 5-7. FPN representation of a sub-module of the Prognosis reasoning process.
CHAPTER 6: A PATIENT OR POPULATION-LEARNING MODEL

When it comes to making specific decisions about a patient, such as detecting health alarming trends or even diagnosing a disorder, a doctor needs to consider the individual physiology and medical history along with general medical diagnosing guidelines. In addition to that, monitored physiological parameters such as blood pressure, heart rate, ECG, respiration rate, oxygen saturation etc, may have significantly varying normal or healthy levels (or morphologies when talking about the ECG) amongst individuals. In that direction, we propose a patient-learning approach in that the system first requires a supervised learning period in order to learn the user’s “healthy physiological behavior” by using a neural-fuzzy-based scheme, which builds on the FPN model described in the previous chapter. After learning the parameters for the weights and the membership functions used in the reasoning process, the system can operate autonomously in order to provide individualized user-adapted detection of health abnormalities and risky conditions.

However, before moving further, it should be noted that our efforts are not aiming towards a system that will replace the doctor in any way, but towards a healthcare solution that will ensure patient safety round-the-clock via early detection of individual health hazards and that will thus potentially improve the quality of life of people at risk.
6.1 The concept of the health-history database

As it has already been discussed, a multi-parametric WHMS is capable of continuously measuring, recording and transmitting a wide variety of physiological signals. However in our system we seek to integrate additional functions instead of just data logging. Namely, the system should be able to extract higher level of information from the collected data by itself and transmit this information to the medical center’s database and/or to the supervising physician if required. This concept was first introduced in Chapter 3, where an ECG analysis algorithm is employed to extract characteristic signal attributes and to derive a per-beat signal characterization. Then in chapters 4 and 5, we described how symptoms or risk-levels are extracted from other types of physiological measurements. Both types of physiological information as well as voice feedback constitute the overall evidence about the user’s occurring symptoms, which are examined according to certain medical rules to possibly detect a health hazard. However the rules according to which symptoms are extracted are not universal. To the contrary, these rules are strongly dependent upon the individual user’s physiology, his medical history and his current context.

To illustrate the previous point, we can consider again the case of the heart rate level. A measurement between 60 and 100 beats per second (bpm) is generally considered as normal, while a heart rate less than 60 bpm is considered as a bradycardia and a value greater than 100 as tachycardia (provided
that the heart rhythm is regular). However, for a given patient/user a certain
tachycardic heart rate value might be quite acceptable and not endangering at
all, while for another one the same heart rate measurement may indicate a ha-
azardous or unstable health state. Different persons might have considerably
different resting heart rate levels for example, and still individually those levels
might be considered normal, because it is not generally easy to define an op-
timal HR for a given person [151]. Similar reasoning pertains to the rest of the
continuously measured physiological signals.

To summarize this point, as it is shown in Fig. 6-1, we would like for the
system to be able to utilize raw measured data in order to extract basic health
information and then by using an individualized “healthy-history” database to
create actual medical knowledge and individualized decisions about the pa-
tient/user. This extracted higher-level knowledge can save time (and in the
long run also reduce medical costs) when it comes to making quick decisions
and possibly taking emergency actions about a patient and especially when
there are several patients to consider or supervise in the same time frame. In
addition to that, such individualized patient models, may lead to better unde-
standing of long-term physiological trends that might be indicators for specific
health disorders. As a result, such models might be then be utilized to pro-
nose with high likelihood the occurrence of health conditions and also hopefully
prevent them by early medical intervention.
6.2 Adapting the WHMS’ decisional capability to the user

As described in the previous section, we seek to integrate in the Prognosis WHMS the necessary intelligence for it to be able to make patient-specific decisions/estimations regarding the health status of the corresponding patient. In addition to that, when making such decisions, the system must also take into account the context in which the corresponding physiological data were captured.

To further explain this we can consider one more time the case of a heart rate value being measured over 100 BPM, e.g. tachycardia, for a certain period. Such a heart rate value range may be perfectly normal when the corresponding individual is exercising, which the system can “verify” by checking
one or more of the following: if there is or has been immediately before high degree of movement (e.g. by checking the magnitude of the acceleration vector) and/or if the user is showing a decrease in skin temperature (due to perspiration) and increased respiration rate. On the contrary, the same heart rate measurement along with increased blood pressure while the patient is laying on a bed (e.g. the posture value is close to -90 degrees), might have a completely different meaning and might indicate a hazardous health state. A tachycardic heart rate might also still be benign in the previous scenarios, except for example in the case that there is an accompanying ECG abnormality, like an absence of P waves or irregular heart rhythm, or for example in the case that the user is experiencing chest discomfort.

The previous example attempts to point out the level of complexity involved in the task of classifying a user’s physiological data as being either normal (healthy) or pathological. Ideally, the patient’s medical history, his personal attributes (age, height, weight etc), his current context and the rest of the concurrently measured physiological parameters will need to be taken into account. Such a task could be performed by comparing vectors of measured data with the contents of a corresponding database. However there would be two drawbacks in such an approach: Firstly, frequent searching even in a relatively small database can be quite tedious and computationally expensive for a hardware-constrained portable platform and secondly it would be practically
impossible to construct a database including all possible combinations of data vectors and corresponding patient symptoms.

From the previous discussion it is made obvious that it would be preferable to have some type of automatic weight adjustment strategy that would update the parameters involved in the reasoning process according to known and clinically labeled input-output combinations. Artificial neural networks have been used extensively in cases where automatic and adaptive learning of weights is required. However, while neural networks are good at recognizing patterns, they are not good at explaining how they reach their decisions [152]. On the contrary, fuzzy-logic based structures for decision making are capable of reasoning with incomplete and ambiguous information and have a very clear structure for explaining their reasoning results. Their main disadvantage is though that they do not have any way of acquiring rules and optimizing the parameters that are part of their inference logic. The pros and cons of fuzzy systems and neural networks appear to be complementary in a sense, an observation which has lead to the development of various types of fuzzy-neuro and/or neural-fuzzy systems [152]-[154].

Thus, the integration of a neural-fuzzy approach in our system appears to be a promising alternative to address the issue of progressively adapting the parameters involved in the decision making process to the physiology of the individual user. Such an approach has indeed been adopted in the case of the
Prognosis prototype and will be developed in detail through the rest of this chapter. However, before we do that it would be appropriate to include a brief introduction to the area of neural-fuzzy systems.

6.2.1 Neural-Fuzzy Systems

Fuzzy systems and neural networks have been combined in various ways for the purpose of either enabling a neural network to handle fuzzy inputs and possibly generate fuzzy outputs as well or having the neural network implement fuzzy decision making and aid in rule generation and fuzzy parameter estimation [153]. The latter option will be of special interest to us in the following section.

