

Muhammad Mahtab Alam
Matti Hämäläinen
Lorenzo Mucchi
Imran Khan Niazi
Yannick Le Moullec (Eds.)



330

LNICST

Body Area Networks

Smart IoT and Big Data for
Intelligent Health

15th EAI International Conference, BODYNETS 2020
Tallinn, Estonia, October 21, 2020
Proceedings



Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering

330

Editorial Board Members

Ozgur Akan

Middle East Technical University, Ankara, Turkey

Paolo Bellavista

University of Bologna, Bologna, Italy

Jiannong Cao

Hong Kong Polytechnic University, Hong Kong, China

Geoffrey Coulson

Lancaster University, Lancaster, UK

Falko Dressler

University of Erlangen, Erlangen, Germany

Domenico Ferrari

Università Cattolica Piacenza, Piacenza, Italy

Mario Gerla

UCLA, Los Angeles, USA

Hisashi Kobayashi

Princeton University, Princeton, USA

Sergio Palazzo

University of Catania, Catania, Italy

Sartaj Sahni

University of Florida, Gainesville, USA

Xuemin (Sherman) Shen

University of Waterloo, Waterloo, Canada

Mircea Stan

University of Virginia, Charlottesville, USA

Xiaohua Jia

City University of Hong Kong, Kowloon, Hong Kong

Albert Y. Zomaya

University of Sydney, Sydney, Australia

More information about this series at <http://www.springer.com/series/8197>

Muhammad Mahtab Alam ·
Matti Hämäläinen · Lorenzo Mucchi ·
Imran Khan Niazi · Yannick Le Moullec (Eds.)

Body Area Networks

Smart IoT and Big Data for Intelligent Health

15th EAI International Conference, BODYNETS 2020
Tallinn, Estonia, October 21, 2020
Proceedings

Editors

Muhammad Mahtab Alam 
U02B-212
Tallinn University of Technology
Tallinn, Estonia

Lorenzo Mucchi 
University of Florence
Florence, Italy

Yannick Le Moullec 
Tallinn University of Technology
Tallinn, Estonia

Matti Hämäläinen 
Centre for Wireless Communications
University of Oulu
Oulu, Finland

Imran Khan Niazi 
New Zealand College of Chiropractic
Auckland, New Zealand

ISSN 1867-8211 ISSN 1867-822X (electronic)
Lecture Notes of the Institute for Computer Sciences, Social Informatics
and Telecommunications Engineering
ISBN 978-3-030-64990-6 ISBN 978-3-030-64991-3 (eBook)
<https://doi.org/10.1007/978-3-030-64991-3>

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2020

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

We are delighted to preface this book which constitutes the refereed post-conference proceedings of the 15th EAI International Conference on Body Area Networks (EAI BODYNETS 2020). The theme for 2020 was “Smart IoT and Big Data for Intelligent Health Management.”

Initially planned to take place in Tallinn, Estonia, the outbreak of the COVID-19 pandemic meant that the conference (like many others), took place as an online event on October 21, 2020.

We received 30 submissions; after a rigorous double-blind review process, 15 papers were selected for the final program. The main topics covered during the conference included connectivity and radio propagation, secure communication networks for smart-health, and connected wearables sensors for healthcare applications.

In addition to the high-quality technical paper presentations, the conference also featured two keynote speeches. Prof. Mohsen Guizani, a Professor at the Computer Science and Engineering Department at Qatar University, Qatar, gave a keynote speech titled “IoT Security Schemes for Healthcare Systems.” Helena Gapeyeva (MD, PhD), a Physician of Physical and Medical Rehabilitation at the Clinic of Medical Rehabilitation in East Tallinn Central Hospital, Estonia, gave a keynote speech titled “Movement analysis in Physical and Rehabilitation Medicine: Data monitoring.”

Organizing the conference would not have been possible without the support of the Steering Committee: the chair Imrich Chlamtac and members Jun Suzuki, Giancarlo Fortino, Matti Hämäläinen, and Lorenzo Mucchi. We warmly thank them for their support and guidance. We also extend our deep appreciation to the Organizing Committee team for their hard work, and to the technical Program Committee members for carefully reviewing and selecting the papers.

We hope you will enjoy reading and studying the proceedings of BODYNETS 2020, and we look forward to seeing you at the next edition of the conference.

October 2020

Muhammad Mahtab Alam
Matti Hämäläinen
Lorenzo Mucchi
Imran Khan Niazi
Yannick Le Moullec

Organization

Steering Committee

Imrich Chlamtac (Chair)	University of Trento, Italy
Jun Suzuki (Member)	University of Massachusetts Boston, USA
Giancarlo Fortino (Member)	University of Calabria, Italy
Matti Hämäläinen (Member)	University of Oulu, Finland
Lorenzo Mucchi (Member)	University of Florence, Italy

Organizing Committee

General Chair

Muhammad Mahtab Alam	Tallinn University of Technology, Estonia
----------------------	---

General Co-chair

Giancarlo Fortino	University of Calabria, Italy
-------------------	-------------------------------

TPC Chair and Co-chairs

Matti Hämäläinen	University of Oulu, Finland
Lorenzo Mucchi	University of Florence, Italy
Imran Khan Niazi	Centre for Chiropractic Research, New Zealand
Ali Imran	The University of Oklahoma, USA

Sponsorship and Exhibit Chair

Alar Kuusik	Tallinn University of Technology, Estonia
-------------	---

Local Chair

Mart Min	Tallinn University of Technology, Estonia
----------	---

Workshops Chair

Muhammad Ali Imran	The University of Glasgow, UK
--------------------	-------------------------------

Publicity and Social Media Chair

Frederick Rang	Tallinn University of Technology, Estonia
----------------	---

Publications Chair

Yannick Le Moullec	Tallinn University of Technology, Estonia
--------------------	---

Web Chair

Haris Pervaiz Lancaster University, UK

Exhibition/Demos Chair

Qammer H. Abbassi The University of Glasgow, UK

Conference Manager

Kristina Petrovicova EAI

Technical Program Committee

Abdul Saboor	Tallinn University of Technology, Estonia
Ahmed Khorshid	University of California, Irvine, USA
Alar Kuusik	Tallinn University, Estonia
Andrei Krivosei	Tallinn University of Technology, Estonia
Attaphongse Taparugssanagorn	National Electronics and Computer Technology Center, Thailand
Chika Sugimoto	Yokohama National University, Japan
Chitra Javali	National Cybersecurity R&D Lab, National University of Singapore, Singapore
Claire Goursaud	CITI Inria, France
Eryk Dutkiewicz	University of Technology Sydney, Australia
Giancarlo Fortino	University of Calabria, Italy
Girish Revadigar	Singapore University of Technology and Design, Singapore
Gravina Raffaele	University of Calabria, Italy
Hamed Farhadi	Ericsson Research, Sweden
Heikki Karvonen	University of Oulu, Finland
Hirokazu Tanaka	Hiroshima City University, Japan
Huan-Bang Li	National Institute of Information and Communications Technology, Japan
Ilangko Balasingham	Oslo University Hospital and NTNU, Norway
Ildiko Peter	NSTM, Polytechnic University of Turin, Italy
Ilkka Laakso	Aalto University, Finland
Imran Khan Niazi	New Zealand College of Chiropractic, New Zealand
John Farserotu	Swiss Center for Electronics and Microtechnology, Switzerland
Kamran Sayraflan	National Institute of Standards and Technology, USA
Kamya Yekeh Yazdandoost	Aalto University, Finland
Kimmo Kansanen	Norwegian University of Science and Technology, Norway
Ladislau Matekovits	Polytechnic University of Turin, Italy
Lin Wang	Xiamen University, China
Lorenzo Mucchi	University of Florence, Italy
Luca De Nardis	Sapienza University of Rome, Italy

Marc Girod-Genet	Institut Polytechnique de Paris, France
Marcos Katz	University of Oulu, Finland
Mariella Särestöniemi	University of Oulu, Finland
Mehmet Yuce	Monash University, Australia
Minseok Kim	Niigata University, Japan
Mohammad Ghavami	London South Bank University, UK
Muhammad Alam	Tallinn University of Technology, Estonia
Qammer Abbassi	The University of Glasgow, UK
Qiong Wang	Dresden University of Technology, Germany
Raffaele D'Errico	CEA-LETI Minatec Campus, France
Rida Khan	Tallinn University of Technology, Estonia
Rizwan Ahmad	National University of Sciences and Technology, Pakistan
Sen Qiu	Dalian University of Technology, China
Shinsuke Hara	Osaka City University, Japan
Takahiro Aoyagi	Tokyo Institute of Technology, Japan
Tauseef Ahmed	Tallinn University of Technology, Estonia
Timo Kumpuniemi	University of Oulu, Finland
Triin Kask	Tallinn University of Technology, Estonia
Tuomas Paso	University of Oulu, Finland
Valeria Loscri	Inria, France
Ville Niemelä	University of Oulu, Finland
Waqas Ahmed	Pakistan Institute of Engineering and Applied Sciences, Pakistan
Xinrong Li	University of North Texas, USA
Yannick Le Moullec	Tallinn University, Estonia

Contents

Connectivity and Radio Propagation

Providing Connectivity to Implanted Electronics Devices: Experimental Results on Optical Communications Over Biological Tissues with Comparisons Against UWB	3
<i>Senjuti Halder, Mariella Särestöniemi, Iqrar Ahmed, and Marcos Katz</i>	
On the UWB in-Body Propagation Measurements Using Pork Meat	18
<i>Mariella Särestöniemi, Carlos Pomalaza-Raez, Chaiimã Kissi, and Jari Iinatti</i>	
Detection of Brain Hemorrhage in White Matter Using Analysis of Radio Channel Characteristics.	34
<i>Mariella Särestöniemi, Carlos Pomalaza-Raez, Jaakko Hakala, Sami Myllymäki, Joni Kilpijärvi, Jari Iinatti, Matti Hämäläinen, and Teemu Myllylä</i>	
UWB Microwave Imaging for Inclusions Detection: Methodology for Comparing Artefact Removal Algorithms	46
<i>James Puttock, Behnaz Sohani, Banafsheh Khalesi, Gianluigi Tiberi, Sandra Dudley-McEvoy, and Mohammad Ghavami</i>	
BSNCloud: Cloud-Centered Wireless Body Sensor Data Collection, Streaming, and Analytics System	59
<i>Ming Li, Ai Enkoji, Matthew Key, Aaron Marroquin, and B. Prabhakaran</i>	

Secure Communication Networks for Smart-Health

Model-Based Analysis of Secure and Patient-Dependent Pacemaker Monitoring System	77
<i>Leonidas Tsiopoulos, Alar Kuusik, Jüri Vain, and Hayretin Bahsi</i>	
Amplitude Modulation in a Molecular Communication Testbed with Superparamagnetic Iron Oxide Nanoparticles and a Micropump	92
<i>Max Bartunik, Thomas Thalhofer, Christian Wald, Martin Richter, Georg Fischer, and Jens Kirchner</i>	
An Enhanced DNA Sequence Table for Improved Security and Reduced Computational Complexity of DNA Cryptography	106
<i>Maria Imdad, Sofia Najwa Ramli, Hairulnizam Mahdin, Boppana Udaya Mouni, and Shakira Sahar</i>	

Solving Generic Decision Problems by in-Message Computation in DNA-Based Molecular Nanonetworks	121
<i>Florian-Lennert Adrian Lau, Regine Wendt, and Stefan Fischer</i>	
A Model for Electro-Chemical Neural Communication	137
<i>Maryam Hosseini, Reza Ghazizadeh, and Hamed Farhadi</i>	
Connected Wearables Sensors for Healthcare Applications	
Activity Monitoring Using Smart Glasses: Exploring the Feasibility of Pedometry on Head Mounted Displays	153
<i>Zhiquan You, Farnaz Mohammadi, Emily Pascua, Daniel Kale, Abraham Vega, Gian Tolentino, Pedro Angeles, and Navid Amini</i>	
Real-Time Human Activity Recognition Using Textile-Based Sensors	168
<i>Uğur Ayvaz, Hend Elmoughni, Asli Atalay, Özgür Atalay, and Gökhan Ince</i>	
Extraction of Respiratory Signals and Respiratory Rates from the Photoplethysmogram	184
<i>Shenglang Xiao, Pengfei Yang, Luyao Liu, Zhiqiang Zhang, and Jiankang Wu</i>	
An Ultra-low-Power Integrated Heartbeat Detector for Wearable Sensors	199
<i>Antoine Gautier, Marine Dael, Robin Benarrouch, Benoit Larras, and Antoine Frappé</i>	
Anxiety Detection Leveraging Mobile Passive Sensing	212
<i>Lionel M. Levine, Migyeong Gwak, Kimmo Kärkkäinen, Shayan Fazeli, Bita Zadeh, Tara Peris, Alexander S. Young, and Majid Sarrafzadeh</i>	
Author Index	227

Connectivity and Radio Propagation



Providing Connectivity to Implanted Electronics Devices: Experimental Results on Optical Communications Over Biological Tissues with Comparisons Against UWB

Senjuti Halder^(✉), Mariella Särestöniemi, Iqrar Ahmed, and Marcos Katz

Centre for Wireless Communications, University of Oulu, Oulu, Finland
senjutihalder@yahoo.com, marcos.katz@oulu.fi

Abstract. Radio and acoustic waves have been conventionally used for transmitting information through biological tissues. However, some radio-based communications often suffer from several drawbacks like security, safety, privacy, and interference. In this paper, we demonstrate that optical wireless communications can be practically used for communications through biological tissues, particularly to transmit information to and from implanted devices. In the experiment, *ex vivo* samples of pork meat were used as the optical channel. Initial results show that information can be optically transmitted through biological tissues to distances of several centimeters, a range of practical interest as many implants today are placed within this extent. Optical links are inherently secure, and interference to and from other equipment is not an issue. With numerous potential benefits, optical wireless communication can be considered as a complementary approach to the existing radio frequency (RF) communications. In this paper, a comparison between the measurement results of ultra-wideband (UWB) and optical communications through the biological tissues is presented. Both experiments have been taken place in a similar environment, with the same meat samples. We have also explored the effect of tissue temperature on successful communications through biological tissues. These initial results are very promising and indicate various potential benefits for in-body communication in the future.

Keywords: Medical implants · Optical communications · Ultra-wideband · Biological tissues · In-body communications

1 Introduction

Improving the quality of life is one of the ultimate goals of technology, and in this respect, wireless and mobile communications play a key role. In particular, remote health monitoring is a cost-effective approach to the prevention of diseases as well as the provision of health care. The aging population, chronic diseases, and the desire to live a healthy lifestyle are present-day healthcare challenges that can be addressed by wireless technology. The concept of wireless body area networks (WBAN) is a well-known example

of this technology, where a network of nodes (e.g., sensors) inside, outside, and on the patient's body collects, processes, and transmits physiological data. Last decades have witnessed a remarkable development in electronic body implant technology. Today, we rely on a great variety of these implants, such as heart pacemakers, cardioverter defibrillators, medicine dispensers, and brain stimulation implants, for example. In addition, in-body sensors and smart pills have been developed to monitor physiological parameters, imaging, drug delivery, etc. Furthermore, there is currently an increasing interest in brain implants beyond conventional brain stimulation and deep brain stimulation. These include monitoring and recording brain activity and ultimately interfacing neural systems with computer systems (e.g., brain-computer interfaces). Recently, the idea of the internet of things (IoT) has been extended to consider in-body, on-body, and out-body nodes networked to provide medical care. New concepts such as the internet of medical things (IoMT) and the internet of the human body (IHB) have been lately proposed.

One of the key enablers in the above-mentioned systems is human centric communications (HCC), a field that has been and continues to be widely investigated. HCC defines wireless communications taking place inside the human body, communication between internal and external nodes as well as communication across nodes on the body. Figure 1 depicts the possible wireless links in HCC, where three types of nodes are shown, a) in-body nodes (IBN), b) on-body (OBN) nodes, and c) out-body or external nodes (EN). In this paper, the focus is on in-body communications, where wireless links are established between IBNs or between IBNs and OBNs. Typical nodes include electronic implants as well as ingestible or implanted sensors, for instance. Radio has been the dominant technology for in-body wireless communications, exploiting narrowband, wideband, and ultra-wideband concepts. Acoustic communication is also an interesting approach to communicating through biological tissue. Human tissue is a challenging medium for the propagation of radio waves. Radio frequency attenuation is typically high, due to the high conductivity and dielectric constant of the biological tissues. These characteristics strongly depend on the frequency used for communications. Moreover, the medium could be very complex, as the signal may need to propagate through many layers of different tissues, each with different characteristics. In-body communication must be highly secure and private, and this is a challenge for radio systems. Several cases have emerged in recent years, showing the vulnerability of commercial implants to hacks, leading even to a massive recall of implantable devices. Besides, the radio is prone to be affected by non-intentional interference. Ultrasound-based communication systems do not face these challenges, though they might be limited by low data support and considerably high-power consumption. In this paper, we propose the use of optical communications for establishing in-body wireless links. Previous work has shown that optical communication is viable for very short optical links of a few millimeters, particularly for reaching electronic devices placed under the skin. This is the case of superficial communications, indicated with "a" in Fig. 1. However, our initial experiments have shown that considerably longer ranges can be easily achieved using optical links based on near-infrared (NIR) light. This case, depicted with "b" in Fig. 1, provides connectivity to in-body electronic nodes (IBNs). Moreover, connecting in-body and out-body nodes (e.g., EN) is also possible using optical links. The prospect of using light for wireless

communications inside the human body is highly relevant. Optical wireless communications have unique advantages such as high security, privacy, safety as well as low complexity and power consumption. Furthermore, light sources are used for a variety of novel purposes inside the human body, such as for diagnosis and therapy. Control of cellular activity (e.g., in neurons) by means of light to activate, inhibit, or monitor certain functions is another highly relevant application of light. The vision of having key health functionalities such as wireless communication, diagnosis, treatment, and monitoring working with under the same light-based infrastructure is truly appealing. In this paper, we report the design, implementation, and testing of an optical communication system, with a testbed using realistic *ex vivo* samples as the optical channel. The optical system is compared to an ultra-wideband communication system working in a similar scenario. The impact of the temperature on the optical and radio communication links is also investigated. There are several studies presenting the impact of the temperature on the dielectric properties of the animal tissues in different frequency ranges [1, 2]. However, up to the authors' knowledge, there is a lack of studies presenting the impact of the temperature on the radio channel characteristics. In ref. [3], authors have presented results on the impact of the temperature on the antenna and channel characteristics. Frequency domain channel results are repeated in this paper to ease the comparison with the measurement results obtained with the optical technique at the same temperature.

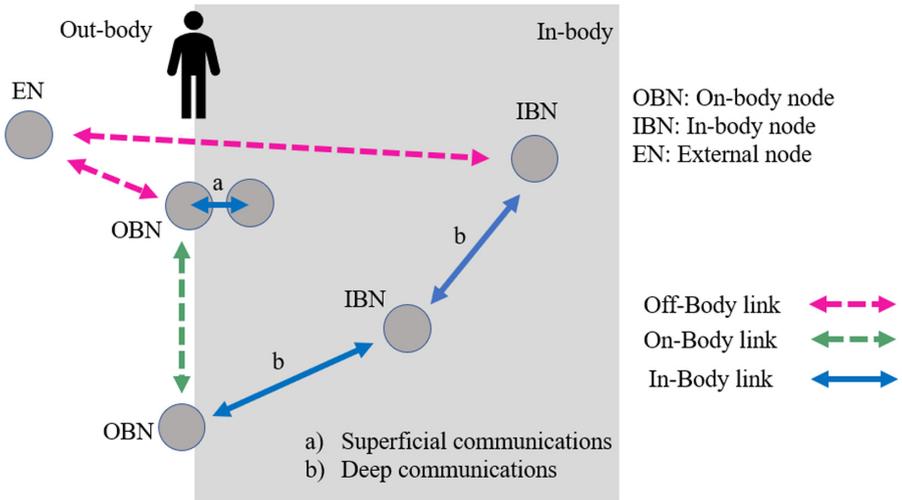


Fig. 1. HCC: In-body, out-body, and off-body links.

This paper is organized as follows. Section 2 introduces optical communications through biological tissues, while Sect. 3 provides a comparison among radio, acoustic, and optical wireless systems for in-body communications. Section 4 presents measurement results from a realistic setup justifying the optical communication approach. Discussions on opportunities, challenges, and novel applications are presented in Sect. 5. Finally, Sect. 6 concludes the paper.

2 Optical Communications Through Biological Tissues

Visible and infrared light is widely used in medical information and communication technology (ICT), mostly for diagnosis and therapeutics. In this section, we briefly introduce the basic ideas behind optical communications in biological tissues, in other words, how to exploit light to transmit information through biological tissues. Unlike conventional uses of optical wireless communications (OWC), where the medium is air or water, biological tissues are a challenging medium. Interaction of light with biological tissues is a well-studied field, particularly aiming at analyzing properties of the tissue and eventually making the diagnosis. However, biological tissues have not been studied from the wireless communications standpoint. Light is reflected, absorbed, scattered, and transmitted through biological tissues. Biological tissues are characterized by high anisotropy and strong absorption and scattering. The near-infrared window in the region (700–1100) nm is commonly used as light is less attenuated than in other wavelengths. Due to scattering, light is dispersed, reducing rapidly its energy as light propagates across the tissue. Light is thus highly attenuated, and pulses are spread in time. The receiver needs to cope with the signal impairments imposed by this severe medium in order to recover the transmitted signal. Contrasting air and water, one can expect that the range of optical links in tissues is quite limited, a fact that is in line with the geometry of the human body and the ranges demanded by typical in-body medical applications. Increasing the optical power of the transmitted signal can be used to increase the range or data rate. On the other hand, to avoid damages to the biological tissues, the energy and power density of the transmitted signal need not to exceed certain limits, defined by the International Commission on Non-Ionizing Radiation Protection [4].

Rayleigh and Mie scattering are commonly used to model the optical scattering process in biological tissues, where the former takes place when the scattering is caused by particles much smaller than the wavelength of light, and the latter when the sizes of particles are in the order or larger of the light wavelength.

3 Radio, Acoustic and Optical Wireless Communications in Biological Tissues: A Comparison

3.1 Radio

Radio technology has been widely used in medical applications for several decades. It has several advantages for implant communications. It can provide high-data-rate communications and, also high-resolution images using low transmit power. Especially with lower frequencies, power losses in the tissues are moderate. Radio is a well-known technique for which inexpensive components are widely available. Besides, the in-body data can easily be transmitted out of the body since most of the medical devices use RF [5, 6].

Narrowband (NB) and UWB technologies are the most commonly used radio technologies in the medical applications based on the WBAN, which is considered as a collection of miniaturized low-power, wireless sensors for monitoring the human body functions [6]. For the successful implementation of WBAN applications, the IEEE published the standard 802.15.6 [7] for in-body/on-body node communications. In addition

to the main technologies NB and UWB, WBAN covers also human body communications (HBC) technology, which is based on galvanic coupling or capacitive coupling. From these techniques, UWB is chosen to be compared with the optical communication technique.

UWB technology has been recognized as a promising candidate for implant communication systems. It provides several advantages: high data rates, low power consumption, high security, simple transceivers, etc. [6]. The main drawback of the UWB is the higher propagation loss within the tissues. However, the propagation loss can be dealt with the use of a lower UWB band (3–5 GHz) in the applications where the signal is required to propagate deep inside the body tissues and with properly directive antennas. UWB has been used in several medical monitoring applications, such as capsule endoscopy, tumor detection, heart and breath rate monitoring, glucose level monitoring, etc.

3.2 Acoustic

Mechanical movements of particles in a substance form the ultrasounds (US) wave [8]. Ultrasonic waves are generally defined by the following parameters: frequency, pressure, amplitude, propagation speed, and intensity [9]. US is an extensively popular technology in healthcare due to its various applications in imaging. Ultrasound consists of all the acoustic waves having a frequency above 20 kHz, which is non-audible for humans [10]. Apart from the numerous imaging applications, US has been considered to have great potential for in-body communications. Acoustic waves, typically generated through piezoelectric materials are preferred to be used for underwater communications, as they propagate better in media mostly made up of water when compared to RF. Up to 60% of the human adult body consists of water [11] and blood is a fluid that provides cells with different resources. This makes ultrasound an eligible technology to support in-body communications.

3.3 Optical

Over the years, there has been a growing interest in utilizing optical communications for the betterment of healthcare technologies. Light has been widely used in both therapeutic and diagnostic medical applications. Light-based communications typically require line-of-sight (LOS) configuration [12]. However, non-line-of-sight (NLOS) configuration is also possible, though the signal strength could be greatly affected in that scenario. In recent years, visible light communications (VLC), a case of OWC, has become an extremely popular research topic. In the electromagnetic spectrum, the visible spectrum expands over a wavelength of 400 nm to 700 nm [13]. In VLC, light-emitting diodes (LEDs) are employed due to their advantages of simultaneously providing illumination and transmitting of data. Optical communications possess numerous advantages like safety, large and unregulated spectrum, data security, less interference, zero-radio-exposure as well as cost and energy effectiveness. These advantages could lead these light-based communications a suitable choice for future medical applications. The utilization of optical communications in implants is very safe and secure for patients, a fundamental advantage of this technology. For OWC, IEEE published the standard 802.15.7 in 2011 which mainly aimed at visible light spectrum. Development work has

been going on to achieve further standards so that the ultraviolet, visible light and infrared areas can also be included.

Some key characteristics of UWB, acoustic, and optical communications system are presented in Table 1.

Table 1. General comparison of UWB, acoustic and optical communications for short-range medical applications.

	Radio (UWB)	Acoustic	Optical
Supported data rate	0.5–10 Mbps	<0.5 bps	100 Kbps
Range	<10 cm (directive antennas)	<15 cm	<5 cm
Power requirements	Low	Low/moderate	Potentially low
Complexity	Low	Low	Low
Security/privacy	Low	High	Very high
Safety/exposure	Safe within SAR limits	Safe	Safe
Standard	802.15.6, 802.15.4	None	None
Frequency	3.1–10.6 GHz	1–3 MHz	300 GHz–430 THz

4 Experimental Results

In this section, we present some measurement results carried out with individual experimental testbeds for both optical and UWB. Results, though initial, show the potential of optical wireless communications for in-body applications.

4.1 Optical Measurements

The optical communication measurements have been carried out using mostly commercially available equipment. An experimental testbed was developed to evaluate optical communication through biological tissues [14]. The testbed consists of an optical transmitter, an optical receiver, and the biological tissue as the optical channel. Figure 2 shows the basic block diagram of the testbed used in the measurement setup. The transmitter side comprises a computer, universal software radio peripheral (USRP) where the key blocks of the transmitter were implemented, bias-tee, LED driver, and LED source. Thorlabs DC2200 LED driver [15] and an 810 nm mounted IR LED produced by Thorlabs [16] are implemented in the transmitter side. On the other end, the optical receiver comprises a photodetector, USRP where receiver blocks were implemented and a computer. The photodetector used in the receiver side is a silicon avalanche photodetector APD120A from Thorlabs [17]. For both transmitter and receiver side, NI USRP-2920 have been used which are designed and developed Ettus Research, the daughter company of National Instruments. We have used NIR light to illuminate the biological tissue since NIR light propagates better through tissues. The tunnel between source and receiving nodes has

been created with the help of GNU radio software. For the optical communication link through biological tissues, a Gaussian minimum-shift keying (GMSK) modulation was used. The GMSK modulation scheme has been used due to its constant envelope property. It is difficult for the LED to be able to respond to the prompt phase changes of some other modulation schemes (e.g., quadrature phase-shift keying (QPSK)). Thus, GMSK modulation is considered as a convenient choice for the optical communications as the phase changes occur steadily in GMSK [14, 18].

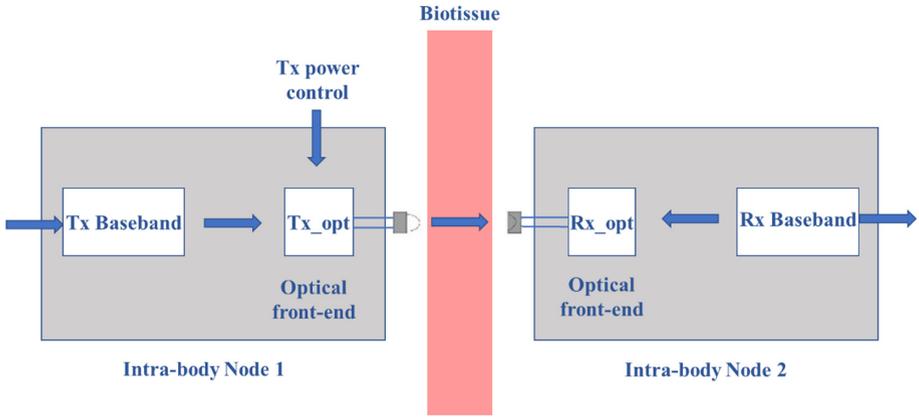


Fig. 2. Simplified block diagram of the experimental testbed.

Here, we have used a 3 cm thick fresh pork meat sample with the skin on. The meat sample consists of both muscle and fat (approximately 0.5 cm of fat and 2.5 cm of flesh). The temperature of the sample was 12 °C. The LED input current has been varied and we have measured the corresponding optical power employed on the meat sample. The optical powers were well in accordance with the standard of laser safety ANSI Z136.1-2007 [19]. Successful optical communication was established through the *ex vivo* sample. Moreover, an image file of 14 MB has been transmitted through the sample and we were able to achieve a data rate of 22 Kbps. The meat sample has been illuminated on the skin surface and the detector has been placed on the flesh side. The experimental setup for optical communications through biological tissues is presented in Fig. 3.

To understand the effect of tissue temperature on optical communication, the same meat sample has been heated to 37 °C. For this heating purpose, we have placed the sample inside a heating box made of acrylic material. A temperature controller was used to set the desired temperature of the sample. Table 2 presents the transmitted and received optical power for both cold and warm meat along with the power losses. Both the transmitted and received optical powers are measured with the help of PM100D power meter from Thorlabs [20].

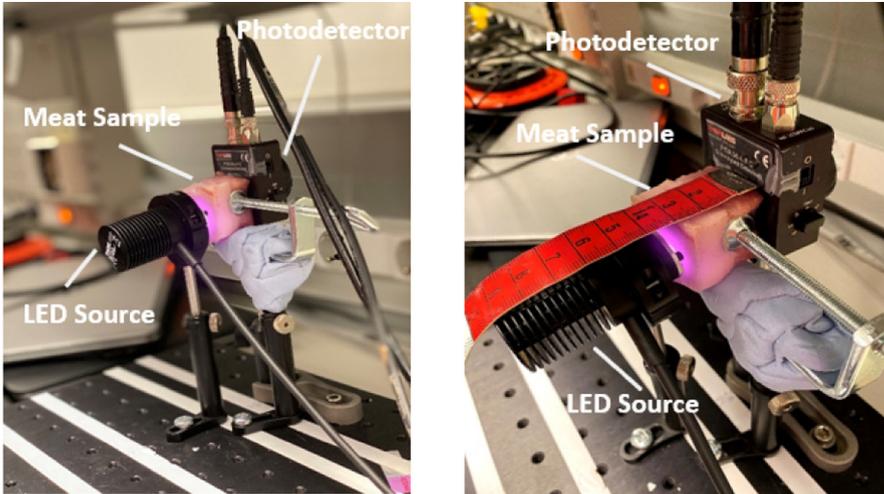


Fig. 3. Experimental setup for optical communications through biological tissue.

Table 2. Transmitted optical power, received optical power and power losses of cold and warm meat.

Transmitted optical power (mW/cm ²)	Received optical power (μW) (Cold meat)	Power loss (dB) (Cold meat)	Received optical power (μW) (Warm meat)	Power loss (dB) (Warm meat)
50.94	98.3	-27.14	128.3	-25.98
101.11	204.2	-26.94	234.2	-26.35
146.26	301.8	-26.85	341.8	-26.31
190.53	389.9	-26.89	419.9	-26.57
231.38	469.2	-26.92	520.2	-26.48

A comparison of the received optical power in both cases (cold and warm meat samples) shows a minor difference between the measured values. The received optical power for both cold and warm meat samples are presented in Fig. 4. The optical properties of the meat sample change with the temperature and this affects the amount of received power. Moreover, as our sample consisted of both flesh and fat, the fat portion has become more transparent due to the rise in the temperature. This transparency also leads to better light propagation through the tissues, as seen in Fig. 4.

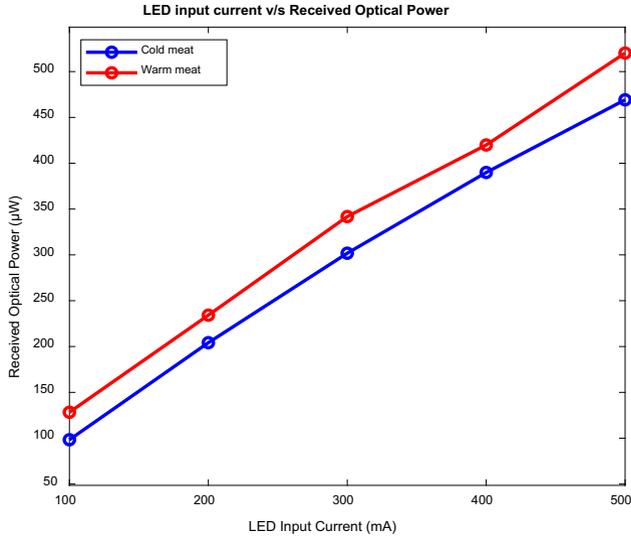


Fig. 4. LED input current v/s received optical power.

4.2 UWB Measurements

For a fair comparison, the UWB testbed used the very same meat sample used for the case of optical communication. UWB measurements were conducted using the Agilent 8720ES Vector Network Analyzer (VNA). Two different UWB-antennas were connected to the VNA's ports with coaxial cables. Antennas were placed on the top and below the biological tissue. The block diagram of the UWB measurement set up is presented in Fig. 5.