As it was mentioned before, the main issue with fuzzy systems is knowledge acquisition and rule-base construction. This has been done traditionally in the case of expert systems (like the ones mentioned in Chapter 4), either via expert knowledge acquisition or statistical methods (whenever such data are readily available). Additionally, since the reasoning process is based on production rules, the universe of discourse (e.g. the input space) of each input variable needs to be partitioned appropriately (e.g. as it was done with the fuzzy sets which were defined heuristically in section 4.2.1). As a result, applications of fuzzy systems tend to be restricted to occasions where expert knowledge is available and the number of input variable is small.
Neural-fuzzy systems constitute extensions of neural networks and can be used to automatically extract fuzzy rules from numerical data and to compute optimal shapes and parameters for membership functions. Tuning of membership function is considered particularly important, because as it was mentioned above the effectiveness of the employed fuzzy model depends greatly on the fuzzy partition of the input-output space. Since this process can be regarded as an optimization problem, aNNs constitute a possible solution to this problem.

ANFIS (Adaptive-Network-Based Fuzzy Inference System) proposed by Jang [154] in 1993 is a neuro-fuzzy model capable of learning the shape parameters of membership functions of the antecedent and consequent part of fuzzy rules. By employing a hybrid learning scheme ANFIS can compute complex nonlinear input-output mappings using both experts’ knowledge and labeled examples. ANFIS is particularly suitable for the case where several IF-THEN rules (with more than 1 antecedents ANDed together) are evaluated on the same output variable and the results of all the rules are combined together using the center of gravity method. This type of network is particularly suited for the case of fuzzy controllers, where the confidence of several input-output mappings need to be evaluated in parallel and then the overall output is acquired by applying a defuzzification scheme which weights every rule output.
accordingly. The way the learning mechanism works is by a) computing some measure of the overall network error after each pattern has been presented:

\begin{equation}
E_k = \frac{1}{2} (o_k - d_k)^2,
\end{equation}

where \( o_k \) is a vector representing the actual output of the network when pattern \( x_k \) was presented to the system and \( d_k \) is the desired output. Then the steepest (or gradient) descent method can be applied to learn all the unknown parameters in the model, e.g. consequent part of the fuzzy rules, shape parameters of antecedent’s membership functions etc:

\begin{equation}
z_i(t + 1) = z_i(t) - \eta \frac{\partial E_k}{\partial z_i},
\end{equation}

where \( z_i \) is the attribute that we want to update and \( \eta > 0 \) is the learning rate parameter.

Fig. 6-2. An ANFIS type-3 neuro-fuzzy network
Jang has also modified the training algorithm by defining a hybrid learning technique that combines the gradient descent method and the Least Squares Estimate (LSE) to estimate the network’s parameters. A neuro-fuzzy network, on which this learning approach can be applied, is shown in Fig. 6-2.

The purpose of the 5 layers in the network is to provide the following:

Layer 1: How much does the input satisfy the linguistic label of the node?
Layer 2: Firing strength of the associated rule (rule nodes)
Layer 3: Normalization of the firing levels
Layer 4: Individual rule output computation
Layer 5: Overall system output (combination)

Another application of the neuro-fuzzy approach is in constructing adaptive-network-based classifiers, where the degree of firing of a rule indicates the degree of membership of the input pattern to the class represented by the consequent part of the rule in question [152]. Such a neuro-fuzzy classifier can be seen in Fig. 6-3.

Fig. 6-3. An adaptive neuro-fuzzy classifier (image taken from [152]).
In this case Layer 1 represents the fuzzification of the inputs and Layer 2 holds the rule nodes. In Layer 3 the outputs are linearly combined and in the final Layer the linear sums are passed through sigmoidal activation functions to compute the degree of membership of the input in each class.

One basic assumption that has been made in the above neuro-fuzzy systems is that the input space has been already partitioned into fuzzy sets and that the fuzzy rules have all been provided a priori. Wang et al. [155] described in 1992 a heuristic method for fuzzy rule generation, whereby the input space of each variable is divided in fuzzy intervals and then input-output pairs are utilized to determine the fuzzy rules, along with their attributes. Their approach however does not give any recommendation on how to choose the fuzzy sets in the inputs, while it might also allow a very large number of rules to be generated.

A conceptually similar approach was given by Abe et al. [156] in 1995. According to their method, fuzzy rules with variable fuzzy regions are defined by activation hyperboxes which show the existence region of data for a class and inhibition hyperboxes which inhibit the existence of data for that class. Their approach results in trapezoidal membership functions and it produces fuzzy decision boundaries that yield very accurate performance. However, as in the previous case, the number of generated rules can grow very high while it
also assumes that data are clustered in hyperboxes in higher dimensional spaces.

Mitra et al. [157] defined a fuzzy multi-layer perceptron model in 1992. According to their model, input data are first transformed into 3 basic fuzzy sets that divide the universe of discourse of every variable. Then the inputs are propagated through multiple hidden layers that retain their sigmoidal activation functions. As an input pattern may have fuzzy memberships to more than one cluster (class), the network’s outputs are in terms of fuzzy belonging values, which are then used to create the error vector during training (instead of crisp, e.g. 0 or 1, classification results). This allows the net to more efficiently classify data with overlapping class boundaries and to avoid oscillations of the decision surface when almost identical pattern are clamped to different classes. They used that model to construct a connectionist expert system, which can handle linguistic inputs and hedges (very low, kind of high, about, etc). Finally, the difference between the output of a neuron (class membership value) and the sum of the outputs of the rest of the neurons is taken as a certainty measure on that decision.

Another approach presented by Figueiredo et al. [158] in 1999, consists of a neurofuzzy network that learns in two phases: 1) A self-organizing phase that lets the input layer neurons to cluster and to perform weight adaptation (competitive learning) and 2) a second phase which uses a supervised learning
scheme for rule consequents adaptation. This approach can be considered as more generic and allows data to cluster more freely than to impose certain shapes on the membership function, but it also assumes that there is a large and representative dataset available.

Finally, a similar approach to [158] was published by Juang et al. [159] in 1998. They proposed an On-Line Self-Constructing Neural Fuzzy Inference Network (SONFIN), which used a structure learning phase (unsupervised) in order to determine the partition of the input space, construct the fuzzy rules and identify the optimal consequent structure. Then using a supervised parameter learning phase, the attributes of the membership functions and the weights of the rules were fine-tuned.

6.2.2 Adaptive parameter estimation for user-learning

The way we have defined the MFPN shown in Fig. 5-6, shows that the Prognosis reasoning process has a similar structure to the neuro-fuzzy systems that were discussed in the previous section. This direct analogy motivates us to employ in our case some of the parameter estimation techniques that were mentioned before. However, the main problem that we face here (as it was also pointed out in section 4.2.4.2) is the lack of a large-scale evidence-base containing a variety of labeled multi-dimensional physiological recordings. As a result, in our case we must concentrate on a more conservative approach, which
will start off with some expert system structure as the one that we described in the previous chapter. Then this network can be continuously improved during a supervised patient-learning phase in order to fine tune the system’s attributes. Ideally such a dataset could be collected and then processed offline in order to determine and label the important physiological changes in the monitored users.

Given a new input vector $x_k$ the inputs will be first fuzzified (e.g. update the fuzzy values in Layer 2), then they will be propagated to Layer 3 where the smoothed confidence of the symptoms is computed (using also the momentum terms of the places). Finally the fuzzy symptom truths are used to fire the production rules that constitute the system’s knowledge base and generate the confidence level of the disorders that are currently considered in the evidence-base. These values will also be aggregated to generate a value that indicates the overall healthiness of the user.

In the system’s training phase, the outputs of Layer 4 will be used to generate the errors that will be propagated back to the network to update its parameters. As the MFPN’s outputs are in terms of fuzzy values [0,1] that indicate the degree of confidence that a specific disorder is present, the target values according to which the error will be computed need to be in the same form. As a result, with each input vector we need a set of output labels describing the
degree of membership (or better occurrence) of each disorder. These labels can be chosen from a set like:

{not present, highly unlikely, unlikely, slightly unlikely, somewhat possible, possible, very possible, probable, very probable, highly probable, certain}

The above linguistic parameters can be mapped to the interval [0,1] in order to acquire numerical values for computing the network’s errors:

(Eq. 6-3) \( E_k = \frac{1}{2} \sum_{i=1}^{N} (o_{k,i} - d_{k,i})^2 \),

where \( o_{k,i} \) is the output of the \( i^{th} \) rule and \( d_{k,i} \) is the target value for that rule, e.g. a linguistic label taken from the set discussed above and mapped to a real number between 0 and 1.

In order to propagate the error to the input layers and update the value of an adjustable network parameter \( z \) we need to compute:

(Eq. 6-4) \( z(t + 1) = z(t) - \eta \frac{\partial E_k}{\partial z} = z(t) - \eta \frac{\partial E_k}{\partial f} \cdot \frac{\partial f}{\partial z} \)

where \( f \) is the activation function in the previous layer. Since we need to compute partial derivatives in order to evaluate the updating formula for an attribute, we can not employ non-differentiable operators such as \textit{max} and \textit{min}. Instead we may use the \textit{probabilistic sum} and the \textit{probabilistic product} or alternatively employ sigmoidal activation functions. Without loss of generality we can consider that the output of a production rule \( i \) is computed as:

(Eq. 6-5) \( o_{k,i} = a(d_i)_n = m(p_i)_0 \cdot u_{R_i} \cdot o_{R_i} \cdot f(R_i) + m(p_i)_1 \cdot a(d_i)_{n-1} \)

175
In this case if we want to update one of the weights $w_j$ of the normalized weighted sum given by $f$ in the fourth layer we have to compute (where we can omit parameters $u_{R_l}$ and $o_{R_l'}$ since they are equal to 1 in this case and we have also omitted index $k$ of the $k^{th}$ training input-output pair to avoid confusion):

$$\frac{\partial E}{\partial w_{jl}} = \frac{\partial E}{\partial o_l} \cdot \frac{\partial o_l}{\partial f} \cdot \frac{\partial f}{\partial w_{jl}} = (o_l - y_l) \cdot m(p_l)_0 \cdot \frac{\partial f}{\partial w_{jl}} , \quad \text{where:}$$

(Eq. 6-6) $\frac{\partial f}{\partial w_{jl}} = \frac{a(d_j) \Sigma_j w_{jl} - \Sigma_j a(d_j) \cdot w_{jl}}{[\Sigma_j w_{jl}]^2}$

If we now want to update a weighting parameter $w_{rl}$ involved in the computation of the confidence of a place $p_l$ in layer 3 we have to evaluate the following partial derivative:

$$\frac{\partial E}{\partial w_{rl}} = \sum_{j \in C} \frac{\partial E}{\partial o_j^{(4)}} \cdot \frac{\partial o_j^{(4)}}{\partial o_i^{(3)}} \cdot \frac{\partial o_i^{(3)}}{\partial w_{rl}} , \quad \text{where:}$$

(Eq. 6-8) $\frac{\partial E}{\partial w_{rl}} = \sum_{j \in C} \frac{\partial E}{\partial o_j^{(4)}} \cdot \frac{\partial o_j^{(4)}}{\partial o_i^{(3)}} \cdot \frac{\partial o_i^{(3)}}{\partial w_{rl}}$

we denote as $o_j^{(4)}$ the confidence of place $j$ in layer 4, as $o_i^{(3)}$ the confidence of place $l$ in layer 3 and as $C$ the set of places in layer 4 whose confidence is computed by a weighted production rule that has place $p_i^{(3)}$ (and thus proposition $d_l$) in its antecedent part.

The first partial derivative in the sum of (Eq. 6-8) has been computed above as part of (Eq. 6-6). The second derivative is given by:

$$\frac{\partial o_j^{(4)}}{\partial o_i^{(3)}} = \frac{\partial o_j^{(4)}}{\partial (a(d_l))} = m(p_j)_0 \cdot \frac{w_{lj}}{\sum k w_{kj}}$$

(Eq. 6-9)
Also the last partial derivative of (Eq. 6-8) can be computed very easily if we consider that the fuzzy conjunctions of layer 3 are evaluated with the probabilistic t-norm.

Finally, the weight can be propagated in a similar manner up to the input layer, where the shape parameters of the fuzzy sets that divide the universe of discourse of the input features can be updated. Again, it would be favorable to employ continuous functions to describe the fuzzy sets. As a result we can employ Gaussian, bell-shaped or sigmoidally-shaped membership functions instead of the trapezoidal functions that we used to describe the fuzzy sets in section 4.2.1. In Fig. 6-4 and Fig. 6-5, we see fuzzy sets representing the various levels of systolic blood pressure, whereby in the first case bell –shaped membership functions have been used:

\[(\text{Eq. 6-10}) \quad \mu(x; a, b, c) = \frac{1}{1 + \left| \frac{x-c}{a} \right|^2} \]

while in the second case the membership functions are composed as a difference of two sigmoidal functions, e.g.

\[(\text{Eq. 6-11}) \quad \mu(x) = \mu_1(x; a_1, c_1) - \mu_2(x; a_2, c_2), \text{ where:} \]

\[(\text{Eq. 6-12}) \quad \mu_i(x; a_i, c_i) = \frac{1}{1 + e^{-a_i(x-c_i)}} \]
The above process can lead to optimal parameter estimation for the fuzzy inference mechanism described in the previous chapter. The main challenge here is though to determine the set of production rules that will be employed in order to describe the clusters that are created in the multi-dimensional input feature space. To do that, expert’s knowledge needs to be acquired during a patient-learning process, whereby the rules to be included in the individual knowledge-base should be chosen with respect to the individual’s physiology, pathology and medical history.
CHAPTER 7: USER – DEVICE INTERACTION AND SYSTEM SIMULATION

As it was stated in the previous two chapters, the process of assessing the patient’s health status could be significantly enhanced if the user of the Prognosis wearable platform could interact with the system. The reason behind this is that a) it would help and hopefully also motivate the user to have an even more active role in his personal health management and b) most importantly it would provide a greater context for physiological measurements which would have been otherwise impossible to capture in a remote and unsupervised health-monitoring scenario.