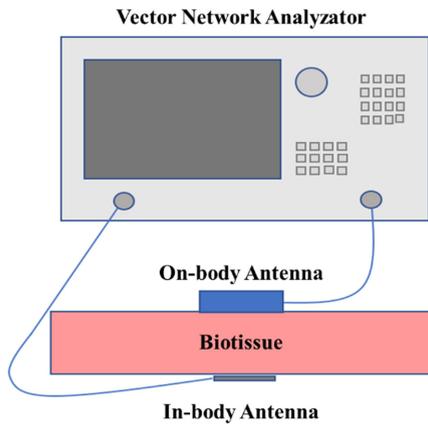


Fig. 5. Simplified block diagram of the UWB measurement setup.

The UWB measurement results are analyzed in more detail in [3] presenting a study on the UWB in-body measurements using pork meat. Ref. [3] evaluates also the impact of the meat's fat and muscle composition as well as the impact of meat's temperature on the UWB antenna and channel characteristics and compares the UWB channel characteristics on humans and pork tissues by simulations. One of the channel results of [3] is represented in this paper to ease the comparison with the results obtained using optical communication.

Figure 6 presents the meat used in the measurements. The size of the meat is 22 cm \times 20 cm, which is much larger than that used in the optical measurements due to antenna sizes and the necessity to avoid propagation flow from the meat sides. The thickness of the meat is 3 cm as in optical measurements. The frequency-domain channel characteristics are evaluated for the UWB range of 0.5–5 GHz. Two different UWB antenna prototypes are used in the measurements: 1) a cavity-backed antenna, which is used as an on-body antenna and is located above the skin layer and 2) a loop antenna, which resembles as an in-body antenna. The cavity-backed antenna is a directive on-body antenna designed for implant communications at the frequency band 3.75–4.25 GHz, which is the frequency range of interest in this UWB measurement study. A detailed description of the antenna can be found in [21]. The loop antenna is an omnidirectional antenna designed to work for the whole UWB band 3.1–10.6 GHz. Details of the loop antenna can be found in [22].



Fig. 6. The meat piece used in the measurements.

The meat piece was set inside a cube form area made by absorber pieces. The loop antenna was set below the meat muscle and fat tissue side. The cavity-backed antenna was set to the top of the tissues on the skin side. Figure 7 presents the measurement setup, which is explained in detail in [3]. Ref. [3] also analyses in more detail the frequency and time domain channel characteristics as well as antenna characteristics with different pork meat pieces. This section summarizes the channel results for the pork meat sample which was also used in optical measurements to ease the comparison with optical measurements. The measurements were taken as the meat was cold (12 °C) and warm (37 °C), as in the optical measurements.

Frequency domain channel characteristics, i.e., S21 parameters, are presented in Fig. 8 for the pork meat with cold and warm temperatures. It can be seen, power loss on the frequency range of interest, i.e., 3.75–4.25 GHz, varies from –48 dB to –58 dB with the cold meat. The power loss difference at the 3.75 GHz and 4.25 GHz is noteworthy. The higher the frequency, the larger is the power loss due to changes in the dielectric



Fig. 7. UWB measurement setup inside the cube made from absorber pieces [3].

properties [23]. Nevertheless, the power loss is at a reasonable level for successful data communications in this case. For the warm meat, the power loss is approximately 6 dB smaller compared to the cold meat. The variation within the frequency range of interest is -42 to -52 dB. These results show clearly, how the temperature has a clear impact on the channel characteristics since the dielectric properties of the tissues change with the temperature.

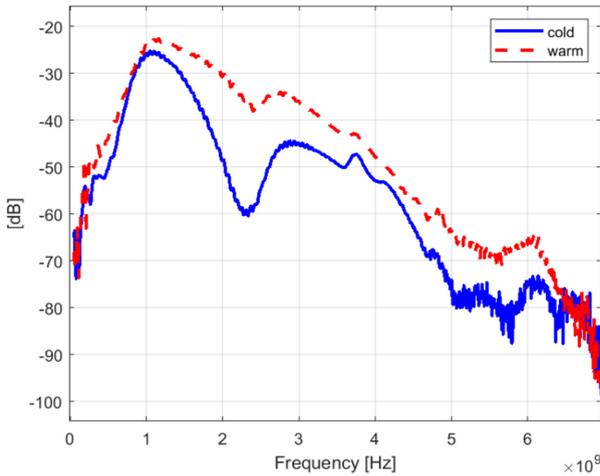


Fig. 8. The S21 parameter obtained in UWB measurements with the meat at the temperature of 12°C and 37°C [3]

It can be concluded that the power loss for 3 cm thick meat remains at a reasonable level for a data communication link. With lower UWB frequencies and directive antennas, data communications might be possible even with double meat thickness in the realistic scenario of 37 °C meat temperature.

The measurements have been taken with a meat piece having interlaced muscle and fat layers. Fat is known to be among the easiest tissues for radio propagations [24, 25] and thus the “fat tunnels” within the muscle layers help to achieve reasonable power loss level. With meat pieces having separate muscle and fat layers, the power loss is worse [3].

With the UWB technique, better propagation depth is obtained. The best propagation depth is achieved with directive on-body antennas. However, such antennas usually require relatively large cavity structures which might be clumsy for wearable devices.

5 Discussion

This study investigates the potential of transmitting information through biological tissues by utilizing optical communications. Results show that optical communications through biological tissues are possible, in a range of practical interest. A comparison of the optical communications through biological tissues with the UWB communication is also presented in this paper. Here, we have used the same piece of meat sample with 3 cm thickness as the medium for both UWB and optical measurements. Individual experimental testbeds have been used to carry out the measurements. Light propagation through biological tissue is an extremely complex process, where phenomena like absorption, scattering, and reflectance take place during the propagation. The meat sample we have used here contains both fat and muscle. The optical power applied to the sample during the optical communications is well below the regulated safety limits. First, we have demonstrated that optical communications through biological tissues are feasible through ranges that are of practical interest, e.g., wireless connectivity to different implanted devices can be provided. In fact, in the future, very secure and safe communications could be provided to several possible implanted electronic devices such as pacemakers, defibrillators, insulin pumps, and others. The simple optical communications system was not matched to the optical channel, and thus, the supported data rate was relatively low, of the order of several tens of Kbps. Nevertheless, in most of the cases involving implanted devices, the required data throughput is relatively small, as typical applications require the sporadic transmission of sensor information, changing device settings, etc. High-order modulation schemes as well as multiple optical beams can be used to increase the data rate.

An initial comparison between optical and UWB communications through biological tissues has been presented here. The comparison here focuses only on the attenuation of the two systems for a similar piece of sample. It was found that power loss with the optical system is minor than with UWB. However, also with UWB the loss is moderate and still far from the UWB receiver’s sensitivity limits. Besides, the comparison is not straightforward since in the UWB measurement setup, only a pure sinusoid signal is inserted directly to the transmitter antenna without any modulation.

Optical and UWB measurements on the meat sample were carried out at 12 °C and 37 °C to observe the effect of tissue temperature during the communications. We have

found that the received optical power of the heated meat sample is higher than that of the cold meat sample. Moreover, the power loss for the heated meat has been found to be smaller than the cold meat during UWB measurements. On the other hand, we have noticed minor differences in the power losses for the cold and warm meat during the optical measurements. So, it can be considered that optical communication is less sensitive to the tissue temperature compared to UWB communications. The measured power losses for both cold and warm samples are such that the signal can be decoded at the receiver.

Various individual advantages of optical and radio communications can be exploited together to implement a flexible hybrid optical-radio wireless network. In this hybrid optical-radio wireless network, both the optical and radio communications complement each other. An optical-radio hybrid network has several advantages, including increased link connection reliability, security, and support of higher data rates. Instead, the radio enables implant communication in the cases where the implant is deeper inside the tissues. The hybrid optical-radio network can be considered as a suitable approach for future medical applications due to these potential benefits.

6 Conclusion

In this paper, the feasibility of optical communications through biological tissues is investigated. A realistic scenario exploiting *ex vivo* samples is used in the study. A conventional UWB communication system is used as a baseline for comparison. The initial results regarding optical communications through biological tissues are very encouraging. Although the data rate achieved for optical communications is relatively low, it could well be sufficient for most of the electronic medical implants in use today. A range of three centimeters was obtained for the optical link, a distance well suited to provide connectivity to prevailing implants such as pacemakers, defibrillators and implanted drug dosing systems, for instance. Conventional narrow-band radio-based systems often face challenges like security, privacy, and safety, which need to be seriously tackled in any medical-related application. Both optical communication system as well as UWB based radio technique inherently overcome these drawbacks. The benefit of optical communication system over UWB is smaller device size and minor power loss. However, despite of minor power loss, optical communication is restricted only to few centimeters whereas with directive antennas at lower parts of the UWB range, the reliable communication can be achieved even up to 10 cm.

Optical communication can either be utilized as a standalone approach or in combination with other conventional communications (e.g., acoustic/radio). These measurements will be carried out to achieve a more comprehensive comparison advantages of optical communications indicate that it can be considered as a suitable approach for different future medical applications. In the future, extensive between the optical and UWB systems. The connectivity to the implanted electronic devices using optical and UWB communications will be further investigated where both the communication system would transmit data through biological tissues.

References

1. Ley, S., Schilling, S., Fiser, O., Vrba, J., Sachs, J., Helbig, M.: Ultra-wideband temperature dependent dielectric spectroscopy of porcine tissue and blood in the microwave frequency range. *Sensors* **19**(7), 1707 (2019)
2. Gabriel, C., Gabriel, S., Corthout, Y.E.: The dielectric properties of biological tissues: I. Literature survey. *Phys. Med. Biol.* **41**(11), 2231 (1996)
3. Särestöniemi, M., Pomalaza-Raez, C., Kissi, C., Iinatti, J.: On the UWB in-body propagation measurements with a pork meat. Submitted to *BodyNets2020*
4. International Commission on Non-Ionizing Radiation Protection: ICNIRP statement on light-emitting diodes (LEDs) and laser diodes: implications for hazard assessment. *Health Phys.* **78**(6), 744–752 (2000)
5. Teshome, A.K., Kibret, B., Lai, D.T.: A review of implant communication technology in WBAN: progress and challenges. *IEEE Rev. Biomed. Eng.* **12**, 88–99 (2018)
6. Movassaghi, S., Abolhasan, M., Lipman, J., Smith, D., Jamalipour, A.: Wireless body area networks: a survey. *IEEE Commun. Surv. Tutor.* **16**(3), 1658–1686 (2014)
7. IEEE Standard for Local and metropolitan area networks - Part 15.6: Wireless Body Area Networks. *IEEE Std 802.15.6-2012*, pp. 1–271 (2012)
8. Davilis, Y., Kalis, A., Ifantis, A.: On the use of ultrasonic waves as a communications medium in biosensor networks. *IEEE Trans. Inf Technol. Biomed.* **14**(3), 650–656 (2010)
9. Galluccio, L., Melodia, T., Palazzo, S., Santagati, G.E.: Challenges and implications of using ultrasonic communications in intra-body area networks. In: 2012 9th Annual Conference on Wireless On-Demand Network Systems and Services (WONS), pp. 182–189. IEEE (2012)
10. Carovac, A., Smajlovic, F., Junuzovic, D.: Application of ultrasound in medicine. *Acta Informatica Medica* **19**(3), 168 (2011)
11. Jéquier, E., Constant, F.: Water as an essential nutrient: the physiological basis of hydration. *Eur. J. Clin. Nutr.* **64**(2), 115–123 (2010)
12. Cevik, T., Yilmaz, S.: An overview of visible light communication systems. arXiv preprint [arXiv:1512.03568](https://arxiv.org/abs/1512.03568)
13. Zwinkels, J.C., Canada, C.: *Encyclopedia of Color Science and Technology*, pp.1–8 (2019)
14. Ahmed, I., Bykov, A., Popov, A., Meglinski, I., Katz, M.: Optical wireless data transfer through biotissues: practical evidence and initial results. In: Mucchi, L., Hämmäläinen, M., Jayousi, S., Morosi, S. (eds.) *BODYNETS 2019*. LNICST, vol. 297, pp. 191–205. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-34833-5_16
15. Thorlabs DC2200 LED Driver. <https://www.thorlabs.com/thorproduct.cfm?partnumber=DC2200>
16. Thorlabs Mounted LEDs. https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_ID=2692
17. Thorlabs APD120A/M - Si Avalanche Photodetector, 400–1000 nm, M4 Taps. <https://www.thorlabs.com/thorproduct.cfm?partnumber=APD120A/M>
18. Saud, M.S., Ahmed, I., Kumpuniemi, T., Katz, M.: Reconfigurable optical-radio wireless networks: meeting the most stringent requirements of future communication systems. *Trans. Emerg. Telecommun. Technol.* **30**(2), 1–15 (2019)
19. American National Standard for Safe Use of Lasers. <https://bit.ly/2VUXsXr>
20. PM100D - Compact Power and Energy Meter Console, Digital 4" LCD. <https://www.thorlabs.com/thorproduct.cfm?partnumber=PM100D>
21. Kissi, C., Särestöniemi, M., Pomalaza-Raez, C., Sonkki, M., Srifi, M.N.: Low-UWB directive antenna for wireless capsule endoscopy localization. In: Sugimoto, C., Farhadi, H., Hämmäläinen, M. (eds.) *BODYNETS 2018*. EICC, pp. 431–442. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-29897-5_38

22. Tuovinen, T., Yazdandoost, K.Y., Iinatti, J.: Comparison of the performance of the two different UWB antennas for the use in WBAN on-body communication. In: 2012 6th European Conference on Antennas and Propagation (EUCAP), pp. 2271–3374. IEEE (2012)
23. Dielectric Properties. <https://www.itis.ethz.ch/virtual-population/tissue-properties/database>
24. Asan, N.B., et al.: Characterization of the fat channel for intra-body communication at R-band frequencies. *Sensors* **18**(9), 2752 (2018)
25. Särestöniemi, M., et al.: Fat in the abdomen area as a propagation medium in WBAN applications. In: Mucchi, L., Hämäläinen, M., Jayousi, S., Morosi, S. (eds.) BODYNETS 2019. LNICST, vol. 297, pp. 175–187. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-34833-5_15



On the UWB in-Body Propagation Measurements Using Pork Meat

Mariella Särestöniemi¹✉, Carlos Pomalaza-Raez², Chaïmaâ Kissi³, and Jari Iinatti¹

¹ Centre for Wireless Communications, University of Oulu, Oulu, Finland

mariella.sarestoniemi@oulu.fi

² Department of Electrical and Computer Engineering, Purdue University, West Lafayette, USA

³ Electronics and Telecommunication Systems Research Group, National School of Applied Sciences (ENSA), Ibn Tofail University, Kenitra, Morocco

Abstract. This paper presents a study on the in-body propagation using pork meat at the lower ultrawideband (UWB) frequency band 3.74-4.25 GHz of the wireless body area network (WBAN) standard 802.15.6. Pork meat in terms of the dielectric properties is one of the most similar to human tissues and thus is commonly used in in-body propagation studies. Nevertheless, there are differences in the dielectric properties, creating some differences also in the radio propagation. The first objective of this paper is to investigate by simulations the propagation differences between human and pork tissue layer models. The simulations results show clear differences between the channel characteristics obtained using a human tissues and pork tissues: within the frequency range of interest, the path loss with pork meat can be up to 5 dB less than with the human meat. The second objective of this paper is to study, by measurements, the in-body channel characteristics using different types of pork meat piece having different fat and muscle compositions. It was found that path loss is clearly higher with the pork meat having separate skin, fat, and muscle layers compared to the pork meat having interlaced fat and muscle layers. Furthermore, the third objective of this paper is to study the impact of the meat temperature on the measured channel characteristics by comparing the channels obtained with the meat at the temperatures of 12 °C and at 37 °C. Also, in this case clear differences were observed in path loss: within the frequency range of interest, the path loss was maximum 5 dB lower with meat at 37 °C than with a colder meat. The results presented in this paper provide useful information and relevant aspects for the in-body propagation studies conducted with pork meat.

Keywords: Dielectric properties of tissues · Fat layer propagation · In-body propagation · Radio channel measurements · Temperature impact · Ultra wideband · Wireless body area networks

1 Introduction

Wireless medical monitoring has increased interest in recent years due to the several benefits it may provide for the healthcare of human beings [1–3]. Capsule endoscope is

one example of the implant monitoring systems which has become popular method of investigating the gastrointestinal (GI) tract [4, 5].

Smooth design of monitoring devices requires deep knowledge of radio channel characteristics. Channel characteristics can be studied with electromagnetic simulations, which are based on solving Maxwell's equations in their differential or integral forms, or measurements using phantom models, anaesthetized living animals, or tissues of the animals [5–8].

Measurements with the anaesthetized animals provide the most realistic results of the measurements since dielectric properties of the tissues start to change immediately after the death of the animal [9]. However, measurements are challenging with anaesthetized animals since such measurements require hospital environment [10, 11]. Instead, measurements with meat pieces are more feasible. Since some of the tissue dielectric properties of adult pig is known to be close to the those of human beings, pig meat is commonly used in the on-body/in-body antenna studies as well as in in-body propagation studies [12–14].

However, the use of meat pieces has some shortcomings and restrictions which should be considered. For instance, dielectric properties of animals can be different compared to those of human beings though dielectric properties of adult pork tissues are found to be close to those of humans [15–17]. Besides, dielectric properties of tissues change with the age of the animal [18] or as the time passes from the death of the animal [9]. Moreover, temperature impacts the dielectric properties of the tissues [16], which obviously has clear impact on the in-body channel characteristics. Finally, composition of fat and muscle layers in the meat piece is assumed to have strong impact on the results since fat is known to be a good propagation channel [19, 20].

Up to the author's knowledge, there is no studies presented on comparing channel characteristics evaluated with meat pieces having different fat and muscle composition with realistic antennas. Furthermore, up to the author's knowledge there are no studies presenting impact of the meat temperature on the channel characteristics.

The aim of this paper is to address the aforementioned aspects. First, channel characteristics obtained using human tissues and pork tissues in the layer simulation models are compared. Secondly, the channel characteristics are evaluated using two different meat pieces: the first one having interlaced muscle and fat layers, the second one having separate muscle and fat layers. Finally, the impact of the meat's temperature is evaluated on the antenna performance and channel characteristics.

The paper is organized as follows: Sect. 2 presents study case including description of the simulations, antennas, and measurements. Section 3 presents simulation-based antenna verification. Section 4 compares channel characteristics obtained using human tissues and pork tissues in the layer simulation model. Measurement results for meat pieces with different fat and muscle composition are evaluated in Sect. 4. Furthermore, the impact of the temperature on the antenna and channel characteristics are studied in Sect. 4. Finally, Conclusions with future work perspectives are discussed in Sect. 5.

2 Study Case

2.1 Simulations

Simulations were conducted with a 3D electromagnetic simulation tool CST Studio Suite software [21], which uses the finite integration technique. A planar layer tissue model, which is presented in Fig. 1, was used in the simulations. The thicknesses of the tissue layers are selected according to the thicknesses of tissue layers of the meat piece 1 used in the measurements. These thicknesses are presented in Table 1 in Subsect. 2.2

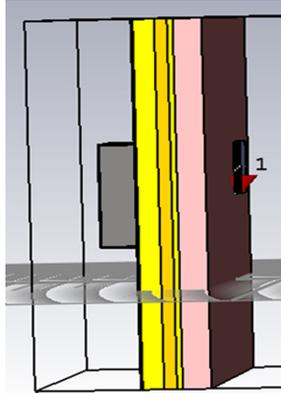


Fig. 1. Layer model used for the verification of the antennas.

Table 1. Tissue thicknesses of *Meat1* and *Meat2* in the location of antennas.

Tissues	<i>Meat1</i>	<i>Meat2</i>
Skin	2 mm	2 mm
Outer fat	10 mm	7 mm
Muscle	15 mm	18 mm
Inner fat	3 mm	3 mm

In this study, we use two types of antennas. The first one is a directional cavity backed UWB on-body antenna, shown in Fig. 2a, which was introduced in [22]. It is designed for on-in body communications for the frequency band 3.75–4.25 GHz which meets the IEEE 802.15.6 standards requirements [23]. Its dimensions with the cavity are $x = 83$ mm, $y = 49.5$ mm, and $z = 19.62$ mm, where x and y are width and length, respectively and z is towards the body. The size of the antenna itself is $x = 47.5$ mm, $y = 47.5$ mm. The second antenna is a loop antenna, shown in Fig. 2b, which is introduced in [24]. The loop antenna is omni-directional antenna UWB antenna, having large bandwidth 3.1–10.6 GHz and it is originally designed for on-body communications. The size of the antenna is 43 mm x 46 mm. In this initial study, the loop antenna is used to resemble the in-body antenna, since in the simulations, it was noticed to achieve similar path loss

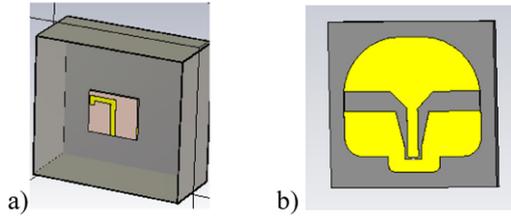


Fig. 2. Antennas used in this study: a) cavity backed low-UWB on-body antenna, b) loop antenna. values as the implant (capsule endoscope) antenna [25], for which a prototype was not available.

2.2 Measurements

Measurements were conducted using two different types of porcine pieces which are presented in Figs. 3a–b. Both pieces had skin, outer fat, muscle and inner fat layers. In

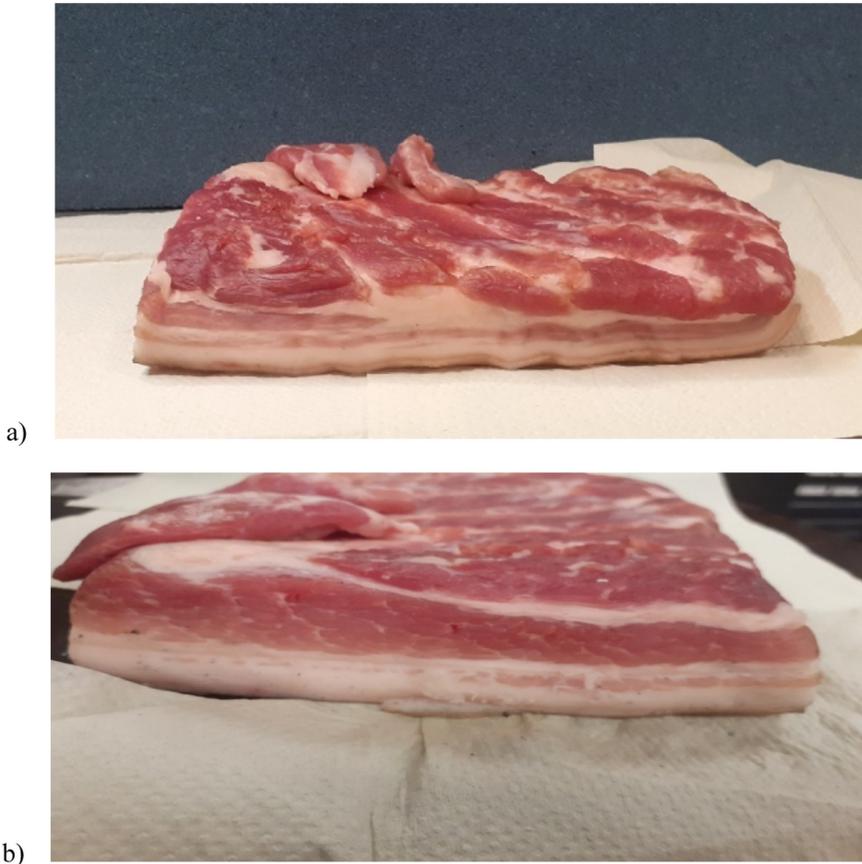


Fig. 3. a) Meat 1 with interlaced fat and muscle layers, b) Meat 2 with distinct muscle and fat layers.

the first meat piece, which is referred to as *Meat1*, the fat and muscle layers are interlaced so that there are “fat tunnels” going through the muscle layer. *Meat2* has distinct fat and muscle layers. Both meat pieces have thickness of 3 cm. The thickness of the tissues layers for *Meat1* and *Meat2* below the antennas are presented in Table 1.

The measurements were conducted with a Vector Network Analyzer (VNA) 8720ES in measurement laboratory premises at the University of Oulu, Finland. The meat pieces were set individually inside a cube form area made by absorber pieces to avoid signal interference from the surrounding environment as well as minimize the propagation overflow from the sides of meat pieces. Meat piece was set inside a thin plastic bag to protect the antennas and the absorbers. The picture of the measurement setup is presented in Fig. 4.

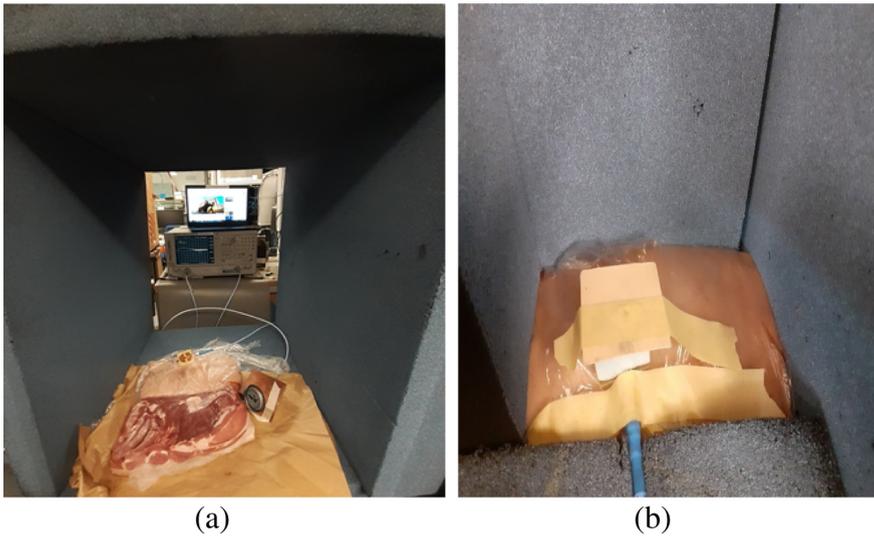


Fig. 4. a) Measurement setup before closing the back and front walls, b) on-body antenna with meat piece

3 Simulation Based Antenna Verification

The aim of this section is to verify usability of the loop antenna on the measurement based in-body propagation studies. This verification is conducted by comparing simulated channel characteristics obtained using an implant antenna [25] and a loop antenna [24] as the in-body antenna. The simulations are conducted using the planar tissue layer model since the pork meat piece also resembles planar layer model. The aim of this comparison is to show that loop antenna can be considered as an alternative in the measurements when using layer meat pieces. Figure 5 presents comparison between the frequency and time domain channel characteristics obtained using a capsule antenna and the loop antenna in the simulations. Figure 5a presents the S21 for the whole simulated frequency range and Fig. 5b the zoomed version of Fig. 5a for the frequency band of interest 3.75–4.25 GHz. Figure 5c presents channel impulse responses obtained applying inverse fast

Fourier transform (IFFT) for the simulated band. As one can note from Figs. 5a-b, the path loss difference is minor between the results obtained using the capsule antenna and the loop antenna within frequency range of interest. However, outside the frequency range of interest the differences are larger.

In time domain, the main peaks of the IRs are at the same level, though the main peak is clearly wider with the loop antenna. From the results it can be concluded that loop antenna can be used as in-body propagation studies in the measurements.

4 Simulation Based Comparison Between the Channel Characteristics Obtained Using Human and Pork Tissues

This section compares simulation results obtained using tissue layer models having dielectric properties of human tissues to those obtained using pork tissues. Dielectric properties (relative permittivity and tangent loss) for human and pork tissues at 4 GHz are presented in Table 2. The dielectric properties of human tissues are found from e.g. [26]. For dielectric properties of pork tissues there are several studies available in the literature [9, 10, 15, 16]. However, the challenge is there is variation in the values presented for pork. The reason for this is that several factors affect the dielectric properties: age and size of the pig, time from the death of the pig, temperature of the meat. For instance, there is wide variation in the reported results for tangent loss values of the pork muscle tissue. Thus, for this study, we decided to use tangent loss value of the human tissues also for pork tissues and we left study of the impact of the reported tangent loss variation for future work.

First, the antenna matching with human and pork tissues are evaluated by studying the S_{11} parameters presented in Figs. 6a–b for the on-body antenna and the loop antenna, respectively. It is found that the difference between the dielectric properties of human and pork tissues cause approximately 3 dB difference in the antenna's notch area so that S_{11} values is slightly higher with the human tissues. Same phenomenon can be found with both antennas.

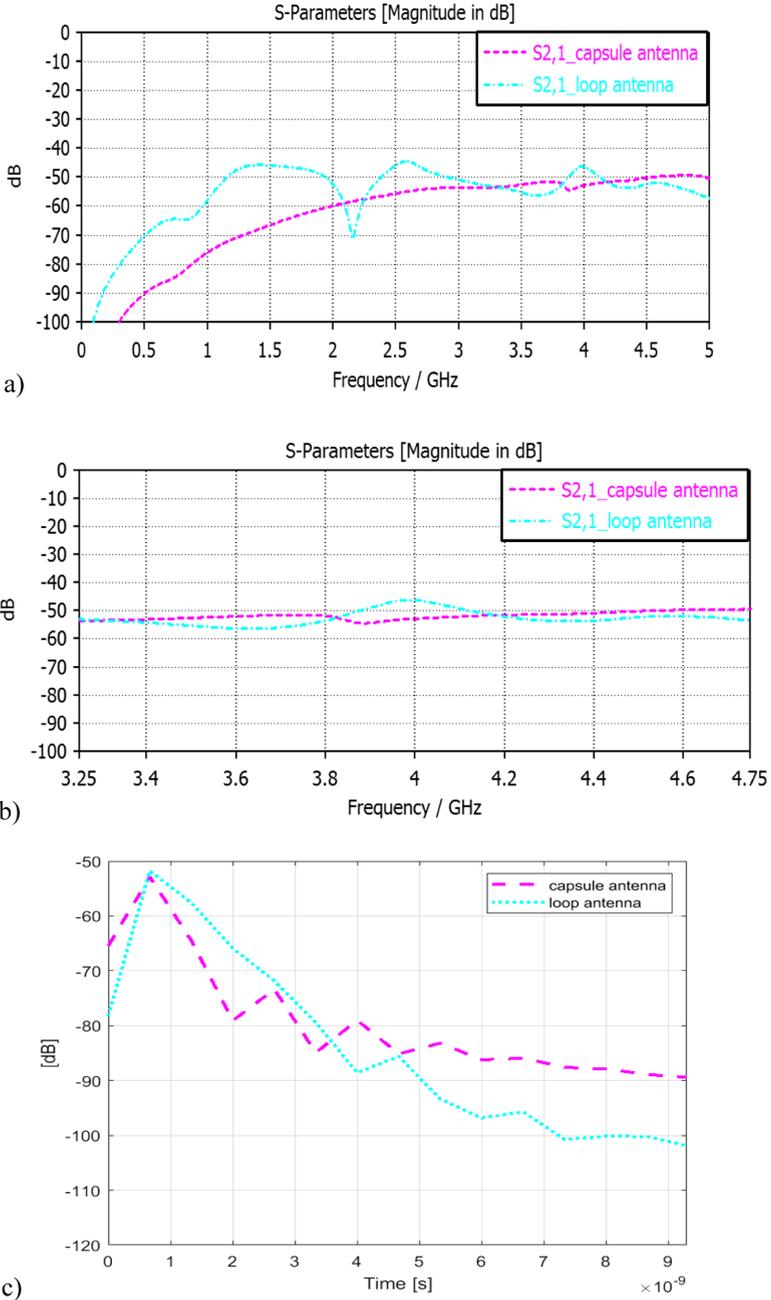
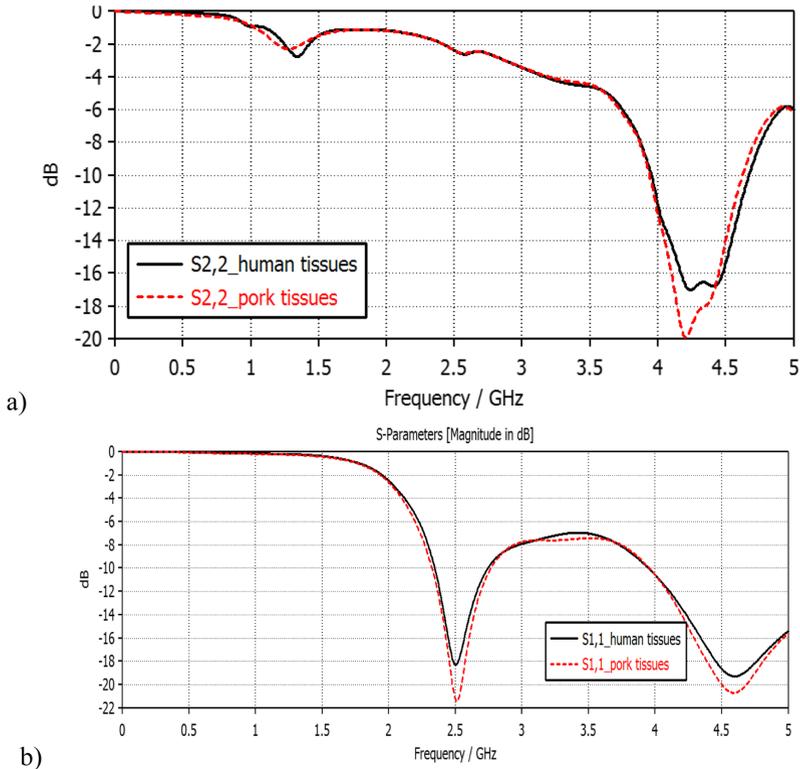


Fig. 5. Simulation based comparison of channel characteristics obtained using a loop antenna and a capsule endoscope implant antenna [25]: a) S₂₁s for the whole simulated bandwidth, b) S₂₁ for the frequency range of interest, and c) impulse responses.

Table 2. Dielectric properties of human and pork tissues at 4 GHz.

Tissues	Human tissues Relative permittivity/tangent loss	Pork tissues Relative permittivity
Skin	36.6/0.29	37.5/0.32
Fat	5.12/0.16	5.78/0.23
Muscle	50.8/0.27	48.0/0.27

**Fig. 6.** S_{11} parameters with a) on-body antenna, b) loop antenna (in-body antenna) obtained with simulation model having dielectric properties of human tissues and pork tissues.

The frequency and time domain channel characteristics are presented in Figs. 7a–b, respectively. As it is noted, difference in the dielectric properties of human and pork tissues have clear impact on the channel characteristics: within the simulated bandwidth, the path loss is up to 8 dB higher with pork tissues until 3.25 GHz than with the human tissues. Instead, from 3.5 GHz onwards, i.e. also within the frequency range of interest 3.75–4.25 GHz, the path loss is higher with human tissues.

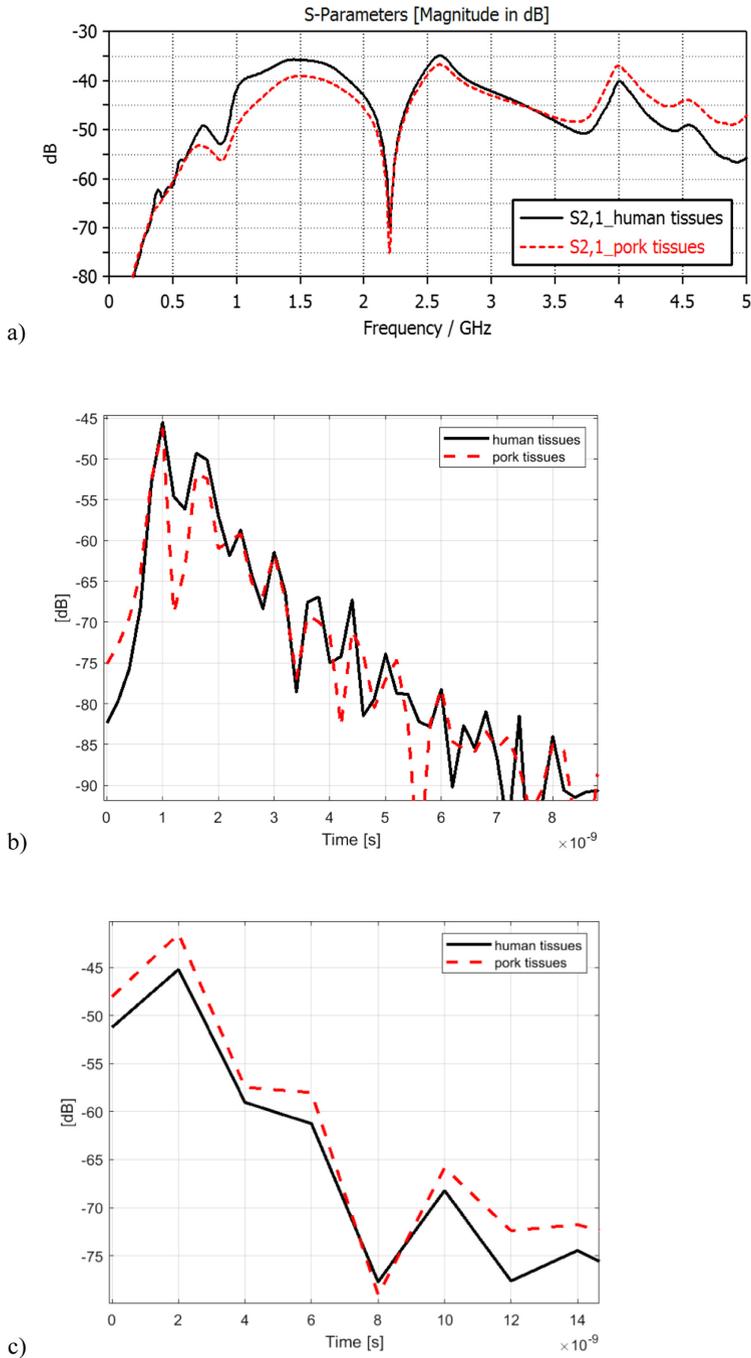


Fig. 7. Comparison between the differences of human and pork tissues: a) S21 parameters, b) IRs obtained by performing IFFT for whole the frequency bandwidth, and c) IRs obtained by performing IFFT for whole the frequency band of interest 3.75–4.25 GHz.

In time domain, as the IFFT is performed for the whole simulated bandwidth, the difference can be seen mainly in within the time range of 1–2 ns, where the level of the channel taps is higher with the human tissues than with the pork tissues. However, as it was pointed out that S_{21} is higher with pork tissues at the frequency range of interest, it is important to study impulse response also with the filtered S_{21} . The IR obtained by performing IFFT only for the frequency range 3.75–4.25 GHz, is presented in Fig. 7c. In this case, it can be clearly seen that the IR peaks are 1–3 dB stronger with pork tissues. Even the difference in the main peaks in almost 3 dB.

5 Measurement Results

5.1 Channel Comparison with Different Meat Pieces

In this section, the frequency and time domain channel characteristics are evaluated and compared using *Meat1* and *Meat2*, which both have thickness of 3 cm. The aim is to

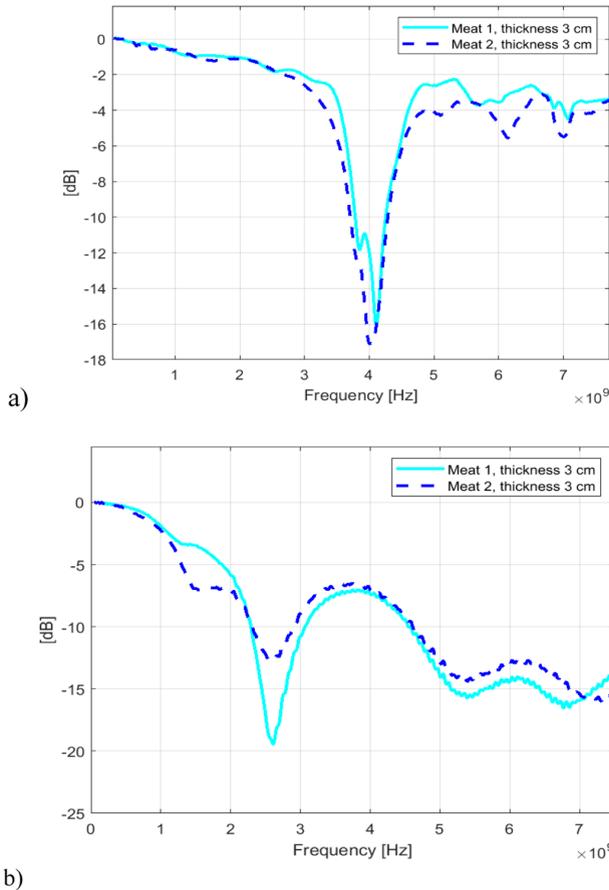


Fig. 8. Reflection coefficient for a) on-body antenna, b) the loop antenna within the tissues with *Meat1* and *Meat2*.

observe how the meat constitution, i.e. whether the fat and muscle layers are interlaced or separated, have impact on the channel characteristics.

Firstly, the antenna reflection coefficients are on-body and the loop antenna are studied when located on or below the pork pieces *Meat1* and *Meat2*. The reflection coefficients for the on-body antenna is presented in Fig. 8a and for the loop antenna in Fig. 8b. Antenna reflection coefficients for the on-body antenna are very similar with both meat pieces, since the skin thickness and outer fat thickness are same with both meats. Instead, one can note differences in the reflection coefficients of the loop antenna, which is located below the meat pieces. This is due to the different tissue constitution of these different pork meats: there is more fat on the inner surface of *Meat1* whereas more muscle tissues on the inner surface of *Meat2*. The differences in the antenna matching are noteworthy only at 1.5 GHz and 2.5 GHz, which however, are out of the frequency range of interest. The antenna matching at 3.75–4.25 GHz is relatively same with both meat pieces.

Next, the channel characteristics are studied. Frequency domain channel characteristics, i.e. S21 parameters, and time domain channel characteristics, i.e. channel impulse responses, are presented in Figs. 9a–b. As one can note, there is a clear level difference in the frequency and time domain channel characteristics obtained using Meat 1 and Meat 2 in the measurements. In frequency domain, the level difference is clearest at 1.5 GHz and 3 GHz which, however, are out of the frequency range of interest. At the on-body antenna's operational frequency range, 3.75–4.25 GHz, the level difference is at highest 10 dB, except at 3.75 GHz there is a large notch with *Meat1*. In time domain, the main difference is observed in the level and width of the main peak. The difference is also in this case at highest 10 dB.

The differences in the channel characteristics obtained using *Meat1* and *Meat2* are partly explained by differences in the antenna matching. However, the main difference comes from the different propagation paths due to different tissue composition. Fat is known to be one of the “easiest” tissue for UWB signal propagation in terms of propagation time and power loss [19, 20], hence, *Meat1* having “fat tunnels” through the muscle layer appear to be clearly easier meat piece to propagate through.

6 Impact of the Temperature

The channel results presented in the previous section were evaluated for the meat pieces which was measured to have temperature of 12 °C. However, temperature of the humans is approximately 37 °C. The temperature is known to have a clear impact on the tissues' dielectric properties [16], which is assumed to change also channel characteristics significantly. Thus, next, the impact of the temperature is evaluated for channel characteristics with *Meat1*.

The *Meat1* was heated in a heating box up to 37 °C. Then the S11, S21, and IR values are evaluated in the same antenna positions as in Section A. The obtained results are presented in Figs. 10a–c. For the comparison, S11, S21, and IR values measured in the cold (12 °C) temperature are included in the figure.

As noted from Fig. 10a, temperature changes the antenna matching slightly at the lower frequencies (up to 4.1 GHz). Due to these changes, the notch of the S11 is transferred 0.2 dB to the lower frequency, to 3.9 GHz. The changes in the S21 values are also

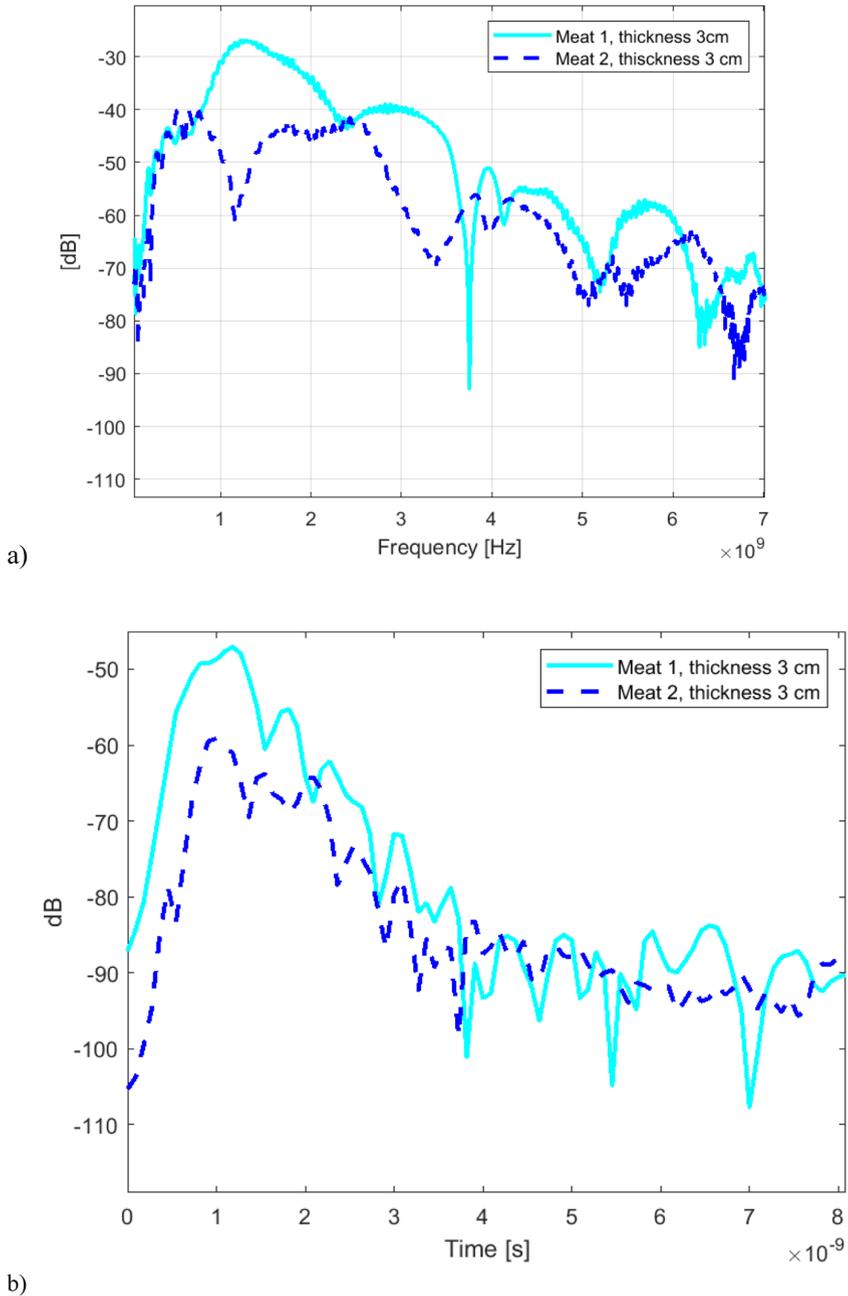


Fig. 9. a) Frequency domain and b) time domain comparison of channel characteristics obtained using *Meat1* and *Meat2*.

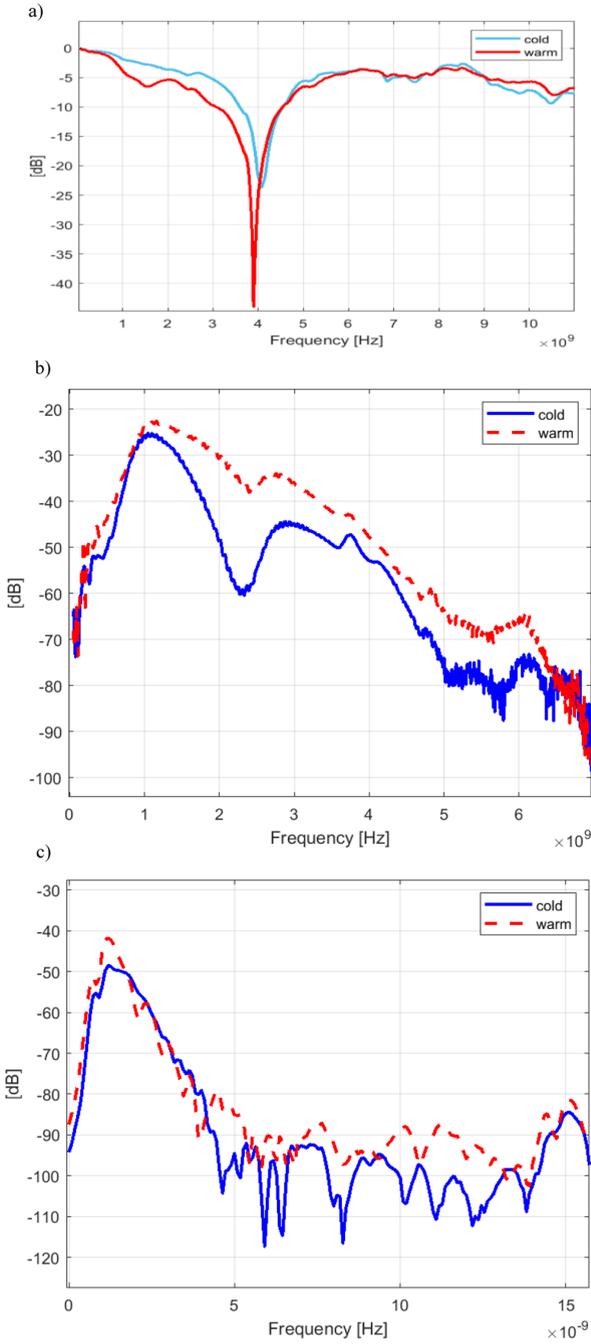


Fig. 10. Comparison between the antenna and channel characteristics at 12 °C and 37 °C temperature: a) S11 parameters for on-body antenna, b) S21 parameters, and c) IRs obtained performing IFFT for the whole frequency bandwidth.

noteworthy at lower frequencies, as presented in Fig. 10b. However, at the frequency range of interest, the difference is maximum 5 dB. In time domain results, presented in Fig. 10c, the main difference can be noted in the level of the CIR's main peak: with warm meat, the main peak is 9 dB at higher level than that with the cold meat.

7 Conclusions

This paper presented a study on UWB in-body propagation measurements conducted using pork meat. The first aim of this paper was to investigate by simulations the propagation differences between human and pork tissue layer models. The simulation results showed clear differences between the channel characteristics obtained using human tissues compared to pork tissues: within the frequency range of interest at 3.75–4.25 GHz, the path loss difference is up to 5 dB. The second aim of this paper was to study, by measurements, the propagation using different types of pork meat piece having different fat and muscle compositions. It was found that path loss is clearly higher with the pork meat having separate skin, fat, and muscle layers compared to the pork meat having interlaced fat and muscle layers. Furthermore, the third objective of this paper was to study the impact of the meat temperature on the measured channel characteristics by comparing the channels obtained with the meat at the temperature of 12 °C and at 37 °C. Also, in this case clear differences were observed in path loss: within the frequency range of interest, the path loss was maximum 5 dB lower with meat at 37 °C than with a colder meat.

The results presented in this paper provide information and insights on the use of pork meats in the in-body propagation studies. It is shown that selection of the meat pieces used in the measurements is crucial: meat composition may have strong impact on the channel characteristics. Besides, it would be important to heat the meat up to 37 °C if more realistic scenario is aim to be evaluated. Furthermore, although the dielectric properties of the human and adult pork tissue are similar, the differences in the dielectric properties cause clear impact on the channel characteristics, which should be taken into the account in monitoring device design.

Our future work plans include extensive measurements using different meat pieces having different thicknesses with different fat and muscle composition. Aim is to perform propagation path calculations for different propagation path options with the corresponding dimensions and compare them with the measured data. Furthermore, different on-body and implant antennas having different operational frequencies will be used in the evaluations.

Acknowledgement. This work is supported in part by the projects WBAN Communications in the Congested Environments (MeCCE), the Academy of Finland 6Genesis Flagship (grant 318927) and the European Union's Horizon 2020 programme under the Marie Skłodowska-Curie grant agreement No. 872752. Dr. Marko Sonkki is acknowledged for his participation on the on-body antenna design.

References

1. Pathinarupothi, R.K., Durga, P., Rangan, E.S.: IoT based smart edge for global health: remote monitoring with severity detection and alerts transmission. *IEEE Internet of Things J.* **6**, 2449–2462 (2019). Early Access
2. Schires, E., Georgiou, P., Lande, T.S.: Vital sign monitoring through back using an uwb impulse radar with body coupled antennas. *IEEE Trans. Biomed. Circuits Syst.* **12**, 292–302 (2018)
3. Leelatien, P., Ito, K., Saito, K., Sharma, M., Alomainy, A.: Channel characteristics and wireless telemetry performance of transplanted organ monitoring system using ultrawideband communication. *IEEE J. Electromagn. RF Microw. Med. Biol.* **2**, 94–101 (2018)
4. Ciuti, G., Menciassi, A., Dario, P.: Capsule endoscopy: from current achievements to open challenges. *IEEE Rev. Biomed. Eng.* **4**, 59–72 (2011)
5. Bchir, O., ben Ismail, M.M., AlZahrani, N.: Multiple bleeding detection in wireless capsule endoscopy. *Signal Image Video Process.* **13**(1), 121–126 (2018)
6. Demir, A.F.: Anatomical region-specific in vivo wireless communication channel characterization. *IEEE J. Biomed. Health Inform.* **21**(5), 1254–1262 (2017)
7. Garcia-Pardo, C., et al.: Experimental ultra wideband path loss models for implant communications. In: 2016 IEEE 27th Annual International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), pp. 1–6 (2016)
8. Andreu, C., Castelló-Palacios, S., Garcia-Pardo, C., Fornes-Leal, A., Vallés-Lluch, A., Cardona, N.: Spatial in-body channel characterization using an accurate UWB phantom. *IEEE Trans. Microw. Theory Tech.* **64**(11), 3995–4002 (2016)
9. Surowiec, A., Stuchly, S., Swarup, A.: Radiofrequency dielectric properties of animal tissues as a function of time following death. *Phys. Med. Biol.* **30**(10), 1131–1141 (1985)
10. Karacolak, T., Cooper, R., Unlu, E.S., Topsakal, E.: Dielectric properties of porcine skin tissue and in vivo testing of implantable antennas using pigs as model animals. *IEEE Antennas Wirel. Propag. Lett.* **11**, 1686–1689 (2012)
11. Shimizu, Y., Anzai, D., Chavez-Santiago, R., Floor, P.A., Balasingham, I., Wang, J.: Performance evaluation of an ultra-wideband transmit diversity in a living animal experiment. *IEEE Trans. Microw. Theory Tech.* **65**(7), 2596–2606 (2017)
12. Floor, P.A., et al.: In-body to on-body ultrawideband propagation model derived from measurements in living animals. *IEEE J. Biomed. Health Inform.* **19**(3), 938–948 (2015). <https://doi.org/10.1109/JBHI.2015.2417805>
13. Thotahewa, K.M.S., Redoutè, J., Yuce, M.R.: Propagation, power absorption, and temperature analysis of UWB wireless capsule endoscopy devices operating in the human body. *IEEE Trans. Microw. Theory Tech.* **63**(11), 3823–3833 (2015). <https://doi.org/10.1109/TMTT.2015.2482492>
14. Novak, O., Brown, R.B.: An empirical model of UWB large-scale signal fading in neocortical research. In: 2016 IEEE International Symposium on Circuits and Systems (ISCAS), Montreal, QC, pp. 2439–2442 (2016). <https://doi.org/10.1109/ISCAS.2016.7539085>
15. Gabriel, C., Gabriel, S., Corthout, E.: The dielectric properties of biological tissues: I. Literature survey. *Phys. Med. Biol.* **41**, 2231–2249 (1996)
16. Ley, S., Schilling, S., Fiser, O., Vrba, J., Sachs, J., Helbig, M.: Ultra-wideband temperature dependent dielectric spectroscopy of Porcine tissue and blood in the microwave frequency. *MDPI Sensors* **19**(7), 1–21 (2019). Special Issue on Microwave Sensors for Biomedical Applications
17. Peyman, A., Gabriel, C.: Cole-cole parameters for dielectric properties of porcine tissues as a function of age at microwave frequencies. *Phys. Med. Biol.* **55**, N413 (2010)

18. Ngadi, M., Dev, S., Raghavan, V.G.S., Kazemi, S.: Dielectric properties of pork muscle. *Int. J. Food Prop.* **18**, 12–20 (2015)
19. Särestöniemi, M., et al.: Fat in the abdomen area as a propagation medium in WBAN applications. In: Mucchi, L., Hämäläinen, M., Jayousi, S., Morosi, S. (eds.) *BODYNETS 2019*. LNCS, vol. 297, pp. 175–187. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-34833-5_15
20. Asan, N.B., Hassan, E., Shah, J., Noreland, D., Blokhuis, T., Augustine, R.: Characterization of the fat channel for intra-body communication at R-band frequencies. *Sensors* **18**(9), 2752 (2018)
21. CST Microwave Studio. <http://www.cst.com>
22. Kissi, C., Särestöniemi, M., Pomalaza-Raez, C., Sonkki, M., Srifi, M.N.: Low-UWB directive antenna for wireless capsule endoscopy localization. In: Sugimoto, C., Farhadi, H., Hämäläinen, M. (eds.) *BODYNETS 2018*. EICC, pp. 431–442. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-29897-5_38
23. IEEE Standard for Local and metropolitan area networks - Part 15.6: Wireless Body Area Networks. *IEEE Std 802.15.6-2012*, pp. 1–271 (2012)
24. Tuovinen, T., Yazdandoost, K.Y., Iinatti, J.: Comparison of the performance of two different UWB antennas for the use in WBAN on-body communications. In: *European Conference on Antennas and Propagation (EUCAP2012)* (2012)
25. Särestöniemi, M., Pomalaza-Raez, C., Berg, M., Kissi, C., Hämäläinen, M., Iinatti, J.: In-body power distribution for abdominal monitoring and implant communications systems. In: *ISWCS*, September 2019 (2019)
26. <https://www.itis.ethz.ch/virtual-population/tissue-properties/database/dielectric-properties>



Detection of Brain Hemorrhage in White Matter Using Analysis of Radio Channel Characteristics

Mariella Särestöniemi¹(✉), Carlos Pomalaza-Raez², Jaakko Hakala³,
Sami Myllymäki⁴, Joni Kilpijärvi⁴, Jari Iinatti¹, Matti Hämäläinen¹,
and Teemu Myllylä^{3,5}

¹ Centre for Wireless Communications, University of Oulu, Oulu, Finland
mariella.sarestoniemi@oulu.fi

² Department of Electrical and Computer Engineering, Purdue University, West Lafayette, USA

³ Optoelectronics and Measurement Techniques Research Unit, Faculty of Information
Technology and Electrical Engineering, University of Oulu, Oulu, Finland

⁴ Microelectronics Research Unit, Faculty of Information Technology and Electrical
Engineering, University of Oulu, Oulu, Finland

⁵ Research Unit of Medical Imaging, Physics and Technology, Faculty of Medicine, University
of Oulu, Oulu, Finland

Abstract. This paper presents a simulation-based study on detection of stroke/brain hemorrhage even in the white matter using radio channel characteristics analysis. The idea is to utilize the fact that blood has different dielectric properties than brain's white and grey matters and, thus, additional blood areas inside the brain change radio channel characteristics between the transmitter and receiver antennas located on the opposite sides of the head. The antennas should be strongly directive and designed to work attached to the body surface so that hemorrhages even in the white matter could be detected. The study is conducted using the electromagnetic simulation software CST and two different simulation models: a spherical tissue layer model and an anatomical voxel model. The antennas used in this study are bio-matched mini-horn antennas designed for implant communications at 1–4 GHz frequency range. Different sizes of the blood areas are evaluated. This initial study shows how even small sizes of hemorrhage can change radio channel even as the hemorrhage is located in the middle of the brain, in the white matter. The path loss difference is 0.5–10 dB between the hemorrhage and reference cases depending on the size and location of the hemorrhage. A practical solution of this hemorrhage detection technique could be a portable helmet type of structure having several small sized antennas around the internal part of the helmet. Such a helmet would be easy to use e.g. in ambulance, which would enable early detection of hemorrhage in its early phase and, hence, improve prospects of the cure significantly.

Keywords: Bio-matched horn antenna · Brain hemorrhage · Dielectric properties · Early diagnosis · Propagation loss · Radio channel analysis

1 Introduction

Brain hemorrhage/stroke is a fatal physical condition, which may cause severe disability or even death, and thus it is essential to be diagnosed in its early phase. Current method for detecting hemorrhage and stroke in the brain is practically limited to computing tomography (CT) and magnetic resonance imaging (MRI). They are expensive and complex techniques and can be used only in central hospitals. Consequently, there is a strong need for a wearable and easy to use technique for fast detection of hemorrhage and stroke already outside hospital. In the literature, there are microwave-based studies presenting hemorrhage detection [1–12]. Most of them are based on detecting changes in the antenna return loss, which has restrictions. For instance, hemorrhage located deep inside the brain white matter cannot be detected reliably with return loss analysis. Besides, several evaluations in the literature are conducted with the large antennas, which are not suitable for practical portable applications. The idea of helmet type brain hemorrhage detection device was originally presented in [13] which would measure channel parameters between the antennas located in different sides of the head. However, Ref. [13] presents only the idea, not any measurement or simulation results.

Most of the simulation-based studies in the literature are conducted using simplified head models. There is a lack of studies presenting radio channel characteristics-based brain hemorrhage detection, which would have been conducted using a realistic voxel model and directive antennas. Besides, there is lack of studies presenting brain hemorrhage detection in the white matter.

This study focuses detecting brain hemorrhage in the white matter in the middle part of the brain, since it is considered the most challenging part for hemorrhage detection. The paper involves propagation studies in the head area aiming at detection of hemorrhage and stroke using radio channel characteristics analysis. Hemorrhage detection could be done by using a simple, portable helmet type of monitoring device in which several transmitter and receiver antennas are located inside the helmet. The diagnosis could be done by setting the helmet in the patient's head, measuring and analysing the radio channel characteristics, and hence the diagnosis could be done outside the hospital in ambulance or even in patient's home which is important for rural area health care. Hence, the medical treatment could be started immediately, and prospects of cure could be significantly improved.

The physical phenomenon behind the proposed technique is that blood have different dielectric properties than the brain matter (grey and white), as shown in Table 1 [14]. Especially the relative permittivity of the blood is significantly higher than that of the brain matters. Thus, blood area inside brain change propagation and channel characteristics between the transmitter and receiver antennas, which are located in different sides of the head. Since the propagation losses in the tissues are relatively high, strongly directive antennas should be used so that the received signal is on the level in which the changes caused by hemorrhage can be still detected.

The objective of this paper is to present electromagnetic simulations-based study on the detection of brain hemorrhage in white matter using radio channel characteristics analysis. CST Simulia Studio Suite [15], which is based on finite integration technique, is used in the simulations. CST's anatomical voxel model Hugo is chosen for these evaluations since it has detailed brain structure with separate grey and white matters.

Table 1. Dielectric properties of different tissues in the head.

Tissue	Dielectric properties at selected frequencies		
	1 GHz Relative Permittivity/Conductivity [S/m]	2 GHz Relative Permittivity/Conductivity [S/m]	4 GHz Relative Permittivity/Conductivity [S/m]
Skin	40.9/0.9	38.6/1.27	36.6/2.34
Fat	11.3/0.12	11.0/0.12	10.4/0.50
Muscle	54.8/0.98	53.3/1.45	50.8/3.02
Skull bone	20.6/0.36	19.1/0.65	16.9/1.40
Cerebrospinal fluid	68.4/2.46	66.9/3.07	63.7/5.20
Brain grey matter	52.3/0.99	49.7/1.51	46.6/3.09
Brain white matter	38.6/0.62	36.7/1.00	34.5/2.14
Blood	61.1/1.58	59.0/2.19	55.7/4.13

The aim is to show how brain hemorrhage with different sizes and different locations can be detected using a directive on-body antenna designed for in-body communications. Even hemorrhages in the middle of the white matter can be detected with this method. 2D power flow illustration and power values at different locations inside the brain of the voxel show how the blood areas change the propagation. Besides, power loss calculations are presented to show the difference between the hemorrhage and reference cases.

The paper is organized as follows: Sect. 2 presents Study case by describing the antenna and simulation model used in this study. Power flow presentation are illustrated in Sect. 3. Channel evaluations are presented in Sect. 4, and Conclusions are given in Sect. 5.

2 Study Case

2.1 Directive on-Body Antenna

In this study, a directive bio-matched mini-horn antenna illustrated in Fig. 1, is used in the evaluations. The antenna is originally presented in [16] and based on the documentation it was remodeled for this study case. The dimensions of the antenna were further modified slightly to get better matching with the voxel model. The modified structure has dimension $h = 2.7$ cm and $d = 1.8$ cm. The bio-matched horn antenna is composed of water-filled holes, which mimic the frequency-dependent relative permittivity of the underlying tissue over its entire bandwidth. Details of the antenna structure can be read in [16].

The modified antenna structure is tuned to operate at 1.7–3.3 GHz. The antenna's S_{11} parameter is presented in Fig. 2.

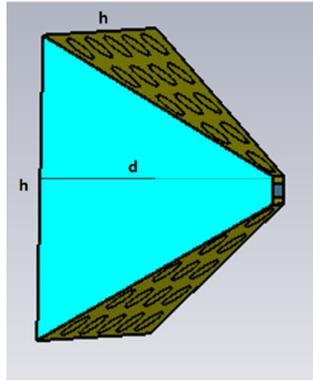


Fig. 1. A mini horn antenna designed for in-body communications.

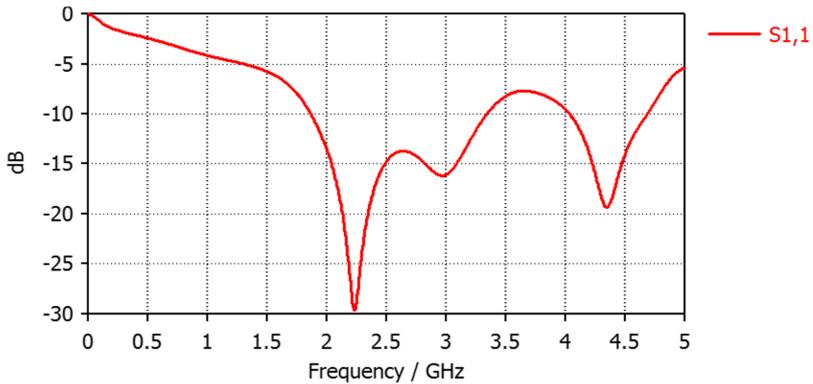


Fig. 2. Reflection coefficient S_{11} of the mini horn antenna as located on the skin.

2.2 Simulation Models

The simulations were carried out using the CST studio suite, which is based on the finite integration technique. Two different simulation models were used in this study: a spherical tissue layer model and CST's anatomical Hugo voxel model, which are presented in Figs. 3a–b, respectively. The spherical layer model is a full sphere, which consists of several different nested tissue spheres, each layer having realistic thicknesses. The thicknesses are presented in Table 2. The anatomical voxel model Hugo has the detailed brain model with separate grey and white matters having realistic shapes. The resolution of the Hugo voxel is $1 \text{ mm} \times 1 \text{ mm} \times 1 \text{ mm}$.

Antennas are located on the skin in the middle of the head, as shown in Fig. 4. For this study case, we chose the widest part in the head where the muscle layer is thickest since it can be seen as to the most challenging location in terms of radio signal propagation. The diameter of the voxel model's head is 16 cm in this location and thus also the diameter of the spherical model is set 16 cm.