In this chapter we will briefly discuss the interaction mode that is integrated in the Prognosis system, by introducing a well-known speech recognition platform that we employed in the system design and by providing simple examples of user-device dialogue. Finally we will describe a simulation framework that has been developed to simulate the operation of the Prognosis system according to the overall operational model of the system described in the current chapter and in chapters 4 and 5.
7.1 Prototyping the Human-Device Interaction

In order for the wearable system to be capable of capturing important diagnostic information from the user’s voice recording, the interaction scheme between the user and the system must be well-defined. In addition to that, we need to take two more criteria into consideration: Firstly, the fact that the feedback mechanism must not be allowed to grow very high in complexity since it is either going to be implemented directly on a hardware-constrained mobile platform, e.g. on a dedicated processing board, or an implementation approach such as the one adopted by Google search [160] or by other voice recognition applications for mobile phones such as Vlingo [161], may be taken. In the latter case, a voice recording is being taken on the mobile device and then that recording is sent via a network connection to a dedicated speech recognition server, which attempts to identify the uttered phrase and then returns the identified sentence in the form of a string to the mobile application.

Secondly the interaction process must be user-friendly to ensure usability and thus also user acceptance. In order to achieve those goals as well as real-time performance it is necessary for the system to be able to respond to specific user feedback, e.g. to request more information, to ask the user to repeat something he said, to inform him about an event etc in a well defined manner.

There is a variety of speech recognition software available, most of which being proprietary. The most important freely available software in this
field are HTK developed at Cambridge University [162] and Sphinx-4 [163] developed from the CMU Sphinx Group at Carnegie Mellon University in collaboration with Sun Microsystems Laboratories, Mitsubishi Electric Research Labs and Hewlett Packard. HTK’s source code has been written in C, while Sphinx-4 has been entirely developed in Java. Both tools are considered to be state-of-the-art speech recognition software and both of them are based on Hidden Markov Models (HMMs) for statistical speech recognition. For the prototyping of the Human-Device Interaction (HDI) in the Prognosis wearable system we have used Sphinx-4. The reason for choosing Sphinx-4 over HTK is simply the fact that Sphinx-4 has been developed in Java and since this is the language in which the system prototype has been programmed (as it will be discussed in the next chapter which describes the system’s implementation), it will be easier to integrate and interface it with the rest of our work.

The first step in defining the interaction scheme between the user and the system consists of specifying what the vocabulary to be used is going to be. The vocabulary includes all possible words that the system needs to be able to understand. Examples of words in the vocabulary could be simple answers like “Yes” or “No”, a set of nouns or phrases describing the possible symptoms of the patient like “headache”, “malaise”, “chest pain” etc and a set of adjectives to describe the type of a certain symptom like “continuous”, “acute” etc.
Another important step is the matching of the vocabulary words to their phonetics, e.g. describing how each word is pronounced. Examples of some of the words in the dictionary are given below:

<table>
<thead>
<tr>
<th>Word</th>
<th>Pronunciation</th>
</tr>
</thead>
<tbody>
<tr>
<td>YES</td>
<td>y eh s</td>
</tr>
<tr>
<td>NO</td>
<td>n ow</td>
</tr>
<tr>
<td>HEADACHE</td>
<td>hh eh d ey k</td>
</tr>
<tr>
<td>ACUTE</td>
<td>ah k y uw t</td>
</tr>
<tr>
<td>CHEST PAIN</td>
<td>ch eh s t . p ey n</td>
</tr>
<tr>
<td>PERSISTENT</td>
<td>p er s ih s t ah n t</td>
</tr>
<tr>
<td>COUGH</td>
<td>k aa f</td>
</tr>
<tr>
<td>DIZZINESS</td>
<td>d ih z iy n ah s</td>
</tr>
</tbody>
</table>

The next step consists of defining the grammar which specifies the constraints of the expected utterance. Here we have to consider what type of feedback the system should be expecting from the user, e.g. phrases like “I feel intense pain on my chest” or “I am dizzy and weak”. The more complex the phrases that the system should be able to recognize, the more complex the grammar will get.

It should be noted here that in general the performance of speech recognition software in terms of accuracy and speed is strongly dependent on the size of the employed vocabulary. Furthermore by trying to limit the grammar to one-word phrases and thus requiring mostly isolated word recognition the performance of the speech recognizer can be further increased (or not allowed to deteriorate). In accordance to these guidelines, we had initially decided to include only one-word phrases in the vocabulary for the HDI in Prognosis. The reason is that there is a considerable amount of symptoms and symptom de-
criptions that we need to guarantee that the system will be able to comprehend and also that the interaction should be as fast as possible, while remaining simple and easy to program and utilize. However, we can, without too much extra computational burden, define a somewhat more complex grammar that would also allow the recognition of short phrases like the ones mentioned above.

Sphinx-4 provides a grammar definition language for specifying grammars for phrases like these. The language consists of a set of variable definitions followed by regular expressions that describe the phrases to recognize. An initial grammar for the Prognosis HDI is given by (where <> denotes a non-terminal symbol and λ is the null variable):

```plaintext
<START> → <SENTENCE_START><QUANTIFIER><INTENSITY><FREQUENCY><DURATION>
<SYMPTOM><PROP><LOCATION><PROP><SIDE>;
<SENTENCE_START> → I feel | λ;
<QUANTIFIER> → very | slightly | λ;
<INTENSITY> → weak | medium | strong | λ;
<FREQUENCY> → continuous | intermittent | λ;
<DURATION> → instantaneous | persistent | λ;
<SYMPTOM> → malaise | dizziness | pain | numbness | cough | headache | weakness | stomach ache | fatigue | nausea | discomfort;
<PROP> → on | over my | on my | λ;
<LOCATION> → arm | leg | torso | head | λ;
<SIDE> → left | right;
```

After defining the vocabulary and the grammar of the interaction scheme, an HMM (or a set of HMMs) needs to be trained for every word or phrase and thus for their corresponding phonemes by using a number of examples for that spoken word or phrase. Training of the HMM refers to the
process of estimating the parameters of each HMM and thus constructing acoustic models for words and phrases. In the case of continuous speech recognition each word will be defined as a sequence of phoneme-based HMMs. Finally to recognize some unknown word or phrase the likelihood of each model generating that word or in the more complex case the likelihoods of all the paths in the HMM interconnected network will be calculated and the most likely model or path will be identified.