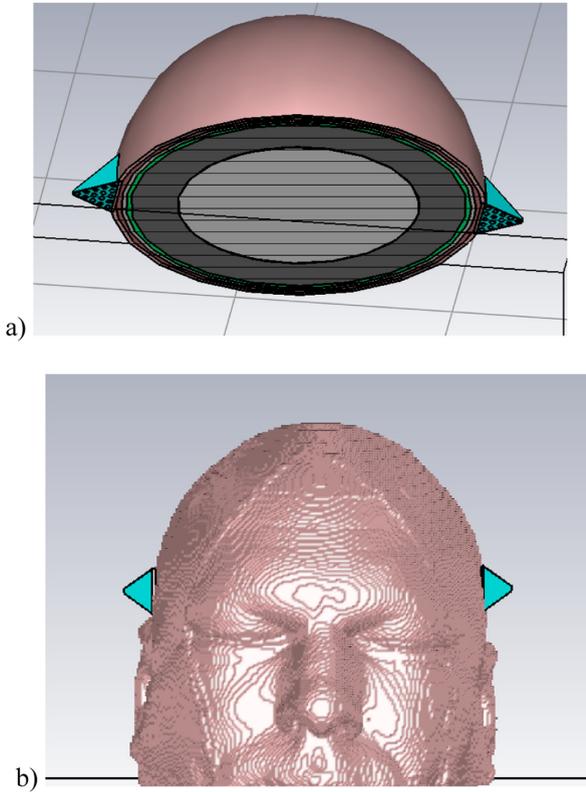


Fig. 3. a) Spherical layer model and b) Hugo voxel model with horn-antennas placed on the skin.

Bio-matched mini-horn antennas are designed to work so that they are completely attached to the skin [16]. Although the resolution of Hugo voxel is small, pixelization may have strong impact on the antenna and channel characteristics, as explained in [17]. Thus, the non-smoothness of the voxel skin surface is compensated by inserting thin skin layer in front of the antenna so that it overlaps with the voxel's non-smooth skin surface.

The possibility to detect brain hemorrhage is studied by inserting blood pieces of different sizes in the voxel model's brain white matter. The sizes with the corresponding dimensions are summarized in Table 2. The blood pieces and their locations are presented in Fig. 4 for the spherical layer model and in Fig. 5 for the voxel model. The spherical model is evaluated with the large piece in two different locations, whereas the voxel

Table 2. Dimensions of different sizes of brain hemorrhages.

Tissues	Small (S)	Medium (M)	Large (L)
Dimensions (x, y, z)	(2, 2, 1) cm	(3, 2, 1) cm	(5, 4, 2) cm

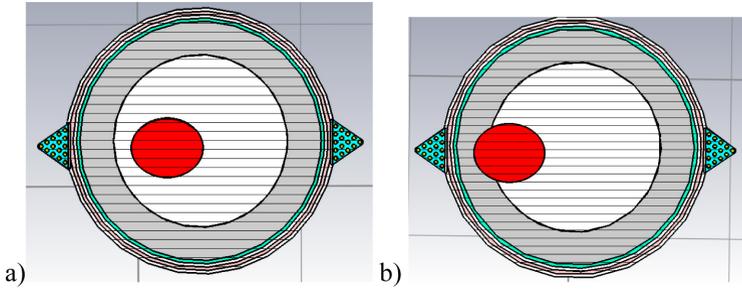


Fig. 4. Blood areas on the spherical model a) middle, b) left.

model is evaluated with large, medium, and small pieces, which are located in the middle of the brain.

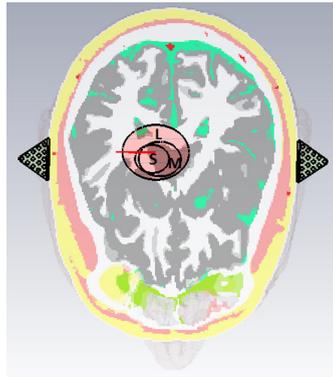


Fig. 5. Different sizes of brain hemorrhages in the cross-cut of the voxel model.

3 Power Flow Analysis

In this section, the power flow from the transmitter antenna to the receiver antenna is presented to understand how the hemorrhage change the propagation inside the brain tissues. Power flow representation at 1.7 GHz with and without the hemorrhage are presented in Figs. 6a–b, respectively. The dB range for the power flow presentation is 0–65 dB, where the reference 0 dB level is set to the antenna. Power flow on the skin is –26 dB.

From Figs. 6a–b one can note clear differences between the cases with and without the hemorrhage. In the presence of hemorrhage, the power flow is narrower. Within the plotted dB range, the Rx antennas is achieved easily in the reference case, whereas in the case of hemorrhage, the Rx antenna is hardly achieved. The differences in the power flows are due to the differences between the dielectric properties between the blood and

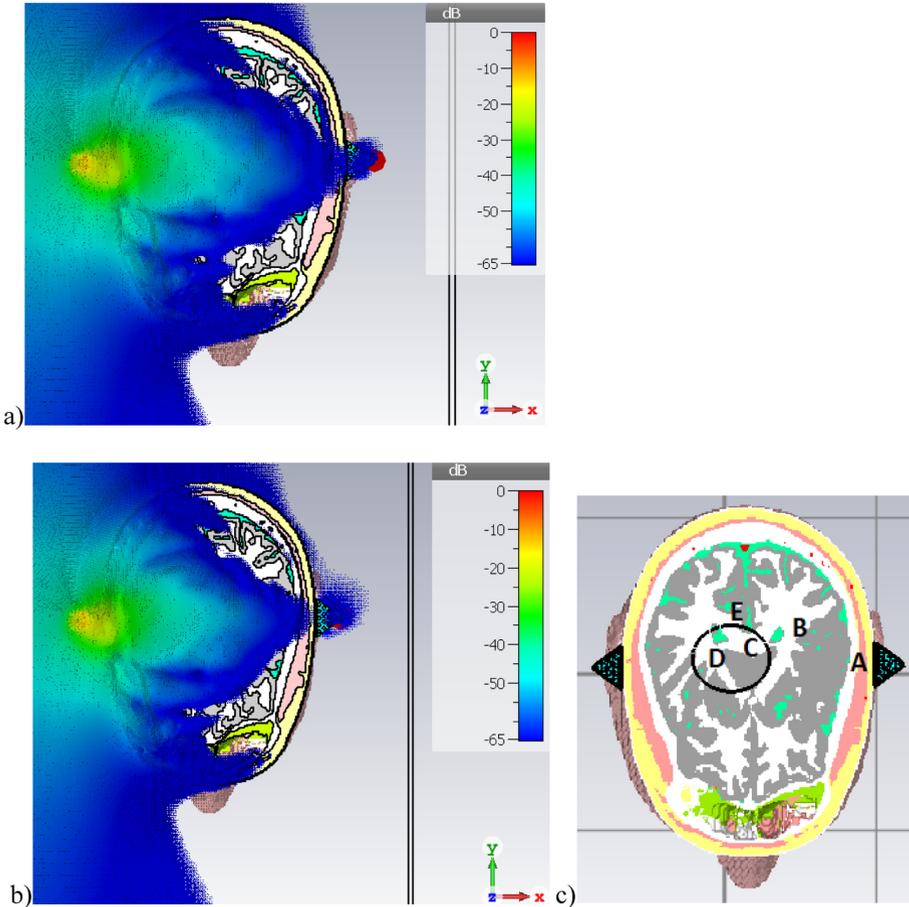


Fig. 6. Power flows at 1.7 GHz a) in the reference case, b) in the hemorrhage case, c) points A-E where power is measured.

brain tissues. Blood has clearly higher permittivity than the brain matters (grey/white) and thus, the power loss is larger in the presence of hemorrhage.

Table 3 presents the comparison between the power loss values obtained with the reference model and the hemorrhage model in the locations depicted in Figs. 6c. In the location A, which is close to the Rx antenna, the instant power flow value for the reference model at the selected cross-cut is 2 dB higher than that of the hemorrhage model. In point B, the power loss with the hemorrhage model is 6.5 dB higher than with the reference model. Interestingly, at point C, the reference model has higher power loss although the power inside the blood area is expected to be smaller. However, at point D, there is no difference between the hemorrhage and reference cases. Instead, at the point E, the difference is the power loss being higher with the hemorrhage model. It is interesting to note that the blood properties affect the power inside the hemorrhage, and this requires further studies.

Table 3. Power measured in points A–E at 1.7 GHz.

Power at points	A [dB]	B [dB]	C [dB]	D [dB]	E [dB]
Reference	−71.6	−71.5	−54	−48	−53
Hemorrhage	−73.6	−77.1	−52	−48	−56

As a comparison, power loss is calculated with Matlab based on planar model propagation. The calculation utilises an extension of the algorithm that includes lossy media proposed by Orfanides [16]. The tissue thicknesses as well as calculated power loss values in the presence of large size hemorrhage are presented in Table 4. As one can note, the power loss difference is significant between the hemorrhage and the reference case: in the presence of the hemorrhage having a large size, the loss is 9 dB higher at 1.7 GHz. However, in the power flow evaluations presented above, the difference is milder. The difference is due to the more complex structure of the voxel model, which enables alternative propagation paths than just a direct path through the tissues, which are on the direct line between the transmitter and the receiver antennas. However, power

Table 4. Thicknesses of the layers used in the power loss calculations.

Tissues	Relative permittivity	Loss tangent	Thickness [mm]
Air	1	0	0
Skin Dry	39.04	1.146	0.002
Muscle	53.68	1.288	0.006
Skull	15.665	0.408	0.002
CSF	67.34	2.853	0.002
Grey-Matter	50.29	1.334	0.015
White-Matter	37.48	0.8737	0.04
<i>Blood</i>	59.55	1.976	0.05
White-Matter	37.48	0.8737	0.016
Grey-Matter	50.29	1.334	0.015
CSF	67.34	2.853	0.002
Skull	15.665	0.408	0.002
Muscle	53.68	1.288	0.006
Skin Dry	39.04	1.146	0.002
Air	1	0	0

Power loss in the reference case: 50 dB

Power loss with blood: 59 dB

loss values obtained using Matlab provides good reference showing clearly the impact of the blood on the channel characteristics.

4 Channel Characteristics Analysis

First, the channel characteristics are evaluated with the spherical model in the presence and absence of blood. Channel parameters S_{21} s obtained using the reference model and the hemorrhage model are presented in Fig. 7. Hemorrhage location middle and left are included. As it can be seen, blood in both brain areas cause clear differences in the channel characteristics, especially in the frequency range 2.5–4 GHz. However, at lower frequencies 1.3–2 GHz one can note differences in the channel responses as well though the differences are milder, approximately 1.5 dB in this case. The closer the modelled hemorrhage area is to the antenna, the clearer is the difference. Interestingly, S_{21} obtained with the reference model is at clearly higher level than S_{21} of the models with blood pieces only in the frequencies 2.5 GHz onwards. At lower frequencies it is vice versa at certain frequencies. When comparing the dielectric properties of the blood, grey and white matter, one can note that within the simulated frequency range, the relativity permittivity of the blood is clearly higher than those of brain matters. Instead, the conductivity values are higher for the blood area.

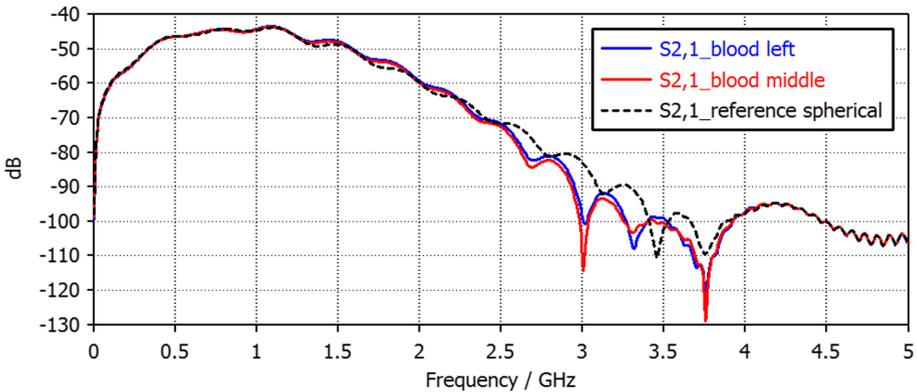


Fig. 7. S_{21} comparison with the spherical layer model as the hemorrhage is in the middle of the brain and in the left part of the brain.

Next, the S_{21} s obtained using the voxel model in the reference case and hemorrhage case are compared. The results are presented in Fig. 8. As it can be seen, the S_{21} of the reference model is at higher level than the S_{21} of the hemorrhage model at most of the frequencies, as expected from the power loss calculations. However, at e.g. 2.6 GHz, the S_{21} of the reference case is at lower level and the case of largest hemorrhage is at highest level. Obviously, the difference on dielectric properties of blood and brain matters have variation depending on the frequency. Thus, it is important to select carefully the frequency range used for this technique.

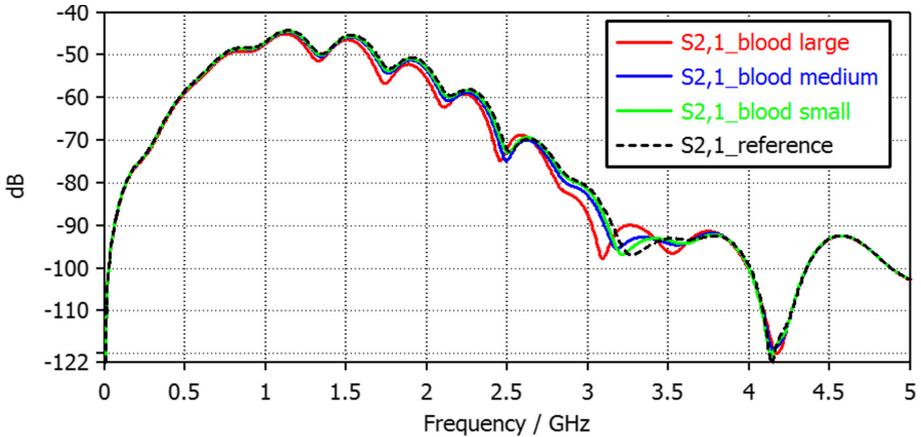


Fig. 8. S_{21} channel parameters obtained using the voxel model with different sizes of hemorrhage.

The zoomed version of the S_{21} results at lower frequencies are presented in Fig. 9. As one can note, even at lower frequencies the difference is 1–5 dB, which is clearly detectable in the practical solutions.

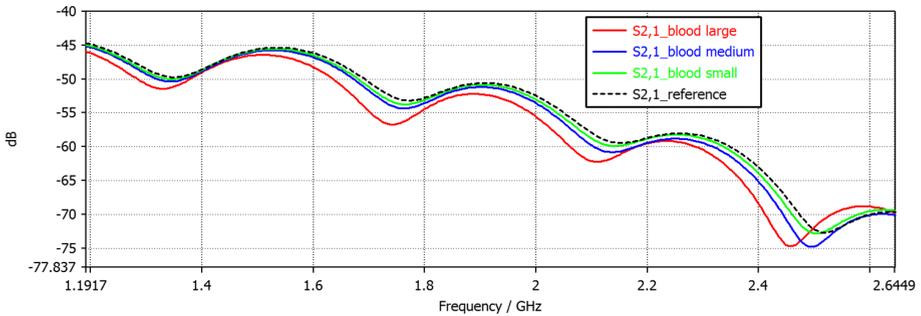


Fig. 9. Zoomed version of S_{21} at lower frequencies.

5 Conclusions

Detection of the brain hemorrhage in its early face is essential for the better prospects of cure. This paper presented a simulation-based study of detecting brain hemorrhage in the white matter using radio channel characteristics analysis, which could be realized in a portable and compact helmet type of structure with several small mini-horn antennas inside. The study is conducted by using the electromagnetic simulation software CST and two different simulation models: a spherical tissue layer model and an anatomical voxel model. The bio-matched mini-horn antennas designed for implant communications were used in this study.

The idea of this technique was to utilize the fact that blood has different dielectric properties than brain's white and grey matters and, thus, additional blood areas inside the brain change the radio channel characteristics between the transmitter and receiver antennas located on opposite sides of the head. This initial study showed how even small sizes of hemorrhage can change the radio channel characteristics even when hemorrhage is located in the middle of the brain, in the white matter. The path loss difference was 0.5–15 dB between the reference and hemorrhage cases depending on the frequency.

Power flow studies at 1.7 GHz revealed interesting phenomenon: although the power loss is expected to be larger in the presence of blood, the power inside the certain part of the hemorrhage area was noted to be higher than in the respective location without the blood. This is one of the future topics we are aiming to solve. Besides, other future work objectives include more detail studies on the impact of the hemorrhage in different frequencies and with several antennas to maximize visibility of the difference. Optimal antenna locations will be determined as well. Furthermore, different sizes of the head models will be used in the studies. Moreover, radio channel measurements will be conducted using a real human skull filled with phantom liquids [17] for brain matter and blood and prototypes of the bio-matched mini-horn antenna.

A practical solution of this hemorrhage detection technique could be a helmet type of structure having several small sized antennas around the internal part of the helmet. The brain hemorrhage could be detected analyzing channel characteristics of different antenna combinations. The benefit of helmet type of brain hemorrhage detector would be its easy use e.g. in ambulance. This would enable the early detection of hemorrhage in its early phase and hence, improve prospects of the cure significantly even for the patients living in rural areas far away from the hospitals.

The side effects and risks related to the radio channel analysis-based hemorrhage detection techniques are minor since the measurement itself is very fast and thus, the exposure for radiation is minor. Besides, input power can be kept low to meet the Specific Absorption Requirements (SAR) for the antennas.

Acknowledgement. This research has been financially supported by the project WBAN Communications in the Congested Environments and in part by Academy of Finland 6Genesis Flagship (grant 318927) and Academy of Finland grant 318347. This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 872752

Bio-matched mini-horn antenna was re-modeled by Mikko Parkkila and Uzman Ali from Radientum, which is greatly acknowledged.

References

1. Wang, D.-D., Chen, D.-Y., Luan, H.-S., Peng, H.-L.: A new method for transcranial vortex microwave beam imaging. In: 2018 International Conference on Microwave and Millimeter Wave Technology (ICMMT), pp. 1–3 (2018)
2. Abbosh, A.: Microwave systems for head imaging: challenges and recent developments. In: 2013 IEEE MTT-S International Microwave Workshop Series on RF and Wireless Technologies for Biomedical and Healthcare Applications (IMWS-BIO), Singapore, pp. 1–3 (2013). <https://doi.org/10.1109/IMWS-BIO.2013.6756184>

3. Zamani, A., Abbosh, A.M., Mobashsher, A.T.: Fast frequency-based multistatic microwave imaging algorithm with application to brain injury detection. *IEEE Trans. Microw. Theory Tech.* **64**(2), 653–662 (2016). <https://doi.org/10.1109/TMTT.2015.2513398>
4. Mobashsher, A.T., Mahmoud, A., Abbosh, A.M.: Portable wideband microwave imaging system for intracranial hemorrhage detection using improved back-projection algorithm with model of effective head permittivity. *Sci. Rep.* **6**(20459), 1–17 (2016)
5. Li, J., Peng, H.L., Zong, B.Q., Yu, Z.: A dual-polarized impulse radiation antenna for cerebral hemorrhage detection applications. In: 2014 IEEE MTT-S International Microwave Workshop Series on RF and Wireless Technologies for Biomedical and Healthcare Applications (IMWS-Bio2014), London, pp. 1–3 (2014)
6. Huang, D., Peng, H.L., Li, X., Jin, W., Mao, J.: A near-field spherical layered human head model for stroke detections. In: Asia-Pacific Microwave Conference (APMC), Nanjing, pp. 1–3 (2015)
7. Fear, E.C., Li, X., Hagness, S.C., Stuchly, M.A.: Confocal microwave imaging for breast cancer detection: localization of tumors in three dimensions. *IEEE Trans. Biomed. Eng.* **49**(8), 812–822 (2002)
8. Ireland, D., Bialkowski, M.: Microwave head imaging for stroke detection. *Prog. Electromagn. Res.* **21**, 163–175 (2011)
9. Tang, Z., Peng, H.L., Zhong, Y., Mao, J.F.: A novel antenna array for stroke enhanced detection. In: 2015 Asia-Pacific Microwave Conference (APMC), Nanjing, pp. 1–3 (2015)
10. Mesri, H.Y.: Localization of hemorrhage site in stroke patients using multichannel microwave measurements. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, pp. 5927–5930 (2012). <https://doi.org/10.1109/embc.2012.6347343>
11. Stancombe, A.E., Bialkowski, K.S., Abbosh, A.M.: Portable microwave head imaging system using software-defined radio and switching network. *IEEE J. Electromagn. RF Microw. Med. Biol.* **3**(4), 284–291 (2019). <https://doi.org/10.1109/JERM.2019.2901360>
12. Mobashsher, A.T., Abbosh, A.M., Wang, Y.: Microwave system to detect traumatic brain injuries using compact unidirectional antenna and wideband transceiver with verification on realistic head phantom. *IEEE Trans. Microw. Theory Tech.* **62**(9), 1826–1836 (2014). <https://doi.org/10.1109/TMTT.2014.2342669>
13. Fhager, A., Persson, M.: Stroke detection and diagnosis with a microwave helmet. In: 2012 6th European Conference on Antennas and Propagation (EUCAP), Prague, pp. 1796–1798 (2012). <https://doi.org/10.1109/eucap.2012.6206715>
14. <https://www.itis.ethz.ch/virtual-population/tissue-properties/database/dielectric-properties>
15. Blauert, J., Kiourti, A.: Bio-matched horn: a novel 1–9 GHz on-body antenna for low-loss biomedical telemetry with implants. *IEEE Trans. Antennas Propag.* **67**(8), 5054–5062 (2019)
16. Orfanidis, J.: *Electromagnetic Waves and Antennas* (2002). Revised 2016. <http://www.ece.rutgers.edu/~orfanidi/ewa/>
17. Hakala, J., Kilpijärvi, J., Särestöniemi, M., Hämäläinen, M., Myllymäki, S., Myllylä, T.: Microwave sensing of brain water - a simulation and experimental study using human brain models. *IEEE Access* **8**, 111303–111315 (2020)



UWB Microwave Imaging for Inclusions Detection: Methodology for Comparing Artefact Removal Algorithms

James Puttock¹(✉), Behnaz Sohani¹, Banafsheh Khalesi¹, Gianluigi Tiberi^{1,2},
Sandra Dudley-McEvoy¹, and Mohammad Ghavami¹

¹ Department of Electrical and Electronic Engineering,
London South Bank University, London, UK
puttockj@lsbu.ac.uk

² UBT – Umbria Bioengineering Technologies, Perugia, Italy

Abstract. An investigation is presented on Artefact Removal Methods for Ultra-Wideband (UWB) Microwave Imaging. Simulations have been done representing UWB signals transmitted onto a cylindrical head-mimicking phantom containing an inclusion having dielectric properties imitating an haemorrhagic stroke. The ideal image is constructed by applying a Huygens' Principle based imaging algorithm to the difference between the electric field outside the cylinder with an inclusion and the electric field outside the same cylinder with no inclusion. Eight different artefact removal methods are then applied, with the inclusion positioned at π and $-\frac{\pi}{4}$ radians, respectively. The ideal image is then used as a reference image to compare the artefact removal methods employing a novel Image Quality Index, calculated using a weighted combination of image quality metrics. The Summed Symmetric Differential method performed very well in our simulations.

Keywords: UWB microwave imaging · Image quality metric · Artefact removal

1 Introduction

In recent years, there has been considerable interest into research in the field of medical imaging. Current imaging techniques are advanced and can produce images of high clarity within a variety of tissue mediums. Medical professionals can draw on a range of technologies to assist with diagnosis depending on the suspected inclusion to be detected or the body part being imaged. Each technology has its own advantages and disadvantages. There are several techniques currently used for diagnosis purposes. Ultrasound scanners are cost-effective and successful in the medical diagnosis of areas of high contrast in soft tissues, hence their use in prenatal imaging and detection of various cancers. However, ultrasound is not a reliable technique for imaging air spaces, hard tissue such as bone/skull and providing definition in similar contrasting tissues. Computed Tomography (CT) is very good at imaging hard tissues but requires a substantial dose of ionising radiation. Magnetic Resonance Imaging (MRI) does not expose the patient to ionising

radiation but does require them to stay still inside a claustrophobic space for a long time (up to hours). MRI and CT Scan devices are also expensive and not portable (owing to the pertinent dimensions of the devices), nor are they usable at rural medical centres, or carried by First Response Services (FRS) which are seen as essential for saving a person experiencing stroke. There is therefore an opportunity for research into a non-ionising, non-invasive, portable and cost-effective alternative.

Ultra-Wideband (UWB) technology has already proved successful in a number of areas related to health monitoring. The non-invasive nature of the signals and the success at detecting changes quickly in the wireless medium has proven effective in health monitoring through movement detection [1]. Recent research has also shown the successful detection of lesions which have different dielectric properties to the surrounding medium using UWB Microwave imaging [2]. This holds the potential for detections of cancer and stroke [3–5].

Current UWB microwave imaging methods rely on algorithms to process the electric field measured at various points around the perimeter of an object. With all the algorithms used there is the risk of inaccurate results due to reflections of the transmitting signal and unwanted reflections of signals from the surrounding tissues. These unwanted signals are known as ‘artefacts’ [6]. For the successful application of any imaging apparatus, a reliable artefact-removal algorithm is necessary.

This study will explore a methodology for analysing and comparing a variety of methods for removing artefacts, using several imaging and signal processing metrics to provide a weighted Image Quality Index.

2 Theoretical Framework

Previous study focussing on breast cancer imaging [7] has identified that microwaves respond differently if they hit tissues which have different dielectric properties. By using UWB across microwave frequencies, it is possible to produce images with enough resolution to show inclusions. The same principle has been used to determine the contrast between blood and brain matter to identify stroke in head-mimicking phantoms [8].

This study uses a technique explained in [9] which uses Huygens’ Principle (HP) to forward propagate the waves [10]. This avoids having to solve complex inverse problems. A simulated waveform is constructed using the principles laid out by Parrikar et al. [11], which is transmitted from a line source external to the cylinder and received at a point on the radius, external or at the edge of the cylinder. The electric field E can be calculated by summing the known Electric Field E_{known} at N_{PT} points np on the perimeter using Eq. (1). HP indicates the Huygens’ Principle method used and G refers to the use of Green’s function.

$$E_{HP}(r, \phi; \theta; tx_m; f) = \Delta_s \sum_{np=1}^{N_{PT}} E_{np,tx_m}^{known} G(k_1 |\vec{r}_{np} - \vec{r}|) \quad (1)$$

where $(r, \phi, \theta) \equiv \vec{r}$ is the observation point, k_1 represents the wave number of the media constituting the cylinder, Δ_s is the spatial sampling, and tx_m is the transmitting line source operating at frequency f .

An image is obtained by summing the solutions and mapping the intensity values. With N_f frequencies f_i , the intensity of the final image I can be obtained using Eq. (2).

$$I(r, \phi, \theta; tx_m) = \frac{1}{B} \sum_{i=1}^{N_f} \Delta_f |E_{HP}(r, \theta, \phi; tx_m; f)|^2 \quad (2)$$

where Δ_f and B are the frequency sampling and Bandwidth, respectively.

By subtracting, before applying Eq. (1), the electric field with no inclusion from the electric field with an inclusion to get the difference in electric fields E^{diff} :

$$E_{np,tx_m}^{diff} = E_{np,tx_m}^{known(WithInclusion)} - E_{np,tx_m}^{known(NoInclusion)} \quad (3)$$

the image obtained through Eq. (2) will show the inclusion.

For the purposes of real scenario medical imaging, Eq. (3) cannot be used. In [6], it has been shown that this problem can be solved by producing a matrix of average values of the electric field obtained when the inclusion is present. Measurements are taken from multiple transmission sources and a mean value generated, which is subtracted from the single transmitter data. This is explained mathematically in Eq. (4).

$$E_{HP}(r, \phi, \theta; tx_m; f) = \Delta_s \sum_{np=1}^{N_{PT}} \left(E_{np,tx_m}^{known} - \text{avg}_M \left\{ E_{np,tx_m}^{known} \right\} \right) \times G(k_1 |\vec{r}_{np} - \vec{r}|) \quad (4)$$

The Average subtraction method represents just one method of obtaining an image of the inclusion. Several methods have been proposed which replace the average matrix with an alternative, such as the measurements from a neighbouring transmitter in the case of Rotation Subtraction [6, 12], or by using a differential method of obtaining the resulting image, as is the case in [2, 13, 14]. Variations on these techniques will be simulated and compared in this study.

3 Methodology

3.1 Ideal Image Construction

For the simulations, an external cylinder was simulated with radius 7cm, relative dielectric constant $\epsilon_r = 10$ and conductivity $\sigma = 0.2$ S/m. An internal cylindrical inclusion was constructed with radius 0.5cm, relative dielectric constant $\epsilon_r = 60$, conductivity $\sigma = 2$ S/m and located 2 cm from the centre of the external cylinder with an eccentricity angle of $-\frac{\pi}{4}$ radians. A simulation was run to construct the electric field at the perimeter of the external cylinder. This field was simulated using MATLAB and generated a value for the electric field at 80 points around the circumference for 1101 discrete frequencies between 1 and 6.5 GHz. The normalised microwave image of such cylinder with an inclusion can be seen in Fig. 1 (a). x and y values are in metres.

As can be seen in Fig. 1 (a), the electric field displays the reflection of the transmitter signal on the right of the cylinder. The transmitting signal has been set as a line source external to the cylinder at $x = -0.2, y = 0$. No evidence of the inclusion can be seen in this image. This is because the reflected transmitter signal is greater by a significant order of magnitude than the reflections from the inclusion.

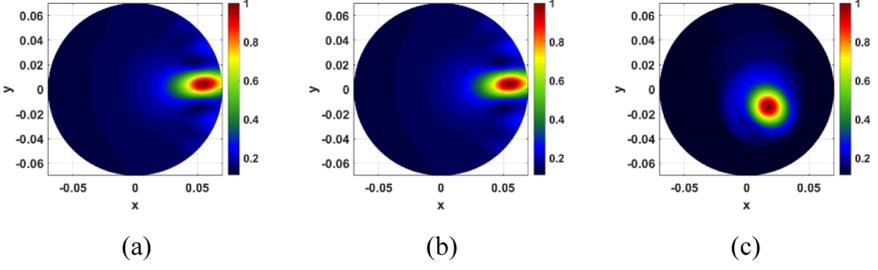


Fig. 1. (a) Image of cylinder with an inclusion, and (b) without inclusion. (c) Ideal image (constructed after employing Eq. (3)).

Figure 1(b) shows instead the image of the cylinder without the inclusion. Figure 1 (a) looks very similar to Fig. 1 (b). To detect the inclusion, Eq. (3) should be used before performing imaging: such an image is given in Fig. 1 (c) and is used as an ‘Ideal Image’ for reference and comparisons.

This process is repeated with an inclusion at an eccentricity angle of π radians and the images used as reference images for comparing subsequent experiments using different pre-processing algorithms.

3.2 Artefact Removal Methods

This study involves a comparison of various artefact removal algorithms operating in the frequency domain.

Average Subtraction (AS). The average subtraction method works by taking the results from multi-transmitting sources (3 in this case) positioned slightly apart from each other (4.5° in this case) on the perimeter of the cylinder, with the first transmitter x_1 positioned at $x = -0.07, y = 0$. When tabulating the (known) Electric Field at np points on the perimeter, the data from transmission point x_1 are placed into matrix A_1 which has dimensions made up by the number of frequencies \times number of observation points. For this series of experiments, 1101 discrete frequencies f are used at 80 observation points, here denoted with ϕ , giving matrix A_1 in Eq. (5).

$$A_1 = \begin{bmatrix} f_1\phi_1 & \cdots & f_1\phi_{80} \\ \vdots & \ddots & \vdots \\ f_{1101}\phi_1 & \cdots & f_{1101}\phi_{80} \end{bmatrix} \quad (5)$$

The results from transmission points x_2 and x_3 are placed into matrices A_2 and A_3 respectively in a similar fashion. Next, the mean of each point is calculated, i.e. A_{avg} .

The resulting average matrix A_{avg} is then subtracted from the matrix of the first transmitter A_1 before performing imaging:

$$E_{AS} = A_1 - A_{avg} \quad (6)$$

Rotation Subtraction (RS). The simulation is set up to replicate a signal from 2 transmitters, placed 4.5° apart on the perimeter of the cylinder. The receivers, frequencies and other input parameters remain the same as for the previous experiments. The image is then constructed using matrix E_{RS} , given as the following:

$$E_{RS} = A_1 - A_2 \quad (7)$$

where A_1 and A_2 are the results matrix from transmitter 1 and 2, respectively.

Differential Neighbouring Receiver Type (DNR). This method is adapted from Klemm's Differential method [2]. Instead of using the raw results to build a matrix from, the input matrix is instead built using the difference in value between neighbouring receivers. The input matrix S is thus built using the following calculation.

$$S(f, \phi; n) = A(f, \phi; n) - A(f, \phi - 1; n) \quad (8)$$

$$\text{for } \phi = 1 \text{ to } \frac{N_a}{2} \text{ with } \phi - 1 = N_a \text{ for } \phi = 1$$

$$S(f, \phi; n) = A(f, \phi, n) - A(f, \phi + 1; n) \quad (9)$$

$$\text{for } \phi = \frac{N_a}{2} + 1 \text{ to } N_a \text{ with } \phi + 1 = 1 \text{ for } \phi = N_a$$

where N_a is the number of receiving antennas, n is the transmitter index and A is the original results matrix. This results in a Differential (Neighbouring Receiver Type) matrix S . To such a matrix is then applied the Average Subtraction or Rotation Subtraction methods.

Differential Symmetric Receiver Type (DSR). By exploiting the (eventual) object symmetry, it may also be possible to construct a differential matrix using the difference between the receivers placed symmetrically opposite. This is adapted from a method used by Mustafa et al. in [13]. The differential matrix S is built by subtracting each receiver value from its symmetrically opposite receiver as in Eq. (10).

$$S(f, \phi; n) = A(f, \phi; n) - A(f, N_a + 2 - \phi; n) \quad (10)$$

$$\text{for } \phi = 1 \text{ to } N_a \text{ with } N_a + 1 = \frac{N_a}{2} + 1 \text{ for } \phi = 1 \text{ and}$$

$$\frac{N_a}{2} + 1 = 1 \text{ for } \phi = \frac{N_a}{2} + 1.$$

This results in a Differential (Symmetric Receiver Type) matrix S . To such a matrix is then applied the Average Subtraction or Rotation Subtraction methods.

Summed Symmetric Differential (SSD). The Symmetric Differential method above relies on the natural symmetry of some objects, such as the brain, across the left and right halves. However, there is a risk with the symmetric method of mirrored artefacts appearing in the images. The ellipsoid shape of the skull and brain have a distinct left-right line of symmetry. The front-back sections of the brain also contain similar densities of tissue. Whilst not completely symmetrical, the similarity in shape and density could be utilised to provide an artefact removal method by summing a differential matrix formed from the left-right differential and a second matrix formed from a front-back differential. This should provide a more intense peak at the area of inclusion and mirrored artefacts should have a reduced intensity. As before, a differential matrix S is constructed as in Eq. (10). A second matrix R is constructed across the front-back receivers as follows.

$$R(f, \phi; n) = A(f, \phi; n) - A\left(f, \frac{N_a}{2} + 2 - \phi; n\right) \quad (11)$$

$$\text{for } \phi = 1 \text{ to } \frac{N_a}{2} + 1$$

$$\text{with } \frac{N_a}{2} + 2 - \phi = \frac{3N_a}{4} + 1 \text{ for } \phi = \frac{N_a}{4} + 1$$

$$R(f, \phi; n) = A(f, \phi; n) - A\left(f, \frac{3N_a}{2} + 2 - \phi; n\right) \quad (12)$$

$$\text{for } \phi = \frac{N_a}{2} + 2 \text{ to } N_a$$

$$\text{with } \frac{3N_a}{2} + 2 - \phi = \frac{N_a}{4} + 1 \text{ for } \phi = \frac{3N_a}{4} + 1$$

where N_a is the number of receiving antennas, n is the transmitter index and A is the original results matrix. The combined matrix C is then constructed by summing matrices S and R .

$$C(i, j; n) = S(i, j; n) + R(i, j; n) \quad (13)$$

This results in a Differential (Summed Symmetric Receiver Type) matrix. To such a matrix is then applied the Average Subtraction or Rotation Subtraction methods.

3.3 Comparison Methods

To compare the proposed artefact removal methods, it is necessary to construct a quantifiable measurement system that can be used to compare images. It is often difficult to construct a useful quantifiable number to measure an image. Whilst humans are good at recognising patterns and contrasts in an image, a machine must be taught each process, and this uses considerable computing power. For this experiment, basic comparison metrics have been calculated to compare the results. These are explained in greater detail below. Some of the below metrics rely on a reference image. For the purposes of this experiment, an ‘Ideal Image’ has been used as shown above.

Error Image. An Error image M_{Err} is constructed by subtracting the ideal image X_{Ideal} , from the image to be tested Y_{Test} .

$$M_{Err} = Y_{test} - X_{Ideal} \quad (14)$$

Mean Square Error (MSE). The mean square error is the mean of all the squared values of the errors in the error matrix. Squaring the values means that any negative values become positive, so the absolute value is important. The average gives a single value which is an indication of the error across the whole matrix.

$$MSE = \frac{1}{N} \sum_{i=1}^N M_{Err_i}^2 \quad (15)$$

where N is the number of elements in the Error matrix M_{Err} . The mean square error can be calculated using MATLAB's imaging toolbox and the command `immse` [15].

Polyshape Construction. To evaluate the shape of the inclusion, we set 0.75 as a threshold on the normalised image, assigning every value above 0.75 to 1, and all others to 0. The resulting shape can then be obtained using MATLAB's `polyboundary` and `polyshape` functions [15].

Area Difference (ArD). This metric is related to the comparison between the size of the target area for an 'Ideal' image and the size of the target area in the test image.

Centroid Difference (CD). To test the accuracy of the image at locating an inclusion, the above Polyshape method was combined with MATLAB's `centroid` function. Comparison of accuracy can be made by assessing the Euclidean difference between the centroid of an ideal image polyshape and the test image. This will assess how accurate the location of the inclusion in the test image is. This is done using the MATLAB `pdist` function [15].

Signal-to-Noise Ratio (SNR). The Signal-to-Noise Ratio (SNR) is a useful metric in determining how clear any detected inclusion is by providing an assessment of the ratio between the background noise and the desired signal. To calculate the Signal-to-Noise Ratio (SNR) in dB, the above threshold is used to calculate the Polyshape to determine the target and background areas. SNR calculations are performed based on this result. This method can be calculated using Eq. (16).

$$SNR = 10 \log_{10} \left(\frac{Q_t - Q_b}{D_b} \right) \text{dB} \quad (16)$$

where Q_t and Q_b are the mean values of the detected target and background regions, respectively, and D_b is the standard deviation of the background.

Structural Similarity Index Metric (SSIM). The SSIM is an image quality metric which gives a value between 0 and 1 which indicates the similarity between two images (with 1 meaning the images are identical) [16]. This is calculated using the following equation.

$$SSIM = \frac{(2 \times \bar{x} \times \bar{y} + C1)(2 \times \sigma_{xy} + C2)}{(\sigma_x^2 + \sigma_y^2 + C2)(\bar{x}^2 + \bar{y}^2 + C1)} \quad (17)$$

where x is the reference image, y is the test image, \bar{x} and \bar{y} represent the corresponding mean, σ_x and σ_y represent the corresponding variance, σ_{xy} is the covariance of the reference and test image and $C1$ and $C2$ are small constants. MATLAB can calculate the SSIM based on two input images using the `ssim` function [15]. This will output both a value and a monochrome mapping which is a useful visual assessment of the quality of the image. An example is shown in Fig. 2(e).

Average Difference (AvD). The average difference is a measure of the mean difference in value between the Ideal Image and the test image. It is calculated by summing the elements of the Error Matrix and dividing by the number of elements.

Image Quality Index (IQI). The above metrics provide several ways of quantifying the precision, accuracy, and quality of the images to be constructed. Whilst these metrics are useful, to aid in comparing the methods an overall quality index will be calculated. This will be constructed by giving each of the metrics a score between 1 and 0 (with 1 being a perfect image or match with the Ideal image). For many of these values, as they are already being based on normalised results which will be between 0 and 1, it is relatively simple to produce an appropriate score. For SNR, instead a comparison is made with the Ideal Image. The value will approach 1 as it approaches the SNR value of the Ideal image. The full metric Indexes I are shown in Eq. (18–23) with R representing the test result value.

$$I_{MSE} = 1 - MSE_R \quad (18)$$

$$I_{SNR} = \frac{SNR_R}{SNR_{Ideal}} \quad (19)$$

$$I_{SSIM} = SSIM_R \quad (20)$$

$$I_{AvD} = 1 - AvD_R \quad (21)$$

$$I_{CD} = 1 - \frac{CD_R}{\text{External Cylinder Radius}} \quad (22)$$

$$I_{ArD} = 1 - \frac{ArD_R}{\text{External Cylinder Area}} \quad (23)$$

The overall Image Quality Index is then calculated by taking a weighted average of all the indexes. The most useful metrics for our study will be ones that measure the accuracy, precision, and quality of the image. Therefore, Area Difference Index I_{ArD} , Centroid Difference I_{CD} and Signal-to-Noise Ratio I_{SNR} are each given a weighting of 0.25. The SSIM value I_{SSIM} provides a very useful measure of the contrast differences between

the test image and ideal image. This has been given a weighting of 0.15. The Average Difference I_{AvD} and Mean Square Error I_{MSE} provide useful additional information but are considered a less reliable assessment of quality as error value can easily be influenced by the power of received signals without necessarily affecting the ability to detect inclusions. These have therefore been given a weighting of 0.05 each. These values have been chosen arbitrarily based on the measurement requirements of this project.

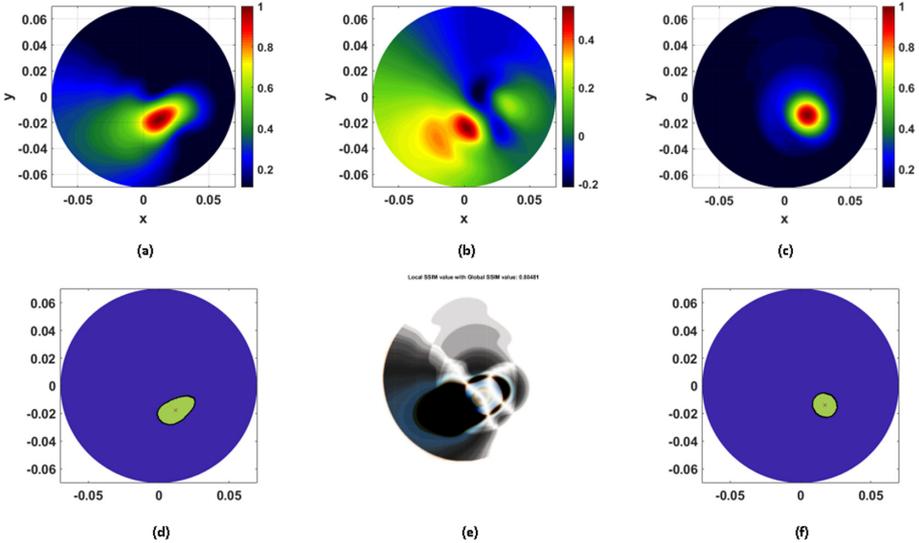


Fig. 2. Average Subtraction Results Images (a) Results Image, (b) Error Image, (c) Ideal Image, (d) Results Polyshape, (e) SSIM Image and (f) Ideal Polyshape.

4 Results

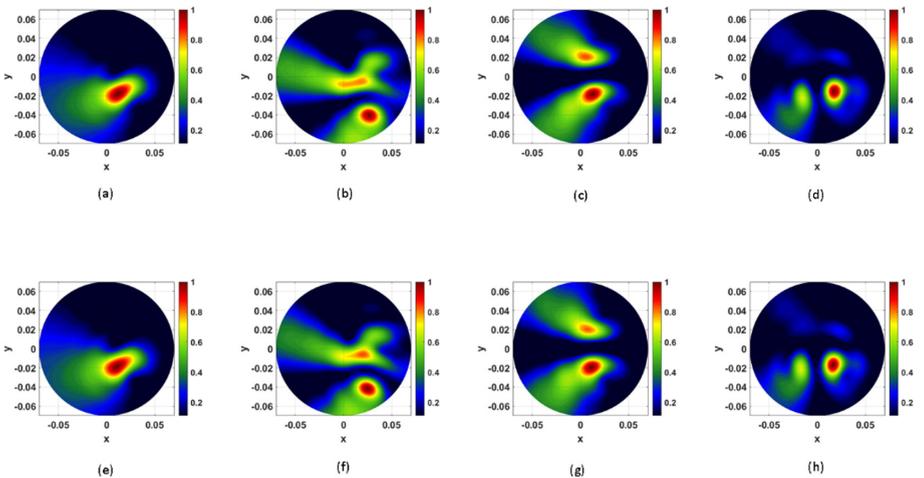
The simulation was run using each of the 5 Artefact removal methods, with the differential methods being calculated using an Average Subtraction or Rotation subtraction sub-method on the differential matrix. The results are presented for an inclusion at π radians (Table 1) and for an inclusion at $-\frac{\pi}{4}$ radians (Table 2). The experiment names and metrics are expressed in abbreviated form. The Ideal SNR value was calculated following the construction of the Ideal Images and had a value of approximately 7.4 dB. The results images for an inclusion at $-\frac{\pi}{4}$ radians are shown in Fig. 3. A subset of the results images for Average Subtraction (AS) methods on an inclusion at π radians are shown in Fig. 4.

Table 1. Artefact Removal Comparison with Inclusion at π radians.

Experiment	Subtraction Method	MSE	SNR [dB]	SSIM	AvD	CD [mm]	ArD	IQI
AS	N/A	0.09729	5.36519	0.70400	0.25517	1.21322	0.0018	0.90422
RS	N/A	0.09937	5.44750	0.70809	0.25663	1.40726	0.0005	0.90775
DNR	Average	0.09814	6.06145	0.71389	0.25937	13.06835	0.0233	0.88642
DNR	Rotation	0.10098	6.04735	0.71476	0.26439	11.70547	0.0179	0.89186
DSR	Average	0.09974	5.40451	0.69871	0.25736	2.43316	0.0008	0.90070
DSR	Rotation	0.10008	5.44746	0.69865	0.25737	1.15772	0.0000	0.90726
SSR	Average	0.09462	6.03662	0.73537	0.23888	3.33967	0.0061	0.92886
SSR	Rotation	0.09523	6.06927	0.73697	0.23870	2.94882	0.0072	0.93161

Table 2. Artefact Removal Comparison with Inclusion at $-\frac{\pi}{4}$ radians.

Experiment	Subtraction Method	MSE	SNR [dB]	SSIM	AvD	CD [mm]	ArD	IQI
AS	N/A	0.02934	5.66908	0.80481	0.13335	6.14234	0.0143	0.87828
RS	N/A	0.03214	5.59032	0.80104	0.13886	7.21017	0.0164	0.86969
DNR	Average	0.06838	3.55544	0.73731	0.20546	9.19064	0.0159	0.78002
DNR	Rotation	0.05846	3.94577	0.74968	0.19013	11.2011	0.0000	0.79266
DSR	Average	0.06179	4.60431	0.75460	0.19962	14.3283	0.0141	0.80071
DSR	Rotation	0.06317	4.62941	0.75185	0.20243	12.8656	0.0118	0.80623
SSR	Average	0.02489	6.99660	0.86229	0.11733	1.62966	0.0082	0.95034
SSR	Rotation	0.02589	6.92966	0.86047	0.12060	2.86084	0.0080	0.94248

**Fig. 3.** Results Images for an Inclusion at $-\frac{\pi}{4}$ radians. Top Row: Average Subtraction – (a) AS, (b) DNR, (c) DSR, (d) SSR. Bottom Row: Rotation Subtraction – (e) RS, (f) DNR, (g) DSR, (h) SSR.

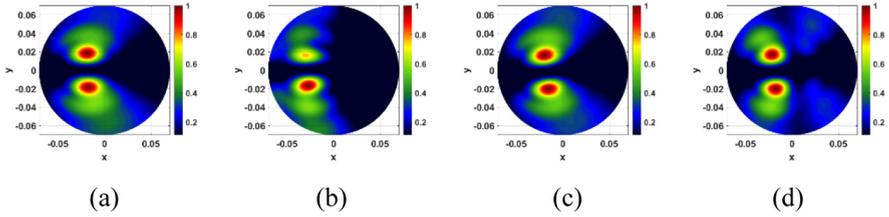


Fig. 4. Average Subtraction results for an inclusion at π radians: (a) AS, (b) DNR, (c) DSR, (d) SSR.

5 Discussion and Conclusion

Summed Symmetric Differential method had the best Image Quality Index. The Signal-to-Noise Ratio was the highest for this method in both positions and the contrast similarity meant that this method also had a high SSIM in comparison to the other methods.

For an inclusion at π radians the Average Subtraction, Rotation Subtraction and Differential (Symmetric Receiver Type) methods all scored highly in the Centroid Distance metric. When the inclusion was at $-\frac{\pi}{4}$ radians, the Summed Symmetric method had the best Centroid Distance score.

The results show that the position of the inclusion can vastly influence the resulting image, with every artefact removal method exhibiting differences in image quality when the inclusion was moved. When the inclusion is at π radians, all images show a symmetric split inclusion image. This is likely due to the inclusion being directly in front of the transmitting source. As the image is split perfectly either side of the inclusion position, the centroid distance (and to a lesser extent the Area Difference) remain accurate. The SSIM is greatly affected though due to the difference in inclusion shape between the test image and ideal image. In simulation, the source signal will hit the inclusion and the highest peaks are observed as the signal reflections are ‘split’ by the inclusion. With Average Subtraction and Summed Symmetric Differential methods, there is far greater definition of the inclusion position. To avoid the inclusion position having an impact on the result, multiple transmitter groups could be used. If five transmitter triplets are used, such as in [4] with each transmitter within a triplet placed 4.5° apart and the triplets placed equally around the perimeter of the cylinder to be imaged, the inclusion will only ever be opposite a maximum of one transmitter triplet. If the results for each triplet are summed, or a mean value taken, and then imaged the resulting image should have improved the definition of the inclusion image, reflected artefacts will be minimised and the dependency on inclusion position will be removed. However, as the Summed Symmetric Differential methods had reflected artefacts which were separate from the inclusion image, the effect of summing the results from multiple transmitters could increase the quantity of artefacts seen for this method. Concerning computational time, we found no significant difference among the algorithms used.

Whilst the Summed Symmetric method scored highly in this study, the experiments were performed using simulated cylinders with significantly contrasting electrical properties. This was performed to prove the algorithms as a proof of concept against contrasting mediums. In a realistic scenario, such as brain stroke detection, the contrast will

be less significant, thus artifacts could be more prominent, lowering SNR values. Future studies should aim to use simulated data which equates the electrical properties closer to that of human tissues. Further study using measured data and head-mimicking phantoms would be an obvious progression from this study to assess the value in these methods.

It is also important to note that this study has also focused only on cylindrical inclusions inside a cylinder. Further studies could explore the effect of changing the shape and size of the inclusion.

The Image Quality Index provides a good metric for quantifying the quality of images. By visually comparing the results images and the IQI, the score seems representative of the quality of the image. For a more rigorous assessment of the IQI method, future study could incorporate a comparison with subjective scoring by a sample of medical professionals.

Acknowledgments. This project leading to this application has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 793449.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 872752.

References

1. Taylor, W., Shah, S.A., Dashtipour, K., Zahid, A., Abbasi, Q.H., Imran, M.A.: An intelligent non-invasive real-time human activity recognition system for next-generation healthcare. *Sensors* **20**(9), 2653 (2020)
2. Klemm, M., Leendertz, A., Gibbins, D., Craddock, I.J., Preece, A., Benjamin, R.: Microwave radar-based differential breast cancer imaging: imaging in homogenous breast phantoms and low contrast scenarios. *IEEE Trans. Antennas Propag.* **58**(7), 2337–2344 (2010)
3. Ghavami, M., et al.: UWB Device for Microwave Imaging: Validation through Phantoms. In: 13th European Conference on Antennas and Propagation, Krakow, Poland (2019)
4. Vispa, A., et al.: UWB device for breast microwave imaging: phantom and clinical validations. *Measurement* **146**, 582–589 (2019)
5. Sohani, B., et al.: Detection of haemorrhagic stroke in simulation and realistic 3-D human head phantom using microwave imaging. *Biomed. Signal Process. Control* **61**, 102001 (2020). <https://doi.org/10.1016/j.bspc.2020.102001>
6. Elahi, M.A., Glavin, M., Jones, E., O'Halloran, M.: Artifact removal algorithms for microwave imaging of the breast. *Prog. Electromagnetics Res.* **141**, 185–200 (2013)
7. Lazebnik, M., et al.: A large-scale study of the ultrawideband microwave dielectric properties of normal, benign and malignant breast tissues obtained from cancer surgeries. *Phys. Med. Biol.* **52**(20), 6093–6115 (2007)
8. Sohani, B., Tiberi, G., Ghavami, N., Ghavami, M., Dudley, S., Rahmani, A.: Microwave imaging for stroke detection: validation on head-mimicking phantom. In: PIERS (Photonics and Electromagnetics Research Symposium), Rome, Italy (2019)
9. Ghavami, N., Tiberi, G., Edwards, D.J., Monorchio, A.: UWB microwave imaging of objects with canonical shape. *IEEE Trans. Antennas Propag.* **60**(1), 231–239 (2012)
10. Enders, P.: Huygens' principle as universal model of propagation. *Latin American J.Phys. Educ.* **3**(1), 19–32 (2009)

11. Parrikar, R.P., Kishk, A.A., Elsherbeni, A.Z.: Scattering from an impedance cylinder embedded in a nonconcentric dielectric cylinder. *IEEE Proc. H: Microwaves, Antennas Propag.* **138**(2), 169–175 (1991)
12. Klemm, M., Craddock, I.J., Leendertz, A., Preece, A., Benjamin, R.: Improved delay-and-sum beamforming algorithm for breast cancer detection. *Int. J. Antennas Propag.* **2008**, 9 (2008). Article ID 761402
13. Mustafa, S., Mohammed, B., Abbosh, A.: Novel preprocessing techniques for accurate microwave imaging of human brain. *IEEE Antennas Wirel. Propag. Lett.* **12**, 460–463 (2013)
14. Zamani, A., Abbosh, A.M.: Clutter rejection techniques for microwave head imaging. In: 2015 International Symposium on Antennas and Propagation (ISAP), Hobart, Australia (2015)
15. MATLAB Documentation Centre. <https://uk.mathworks.com/help/documentation-center.html> Accessed 15 May 2020
16. Zhou, W., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004)



BSNCloud: Cloud-Centered Wireless Body Sensor Data Collection, Streaming, and Analytics System

Ming Li¹✉, Ai Enkoji¹, Matthew Key¹, Aaron Marroquin¹, and B. Prabhakaran²

¹ California State University Fresno, Fresno, CA 93740, USA
mingli@mail.fresnostate.edu

² The University of Texas at Dallas, Richardson, TX 75080, USA

Abstract. Cloud-assisted body area networks have been the focus of researchers in past years as a response to the development of robust wireless body area networks (WBANs). While software such as Signal Processing in Node Environment (SPINE) provide Application Programming Interfaces (APIs) to manage heterogeneous biomedical sensor networks, others have focused on developing tools that address the issue of sensor connection/control, data receiving, and visualization. However, existing software tools lack sufficient flexibility, scalability, and support for complicated biomedical systems. In this paper, BSNCloud, a cloud-centered heterogeneous and comprehensive wireless body sensor data collection, streaming, and analytics framework is proposed. The system combines the sensor control and data aggregator event detection, real-time data analysis, visualization, and streaming into one Android App and incorporated four key components in the cloud server: data repository, algorithm repository, machine learning engine, and web portal. A prototype has been implemented with preliminary performance evaluation. Results show that the system is promising in its full utilization of the high performance computing power as well as the large volume storage capacity.

Keywords: Body sensor networks · Cloud-assisted · Wireless body area networks

1 Introduction

Wearable computing has gained increasing research and development, especially in biomedical applications. For example, several software tools from Shimmer Sensing [1], Qualcomm [2], Intel [3] and Samsung [4] attempted to address the issue of sensor connection/control, data receiving, and visualization. However, these tools are largely limited in flexibility, scalability, and support for complicated biomedical systems. On the other hand, systems have been recently proposed to enable cloud assisted wireless body area networks. Specifically, BodyCloud [5] provides a general-purpose software that covers a wide range of sensors and provides APIs for the creation of new biomedical applications.

Despite of existing research advances, two key issues have not been addressed: (i) medical doctors and practitioners need user-friendly, multiple sensor supported, real-time, and powerful analytics support for high quality patient health monitoring activities. For this purpose, body sensor data should be collected, streamed, processed, analyzed (real-time and on-demand), and archived in cloud server. (ii) biomedical researchers face significant challenges of high cost and overhead performing data analytics. To resolve this issue, both body sensor data repository and data analytics algorithm repository should be created in the cloud server so that researchers can complete the task of large volume data analysis at minimum cost without the overhead of purchasing computing facility, recruiting algorithm developers, as well as obtaining results within a few days instead of months.

We propose BSNCloud, a heterogeneous and comprehensive wireless body sensor data collection, streaming, and analytics framework. BSNCloud is unique in the sense that it provides both real-time Android App based analysis as well as on-demand web portal based data analysis with the potential of supporting a large set of body sensor data and analytics algorithms. The system operates across a set of wireless body sensors, Android phone/tablets, PC desktops/laptops, and a cloud cluster. For implementation, we adopted the widely used Shimmer3 IMU motes that is able to sample multiple signals such as accelerometer. Then, the basic MultiShimmer Template for Android was repackaged to include additional key features such as signal statistics, event detection, real-time data analysis, and multi-signal streaming (using WebSocket [6]). The cloud server is setup with OpenStack operating system, a PHP web server, and MySQL database. Users can perform data analysis using existing data sets and algorithms, upload live data through the Android App, upload their own data on the Web Portal, upload their own algorithms, or modify contributed algorithms for validation and performance evaluation. Experimental results show that the proposed BSNCloud is very promising. Specifically, it achieves several desirable features:

- *Usability*: given the reasonable cost of Shimmer motes and user friendly web portal, medical researchers and doctors can easily collect data and archive them securely in the cloud server without worrying about storage management.
- *Flexibility*: users have the option of using existing shared data and contributed algorithms as well as experimenting with their own collected data and algorithms, making performance evaluation a fairly easy task. The Android App supports multiple users and multiple signal streaming and therefore fits well for both hospital and home rehabilitation setting.
- *Efficiency*: both real-time and on-demand web portal data analysis are performed in the high performance cluster server, it is possible to run large data volume without significant delay.
- *Scalability*: BSNCloud can be extended to include other Bluetooth enabled sensors. Development of iOS App is also feasible so that data from iPhone sensors can be collected too. The proposed algorithm repository has the potential to grow as more researchers contribute new algorithms.

This paper is organized as follows. Section 2 reviews related works. Section 3 introduces and describes the proposed BSNCloud architecture. Section 4 describes the data

aggregator side design and implementation with an Android App. Section 5 describes the cloud server, which includes a web server, machine learning engine, and MySQL database. Section 6 presents design decisions, implementation details and performance evaluation. Finally, Sect. 7 summarizes the work with future works.

2 Related Works

Among many potential communication technologies, Zigbee [7] and Bluetooth are most widely deployed. Zigbee is a very low power, collision avoidance protocol optimized for lower power sensors. It has developed a health care specific protocol and is compliant with all IEEE 11073 devices as well as most other IEEE 802.15.4 wireless devices. Bluetooth supports high-bandwidth and many existing devices with a health care compliant version defined. However, it has very high power requirements and uptime for the radios. Bluetooth Low Energy [8] is a new proposed system from Bluetooth for lower energy requirements, while being interoperable with Bluetooth Classic.

SPINE [9], a TinyOS based platform, has enabled the implementation of a heterogeneous body sensor network by abstracting the hardware level of multiple sensors such as TelosB and MicaZ and creating an easy to use software. Furthermore, SPINE contains APIs for general-purpose processing functions such as average, median and RMS. Separately, Shimmer Research [10] delivers two sets of software tools: development drivers for LabView, MATLAB, Android and C#; standalone software such as Shimmer Connect, Shimmer Log and Shimmer Plot for easy connection, storage and visualization of sensor data. While these drivers and software provide a good combination of tools for easy direct use or further API developments, the research community needs a scalable, robust instrument to meet their fast research and development needs.

One WBAN architecture that incorporates SPINE is DexterNet [11]. The system takes a real-time approach and seeks to provide an open-source platform that allows indoor and outdoor persistent human monitoring. The Body Sensor Layer, Personal Network Layer and Global Network Layer provide a three-tier architecture to address the different API needs from the sensors, data aggregator and server. This system is both scalable and reconfigurable in real-time due to the versatility of the layered approach. CodeBlue [12], proposed by Fulford-Jones and Malan, presents an ad hoc infrastructure for emergent medical care. In this project, several types of body sensors (e.g., pulse oximeter, ECG/EKG sensor) are individually connected to Zigbee enabled radio transmitters. Due to the ad hoc architecture and the capability of self-organizing, CodeBlue yields scalability for network expanding and flexibility to connect various wireless devices. Jiang and Cao proposed CareNet [13], an integrated wireless environment used for remote health care systems. CareNet offers features such as high reliability and performance, scalability, security and integration with web based portal systems.

Combining Cloud Computing and Wireless Sensor Networks (CC-WSN) [14] and Open Sensor Web Architecture (OSWA) [15] propose architectures that integrate wireless sensor networks (WSN) with the cloud. CC-WSN has base services including sensor data management run on Google AppEngine and Microsoft Azure, filter chain and filter management and user services including visualization and notification. OSWA provides an architecture for integrating sensor networks with various distributed computing platforms like SOA and Grid-Computing. BodyCloud [5] proposes a SaaS approach that

supports cloud-assisted BSN applications. The purpose of the software is to allow fast prototyping of cloud-assisted BSN applications and flexibility among architectural components utilizing web standards-based procedures and scalability aided by the Google App Engine PaaS infrastructure. Their paper discusses the issues of providing body-side and cloud-side context-aware sensing and adaptation, an issue our proposed architecture will address.

A survey of the architectures presented above reveals that several issues still need to be explored [16]. One such issue is providing an efficient collection of body sensor data and implementing an adaptive sensing mechanism. The data transmission throughout the network, including sensor to aggregator and aggregator to cloud, still has potential for further optimization. Contextual adaptation is another area of concern, relating to dynamic adaptation in an array of services depending on the current network and sensing context. Finally, a general software that assists programmers for developing specific healthcare related apps is worth further investigation.

To address data reliability between sensor and data aggregator, an analysis driven adaptive framework [17] was proposed. In this framework, a group data importance based scheme was designed to mitigate the effect of data loss on analysis algorithms. Then, a hybrid transmission protocol was developed to meet analysis needs. Furthermore, a multi-level importance ranking was introduced to allow a fast organization of sensor data into priority and non-priority packets. These packets are transmitted progressively in response to data reliability demands and network constraints. Finally, the concatenation of small data frames from sensors with low sample rates like ECG/EKG, along with an adaptive transmission scheme can provide loss resilience among the sensor network [18].

3 System Architecture

As shown in Fig. 1, our proposed system consists of three basic parts: the sensors, the data aggregator and the cloud server. In a typical scenario, a patient may wear a number of wireless sensors from blood pressure to accelerometer for continuous health monitoring. Depending on the kind of illness and the purpose (rehabilitation vs simple monitoring), some patients may also get scheduled for periodical measurement of specific signals at various frequency in doctor's office. During such scheduled measurement, real-time data analysis is often desired so that more than just visualized signal is displayed. While samples are being collected, a phone/tablet/PC, or the "data aggregator", should be connected with sensors wirelessly (via short range communication such as Zigbee, Bluetooth, or even the newly proposed standard WBAN) so that data can be processed, analyzed, and then streamed to a cloud server for backup. In this case, a software tool (ideally, a tablet App) plays a key role by connecting the sensors and the cloud server. The App should be able to control the data sampling, visualization, analysis, and streaming such that the system can work autonomously and continuously without human intervention.

Different from existing "cloud-assisted" body area network architectures, our proposed BSNCloud system is "cloud-centered", i.e., it fully utilizes the processing power and storage capacity for optimal efficiency and minimum cost. In our design, the cloud server consists of three key components: (i) a web server that keeps receiving data

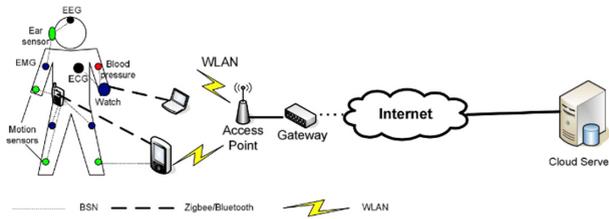


Fig. 1. System architecture of BSNCloud.

streaming and analysis requests from the data aggregator; (ii) a machine learning engine that automatically performs real-time or on-demand analysis as requested and responds with the results; and (iii) a MySQL database that stores various body sensor data sets, specific data analysis algorithms, as well as information about all data analysis sessions.

Essentially, BSNCloud aims at improving sensor-based event monitoring for remote patient care to a new level. Its scalable biomedical body area network computing platform has the potential of eliminating distance barriers and providing health care remotely. Another key aspect of the system is its open source, user friendly, and scalable cloud computing platform for sensor data storage, indexing, and analysis. It also allow researchers to upload, update, and contribute algorithms to the repository and conduct comparative study with the availability of a large number of high volume data sets.

4 Data Aggregator Design

Despite of efforts that enables sensor side programming on various tasks such as data sampling, event detection, data analysis, pre-processing, and data transmission, such approaches are many times limited to specific types of sensor motes. For example, the SPINE open-source software, built on TinyOS [19], provides a set of APIs that allow heterogeneous data collection, multiple forms of data analysis, threshold based event-detection and some forms of data reliability enhancement. However, the widely used TelOsb mote [20] is no longer produced. While Zigbee compatible sensors still exist, they lack enough capability on biomedical sensors. On the other hand, Shimmer Sensing [1] decided not to support TinyOS in their Shimmer 3 IMU motes and instead support Bluetooth only. There are two reasons: (i) there is a significant barrier for mobile devices to communicate with sensor motes in Zigbee while Bluetooth is readily available among all Android and iOS devices; (ii) it is much easier to control sensors at the data aggregator than coding the the sensor itself since vendors may choose different implementation languages such as nesC, C, and Java.

4.1 Data Aggregator Functions

In our system, we adopted Shimmer3 IMU development kit and combined sensor side functions and data aggregator side functions together. Core functions are implemented inside a comprehensive Android App, which is repackaged from the original Multi-Shimmer Template for Android. As shown in Fig. 2, specific data aggregator functions include:

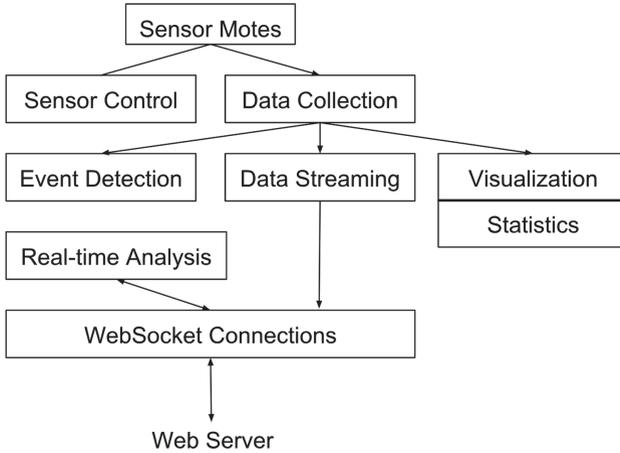


Fig. 2. Data aggregator modules.