7.2 An example of HDI via voice

In this section we present a possible scenario to illustrate the operation of the Prognosis wearable system and especially the interaction between the user and the device via the voice module. Consider the case where the user of the wearable system is an elderly individual with a personal medical history of signs or tendency towards cardiovascular disease. We can imagine this person going about his daily activities either in his own home or even at a care center, while he is wearing the wearable health-monitoring system, for example a chest belt measuring ECG, body temperature and respiration rate, a wrist-based pulse oximeter and a cell-phone placed on his waist with a hands-free or Bluetooth earpiece.

Now consider the following scenario: The patient starts to feel a slight discomfort in the chest along with a feeling of malaise. His heart rate may also
increase and his heartbeats may become more irregular. Furthermore, his respiration effort may also become stronger and quicker and there might be a slight decrease, perhaps in a longer period of time, of his blood oxygen saturation level. The patient decides he should sit down but does not consider his condition alarming or hazardous enough to call for help. However the moment the system detects all the above events, e.g. the raise in the levels of some of the bio-sensors’ measurements along with a possible arrhythmia, it decides to take action and initiates a brief dialogue with the user in order to obtain more information about his condition and then to act accordingly. The first thing the system may ask the user for, is if he needs immediate help or not. Supposing the user answers no to the previous question, the system may continue to inquire about any other possible symptoms that the patient may be experiencing. Then the patient mentions the term chest pain. In that case a small series of chest pain related questions may follow, which can lead to the establishment of such a health-condition context so that the system may indeed choose to send a notification to the medical station or to the supervising physician. Eventually appropriate help and care may be provided on time to the patient in order to possibly avoid the deterioration of his condition.

The previously described scenario is illustrated in the following in terms of the events that occur and the actual dialogue that may take place:
User: <Feels a slight chest discomfort and some of his vitals start to go out of range>

System: Some of your vitals are out of regular range. Do you need immediate help?

User: No

System: Do you have any other symptoms?

User: Yes

System: Please describe your symptoms

User: I have slight chest discomfort.

System: <the system does not acknowledge the user’s phrase>

Please state your symptom again in a brief manner

User: chest pain

System: <Recognizes the symptom>

Is the symptom intermittent or continuous?

User: continuous

System: Is the symptom instantaneous or persistent?

User: persistent

System: Is the pain radiating to your arms?

User: Yes

System: Is the pain aggravated by movement or breathing?

User: No

System: Are you experiencing any other symptoms, such as hemoptysis or sputum?

User: No

System: Describe your pain level on a scale from one until ten

User: five

System: <decides that according to the user’s answers and his recent vital signs the user could be in an alarming health condition>

<notifies the medical center by sending an alarm message via GPRS>

I have notified the medical center about your current condition
As the above example illustrates, the interaction between the system and the user can either be invoked by the user himself or by the system, in case specific changes are noticed in the user’s physiology. This can be done by basically constructing a decision tree like structure, where according to the irregularity detected by the system and the user’s previous responses and medical history, appropriate questions may be invoked to get additional important feedback from the user about his symptoms. Such a process is shown in Fig. 7-1 in the form of a Petri Net, where the system evaluates the answers and other preconditions to decide on posing a certain set of questions.

Fig. 7-1. HDI interaction initiated by the system.
7.3 Simulation Framework

In this section we will describe the simulation framework we have developed according to the SPN and FPN-based models of the WHMS, which were presented in chapters 4 and 5. The purpose of the derived simulation is to illustrate the feasibility as well as the actual operation of the Prognosis WHMS. In addition to that it can be used to study the dependencies between the various components of the system and to highlight the synchronization and concurrency issues that arise in such a system design. The presented framework has been implemented in Java, since the Java programming language can facilitate a modular and hierarchical design approach, in correspondence to the SPN model. Furthermore Java provides the required synchronization and concurrency primitives for simulating the various tasks that can be present in a WHMS model as explained below.

As we have seen so far, a wearable health-monitoring system may comprise a variety of bio-sensors. In the WHMS simulator, the types of sensors that can be selected to be included in the system’s simulation include an ECG monitor, a thermometer, a blood pressure monitor, a respiration rate sensor, a pulse oximeter and a speech recognizer/synthesizer module to capture additional feedback from the user in terms of verbally-described symptoms. In the initial simulator’s design we have assumed that the ECG monitor includes an automated classifier, thus its output is comprised of per beat classification re-
sults along with the raw ECG data. Such a classifier can be implemented via a light-weight and computationally-inexpensive algorithm for ECG classification, like the one discussed in Chapter 3.

Under the previous assumption, the collection of physiological measurements in the central node as it was depicted in Fig. 4-15 can take place according to two different disciplines:

- The central node (e.g. the main thread or the main processing unit) is responsible for polling the distributed biosensors (or the ring buffers dedicated to each type of signal) for data in a round-robin manner and at a constant rate and therefore it enforces explicit synchronization in the way the data are collected.

- Dedicated threads (or even dedicated hardware) are assigned for collecting the measurements from each sensor and for asynchronously signaling the event of new data being collected to the main monitoring thread or process (or to the main processing unit). The main process is synchronized with the event of “data arrival” and wakes up after a complete cycle of data collection to process the new set or the new vector of measurements according to the Prognosis language data fusion scheme.

The former is a more centralized and thus a more straightforward and rather simplified approach to data collection, as there is an explicit synchronous manner of data gathering: the central node continuously and iteratively
addresses each sensor in a round robin fashion and either collects data from the sensor’s buffer or requests that an instant physiological measurement is performed.

The latter case is a quasi-asynchronous distributed approach to data collection and therefore it is more complex and more general than the former one. For every type of bio-signal to be monitored, there is a dedicated process responsible for collecting the data, pre-processing it and passing it over to the main monitoring process. Such a process can be thought of as either a thread or a dedicated hardware module, which signals (interrupts) the main process when new data are available. However, even in this case there is implicit synchronization in the main process, as we are assuming that the distributed processes have been programmed to collect data at a regular rate. Finally, the main process is awoken upon a “new data collected” event (or upon completion of a new cycle of data collection from each sensor) and it immediately starts fusing the data according to the Prognosis scheme, in order to derive an estimation of the user’s health condition.

In the current implementation of the WHMS simulator all the bio-sensors (or bio-signals) are represented by individual classes – threads, which constantly generate new data in a semi-random manner and deposit them in dedicated ring buffers. It is up to the main process to read new data in the buffers before they get overwritten, but this issue is dealt with in this case by ex-
plicitly defining identical data delivery rates for all types of sensors. Furthermore signal samples are not generated in a totally random way, but in a rather biased manner based on previous data and also based on recent values of other bio-signals, e.g. low oxygen saturation is more possible to coincide with high respiration rate etc. For example, heart rate values will be generated according to a normal distribution where the initial mean is set as a generally normal heart value, e.g. 65 BPM. The subsequent heart rate value will be generated from a Gaussian distribution as well, however this time the mean of the normal distribution from which the new value will be generated is the previous heart rate value. That way simulated bio-signals will have a smoother time graph and thus they will be more realistic. Furthermore, the values of other signals’ previous values will control the standard deviation of the signal generator, e.g. “extreme” or “unstable” values in other signals should increase the “instability” of the current signal as well, which means increasing the standard deviation of the corresponding bio-signal and therefore allowing it to possibly take more extreme values as well.