- *Sensor Control* that includes turning sensors on/off for sampling as well as manually (through configuration) or automatically (through algorithm) manipulation of the frequency of data sampling. This allows the system to find the appropriate sampling rate for a specific application dynamically and sensors to be emulated for experimental purposes.
- *Data Presentation* that includes data visualization and signal statistics (e.g. min, max, average, standard deviation, time between peaks, binned distribution, etc.), which allows users to directly view the time series sensor data as well as key measures.
- *Event Detection* that handles lightweight analysis of data samples (e.g. threshold-based detection algorithms) to detect emergent events. It can be for both general and application specific functions. General functions will detect abnormal sample values based on configuration of sensors. Application specific functions will focus on identifying certain simple patterns indicating the occurrence of specific events. In the case when a critical event occurs, an alarm message is triggered for emergency and further real-time data analysis may be automatically initiated for diagnosis.
- *Data Analysis* that perform real-time on-demand analysis from users. It can also be triggered automatically when certain abnormal events are detected. In any case, a request is sent to the cloud server so that reasonably sophisticated, capable, and trained algorithms can be executed to analyze the N number of most recent or upcoming samples. This allows the data aggregator to utilize the processing power of the cloud server to run some complex machine learning algorithms for activity classification such as Fall Detection. A list of algorithms specific to each body sensor data can be available to the App user for selection. This makes the App scalable and eliminates the overhead of implementing a large set of analytics algorithms locally on the App.
- *Data Communication* that includes both sample streaming and real-time data analysis requests. For sensors with multiple signals (e.g., accelerometer), multiple socket threads should be used to transmit different signals of the same samples (e.g., X, Y, Z) simultaneously. Streaming can also be adapted based on battery power and network connectivity (i.e. offline or online with WiFi/3G/4G).

4.2 Data Streaming

Compared with traditional socket implementation that requires a server program stays running and waits for requests, WebSocket has several desirable features: (i) it's full-duplexed and enables streaming on top of TCP, which ensures the reliability of data streaming; (ii) it is supported by all browsers as well as programming languages such as Java, Javascript, and Python; (iii) its secure version is supported by Firefox 6, Safari 6, Google Chrome 14, Opera 12.10, and Internet Explorer 10. Essentially, WebSocket protocol offers a standard and universal implementation for interaction between a web browser (or client applications such as Android Apps) and a web server with lower overheads, facilitating real-time data transfer from and to the server.

In BSNCloud, a XAMPP Apache based PHP web server is setup and configured to receive requests from both Android App side real-time analysis and also Web Portal side on-demand analysis request. Whenever a user decides to stream and select the specific sensor data, a Java WebSocket connection is created for each signal (3 connections for accelerometers). Collected samples are temporarily stored locally in the data aggregator before delivered to the server. If the data aggregator goes offline, samples that have not been streamed yet will stay in local storage and will be streamed once Internet connection is restored. Each data streaming is uniquely indexed by, a stream_id, and then a sensor_id for each sensor, as well as the name of the specific signals (e.g., X/Y/Z for accelerometers).

4.3 Real-Time Data Analysis

Real-time analysis is a key feature of data aggregator. A user may follow a predefined schedule or decide based on certain observation. It is also possible that abnormality is identified by the event detector and therefore an analysis request is automatically initiated. In either case, a message is sent to the web server and then wait for the response. Frequent requests may make a separate WebSocket connection a necessity. Otherwise, existing WebSocket for data streaming can be used. Upon the receipt of the result, a message box is displayed in the App showing the result.

Due to its time sensitivity, real-time data analysis does not use a large data set and instead focus on either some most recent samples or upcoming samples. In addition, the size of such samples may affect the response time. While most recent samples can be analyzed immediately at the server side, upcoming samples will need time to accumulate and then get started.

5 Cloud Server Design

As the most important component of the proposed BSNCloud system, the cloud server aims at achieving large scale, distributed, and even parallel biomedical data management and analytics. In addition to the hardware resources that offer high performance computing power and large volume storage capacity, we designed four key modules for web server:

- *Data Repository* that stores, indexes, and manages body sensor data storage. MySQL is used to manage all data, which consists of both data sets and data analysis sessions. All raw data are stored as files locally in the hard drive with file location stored in the database tables. Data are classified as “Healthy”, “Patient”, and “Synthesis” to indicate their source. Data sets can be streamed from the Android App, uploaded data from lab/hospital or online repositories [21].
- *Algorithm Repository* that includes both contributed and user uploaded algorithms. These algorithms can be public and therefore open to all users or private and only for the owner to modify and update. Private algorithms can be contributed and become public once tested and approved. Sample algorithms include distributed clustering algorithms (e.g. K-means algorithm), parallel EM algorithms, parallel SVM algorithms, distributed random forest algorithms and distributed deep learning algorithms (e.g. Deep Convolutional Network, Auto Encoders, Restricted Boltzmann Machines). Based on the specific application, there may be multiple algorithms corresponding to the same body sensor data. To facilitate the contribution of algorithms, an online domain specific language (DSL) based template can be designed for uploading algorithm modules and test of algorithms. A biomedical DSL with uniform language interface [22] will support multiple data sources such as time series data from accelerometers, image data from vision sensor networks, as well as different types of biomedical body sensors.
- *Machine Learning Engine* that is able to execute various algorithms over selected data in the repository. The machine learning engine works closely with data repository and algorithm repository. It pre-processes raw data so that it fits specific input requirements by data analysis algorithm. For example, when a request is sent to analyze accelerometer data, the machine learning engine combines the raw data of the three signals X, Y, and Z and combine it to a CSV file, before feeding it to a specific algorithm. It should also incorporate major implementation environments such as Python, Java, C ++, and Matlab. Integration with existing frameworks such as TensorFlow is another possibility. The engine may also draw from MPI [23] for computational-intensive algorithms and the MapReduce model [24] for data-intensive algorithms.
- *Web Portal* that offers account management, data management, data analysis, and algorithm development. The web portal provides a software as a service (SAAS) level of user experience so that minimum overhead is achieved. Through the web portal, a user can view data sets by sensor types or algorithms, manage data analysis sessions, upload data sets, upload/contribute analysis algorithms, and perform new analysis.

The modules presented in Fig. 3 provide the basis for the cloud server’s flexible and scalable design. Data received from the data aggregator will be formatted and stored on the server. Through the web portal, a user can request real-time analysis as data is being received, triggering result messages as desired. Furthermore, users will have the opportunity to upload algorithms or select algorithms from the expansive existing libraries and make analysis requests. The results will then be displayed in the web portal and the user can retrieve the data in a standard format. In addition, a user can request raw data from the cloud server. This provides an opportunity not only for researchers to analyze multiple data sets at once, but also for professionals to make real-time judgements based on patient data.

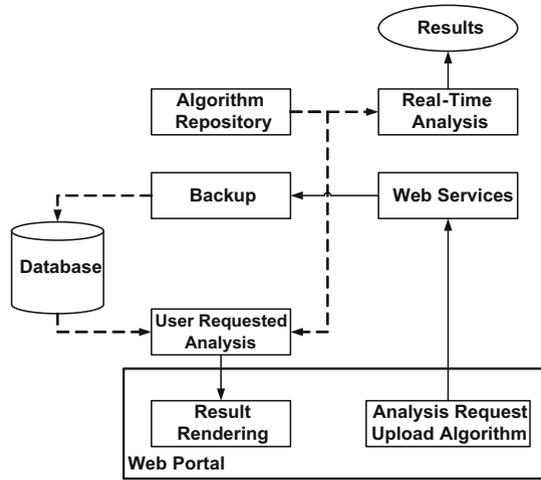


Fig. 3. Cloud server modules.

6 Implementation and Performance Evaluation

We have completed the prototype implementation of the proposed BSNCloud system and performed preliminary study on its performance. has acquired a 5-server cloud cluster with one server being the master and four others as slaves. Each server has 2x EightCore Intel® Xeon® Processor E52640 v2 2.00 GHz 20 MB Cache (95 W) and 8 × 4 GB PC314900 1866 MHz DDR3 ECC Registered DIMM. The total storage space is over 40 TB. Servers run Ubuntu and OpenStack systems. XAMPP-Apache PHP web server is setup for receiving both Android App and Web Portal.

Shimmer3 IMU mote is used for streaming. The mote has the dimension of 51 mm × 34 mm × 14 mm, runs 24 MHz MSP430 CPU with very low power consumption, light weight, and is wearable at different part of human body with a short/long strap. 5 colored LEDs are used to indicate the device status and operating mode. The data is highly accurate and scientifically reliable with a SD card of 7 GB capacity. It offers integrated 9DoF inertial sensing via accelerometer, gyroscope, magnetic and pressure sensors. The supported communication standard is Bluetooth.

A fall detection algorithm is implemented in Python and Java for experimenting the machine learning engine, real-time data analysis, and on-demand web portal data analysis. While multiple sensor data can be used, accelerometer data is the primary focus in this study. The algorithm is firstly trained with labeled data and then read csv files combined from the X, Y, Z raw data stored in the cloud server.

6.1 Android App

The developed Android App supports synchronized login with web portal. Figure 4 illustrates the sensor configuration as well as the main page with multiple functionalities such as event detection, set streaming on/off, enable statistics, perform real-time analysis, etc.

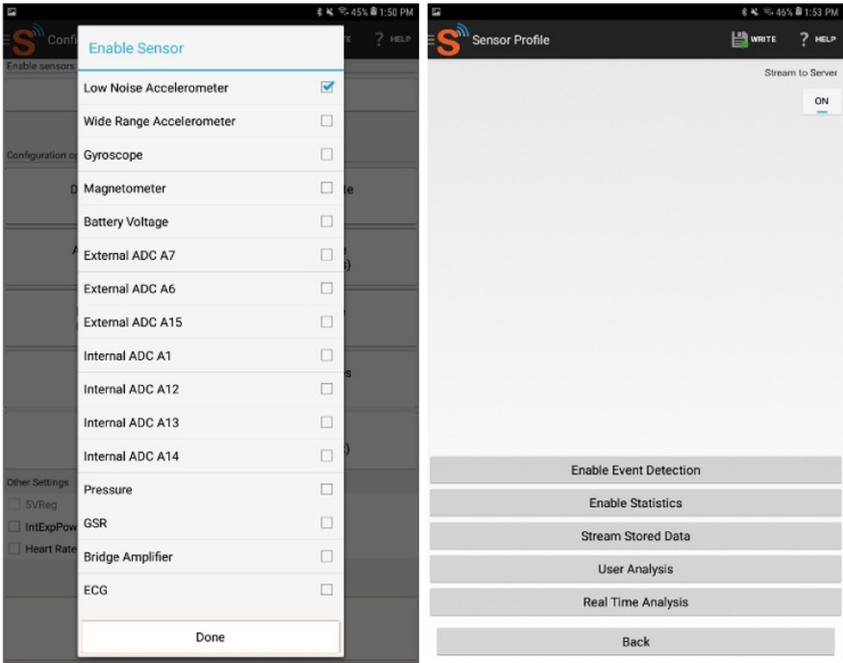


Fig. 4. Android App main page and sensor configuration.

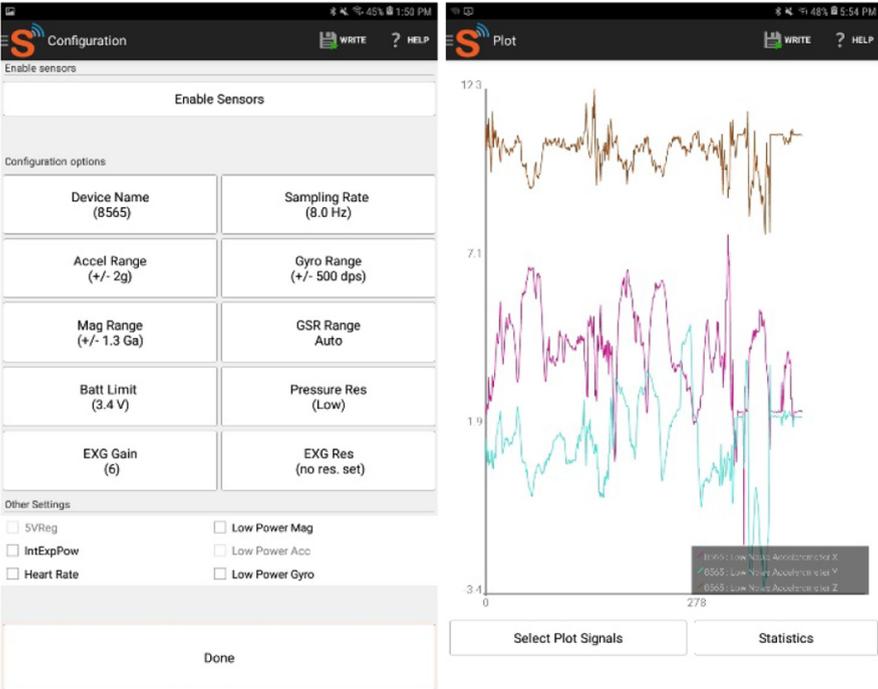


Fig. 5. Android App sensor configuration and data visualization.

Figure 5 illustrates the sensor configuration and data visualization that is provided by the MultiShimmer Template. However, MultiShimmer does not provide functions other than sensor control, data collection, and visualization. We have repackaged the app to include many additional functions, especially on streaming multiple body sensor signals to the cloud server side.

Figure 6 illustrates the real-time data analysis. A user may select one sensor signal, sends a request to the server, and then wait for the response. Please be noted that for accelerometers, choose one signal X also requests Y, Z signals for analysis together.

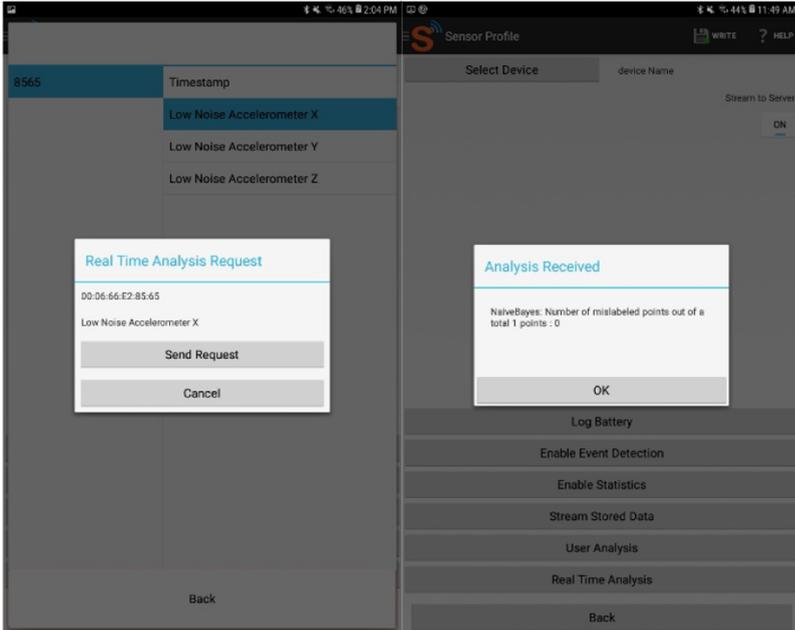


Fig. 6. Android App real-time analysis.

6.2 Web Portal

Figure 7 shows the dashboard page of the web portal. A user enters the system and select data and algorithms for data analysis. First, a specific sensor is chosen, then followed by a specific algorithm, and finally one data set. Each step of selection helps filter non-related algorithms or data sets. Algorithms are labeled by both method and application (e.g., fall detection). Before an analysis is initiated, a summary is presented for verification. The analysis results are displayed in the textbox at the bottom.

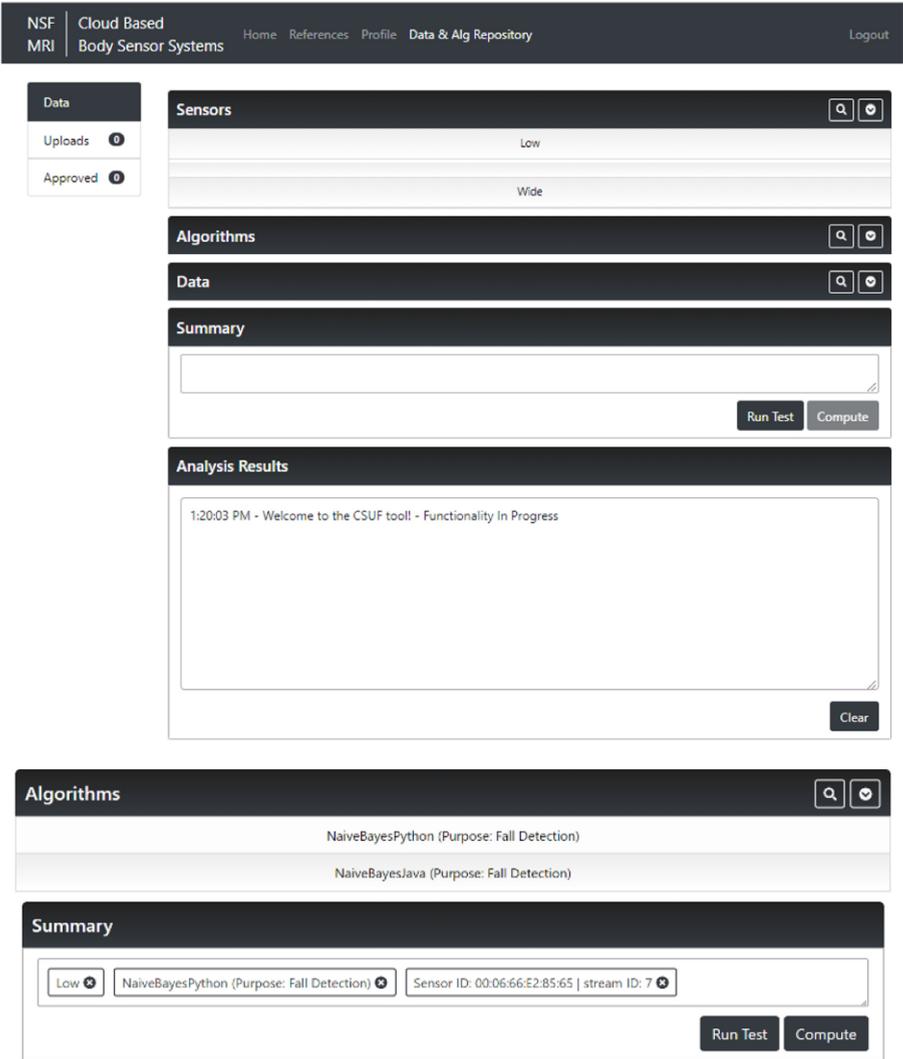


Fig. 7. Web Portal dashboard and data analysis pages.

6.3 Experimental Results

We have also conducted experiments to evaluate some key performance measure for the system. Figure 8 shows the measured throughput of a 8 Hz sampling rate accelerometer data streaming with 3 parallel WebSocket connections by varying the number of samples. It can be seen that the throughput is quite consistent over time for different sample sizes. It should be note that this throughput is not equivalent to the network capacity since it takes time to collect samples.

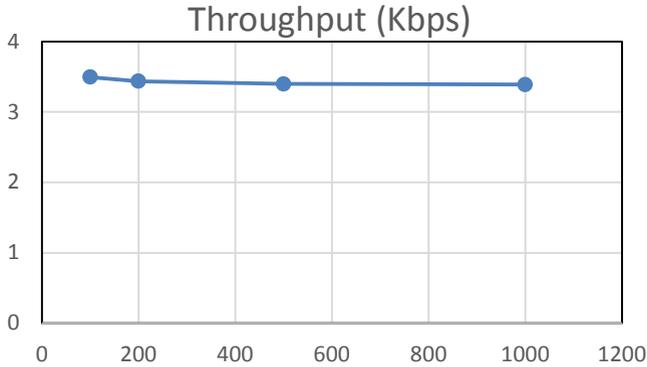


Fig. 8. Data streaming throughput vs sample sizes

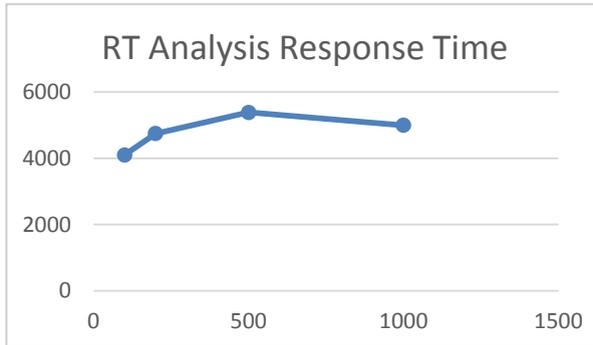


Fig. 9. Real-time data analysis response time

Figure 9 shows the real-time analysis response time in ms. It can be seen the results are sent back from the cloud server within a very short period of time. We then increases the sampling rate from 8 Hz to 16 Hz and 51.2 Hz, results show that the response time increases to over one min. Further analysis reveals that since the real-time analysis request is actually sent over an existing streaming WebSocket, which is already in near saturation status under high sampling rates. We will experiment with a separate WebSocket connection for real-time analysis and/or prioritize the message so that it is sent before queued sample packets.

Figure 10 shows the web portal data analysis response time for accelerometer samples collected for 1, 5, 10, 20, 30, and 60 min, respectively. It can be seen that results come back within 1–2 min. For reference, the data size of 60 min correspond to 2.6 Mbytes. Therefore, the high performance of the cloud server is clearly validated.

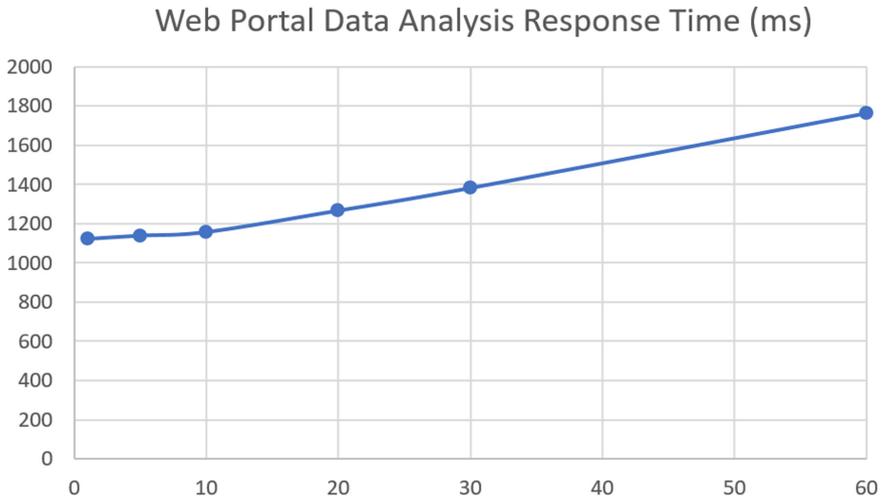


Fig. 10. Web portal data analysis response time for accelerometer samples.

7 Summary

In this paper, we have presented the design, implementation, and performance evaluation of BSNCloud, a cloud centered system for biomedical body sensor data collection, analysis, streaming, and storage. Key design decisions were made to maximize usability, efficiency, and scalability. The system helps facilitate the data analysis research and significantly reduces the cost and overhead by researchers. It is expected to grow with active use and contribution from the research community. In the future, we will address the issue of reducing the real-time data analysis response time under high sampling rate. We will also include more analysis algorithms with the proposed DSL based algorithm contribution template.

Acknowledgments. This work was supported by the National Science Foundation, CNS division (Award No. 1626586). We also would like to thank Rittika Shamsuddin and Barbara Mukami Maweu at University of Texas at Dallas for providing the Naïve Bayes fall detection algorithm.

References

1. Shimmer Enabling Softwares, Available <http://www.shimmersensing.com/research-and-education/applications/data-aquisition-software-sensor-systems/>
2. Qualcomm Wearables, Available <https://www.qualcomm.com/products/wearables/>
3. Intel IoT Developer Kit, Available <https://software.intel.com/en-us/iot/devkit>
4. Samsung Architecture for Multimodal Interactions: An open data platform for innovation, Available http://www.samsung.com/us/globalinnovation/innovation_areas/#digital-health
5. Fortino, G., et al.: BodyCloud: a SaaS approach for community body sensor networks. *Future Gener. Comput. Syst.* **35**, 62–79 (2014)
6. WebSocket. Available <http://www.websocket.org/>

7. Baronti, P., et al.: Wireless sensor networks: a survey on the state of the art and the 802.15.4 and ZigBee standards. *Comput. Commun.* **30**(7), 1655–1695 (2007)
8. Bluetooth Low Energy Core Specification Version 4.0. Available http://www.bluetooth.com/English/Technology/Works/Pages/Bluetooth_low_energy_technology.aspx
9. SPINE Project. Available <http://spine.deis.unical.it/>
10. Shimmer Sensing. Available <http://www.shimmersensing.com/> (2013)
11. Kuryloski, P., et al.: DexterNet: an open platform for heterogeneous body sensor networks and its applications. In: 2009 Sixth International Workshop on Wearable and Implantable Body Sensor Networks (2009)
12. Fulford-Jones, T., et al.: CodeBlue: an ad hoc sensor network infrastructure for emergency medical care. In: Proceedings of International Workshop on Body Sensor Networks (2004)
13. Jiang, S., et al.: CareNet: an integrated wireless sensor networking environment for remote healthcare. In: Proceedings of the ICST 3rd international conference on Body area networks, ICST, Tempe, Arizona (2008)
14. Kurschl, W., Beer, W.: Combining cloud computing and wireless sensor networks. In: Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services, p. 512–518. ACM, Kuala Lumpur, Malaysia (2009)
15. Chu, X., Buyya, R.: Service Oriented Sensor Web, in *Sensor Networks and Configuration: Fundamentals, Standards, Platforms, and Applications*, N.P. Mahalik, Editor. 2007, Springer Berlin Heidelberg: Berlin, Heidelberg, p. 51–74
16. Fortino, G., Di Fatta, G., Pathan, M., Vasilakos, A.V.: Cloud-assisted body area networks: state-of-the-art and future challenges. *Wirel. Networks* **20**(7), 1925–1938 (2014). <https://doi.org/10.1007/s11276-014-0714-1>
17. Li, M., Cao, Y., Prabhakaran, B.: Multi-level sample importance ranking based progressive transmission strategy for time series body sensor data. In: 2015 IEEE 16th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM) (2015)
18. Read, N., et al.: Loss resilient strategy in body sensor networks. In: Proceedings of 2011 ACM/IEEE International Conference on Body Area Networks (BodyNets 2011, accepted), Beijing, China (2011)
19. Levis, P., et al.: TinyOS: an operating system for sensor networks. In: Weber, W., Rabaey, J.M., Aarts, E. (eds.) *Ambient Intelligence*, pp. 115–148. Springer, Berlin, Heidelberg (2005). https://doi.org/10.1007/3-540-27139-2_7
20. TelosB sensor data sheet. DOI=http://www.willow.co.uk/TelosB_Datasheet.pdf
21. Repository, UC Irvine Machine Learning Laboratory. Available <https://archive.ics.uci.edu/ml>
22. Huang, X., Zhao, T., Cao, Y.: PIR: a domain specific language for multimedia retrieval. In: Proceedings of IEEE International Symposium on Multimedia (ISM 2013), Anaheim, California, USA (2013)
23. Forum, M.P.I.: MPI: A Message-Passing Interface Standard - Version 2.2 - Sep.4. 2009, Message Passing Interface Forum (2009)
24. Dean, J., Ghemawat, S.: MapReduce: simplified data processing on large clusters. In: Proceedings of the 6th conference on Symposium on Operating Systems Design & Implementation, San Francisco, California, USA (2004)

Secure Communication Networks for Smart-Health



Model-Based Analysis of Secure and Patient-Dependent Pacemaker Monitoring System

Leonidas Tsiopoulos^{1,2} , Alar Kuusik¹ , Jüri Vain¹ ,
and Hayretdin Bahsi¹ 

¹ Tallinn University of Technology, Ehitajate tee 5, 19086 Tallinn, Estonia
{leonidas.tsiopoulos, alar.kuusik, juri.vain, hayretdin.bahsi}@taltech.ee
² Åbo Akademi University, Tuomiokirkontori 3, 20500 Turku, Finland

Abstract. Pacemakers' safety, security and reliability are of utmost importance for patient's life quality in various daily situations. An integral characteristic of the pacemaker that depends on all of these attributes is its lifetime. In current medical practice the pacemaker's expected lifetime is estimated relying on manufacturer's data sheet and expert knowledge that may result in quite rough approximations if patient's specifics are not taken into account. In this paper we perform a model-based quantitative analysis of pacemaker lifetime that takes into account patient specific factors, including general health condition, acting environment, remote reporting and others. We demonstrate that including these factors in analysis can provide drastically different results compared to that of average approximating estimates.

Keywords: Cardiac implanted electronic devices · Pacemaker · UPPAAL timed automata.

1 Introduction

The embrace of sensors and advances in communication and cloud technology has enabled us to develop effective monitoring applications in different areas including the health sector. Implantable medical devices such as cardiac implanted electronic devices (CIEDs), insulin pumps and gastric stimulators can continuously track a patient and transfer the data to the medical institution regardless of the location. CIEDs, namely permanent pacemakers (PPM), implantable cardioverter-defibrillators (ICDs) and cardiac resynchronization therapy (CRT) devices exceed 500 units per million inhabitants worldwide [21] and their number is increasing in correlation with the expected lifetime and national wealth. In well-developed countries the average prevalence is over 1000 implants today and reaches 10000 devices per 1M elderly people (ibid). For such remarkable deployment rate the reliability and operational lifetime of devices is an important question to prevent fatalities. According to patient support organizations the complete

failure of a modern pacemaker is rare [11] and most frequently devices need just timely reprogramming for proper operation [1]. Acquiring timely information of patient condition changes and the discovery of possible mismatch of device settings requires periodic device operation analysis to discover irregularities in the heart behavior. Regardless the maintenance related usability concerns it is important to underline that remote monitoring of pacemakers and other CIEDs has improved the patient survival rate [25] and not only. The accurate predicting of device maintenance dates not only prevents malfunctioning in conditions when the monitoring is urgently needed but also improves the user trust and convenience of wearing the devices without compromising patient's life quality.

Permanent pacemaker communication is required for device adjustments according to patient's physical condition and its deviations over the time. Due to technological limitations, mainly because of the high energy consumption of legacy wireless communication, frequent data transmissions of pacemakers were not possible till recent times. Inductive link coupling used for legacy pacemakers made data exchange - device programming and event report offloading possible only at hospital environments. Therefore, not much attention has been paid to pacemaker communication security and energy efficiency measures. Bluetooth connectivity for PPMs was proposed already in 2002 by Hsu and Avudaiappan [14]. However, Bluetooth Classic technology available at this time was not suitable for long duration battery operation. The energy efficient Bluetooth Low Energy (BLE) connected pacemakers were actually introduced to the market in 2016 and are considered simple to use and helpful for both sides, patients and cardiologists [24]. Today, there is a selection of modern PPM and ICD devices from companies Medtronic and Abbott that provide BLE connectivity with an acceptable reduction of the device operation lifetime. However, the impact of wireless communication to device energy consumption is still remarkable; an active modern Bluetooth transmitter consumes around one milliamper current while modern microcontrollers in pacemakers can operate with few microamps per pacing in average. Therefore it is important to optimize wireless communication frequency and data amount to be transmitted heavily depending on patient's clinical needs. It is also essential to pay attention to personal data protection and protection against external operational vulnerabilities to CIEDs as life supporting medical devices.

Deployment of advanced wireless communication made possible periodic transmission of heart data to cardiologist which significantly improved the patient monitoring quality but it also added security risks and reduced device battery lifetime. As argued in [20] wireless interfaces of such devices are vulnerable to cyber attacks with even life-threatening impact. Various passive and active attacks targeting the communication medium are possible [17]. The upgrade mechanisms of the PPMs and the vulnerabilities of the cloud server could be also used for compromising the devices remotely [2]. Moreover, pacemakers of St. Jude Medical were found to have insufficient cyberprotection measures that made possible battery drain, firmware downgrade and malfunctioning attacks over the wireless interface [16]. As various nodes such as home routers, mobile

phones, enterprise gateways, would take part in the communication, the range of security threats gets very wide. Therefore, it is essential to establish end-to-end security that covers the whole communication medium encompassing the device, cloud system, and all intermediate nodes. On the other side, the security countermeasures create an additional burden to resource-constrained devices with the additional energy consumption affecting significantly the device lifetime. From the clinical point of view, requirements to communication frequency and payload depend on patient's profile (health condition, daily activities, environment etc.) that in combination with the security countermeasures on the communication can result in largely varying expectations to device lifetime.

To reduce threats and possibly infeasible design decisions (that ignore the factors coming from different use cases) a model-based formal analysis can be used to justify design and runtime configuration decisions. The formal analysis in this paper is targeted to taking into account the effects of pacemaker remote monitoring and secured dynamic episode alerting to the device lifetime, depending on different patient profiles and security countermeasures adjusted to daily living environment. We apply probabilistic modeling, simulation experiments and model-checking with UPPAAL Timed Automata to estimate the impact of various factors to pacemaker expected lifetime. In particular, we focus on factors such as averaged patient-dependent condition episode alerting frequency and the cost of secure communication (in terms of energy consumption) in three possible patient environments during a day. The impact of listed factors is characterized with numerical data extracted from model experiment that justifies the need for refined model-based analysis of pacemaker operational lifetime. The choice of UPPAAL Timed Automata as the formalism for our work is justified by the facts that the checking of behavioral, stochastic and real-time aspects of systems is efficient and the UPPAAL tool family includes several extensions of the standard modeling and verification tool which will allow us in the future to address different development aspects of pacemakers.

2 Related Work

In the context of wireless Body Area Sensor Networks (BASN) extensive research has been done with application of formal methods. Ahmed et al. [6] used higher-order-logic theorem proving to formally analyse energy consumption of BASNs by verifying the mathematical relations for energy, delay and distortion of a given BASN. The results of this analysis can then be used to determine the parameters of optimal energy consumption of BASN algorithms. In a closely related work, Dai et al. [12] proposed synthesizing power management strategies for wireless sensor network nodes with UPPAAL STRATEGO by taking into account the various power states of a node device during the runtime in order to achieve a tradeoff between power consumption and performance.

Regarding verification of wireless body area network protocols, several works exist in the literature. For example, timed automata have been used in [8] to verify a recently proposed Medium Access Control protocol called STDMA (Statistical frame based TDMA protocol). Taking also security into account, Chen

et al. [10] proposed a formal modeling and verification method using the PAT model-checker for wireless BASN-specific authentication security protocols.

However quite few results are available on the verification of pacemaker operation correctness using model-checking tools. An operational model verification with UPPAAL tool is conducted by Pajic et al. [19]. As stated in the Introduction, pacemakers were found to have insufficient cyberprotection measures [2, 16]. This triggered research work to present solutions for protection against cyber attacks under the assumption the communication has been already compromised. For example, Rao et al. [22] presented a model for a multi-modal design approach for risk assessment of pacemaker devices and they propose an adaptive remediation scheme to mitigate security threats. The approach is integrated into the hardware-software development with a middleware for dynamic switching between the modes based on risk values assigned to the different functions each mode has.

The analysis of the recent related work, with some of the example papers discussed above, shows the lack of satisfying solution in the pacemaker literature and motivates the authors for the work presented in this paper. To the best of authors' knowledge, a formal model-based analysis on the effects of pacemaker remote monitoring and secured dynamic episode alerting to the device lifetime, taking into account different patient profiles regarding heart condition and security measures adjusted to daily living environment, has not been presented yet.

3 Pacemaker Background

In general, an implantable pacemaker monitors and regulates the patient's heart rate continuously by providing single or dual chamber rate-responsive corrective bradycardia and atrial tachyarrhythmia electrical pacing. As a reference device of current study we use Medtronic dual chamber permanent pacemaker model Azure XT DR W1DR01 [4].

3.1 Pacing

Dual chamber and single chamber pacing modes address different cardiac conditions. Dual chamber pacing restores AV synchrony by sensing and stimulating two chambers of the heart, the right atrium and right ventricle. Single chamber pacing supports patients with infrequent asystole or patients with chronic AT/AF and for whom dual chamber pacing is not justified [4]. For some of the dual chamber pacing modes pacing occurs at the programmed lower rate and for some modes the pacing is occurring at the sensor rate. Rate-responsive pacing adapts the pacing rate to changes in patients' physical activity due to some patients exhibiting heart rates that do not adapt to changes in their physical activity. The device uses an activity sensor to measure the patient's movement and to determine the appropriate pacing rate.

3.2 Monitoring and Alerting

Data collected by the pacemaker is encrypted and sent to the CareLink network through the MyCareLink Heart mobile app [3], providing clinicians with alerts on clinically-relevant patient events. The app also makes selected pacemaker data, such as transmission success history, pacemaker battery information and updates on physical activity, easily accessible to patients. If a clinical or system performance event occurs and Medtronic CareAlert Monitoring is programmed to respond with an alert, the device automatically attempts to establish wireless communication with the mobile phone. After communication is established, the mobile phone receives the alert data from the device, and then transmits the alert data to the CareLink Network wherefrom the data is accessible to clinicians.

Examples of clinically-relevant alerts are “*Average Ventricular Rate During AT/AF larger than Threshold*” indicating that the average ventricular rate during a selectable duration of AT/AF exceeds the programmed threshold, and “*Monitored VT Episode Detected*” indicating that one or more monitored VT episodes were detected.

The clinician may also configure the device to send periodic reports with the frequency depending on the patient. The information included in such reports is, e.g., “episode data and EGM storage”, providing an arrhythmia episode log that enables to view the summary and detailed diagnostic data, including stored EGM, for the selected arrhythmia episode, and “rate drop response episodes data” displaying beat-to-beat data that is useful in analyzing Rate Drop Response episodes and the events leading up to these episodes.

3.3 Pacemaker Security Countermeasures

We assume that the communication between pacemaker and cloud is secured with end-to-end mechanisms. From the connectivity perspectives, device performs internal data encryption and communicates with smartphone or tablet PC gateway over BLE communication using Medtronic BlueSync technology. Gateway device runs Medtronic MyCareLink Heart application that acts as a pass-through element for the encrypted pacemaker data.

In a typical scenario, three main security tasks should be performed in a system: (1) relevant cryptographic keys should be generated, (2) each device should be authenticated and authorized, (3) based on the generated keys, secure end-to-end communication should be established. In typical end-to-end communication, a gateway acts as an intermediary node between sensor nodes and the cloud. Gateways obtain the sensor data with short-range local area network protocols (i.e., wireless or wired) and relay them to the cloud server over wide area network infrastructures. In addition to relaying responsibility, these gateways could take over some cryptographic operations to relieve the sensors from resource-intensive tasks [13, 15, 18]. However, this functionality comes with a compromise in the end-to-end security property, requiring more trust to the gateway.

In most cases, these gateways do not have resource restrictions like sensors have but may be subject to different threats or security controls. In stationary locations, a patient can use the existing network infrastructures, thus, a pacemaker can benefit from hospital gateway or home router. Hospital gateways could be more trustable as they could be secured by IT staff of the hospitals, and we can assume that more physical security is guaranteed. This can eliminate the cyber threats requiring physical proximity to the target. On the other side, an ordinary user generally relies on the default configuration of the home routers and ignores the security hardening. Thus, it is reasonable to delegate some cryptographic tasks to the hospital gateways but not to home routers.

A mobile phone could be a perfect fit as a gateway while the patient visits non-stationary places. However, it is not trustable for sensitive tasks as mobile malware poses an important threat to the secure communication passing over the mobile phone. These devices are prone to physical loss or theft. Most of the public places may enable the attackers to come close to the target and conduct man-in-the-middle (MitM) or denial of service (DoS) attacks. Comparing with the home environment, despite the existence of similar threats in home environment, the likelihood of threats could be considered higher for communication over the mobile phone in public places. Therefore, we assume that a stronger security configuration should be enabled in this option and any compromise in end-to-end security is not tolerable.

In resource-constraint scenarios such as our pacemaker remote monitoring case, security mechanisms are required to be use-case adaptable and lightweight, meaning that they should operate with less storage, computing, and energy resources. The current study does not follow proprietary BlueSync implementation that, according to our knowledge, supports a single security model for the pacemaker data communications. The strength of lightweight security mechanisms is that they can be deployed according to the patient environments with varying threat profiles. As we assume in this study the patient profile is determined by being in different locations such as hospital, home, and other places during a day, we selected a distinct security configuration setting for each location. Hospital and home are more stationary ones whereas the third option covers all other places a user can visit or stay at (e.g., shops, public transport, etc.).

We assume that the pacemaker application uses Datagram Transport Layer Security (DTLS) which can be considered a UDP-based alternative of transport layer security (TLS) protocol [23]. A DTLS session starts with a handshake for authenticating the parties and exchanging the session keys. The parties use the agreed session key to perform secure communication. However, in the case of using a certificate, large messages should be fragmented into various packets exchanged between parties, ending up with huge energy consumption for the handshaking phase. Therefore, delegation schemes in which gateways conduct handshake operation on behalf of sensors are proposed in the literature [13, 15, 18].

We consider the security configuration options given in the benchmarked study, [18], as the main baseline. The results in this study are very relevant for our cases as it provides a detailed performance analysis of a health-

care IoT system and proposes a solution that uses delegation idea (i.e., namely DTLS session resumption) which is compared with other configuration options. We assume that in a hospital setting, DTLS session resumption using elliptic curve operations for certificates and transferring handshake responsibility to the gateway is enabled. In home environment, symmetric key-based DTLS (DTLS_PSK_WITH_AES_128_CCM_8) is utilized. The pacemaker is assumed to initiate a certificate-based handshake using Elliptic Curve Digital Signature Algorithm in non-stationary places (i.e., security option is DTLS_ECDH_ECDSA_WITH_AES_128_CCM_SHA_256) which has a higher cost but provides more assurance about the authentication of the pacemaker. We omit the details about the key generation algorithms for simplicity.

In our model, as the frequency of the data transmission is so low, we consider that a handshake happens for each transmission of monitoring data or alert. As the data sizes are also low, one or two messages are enough after agreeing on the session key, considering the available payload sizes after excluding all headers. Therefore, we assume that energy consumption is heavily determined by handshake operation and the sensor side energy consumption values given in [18] are considered as the baseline for the calculation and correlation of energy costs of each security configuration for the model (i.e., DTLS session resumption costs 1/6 less than Symmetric key-based DTLS and certificate-based DTLS costs 32 times more than Symmetric key-based DTLS).

4 UPPAAL Probabilistic Timed Automata

UPPAAL Timed Automata (UTA) [7] address the behavioral and timing aspects of systems providing efficient data structures and algorithms for their representation and analysis through simulation and model checking.

An UTA is given as the tuple $(L, E, V, CL, Init, Inv, T_L)$, where L is a finite set of locations, E is the set of edges defined by $E \subseteq L \times G(CL, V) \times Sync \times Act \times L$, where $G(CL, V)$ is the set of constraints in guards, $Sync$ is a set of synchronization actions over channels and Act is a set of sequences of assignment actions with integer and boolean expressions as well as with clock resets. V denotes the set of integer and boolean variables. CL denotes the set of real-valued clocks ($CL \cap V = \emptyset$). $Init \subseteq Act$ is a set of assignments that assigns the initial values to variables and clocks. $Inv : L \rightarrow I(CL, V)$ is a function that assigns an invariant to each location and $I(CL, V)$ is the set of invariants over clocks CL and variables V . $T_L : L \rightarrow \{ordinary, urgent, committed\}$ is the function that assigns the type to each location of the automaton.

UPPAAL Probabilistic Timed Automata (UPTA) [9] is a stochastic and statistical modeling extension of UTA. UPTA preserves the standard UTA constructs such as integer variables, data structures and user-defined C-like functions. Additionally, UPTA support branching edges where weights can be added to define a probability distribution on discrete transitions. The weights may be general expressions that depend on the states and not just simple constants. For the work in this paper we use the branching edges with probability weights.

The requirement specification language (in short, query language) of UTA, used to specify properties to be model checked, is a subset of Timed Computation Tree Logic (TCTL) [7]. The query language consists of path formulae and state formulae. State formulae describe properties that can be interpreted in individual states, whereas path formulae quantify over paths or traces of the model and can be classified into reachability, safety and liveness [7]. For this paper we consider safety properties that are specified with path formula $A\Box\phi$ stating that state formula ϕ should be true in all reachable states.

5 Pacemaker Monitoring and Dynamic Alert Model

For the objective of this paper the model concerns only the continuous monitoring and possible episode alerting from the pacemaker in a probabilistic manner depending on three different patient profiles listed as Type 0, Type 1 and Type 2.

5.1 Patient Profiles

The “best” case patient profile (Type 0) is away from home (traveling to work and back, being at work and any possible shopping) 10 h per day. Hence, he is at home 14 h per day. He visits the hospital once per year and this visit is 2 h long. Regarding dynamic alerts, he has 1 alert per year in any of the three possible environments.

The “medium” case patient profile (Type 1) is away from home 6 h per day. He visits the hospital 2 times per year and each visit is 4 h long. Regarding dynamic alerts, he has 4 alerts per year in any of the three possible environments.

The “worst” case patient profile (Type 2) is away from home 1 h per day. He visits the hospital 4 times per year and each visit is 2 d (48 h) long. Regarding dynamic alerts, he has 10 alerts per day in any of the three possible environments.

5.2 Modeling Alerting Cost, Parameters and Constants

The device battery characteristics are as follows. The mean usable capacity is 1.2 Ah. Depending on the device setting/programming per patient needs, the lifetime of the device varies from 7.4 to 15.8 years (see Tables 4, 5, 8 in technical manual [5]). For our model let us choose the value 13.7 years/5000 d for the lifetime of the device regardless the patient profile. Note that all numbers for the model can be changed accordingly to different patient profiles and concrete living scenarios. One of the main assumptions is that daily reporting is active for all patients. This reduces lifetime of the device battery by 14.4% which translates to 564 d less lifetime. After applying simple calculations in order to have reasonable numbers for the model parameters to work on a day basis for the lifetime reduction, we know that 189 daily reporting sessions are needed to reduce battery lifetime by 1 day.

The dynamic alerts have a different cost in terms of battery lifetime depending on the environment the patient is in when the alert is triggered. This is due to different security protocols applied for communication in different environments. An additional assumption is that the alert triggered when at home costs the same as sending the daily report meaning that the security countermeasure cost is the same for both. Then, we assume that the cost of sending an alert from hospital is 1/6 less than the cost of an alert from home and the cost of the alert while traveling is 32 times higher than the cost of the alert from home, taking into account recommendations in Sect. 3.3. Thus, 220 hospital alerting sessions and 6 travel alerting sessions are needed in order to reduce battery lifetime by 1 day, respectively.

Figure 1 shows the model parameters, variables and constants. P is the patient type index with domain $\{0, 1, 2\}$. N is the number of patient profile types, 0 for “best”, 1 for “medium” and 2 for “worst”. M is the normalization to 1000 of the probability weight in order to accommodate all model probability weights into approximating integer scale from 0 to 1000. On this scale Ph denotes the probability weight of being at home during any hour around the clock. Similarly Pt is the probability weight of traveling and Pv is the probability weight of visiting the hospital during any hour around the clock. Pr is the probability weight of daily reporting and it is the same for all patients. Pa is the probability weight of alerting at any hour. Alerting may occur in any of the three different environments. Variable env is for the environment the patient is currently located in and it can be assigned symbolic values *home*, *travel* and *hospital* for the environment options.

<pre> const int P = 0; const int N = 3; const int M = 1000; const int [0,M] Ph[N] = {583, 748, 938}; const int [0,M] Pt[N] = {417, 250, 41}; const int [0,M] Pv[N] = {0, 2, 21}; const int [0,M] Pr[N] = {42, 42, 42}; const int [0,M] Pa[N] = {1, 4, 417}; int env = 0; const int home = 0; const int travel = 1; const int hospital = 2; </pre>	<pre> int lifetime = 5000; const int deltaLTN = 1; const int deltaLTH = 1; const int deltaLTT = 1; const int deltaLTO = 1; int crD; const int cD = 189; int crO, crT, crH; const int cO = 220, cT = 6, cH = 189; </pre>
---	---

Fig. 1. Pacemaker UPPAAL model parameters, variables and constants.

Variable *lifetime* expresses the battery lifetime estimate in days which is calculated iteratively in every hour (model time) depending on different environment and communication events that can occur with probability weights

specified as described above. The initial value of battery lifetime estimate (without any battery use) is 5000 d. $\mathit{deltaLTN}$ is a constant showing the decrease of battery lifetime by 1 day after 189 daily reporting sessions. $\mathit{deltaLTH}$ is a constant for the decrease of battery lifetime by 1 day after 189 home dynamic alerting sessions. $\mathit{deltaLTT}$ is a constant for the decrease of battery lifetime by 1 day after 6 dynamic alerting sessions during travel. $\mathit{deltaLTO}$ is a constant for the decrease of battery lifetime by 1 day after 220 hospital dynamic alerting sessions. crD is a variable for counting sessions of daily reporting. This counter is reset after reaching constant value cD which, in turn, triggers the subtraction operation $\mathit{lifetime} - \mathit{deltaLTN}$. Similarly reaching values crO , crT and crH with corresponding constants cO , cT and cH trigger resets of variables crO , crT and crH , respectively. Letters O , T and H stand for hospital, traveling and home, respectively.

5.3 Pacemaker Model

Figure 2 shows the complete model composed of interacting automata Patient, Device and HourClock. The hour clock of the system model is depicted in the lower part of the figure. The state updates of Patient and Device are triggered every 1 h with synchronization channel chH . The automaton Patient that models possible moves of the patient between different environments is shown in the upper part of the figure and the pacemaker device automaton is shown in the middle part of the figure. The initial location of the patient is his/her home. Probability Ph indicates that he/she stays at home and probability $\mathit{M} - \mathit{Ph}$ indicates that he/she is traveling. Probability Pt indicates that he/she keeps traveling, while probability Pv indicates that he/she will visit the hospital and probability $\mathit{M} - \mathit{Pt} - \mathit{Pv}$ indicates that he/she will return back home. The value vectors of the probabilities calculated from assumptions can be seen in Fig. 1 where the elements of vectors correspond to different patient profiles 0, 1 and 2. A concrete element is chosen from the vectors Ph , Pt , Pv and Pa depending on what value of Patient template parameter P (denoting profile) is selected for the model experiment. Variable env is updated according to the patient location in current state and is used within branching conditions in Device automaton. Location value affects the type of dynamic alerts which in turn affect differently the battery lifetime as described above and shown in the device model template.

In the device automaton template there are two main locations, one modeling the regular Monitoring and the other for Alerting. From Monitoring with probability Pr daily reporting occurs and whenever variable crD (for counting sessions of daily reporting) reaches constant value cD the battery lifetime is decreased by constant $\mathit{deltaLTN}$. From Monitoring with probability $\mathit{M} - \mathit{Pr}$ daily reporting does not occur. During Monitoring dynamic episode alerting occurs with probability Pa being different for each patient type as it can be seen in Fig. 1 and battery lifetime is decreased depending on the environment the patient is currently located. With probability $\mathit{M} - \mathit{Pa}$ dynamic alerting does not occur in current hour. Since the minimal time step in the model is one hour, sub-hour activities' timing is ignored and modeled using *Committed* locations

in which the model time is not progressing, i.e., all state updates of transitions from *Committed* locations are instantaneous (in model time).

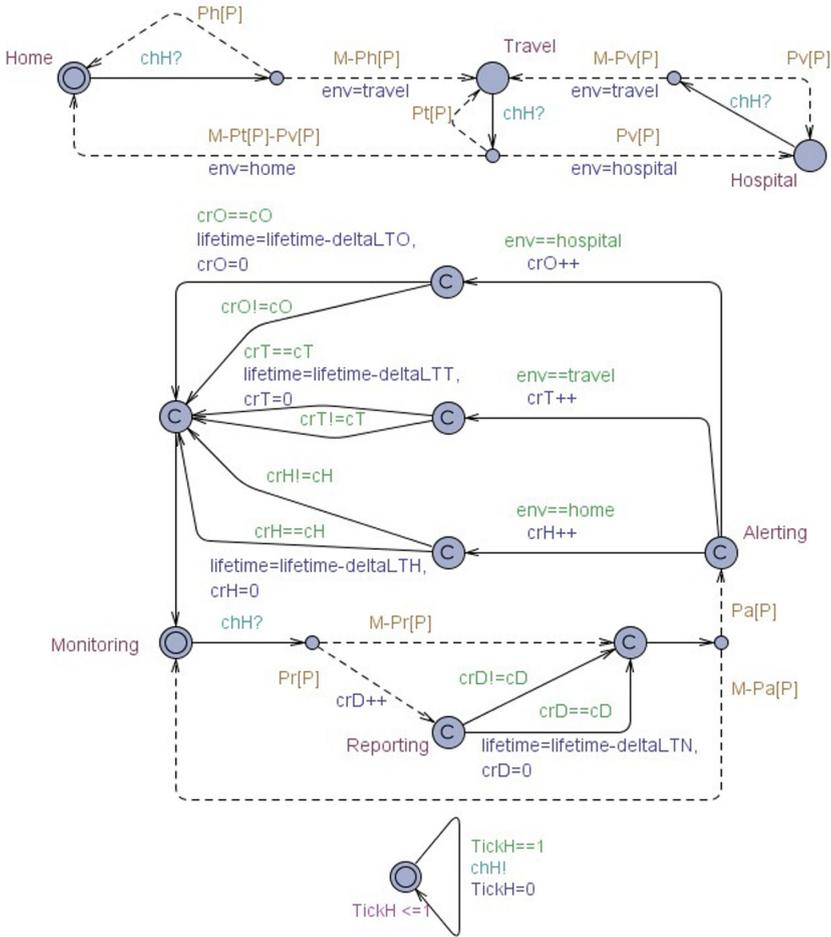


Fig. 2. Pacemaker UPPAAL probabilistic model

5.4 Pacemaker Model Analysis

The model depicted in Fig. 2 allows for checking critical battery lifetime properties depending on the different patient profiles and patient activities. For example, let us prove with model-checking that the expected lifetime is *safe*, i.e., the battery level never gets below a critical value before planned maintenance time, which could be 15% of its initial level. Assuming that the battery nominal lifetime is 5000 d and for simplicity we approximate the depletion rate being

Table 1. Experimental estimates of device lifetime for different patient profiles

Environment	Depletion rates in env. (days per alert)	Impact of the environment	Patient Type 0	Patient Type 1	Patient Type 2
Home	0,12714	Hrs in env. per day	14	18	22,475
		No of alerts in env. per day	0,0016	0,0082	9,3646
		Depletion per day (days per day of operation)	0,0002	0,00105	1,19063
Travel	4,06853	Hrs in env. per day	10	6	1
		No of alerts in env. per day	0,00114	0,00274	0,41667
		Depletion per day (days per day of operation)	0,00012	0,00029	1,69522
Hospital	0,10595	Hrs in env. per day	0,00548	0,0219	0,526
		No of alerts in env. per day	0,0000006	0,00001	0,21918
		Depletion per day (days per day of operation)	0,0000025	0,00004	0,02322
Total no of alerts per day			0,00274	0,01097	10,0004
Total expected lifetime			4435	4431	1239
Lifetime till 15% remaining battery lifetime			3770	3766	1053

constant the critical threshold will be reached by 4250 d (15% of 5000 d is 750 d). Also we use an auxiliary clock variable GCI in the query to refer to the time instances in the interval from 0 to $EXLT$, which stands for the expected lifetime until reaching the critical threshold value. The TCTL formula (1) expresses the property that in all states of all possible scenarios within closed time interval $[0, EXLT]$ the calculated battery level never gets less than the *Critical* value (750 d of remaining lifetime).

$$A \square GCI \leq EXLT \text{ imply } lifetime \geq Critical \quad (1)$$

In addition to model-checking the correctness of pacemaker's maintenance schedule, concrete numerical estimates of battery lifetime can be generated by

UPPAAL simulation experiments under various scenarios. Table 1 exemplifies the results of simulation experiments for all three patient profiles described in Sect. 5.1. As can be seen from Table 1 the pacemaker expected lifetime average estimates for patient type 0 and type 2 differ drastically (over 3 times), and claiming common rough estimate regardless the specifics of patient profile is obvious risk to patient’s safety. The factors, such as, frequency of alerting and security protocols used in different environments have substantial effect on pacemaker expected lifetime. While one would expect that the expected lifetime and maintenance deadline estimates for patient types 0 and 1 can be proved to be correct (Query 1 would be satisfied) and for patient type 2 the Query 1 would be clearly unsatisfiable, it turns out that if more than daily reporting is taken into account the query is not satisfiable for all patient types for this case study. That confirms our main hypothesis that patient profile specific analysis using patient and device models incorporating all influential factors is inevitable.

6 Conclusions

In this paper we have presented a model-based quantitative analysis of pacemaker lifetime that takes into account patient specific factors, including general health condition, acting environment, remote dynamic reporting of vital patient data and alternative security protection measures for these data communications. The study did not follow proprietary pacemaker BlueSync implementation that, according to our knowledge, supports a single security model. Instead, we proposed a multi-level security model that, depending on the security context, allows to select different security levels. Moreover, due to the fact that current patient safety regulations do not foresee remote modifications of the pacemaker firmware, our proposed cryptographic implementation should be applicable for the secure implantable device firmware upgrade as well in the further.

The analysis of the recent related work on development of BANs and specifically on implantable pacemakers motivated us for the work presented in this paper. We performed probabilistic modeling and model analysis with UPPAAL Timed Automata on a fully parameterizable model regarding some of typical patient profiles. For that we took into account averaged patient-dependent data for episode alerting frequency which is additionally assigned different security costs based on three possible patient environments during a day. The focus of our work was on pacemaker operational lifetime prediction with UPPAAL model-checking tool with different remote observation needs according to cardiovascular disease severity and location-dependent security requirements. The proposed approach to fine-grained battery lifetime analysis using probabilistic timed models is flexible in the sense that it can be easily adjusted by incorporating in the model only the factors that have impact to device endurance and reliable maintenance planning.

Acknowledgement. This work has been supported by the ERDF funded centre of excellence project EXCITE (2014–2020.4.01.15-0018), the Estonian Ministry of Education and Research institutional research grant no. IUT33-13 and supported in part by the Estonian Research Council grant PRG 424.

References

1. <https://www.fairview.org/patient-education/116363EN>, Accessed 15 August 2020
2. <https://www.wired.com/story/pacemaker-hack-malware-black-hat/>
3. <https://www.medtronic.com/us-en/healthcare-professionals/products/cardiac-rhythm/managing-patients/improved-outcomes/remote-monitoring-options.html>
4. Medtronic Azure MRI Surescan Pacemaker Reference Manual, <https://www.medtronic.com/us-en/healthcare-professionals/products/cardiac-rhythm/pacemakers/azure.html>
5. Medtronic Azure MRI Surescan Pacemaker Technical Manual. <https://www.medtronic.com/us-en/healthcare-professionals/products/cardiac-rhythm/pacemakers/azure.html>
6. Ahmed, A., Hasan, O., Tahar, S., Mohamed, A.: Formal verification of energy consumption for an eeg monitoring wireless body area sensor network. In: 2016 International Conference on Open Source Systems Technologies (ICOSST), pp. 18–22. IEEE (2016)
7. Behrmann, Gerd., David, Alexandre, Larsen, Kim G.: A tutorial on UPPAAL. In: Bernardo, Marco, Corradini, Flavio (eds.) SFM-RT 2004. LNCS, vol. 3185, pp. 200–236. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-30080-9_7
8. Ben Hamouda, R., Ben Hafaiedh, I.: Formal modeling and verification of a wireless body area network (WBAN) protocol: S-tdma protocol. In: 2017 International Conference on Internet of Things, Embedded Systems and Communications (IINTEC), pp. 72–77. IEEE (2017)
9. Bulychev, P.E., et al.: UPPAAL-SMC: statistical model checking for priced timed automata. In: Wiklicky, H., Massink, M. (eds.) Proceedings 10th Workshop on Quantitative Aspects of Programming Languages and Systems (2012)
10. Chen, T., Yu, Z., Li, S., Chen, B.: From wireless sensor networks to wireless body area networks: formal modeling and verification on security using pat. *J. Sens.* **2016**, 1–11 (2016). <https://doi.org/10.1155/2016/8797568>
11. Cronin, B., Maus, T.M., Khoche, S., Rozner, M.A.: Chapter 4 - cardiovascular implantable electronic device management in noncardiac surgery. In: Kaplan, J.A., Cronin, B., Maus, T.M. (eds.) Essentials of Cardiac Anesthesia for Noncardiac Surgery, pp. 70–99. Elsevier, New York (2019). <https://doi.org/10.1016/B978-0-323-56716-9.00004-7>, <http://www.sciencedirect.com/science/article/pii/B9780323567169000047>
12. Dai, S., Hong, M., Guo, B.: Synthesizing power management strategies for wireless sensor networks with UPPAAL-STRATEGO. *Int. J. Distrib. Sens. Networks* **13**(4) (2017). <https://doi.org/10.1177/1550147717700900>
13. Granjal, J., Monteiro, E., Silva, J.S.: End-to-end transport-layer security for internet-integrated sensing applications with mutual and delegated ECC public-key authentication. In: 2013 IFIP Networking Conference, pp. 1–9. IEEE (2013)
14. Hsu, C.L., Avudaiappan, R.: Bluetooth technology. *Academia Sinica Computing Centre* **18**, (2002)
15. Hummen, R., Shafagh, H., Raza, S., Voig, T., Wehrle, K.: Delegation-based authentication and authorization for the IP-based internet of things. In: 2014 Eleventh Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), pp. 284–292 (2014)
16. Kramer, D.B., Fu, K.: Cybersecurity concerns and medical devices: lessons from a pacemaker advisory. *JAMA* **318**(21), 2077–2078 (2017). <https://doi.org/10.1001/jama.2017.15692>

17. Li, C., Raghunathan, A., Jha, N.K.: Hijacking an insulin pump: security attacks and defenses for a diabetes therapy system. In: 2011 IEEE 13th International Conference on e-Health Networking, Applications and Services, pp. 150–156. IEEE (2011)
18. Moosavi, S.R., Nigussie, E., Levorato, M., Virtanen, S., Isoaho, J.: Performance analysis of end-to-end security schemes in healthcare IoT. *Procedia Comput. Sci.* **130**, 432–439 (2018)
19. Pajic, M., Jiang, Z., Lee, I., Sokolsky, O., Mangharam, R.: From verification to implementation: a model translation tool and a pacemaker case study. In: 2012 IEEE 18th Real Time and Embedded Technology and Applications Symposium, pp. 173–184. IEEE (2012)
20. Pycroft, L., Aziz, T.Z.: Security of implantable medical devices with wireless connections: the dangers of cyber-attacks. *Expert Rev. Med. Devices* **15**(6), 403–406 (2018). <https://doi.org/10.1080/17434440.2018.1483235>
21. Raatikainen, M.P., et al.: Statistics on the use of cardiac electronic devices and electrophysiological procedures in the European Society of Cardiology countries: 2014 report from the european heart rhythm association. *EP Europace* **17**(suppl-1), i1–i75 (2015). <https://doi.org/10.1093/europace/euu300>
22. Rao, A., Rozenblit, J., Lysecky, R., Sameting, J.: Composite risk modeling for automated threat mitigation in medical devices. In: Proceedings of the Symposium on Modeling and Simulation in Medicine, MSM 2017, pp. 1–10. Society for Computer Simulation International, San Diego, CA, USA (2017)
23. Rescorla, E., Modadugu, N.: Datagram transport layer security version 1.2. RFC **6347**, 1–32 (2012)
24. Tarakji, K., et al.: P577 Early experience with the first pacemakers to directly connect with smart devices for remote monitoring. *Eur. Heart J.* **40**(Supplement_1) (2019). <https://doi.org/10.1093/eurheartj/ehz747.0188>, ehz747.0188
25. Varma, N., et al.: The relationship between level of adherence to automatic wireless remote monitoring and survival in pacemaker and defibrillator patients. *J. Am. Coll. Cardiol.* **65**(24), 2601–2610 (2015). <https://doi.org/10.1016/j.jacc.2015.04.033>



Amplitude Modulation in a Molecular Communication Testbed with Superparamagnetic Iron Oxide Nanoparticles and a Micropump

Max Bartunik¹✉, Thomas Thalhofer², Christian Wald², Martin Richter², Georg Fischer¹, and Jens Kirchner¹

¹ Institute for Electronic Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany
max.bartunik@fau.de

² Fraunhofer EMFT Research Institution for Microsystems and Solid State Technologies, Munich, Germany

Abstract. Molecular communication uses molecules or other nanoscale particles to transmit data in scenarios where conventional communication techniques are not feasible. In previous work a testbed using superparamagnetic iron oxide nanoparticles (SPIONs) as information carriers in a fluid transmission channel with constant background flow was proposed. The SPIONs are detected at a receiver as change of a coils inductance. We now improve the testbed by using a piezoelectric micropump as transmitter, making amplitude modulation (AM) with different injection volumes possible. Machine learning is employed at the receiver to differentiate between six different amplitude levels and grey code is used to reduce bit errors. With AM and the designed coding scheme, the achievable effective data rate was doubled to 4.45 bit s^{-1} .

Keywords: Molecular communication · Amplitude modulation · SPIONs · Micropump

1 Introduction

The aim of molecular communication is to achieve data transmission using molecules or other particles in the nanoscale. So far, setups to experimentally investigate molecular communication with various transmission methods, such as air-based transmission with alcohol [3, 8, 18] or encoding information as change of pH-value [4, 6, 9] have been proposed. In [16] a different setup that uses superparamagnetic iron oxide nanoparticles (SPIONs) as information carriers in a fluid transmission channel is presented. [1, 2] provide some induction-based receiver concepts for such SPION-based testbeds.

Investigated data transmission scenarios in [1,2] were confined to binary-state symbol transmissions, encoding a ‘1’ as an injection of SPIONs into the transmission channel and a ‘0’ as no injection. The achievable information rate was restricted due to an increase of inter-symbol interference for reduced symbol intervals.

Schemes for amplitude modulation (AM), or concentration shift keying, in molecular communication have been proposed previously (e.g. [10,14]). However, these focus on a diffusion-based scenario in contrast to the active transport present in the provided testbed and are restricted to theoretical channel observations. In [2,16] a peristaltic pump was used as a transmitter to enable on-off keying (OOK). However, the pump was limited to a fixed transmission volume. We now replace the peristaltic pump with a micropump allowing for use of AM to achieve a higher-level encoding per symbol. Machine learning is used at the receiver to detect the transmitted amplitude levels.

Today, micropumps are used in many applications including drug delivery [5], lab-on-a-chip devices [13] and sensor enhancement [12]. As described in [15,19] many different actuators have been used for micropumps. Of these technologies, piezoelectric actuated micropumps show the most promising features since they are able to create very high forces and actuation speeds, enabling high pumping frequencies and pressures. For this paper a refined version of the steel-foil based micropump introduced in [17] with a diameter of 20 mm, designed by Fraunhofer Research Institution for Microsystems and Solid State Technologies (EMFT) in Munich, was used. The pumps parameters allow a fast injection of a well-defined, discrete volume of SPIONs of up to 11.5 μl per pump stroke. The stroke volume can be controlled by changing the piezo actuator voltage enabling a variety of amplitude levels.

We used the described micropump to achieve AM in a molecular communication testbed with SPIONs for the first time. Together with a comprehensive coding scheme, employing two-dimensional grey code and symbol detection based on machine learning, the achievable data rate was doubled in comparison to previous setups using OOK.

In the following, we will first describe the testbed setup and its individual components. Then we will present the methods used to encode and decode data, as well as the achieved data transmission in the experimental testbed. A discussion of the results and an outline of future improvements concludes the article.

2 Testbed

The testbed can be separated into the four hardware components information carriers, transmitter, channel and receiver. The principle setup was first described in [16].

2.1 Information Carriers

Information is transmitted in the testbed by changing the concentration of SPI-ONs. Due to their superparamagnetic properties they can be detected as change of a coils inductance. In our case we used particles with a hydrodynamic diameter of 50 nm, a specific susceptibility of 8.78×10^{-3} for an aqueous solution with a concentration of 1 mg Fe ml^{-1} and a coating of lauric acid. They were synthesized by the Section for Experimental Oncology and Nanomedicine (SEON) of the University Hospital Erlangen.

2.2 Transmission Channel

A flexible tube with an inner diameter of 1.52 mm is used as the transmission channel. A constant background flow of water is provided by a peristaltic pump (Ismatec ISM596D) with a flow rate of 10 ml min^{-1} . SPI-ONs are injected into the transmission tube as described in Subject. 2.3 and travel a distance of 5 cm to the receiver. Due to laminar flow the injected nanoparticles are spread out axially along the transmission channel, causing inter-symbol interference for repeated injections.