Regarding ECG data, in the WHMS simulator they are not being generated randomly but they are rather taken from real ECG recordings. Namely, the simulator upon start-up will let the user choose among the 48 available ECG recordings in the widely studied MIT-BIH Arrhythmia Database [99]. These ECG recordings have been annotated on a beat-by-beat basis by profes-
sional physicians, which have also provided annotations regarding the heart rhythm, e.g. normal, sinus bradycardia, atrial fibrillation, ventricular bigeminy etc. These annotations have been directly employed in the WHMS simulator as an automated ECG classifier, providing concrete heartbeat and heart rhythm classifications.

In Fig. 7-2 the start-up menu of the WHMS simulator is shown. As explained before, the user can choose one of the available ECG recordings in the MIT-BIH Database to be included in the simulation. Furthermore, the type of sensors that will compose the system to be simulated can be chosen at the same set-up screen. Finally, the size of the simulation step can be manually selected, where 1 sec corresponds to normal or real-time operation (e.g. sensors readings are collected every 1 second).

![Fig. 7-2. WHMS Simulator start-up screen](image)

Choose MIT-BIH Record # and lead
Choose sensors for the simulation
Slower (or even faster) simulation steps may also be chosen in order to study more closely the simulator’s operation and results.

Coming back to the simulator’s actual operation, once a new measurement is acquired, it will first be checked for validity. In the case that there is a redundant number of sensors used to monitor each bio-signal, erroneous values can be relatively easy detected by considering the correlation of the measured data between sensors of the same type. However the fact that a single sensor is employed for each bio-signal makes this task very challenging. A simple approach to this issue is to make use of a linear predictor for each type of signal and then calculate the difference between the predicted value and the actual value that was measured by the sensor. Then if that difference is found to be greater than a predefined threshold the corresponding measurement can be considered erroneous. The drawback of this method is that it requires calibration for each sensor in order to estimate the predictor’s order and parameters as well as the threshold value to be used.

After validating each acquired measurement (or even pattern), the physiological data are fuzzified according to the fuzzy sets defined in section 4.2.1. Then these fuzzy certainty values are used to compute the outputs of layers 3 and 4 in the FPN network shown in Fig. 5-6.

The system re-evaluates the condition of the user at any given time that any new “pathological” or slightly abnormal signal state is detected. According
to the estimated health-threat level, it may choose to either alert the user or to request further information from him/her or even to send an alarm message to a remote medical center or to the supervising physician’s cell-phone.

In Fig. 7-3 the main screen of the WHMS simulator is depicted.

The ECG plot displays real-time ECG data along with a per-beat classification. In addition to that, there is a plot for every one of the rest of the recorded bio-signals, where the signal’s trend and its instantaneous value is being depicted. Furthermore, there is a text box where any possible extracted health symptoms are printed out, for example tachycardia or hypotension or fever etc. An alarm-detection box is used to print out any recognized health conditions, such as hypothermia or left ventricular failure for example,
represented in the local signal database in terms of specific symptom strings or words. A health-threat-level display bar is used, which illustrates the risk level of the user’s health as perceived by the wearable system and which is based on the computation of the fuzzy certainty of the single node in layer 5 of the FPN shown in Fig. 5-6. Finally a pop-up dialogue frame appears (at the bottom of the screen) whenever there is a user-device interaction initiated. This pop-up frame displays the dialogue that takes place between the hypothetical user and the system’s simulator.

Finally we should that the described WHMS framework simulates the ideal operation and performance of an actual system prototype since:

- The system is capable of performing accurate physiological sensing and robust on-site signal processing & feature extraction/pattern recognition.
- Outlier detection is carried out successfully in the measured data.
- There is coherent speech recognition and user-device interaction.
CHAPTER 8: SYSTEM IMPLEMENTATION

In this chapter we will provide a detailed description of the implementation of the prototype platform of the *Prognosis* WHMS. As it was shown in the survey presented in Chapter 2, there are various approaches towards realizing a wearable health-monitoring system: utilizing motes (e.g. tiny wireless sensor nodes), realizing a system based on a custom microcontroller board with wired on-body sensors, employing textile sensors for a garment-based system design and finally making use of off-the-shelf wireless sensors and PDAs or smart-phones as system central nodes. Hybrid designs are also possible, whereby two or more system design approaches are merged. The approach that has been adopted in our research to implement the *Prognosis* prototype is to make use of off-the-shelf wireless-enabled bio-sensors and a state-of-the-art smart-phone as the central node. The reasoning behind this choice is that a) it is not the goal of the current research to develop application-specific hardware, e.g. new wearable sensor technologies, b) wearable wirelessly-enabled bio-sensors constitute the current state-of-the-art in unobtrusive physiological monitoring solutions, c) smart-phones and PDAs comprise a new wave of portable computing devices, that are constantly becoming more powerful and diverse and that will undoubtedly play a significant role in our future every-day life. As a result it is of particular importance to be able to realize intelligent
and interactive health-monitoring capabilities with pervasive hardware technolo-
gies.

8.1 The WHMS’ Hardware Components

The components of the Prognosis WHMS prototype are depicted in Fig. 8-1.

![Fig. 8-1. The Prognosis WHMS prototype](image)
The system is built around the BlackBerry Bold 9000 PDA/smart-phone [164] developed by RIM [165], which has been chosen due to its following properties:

- Powerful processing capabilities (624 MHz processor) for a cell-phone.
- J2ME (Connected Limited Device Configuration - CLDC 1.1 and Mobile Information Device Profile - MIDP 2.0 APIs) and multi-threading/multi-programming support.
- The BlackBerry® Java Development Environment contains a plug-in for Eclipse™ to develop, debug and test BlackBerry device applications from within the Eclipse Integrated Development Environment (IDE) [166].
- Support for standard JAVA APIs, like the JAVA API for Bluetooth®.
- Support for multiple simultaneous Bluetooth connections (maximum 5), thus allowing the formation of Bluetooth PANs (Personal Area Networks) or piconets.
- 3G (HSDPA) and WLAN (802.11a/b/g) support.

Table 8-1 provides a list of the most important hardware and networking specifications of the Blackberry Bold 9000.

The set of Bluetooth-enabled bio-sensors that have been currently employed in the WHMS prototype include the Bioharness BT chest belt by Zephyr™ [81], the 4100 Digital Pulse Oximeter by Nonin Medical [66] and a Blood Pressure monitor from A&D Medical [167].
Table 8-1. Hardware Specs of Blackberry Bold 9000

<table>
<thead>
<tr>
<th>Feature</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU</td>
<td>Intel Xscale 624MHz processor</td>
</tr>
<tr>
<td>Memory</td>
<td>128MB RAM (Flash) &amp; 1GB storage</td>
</tr>
<tr>
<td>Weight</td>
<td>133g</td>
</tr>
<tr>
<td>Operating System</td>
<td>BlackBerry OS (proprietary software platform), supports multi-programming</td>
</tr>
<tr>
<td>Wireless Networks</td>
<td>GSM, GPRS, UMTS, EDGE, HSDPA</td>
</tr>
<tr>
<td>Wi-Fi</td>
<td>802.11a/b/g enabled</td>
</tr>
<tr>
<td>Bluetooth</td>
<td>Bluetooth® v2.0, Serial Port Profile and support for up to 5 simultaneous RFCOMM connections (PAN)</td>
</tr>
<tr>
<td>Battery</td>
<td>1500mAhr Lithium battery, 13.5 days standby time</td>
</tr>
<tr>
<td>GPS</td>
<td>Built-in GPS</td>
</tr>
</tbody>
</table>

The Bioharness BT, which can be seen on the bottom left in Fig. 8-1, is a physiological monitoring fabric-based chest-belt device capable of measuring ECG, heart rate, respiration rate, skin temperature, activity and posture. The communication protocol of the device is protected by a confidentiality agreement and cannot be elaborated upon here. However, the following specs can be given:

- The Bioharness BT can continuously stream data over a single Serial Port Profile (SPP) connection (10m range) to a Bluetooth-enabled device.
- The device is capable of periodically transmitting real-time ECG data, which are sampled at 250Hz and digitized with 10bit resolution (it uses dry-contact skin electrodes).
- It is also capable of periodically transmitting heart rate, respiration rate, skin temperature, activity level and posture measurements.
The wireless pulse oximeter from Nonin® is a wrist-worn device (as it can be seen on the bottom right in Fig. 8-1), which uses an infrared-light-emitting finger clip sensor to capture the photoplethysmograph (PPG) signal, which captures the changes in blood volume. The pulse oximeter can transmit the raw PPG signal or periodically derived values of oxygen saturation and heart rate over a single SPP connection, with 10 meters maximum range.

The Bluetooth-enabled blood pressure monitor from A&D Medical, includes a base-device and a blood-pressure cuff which is placed around the user’s arm to take a measurement. When new data have been acquired by the device, it automatically sends them to any Bluetooth paired device. The data that are communicated are the systolic and diastolic blood pressure levels as well as the pulse rate. The main difference of this sensor with the previous two, is that in this case a measurement needs to be initiated manually by the user and that the device cannot be worn continuously.

Finally as it can be seen in Fig. 8-1, a Bluetooth headset is depicted to represent the voice-interaction that is possible in the system. Future research could investigate the feasibility of simultaneously incorporating a Bluetooth headset for capturing voice feedback. However, in this case the effect of the A2DP connection for the headset on the operation of the PAN should be carefully examined and evaluated.
8.2 Implementation Walkthrough

The operation of the WHMS prototype with the previously described hardware components is based on the formation of a Body Area Network (BAN). In the context of Bluetooth this is referred to as a piconet or more generally as a Personal Area Network (PAN). In the simplest case, there are two devices in the piconet, e.g. there exists a single point-to-point communication link. The device that initiates the connection is called the master and all other devices in the piconet are referred to as slaves. All devices that are part of the same piconet share the same physical radio channel in that they are all synchronized to a common clock and a certain frequency hopping pattern, which is derived from the master’s clock and the master’s Bluetooth device address.

The Bluetooth® specification permits up to seven slaves to be active in a piconet. This constitutes a communication topology referred to as point-to-multipoint, as there is a physical link between each slave and the master. In this formation the PAN’s master (e.g. the BlackBerry smart-phone) assigns dedicated time slots to the peripherals and data are received in a protocol-inherent round-robin manner, although it has been shown that are more efficient intra-piconet scheduling algorithms [168]. The type of Bluetooth connections between the BAN’s master and the slaves are Serial Port Profile (SPP) connections that utilize the Radio Frequency Communication (RFCOMM) protocol to emulate RS-232 serial ports.
The J2ME software architecture of the currently implemented WHMS prototype on the BlackBerry Bold platform can be seen in Fig. 8-2. All user actions in the Prognosis App are made available through a single Screen, in order to avoid pushing more screens onto the stack (and thus increasing memory requirements) and in order to keep the application as simple and easy to use as possible. The “Menu Button” of the BlackBerry device makes the following 4 actions available to the user: a) initiate a client socket connection to a remote network server, b) initiate a connection to the Bioharness BT, c) initiate a connection to the Pulse Oximeter, d) exit the application.

By selecting b) or c), the application will try to locate an already paired Bioharness BT (or Nonin Pulse Oximeter). If such a pairing is found, then the application attempts to establish a socket connection to the corresponding Bluetooth-enabled bio-sensor with the appropriate protocol specification (SPP in this case). After successfully establishing a connection, listener objects for handling input and output data are created and then the appropriate handshaking, for setting up the operational mode of the sensor, is performed. Once the previous processes are completed, a new thread is created which is responsible for monitoring the corresponding serial port connection.
Fig. 8-2. J2ME Software Architecture of the Prognosis prototype.
As it can be seen in Fig. 8-2, a SPP monitoring thread constantly waits and receives a data packet from the sensor, then it un-mashals (de-serializes) the received packet, it verifies the received physiological measurements (currently this is done just by checking the corresponding sensor’s status flags), it performs appropriate pre-processing (in the case of ECG data, the R peaks are detected by implementing in Java the algorithm developed by Patrick Hamilton and which can be accessed at [169]) and finally it delivers the data to the Main Screen thread. Following that, the new data are saved in dedicated ring buffers and then they are plotted on the screen. Fig. 8-3 shows two screenshots of real-time data displayed on the smart-phone.

There is also a thread that listens continuously for incoming data from the A&D BP monitor, since the BP monitor acts as a master device in the Bluetooth communication. Whenever new data are received, the listening thread delivers them to the main thread.

Moreover, one of the side-buttons of the BlackBerry smartphone is utilized to initiate a thread which will take a voice recording from the user. When the user releases the button, the thread handles the recorded data to the main thread (and then it terminates) which will immediately send them via Wi-Fi to the remote workstation for speech recognition. After the speech data have been recognized, the result is sent back to the phone as a simple string.
After receiving and plotting the data the main thread will deliver the new measurements to the Data Fusion thread, which is responsible for parsing the derived Prognosis string and for returning an estimation of the user’s health condition to the main thread. Upon detection of an alarming health condition, the main thread will communicate the alarm via the Wi-Fi socket connection to the remote server.

Finally, in the current prototype we have realized an “intensive” physiological data monitoring mode, whereby all the physiological measurements captured from the WHMS are transmitted to the remote server for real-time data inspection. In this scenario, the server application (running on a conventional laptop computer) continuously receives data packets from the BlackBerry device and plots them on the computer’s screen for inspection. A GUI of this application can be seen in Fig. 8-4. In this screen we can also see a window displaying the current status of the Wi-Fi connection, trend graphs for all measured bio-signals (as well as for the RR derived parameter) and finally also
an indication of packets being lost in the Bluetooth BAN and over the Wi-Fi socket connection.

Fig. 8-4. The GUI of the remote medical server.
CHAPTER 9: CONCLUSIONS AND RESEARCH CONTRIBUTIONS

9.1 A wearable system prototype for health monitoring

This dissertation work describes research related to the design of an intelligent, self-adaptive and interactive wearable system for health monitoring of people at risk. To provide a deeper insight into the current state-of-the-art of WHMS, a comprehensive survey on corresponding research prototypes and commercial products was carried out. This helped us identify the shortcomings of the current technology, but also to define new capabilities that could be integrated in WHMS to improve their overall functionality and also the users’ quality of life.

The developed prototype platform utilizes cheap off-the-shelf components and constitutes a novel paradigm of how multiple Bluetooth-enabled biosensors can be utilized for ubiquitous health monitoring applications. The developed interfaces on both the smart-phone and the remote workstation allow the user of the WHMS and the person supervising the patient to have a complete picture of the user’s health and to have instant access to real-time and past physiological data.

The following full-paper refereed publications support the new scientific contributions of the Prognosis prototype:
**Journals**


**Conferences**


**Book Chapters**


**9.2 Research contributions**

The new research contributions can be synopsized in the following points:

• Investigation of efficient methods for denoising of the ECG signal was carried out, whereby unusable segments are detected by examining the energy
of the various wavelet scales of the DWT transformed signal. Retained signal portions are denoised via the undecimated wavelet packet transform which is utilized for the first time for ECG denoising and yielded promising results in the presence of white and pink noise.

- A novel and efficient methodology for ECG beat classification was proposed. It utilizes features obtained by projecting the ECG templates on selected wavelet packet atoms. Extensive analysis was carried out to identify the optimum decomposition structure, by examining the effect of the wavelet type utilized, the number of samples used and the number of atoms chosen on the accuracy of the classification. The developed method is both simple and fast and yields comparable performance to other more complex classifiers.

- A novel model based on a fuzzy regular formal language was defined, which describes the current state of health of the WHMS user. It considers symptom ambiguity and causal relationships between various disorders and symptoms to derive a thorough estimation with a certain degree of confidence. This approach was inspired by pioneering medical expert systems (MYCIN and CADIAG-2), from which it borrowed some of its concepts and fused them together to yield a new representation. It is hoped that such a solution can lead to early detection and hopefully also to pre-
vention of health episodes by carefully following, capturing and describing the health trends recorded from physiological and contextual sensors.

- An extension of the decision making methodology based on a modified Fuzzy Petri Net (FPN) model was presented, which considers additional concepts such as obligatory and proving symptoms, in its reasoning process. The developed model has the form of a feed-forward fuzzy Petri Net that propagates the beliefs of symptoms to yield the confidence of the disorders described by the rule base. Finally all fuzzy membership values of the last layer of the network are combined together to estimate the overall healthiness of the user and produce an index of his health deterioration, which can be considered as an index of health stability and which can also be utilized to trigger alarms.

- A user-learning strategy based on a neural-fuzzy extension of the FPN was presented. Neuro-fuzzy networks were studied and a novel approach inspired by this domain was integrated with the Prognosis reasoning model in order to derive optimal parameter estimation and thus provide individualized health decision support.

- A system-patient dialogue interaction was defined in order to capture non-measurable patient symptoms such as chest pain, dizziness, malaise etc. This is done for the first time in the context of wearable health monitoring systems and allows out prototype to acquire feedback from the user which
is not observable via physiological sensor readings. This information can greatly enhance the embedded decision ability of the system and provides a more comprehensive user health status evaluation.

- A system prototype was developed which is based on a smart-phone that runs multi-threaded J2ME software for handling multiple simultaneous Bluetooth connections with off-the-shelf wireless bio-sensors. Although there have been several examples in the research community that leverage cell-phones and wireless sensors, the developed platform utilizes multi-threaded code ssin order to communicate with a variety of bio-sensing devices, process their results, acquire voice-feedback from the user and communicate raw and fused information to a remote location.

### 9.3 Future Research

Although this research tried to address several issues that have to do with deploying a wearable system prototype, our focus was mainly on two broad subjects: a) ECG signal processing employing portable devices in an unsupervised scenario and b) designing an overall framework for enabling personalized decision support and interaction with the system’s user. As with any engineering project there are several areas on which the proposed prototype could be improved.
One major point is ECG denoising and feature extraction. ECG denoising needs to consider several types of noise and is generally a very challenging research problem. [100]. Adaptive models can be employed to estimate optimum filters to remove artefact noise from the signal. Other physiological parameters can be used as reference signals. Additionally, the possibility of accurately detecting wave boundaries of an ECG signal being recorded on a portable device can be explored. Although the proposed ECG classification methodology proved very accurate, there is a need for quantification of some important wave attributes (such as QT duration and ST elevation), which might be proven particularly important in the decision making process.

Additionally, the knowledge base of diseases/disorders that system takes into consideration when trying to estimate the user’s health condition can be expanded with more rules and questionnaires can be utilized to acquire more detailed expert knowledge. In addition to that, a study to identify important parameters (mean, max, min, variance, derivatives etc) to improve the performance of health status classification and trend evaluation/healthy history database creation can be performed.

Finally a large scale of multi-dimensional data can be captured from a variety of patients, which can then be later used to define comprehensive statistical models that are capable of detecting even the slightest physiological changes or health trends that might be an indication of future health deteriora-
tion. Such an approach was adopted by Tatassenko et al. [170] which resulted in the BioSign™ system, which fuses five vital signs in real-time to identify patient “departure from normality”.
REFERENCES


216


[135] K.P. Adlassnig, "Update on CADIAG-2: a fuzzy medical expert system for general interna-


ALEXANDROS PANTELOPOULOS Curriculum Vitae

Alexandros Pantelopoulos received the Diploma degree from the Department of Electrical and Computer Engineering, University of Patras (Patras, Greece) in 2007. During his final year he was also an intern for six months at the Fraunhofer Fokus Institute for Open Communication Systems, Smart Environments in Berlin, Germany. He received his Ph.D. in Computer Engineering from Wright State University at Dayton, OH in 2010. His research activities involve signal processing for biomedical data, decision support systems for medical applications and portable personal health systems. Currently he is a post-doctorate fellow at the West Wireless Health Institute in San Diego, CA.