2.3 Transmitter

The transmitter consists of a steel micropump, manufactured by Fraunhofer EMFT in Munich, and a venous cannula.

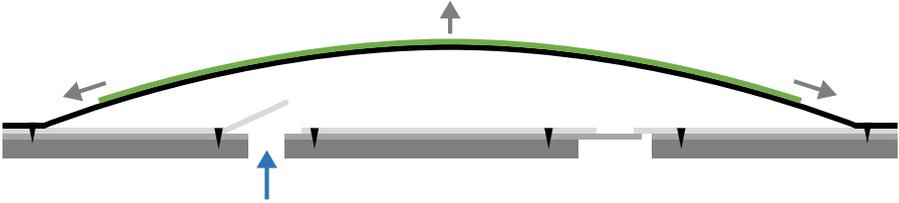
The pump has a height of 1 mm, a diameter of 20 mm and moves a volume of $11 \mu\text{l}$ per full stroke. Due to its high flow rate of up to 20 ml min^{-1} (for water), and high pressure tolerance of 60 kPa, it is ideal for use in the testbed environment. As can be seen in Figs. 1 and 2, the micropump is composed of a coin shaped steel base plate, three steel foils and a piezo actuator disc. Two steel foils are laser-welded on top of the base plate to create passive inlet- and outlet valves and valve seats. The third steel foil closes the top of the pump chamber. The piezo actuator is glued onto the pump chamber with a patented process described in [7].

The piezo actuator is driven with a rectangular voltage signal. The stroke volume can be adjusted by varying the voltage amplitude: A signal alternating from -80 V to $+300 \text{ V}$ results in a full stroke with an approximate volume of $11 \mu\text{l}$. For an input signal with a reduced voltage swing of -80 V to $+150 \text{ V}$ a volume of approximately $6 \mu\text{l}$ is expedited per stroke.

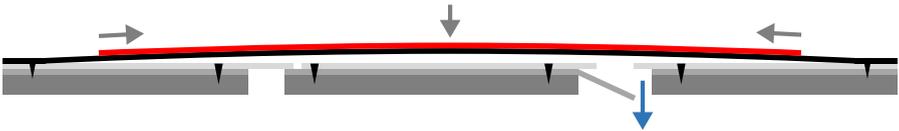
A venous cannula with an inner diameter of 0.8 mm is used to inject the nanoparticles into the transmission channel. The cannula is placed so that the injection occurs in the middle of the channel in the same direction as the background flow.



(a) The pump consists of a steel base plate (1), passive inlet (2) and outlet (3) valves, an actuator diaphragm (4) and the piezoelectric actuator (5). Here the initial state, when no voltage is applied to the actuator, is shown. The mounting process leads to a pre-bent piezo actuator and actuator diaphragm.



(b) Applying a negative voltage of -80 V to the actuator causes the diaphragm to bend upwards and the pump chamber is filled.



(c) With a positive voltage in the range of 80 V to 300 V the diaphragm is bent downwards and the liquid in the pump chamber is ejected through the outlet valve.

Fig. 1. Schematic cross-section of the used steel micropump shown for the three different actuation states.

The amount of SPIONs that are introduced into the background channel in each symbol interval is varied to achieve different amplitude levels. As the micropump is operated at a significantly higher rate (40 Hz) than the symbol frequency (maximally 3.5 Hz), multiple pump strokes are observed as an individual injection with higher volume. Furthermore, the injected volume per pump stroke can be controlled. We therefore chose six different amplitude levels with steps of a half stroke (approx. $6\text{ }\mu\text{l}$) ranging from no injection ($0\text{ }\mu\text{l}$) to two and a half strokes (approx. $28\text{ }\mu\text{l}$) per injection.

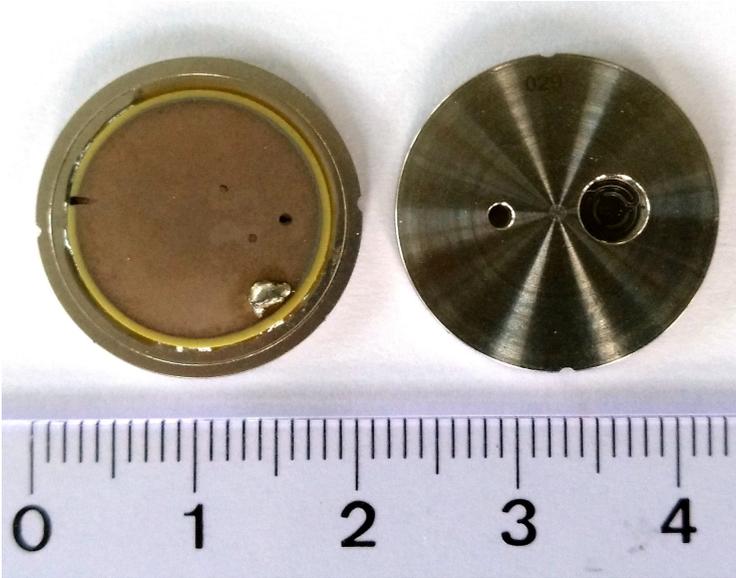


Fig. 2. Pictures of the steel micropump showing the top with the piezoelectric actuator (left) and the steel base plate (right). The pump has a diameter of 20 mm.

2.4 Receiver

The transmission channel is passed through a custom made coil at the receiver. Due to the magnetic properties of the SPIONs the coil inductance changes as the nanoparticles pass through it. In a parallel circuit this change of inductance can be observed as a shift of resonance frequency that is measured using the LDC1612 from Texas Instruments. The receiver setup and capabilities were presented in [2]. The change in resonance frequency depends on the amount of SPIONs inside the detection coil, allowing for differentiation of various amplitude levels.

3 Data Transmission

Data transmission in the testbed is achieved by encoding binary data in six different amplitude levels (as described in Subsect. 2.3) that are detected at the receiver and decoded using machine learning classification.

3.1 Encoding

Each transmitted symbol has six possible amplitude levels (0 to 5). Words are constructed of two symbols with one amplitude level each. Binary data is represented by encoding five bits to one word. When encoding all possible 32 combinations of five bits to two symbol words (36 possible combinations), 4 unused

words remain. In future these may be used for transmission synchronisation or other control commands. In our case these remaining code words are used to increase the distance between similar amplitude levels. Furthermore, to reduce bit errors in scenarios where a detected amplitude level is incorrect but close to the actually transmitted level, a two-dimensional grey code is applied, changing only one bit for neighbouring words. Table 1 shows the complete encoding.

Table 1. Encoding table for the representation of 5 bits using words consisting of two symbols (S_0 and S_1). To reduce bit errors for similar amplitude levels neighbouring words have only a single bit change. The four remaining words in blue are not used for transmission. A neighbouring bit sequence was arbitrarily chosen for decoding.

S_0	S_1					
	0	1	2	3	4	5
0	00000	00001	00011	01011	01001	01000
1	00100	00001	00010	01010	01001	01100
2	00101	00111	00110	01110	01111	01101
3	10101	10111	10110	11110	11111	11101
4	10100	10001	10010	11010	11001	11100
5	10000	10001	10011	11011	11001	11000

To enable simple detection, each transmission is initiated and terminated with a symbol of maximal amplitude (value 5).

3.2 Decoding

The recorded signal is decoded using machine learning in a Matlab (The Math-Works, Inc.) script. First, a moving averaging filter with a length of two samples is applied to reduce noise and the amount of data points is increased to 100 Sa s^{-1} with cubic spline interpolation. Next, a zero baseline is determined as the mean value of two symbol intervals before injection of SPIONs. The initial synchronisation peak is set to the first local maximum after a threshold at half of the maximal signal amplitude is exceeded. This peak is used to determine the fixed symbol intervals with a known duration. To accommodate for transmission timing inaccuracies, the symbol intervals are resynchronised at the highest available peaks that are at least five symbol intervals apart and exceed the initial detection threshold.

For each symbol interval three signal parameters, which were determined by a feature selection algorithm, are calculated. First, local maxima inside the symbol interval are searched. For the peak with the largest amplitude the height of the rising and falling edges are calculated by determining the minimal amplitude within one symbol interval before and after the peak. In addition, the absolute amplitude value is also recorded. If no local maximum is found, the symbol

amplitude is set to the minimal amplitude inside the symbol interval. Figure 3 shows a section with eight symbols of a sample transmission, demonstrating detection of the prediction parameters. As inter-symbol interference causes a changed baseline for each symbol interval, the application of fixed threshold values to differentiate between multiple amplitude levels is not possible.

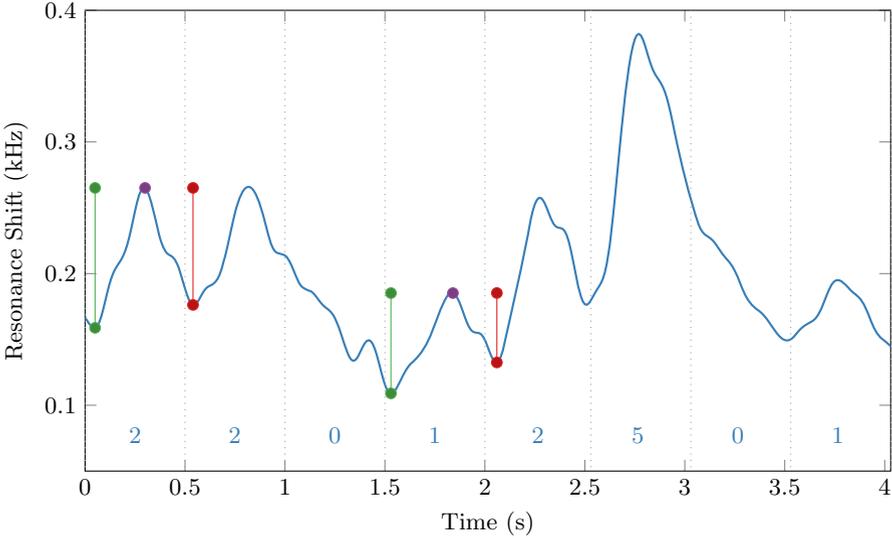


Fig. 3. Sample transmission sequence section with eight symbols. The symbol intervals are shown with dashed lines and are annotated with the transmitted amplitude level. Peaks are detected and the parameters peak amplitude (purple), rising edge height (green) and falling edge height (red) determined for each symbol interval. The parameter detection is shown exemplary for two symbols.

The rising and falling edge heights are normalised to the values of the initial synchronisation peak and then passed to a classification predictor together with the amplitude levels. A discriminant analysis model fitted to a training sample sequence with a total of 204 symbols is used for classification. An individual model was trained for each symbol rate (2 Hz to 3.5 Hz) with the ‘fit discriminant analysis classifier’ algorithm in Matlab. Figure 4 shows the spread of the three prediction parameters in the training sequence for a symbol rate of 2 Hz. The high deviation of the predictor parameters with overlapping areas does not allow for simple threshold detection. By using the combined information of all three parameters in a machine learning model a good prediction of the transmitted symbol can be made.

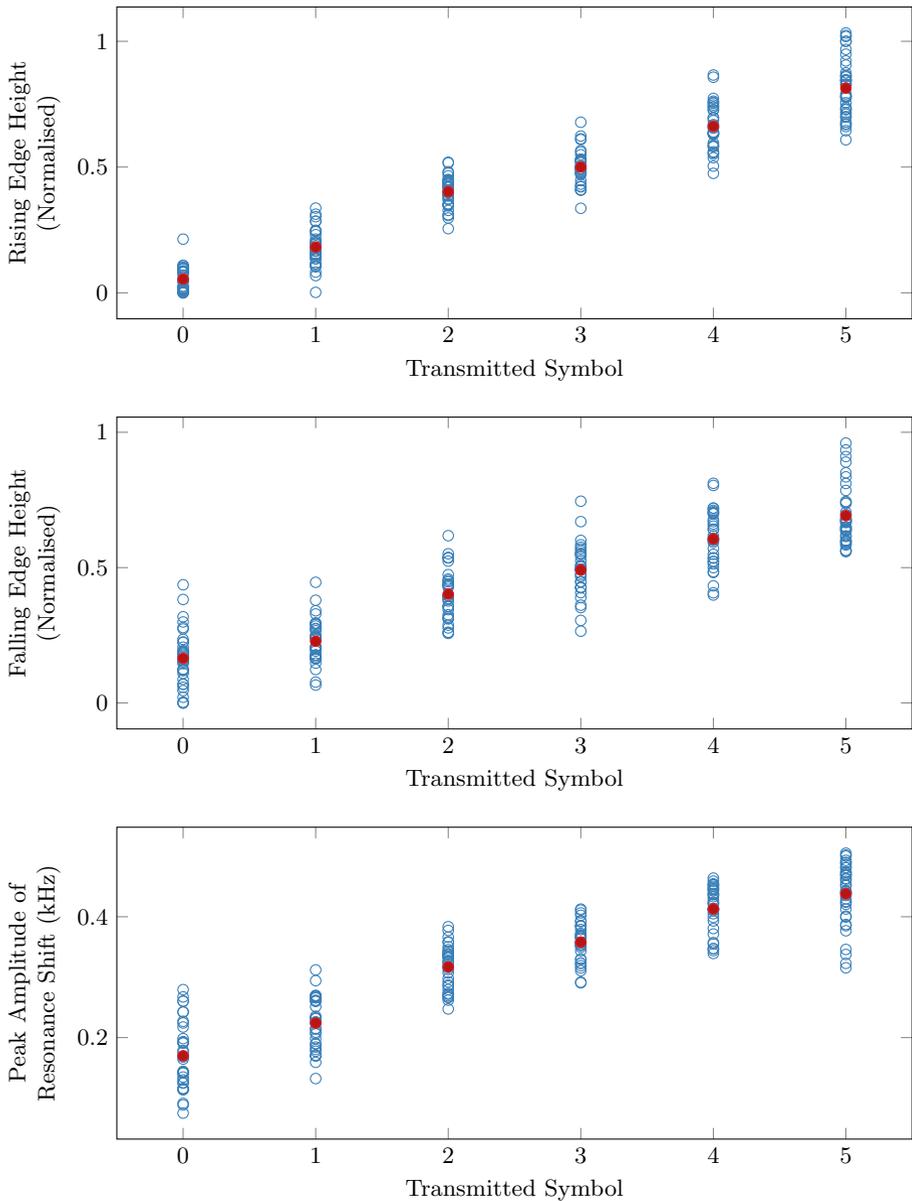


Fig. 4. Distribution of the three predictor parameters rising edge height, falling edge height and peak amplitude for a training set of 204 symbols. Although the mean values marked in red show a clear tendency, definite detection thresholds cannot be set due to the high deviations.

4 Results

4.1 Injection Volume Correlation

In a first step, the correlation between injected volume of SPIONs and detected shift in resonance frequency is of interest. To determine this relation a series of injections were performed with the micropump ranging in volume from one to five pump strokes. Between each injection a pause, sufficient to ensure a return to zero, was made. Figure 5 shows the detected resonance shift, normalised to the maximal value, in relation to the normalised injection volume. As expected, a linear relation between the detection signal and the injection volume can be observed. Therefore, detection of different amplitude levels is possible.

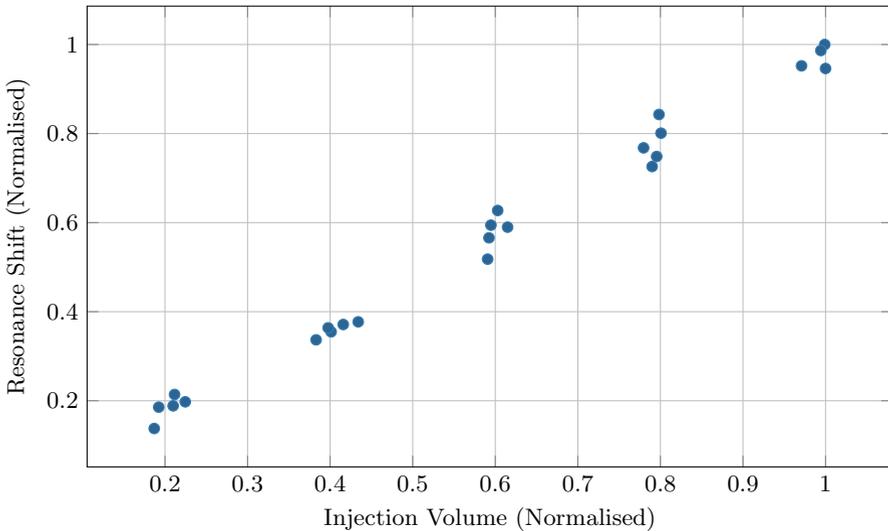


Fig. 5. Resonance shift at the receiver in relation to injection volume for 24 individual injection bursts. Each burst consists of one to five pump strokes. A linear relation between the received signal and the transmitted volume of SPIONs can be observed.

4.2 Data Transmission

Information en- and decoding was evaluated for four different symbol rates ranging from 2 Hz to 3.5 Hz in steps of 0.5 Hz, each with a random sequence of 2000 bits. Due to restrictions for the micropump driver the sequence was split into four transmissions of 100 symbols. Each sequence was decoded as described in Subsect. 3.2 and the binary output evaluated for bit errors.

Figure 6 shows a sample transmission section consisting of eight symbols. All words (i. e. symbol tuples) but one were decoded correctly. Due to the choice of coding only a single bit error results from the incorrectly detected word.

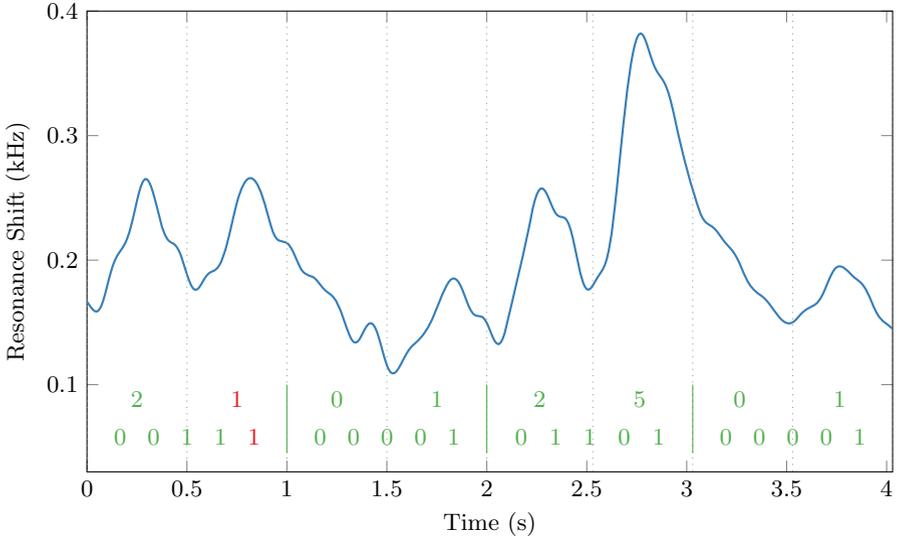


Fig. 6. Sample transmission sequence section with eight symbols. The detected symbol values for each interval and the decoded words are shown. The second transmitted symbol value (2) was incorrectly detected as 1. However, as neighbouring words differ only by a one bit the selected coding restricts the decoded binary word to a single bit error.

In Fig. 7 the average symbol and bit error probabilities are shown for the four different symbol interval durations. No significant correlation between reduced symbol intervals and an increase of decoding errors can be observed. As can be seen in Fig. 8, symbol detection errors are restricted to a single value offset in most cases. Due to the choice of coding, only a single bit error occurs for such a single value detection offset. The result is a significantly lower bit error rate (20% for a single word with a typical symbol error) than symbol error rate (50% for a single word with one detection error).

From an information theory perspective our transmission from encoding to decoding can be regarded as a binary symmetric channel, implying equal probabilities for bit errors independent of bit value. The capacity of such a channel with a given error probability f is

$$C(f) = 1 - \left(f \log_2 \frac{1}{f} + (1 - f) \log_2 \frac{1}{1 - f} \right). \quad (1)$$

The noisy-channel coding theorem provides an upper boundary for achievable effective data rate given a channel capacity and the tolerable bit error rate (BER) [11]. This effective data rate is given by

$$\frac{C}{R} = 1 - \left(\text{BER} \log_2 \frac{1}{\text{BER}} + (1 - \text{BER}) \log_2 \frac{1}{1 - \text{BER}} \right) \quad (2)$$

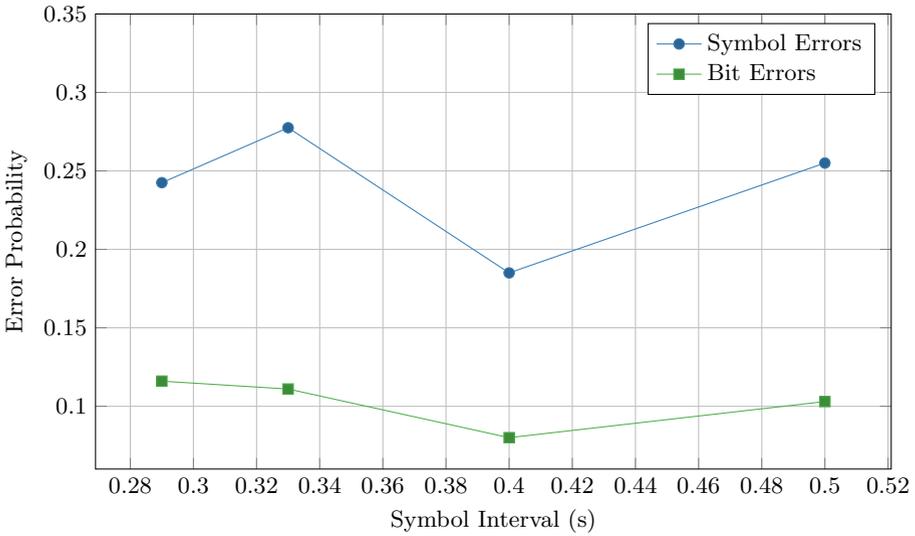


Fig. 7. Probabilities for bit and symbol errors using amplitude modulation and the provided coding scheme. Due to the use of grey code, the average probability for a bit error (10.3%) is significantly lower than the symbol error probability (24%).

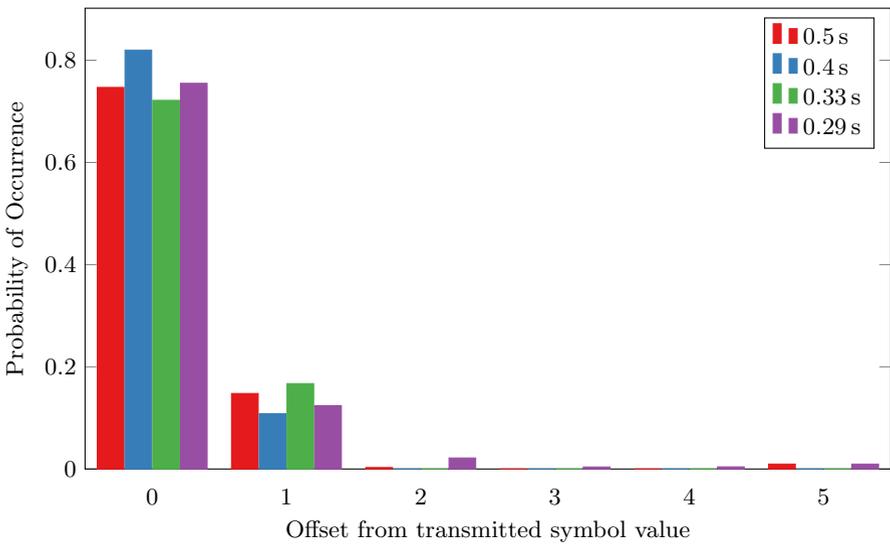


Fig. 8. Probabilities for various receive symbol errors. A detection error with an offset larger than 1 amplitude level occurs in only very few cases, resulting in the coding gain achieved by the use of grey code.

where R is the code rate. Using (2) we can calculate the possible effective data rates for both the transmission scenario using AM and previous results with OOK presented in [2].

Figure 9 shows the calculated values for the various symbol intervals and multiple tolerable BER scenarios. Although different symbol intervals were investigated in [2], the possible effective data rates can be compared to the results presented here. For a symbol interval of 0.4 s and a tolerable BER of 1% an effective data rate of 3.9 bit s^{-1} is achievable using AM. With the same parameters the data rate was restricted to 2.3 bit s^{-1} using OOK. Of the investigated intervals, the maximal data rate using AM is reached at a symbol interval of 0.29 s and is 4.5 bit s^{-1} for a remaining BER of 1%. This is close to double the value achievable with OOK in [2].

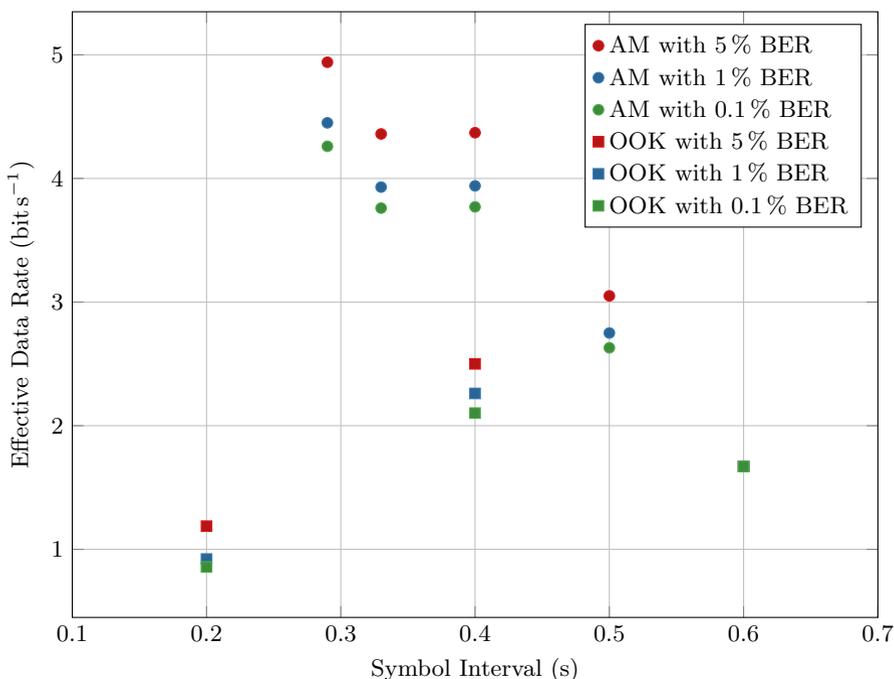


Fig. 9. Achievable effective data rate according to the noisy-channel coding theorem. Comparable values for OOK were taken from [2]. With a tolerable BER of 1% an effective data rate of 4.5 bit s^{-1} can be reached using the proposed AM scheme. The maximal value for OOK is 2.3 bit s^{-1} .

5 Conclusion

A micropump was successfully used as a transmitter in the provided molecular communication testbed. Due to the pumps capabilities, six different amplitude

levels could be differentiated at the receiver by modulating the volume of SPIONs injected in one symbol interval. With machine learning and an adapted coding a bit error probability of 10.3% was achieved.

As no significant increase of bit errors for reduced symbol intervals was observed, a further increase of symbol rate and therefore effective data rate is possible and will have to be investigated in future work. Limiting factors may be the increase of inter-symbol interference due to laminar flow and the sensor sample rate.

In comparison to data transmission using on-off keying presented in [2] the achievable effective data rate was doubled.

In future work improvements may be made on both the coding scheme and the testbed setup. In particular, an increase of the sensor sample rate and an optimisation of the used coils is of interest. Symbol encoding could be optimised to ensure larger amplitude changes and therefore reduce bit errors. Finally, the machine learning model used for detecting amplitude values could be improved with a significantly larger training set.

Acknowledgements. The authors would like to express their sincere gratitude to Harald Unterweger from the Section for Experimental Oncology and Nanomedicine of the University Hospital Erlangen for manufacturing and providing continuous support regarding SPIONs.

This work was supported in part by the German Federal Ministry of Education and Research (BMBF), project MAMOKO (16KIS0913K).

References

1. Bartunik, M., et al.: Novel receiver for superparamagnetic iron oxide nanoparticles in a molecular communication setting. In: Proceedings of the Sixth Annual ACM International Conference on Nanoscale Computing and Communication - NANOCOM 2019. ACM Press (2019). <https://doi.org/10.1145/3345312.3345483>
2. Bartunik, M., et al.: Comparative evaluation of a new sensor for superparamagnetic iron oxide nanoparticles in a molecular communication setting. In: Chen, Y., Nakano, T., Lin, L., Mahfuz, M.U., Guo, W. (eds.) BICT 2020. LNICST, vol. 329, pp. 303–316. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-57115-3_27
3. Farsad, N., Guo, W., Eckford, A.W.: Tabletop molecular communication: text messages through chemical signals. PLoS ONE **8**(12), 1–13 (2013). <https://doi.org/10.1371/journal.pone.0082935>
4. Farsad, N., Pan, D., Goldsmith, A.: A novel experimental platform for in-vessel multi-chemical molecular communications. In: Proceedings of the IEEE Global Communications Conference, GLOBECOM 2017. IEEE (2017). <https://doi.org/10.1109/glocom.2017.8255058>
5. Ginsberg, B.H.: Patch pumps for insulin. J. Diab. Sci. Technol. **13**(1), 27–33 (2019). <https://doi.org/10.1177/1932296818786513>
6. Grebenstein, L., et al.: Biological optical-to-chemical signal conversion interface: a small-scale modulator for molecular communications. IEEE Trans. NanoBiosci. **18**(1), 31–42 (2019). <https://doi.org/10.1109/tnb.2018.2870910>

7. Herz, M., Richter, M., Wackerle, M.: Method for manufacturing a bending transducer, a micro pump and a micro valve, micro pump and micro valve (2016). <https://patents.google.com/patent/US9410641B2/en>
8. Koo, B.H., Lee, C., Yilmaz, H.B., Farsad, N., Eckford, A., Chae, C.B.: Molecular MIMO: from theory to prototype. *IEEE J. Sel. Areas Commun.* **34**(3), 600–614 (2016). <https://doi.org/10.1109/jsac.2016.2525538>
9. Krishnaswamy, B., et al.: Time-elapse communication: bacterial communication on a microfluidic chip. *IEEE Trans. Commun.* **61**(12), 5139–5151 (2013). <https://doi.org/10.1109/tcomm.2013.111013.130314>
10. Kuran, M.S., Yilmaz, H.B., Tugcu, T., Akyildiz, I.F.: Modulation techniques for communication via diffusion in nanonetworks. In: *Proceedings of the IEEE International Conference on Communications (ICC 2011)*, pp. 1–5 (2011)
11. MacKay, D.J.C.: *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, Cambridge (2003)
12. Müller, G., Friedberger, A., Kreisl, P., Ahlers, S., Schulz, O., Becker, T.: A mems toolkit for metal-oxide-based gas sensing systems. *Thin Solid Films* **436**(1), 34–45 (2003). [https://doi.org/10.1016/S0040-6090\(03\)00523-6](https://doi.org/10.1016/S0040-6090(03)00523-6)
13. Pabst, O., Perelaer, J., Beckert, E., Schubert, U.S., Eberhardt, R., Tünnermann, A.: All inkjet-printed piezoelectric polymer actuators: characterization and applications for micropumps in lab-on-a-chip systems. *Org. Electron.* **14**(12), 3423–3429 (2013). <https://doi.org/10.1016/j.orgel.2013.09.009>
14. Singhal, A., Mallik, R.K., Lall, B.: Performance analysis of amplitude modulation schemes for diffusion-based molecular communication. *IEEE Trans. Wireless Commun.* **14**(10), 5681–5691 (2015). <https://doi.org/10.1109/TWC.2015.2441067>
15. Tsai, N.C., Sue, C.Y.: Review of mems-based drug delivery and dosing systems. *Sens. Actuators, A* **134**(2), 555–564 (2007). <https://doi.org/10.1016/j.sna.2006.06.014>
16. Unterweger, H., et al.: Experimental molecular communication testbed based on magnetic nanoparticles in duct flow. In: *Proceedings of the IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2018)*, pp. 1–5 (2018). <https://doi.org/10.1109/SPAWC.2018.8446011>
17. Wald, C., Richter, M.: Low-Cost Edelstahl-Mikropumpe zur Probenzufuhr eines miniaturisierten Gassensorsystems zur Brandfrüherkennung. In: *MikroSystemTechnik Kongress Proceedings. Mikroelemente & Systeme 2015*, pp. 448–451 (2015)
18. Wang, L., Farsad, N., Guo, W., Magierowski, S., Eckford, A.W.: Molecular barcodes: information transmission via persistent chemical tags. In: *Proceedings of the IEEE International Conference on Communications (ICC 2015)*. IEEE (2015). <https://doi.org/10.1109/icc.2015.7248469>
19. Wang, Y.N., Fu, L.M.: Micropumps and biomedical applications - a review. *Microelectron. Eng.* **195**, 121–138 (2018). <https://doi.org/10.1016/j.mee.2018.04.008>



An Enhanced DNA Sequence Table for Improved Security and Reduced Computational Complexity of DNA Cryptography

Maria Imdad, Sofia Najwa Ramli^(✉), Hairulnizam Mahdin, Boppana Udaya Mouni, and Shakira Sahar

Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, 86400 Batu Pahat, Malaysia

maria.imdad123@gmail.com, {sofianajwa, hairuln}@uthm.edu.my, mouniudaya@gmail.com, shakirasahar@gmail.com

Abstract. Recently, DNA cryptography rejuvenates the art of secret writing by combining biological information and cryptography. DNA's double-helical structure serves as a template for encoding decoding information, vast storage and randomness. The structure includes DNA encryption that uses a DNA sequence table to substitute plaintext into the DNA sequence. However, this encoding table can result in leakage of information about the plaintext, character frequency, and key, by carefully examining the ciphertext through frequency analysis attack. Therefore, this paper proposes an enhanced DNA table for all 96 printable ASCII characters which are created to improve the entropy so that the probability of each encoding base (A, T, C, G) is equally likely and to reduce the computational complexity of DNA cryptography. An algorithm has been selected to implement both tables for performance measurement. The results show that encoding and encryption time is reduced, high entropy ciphertext, better frequency distribution ciphertext is obtained. Information leakage in terms of conditional entropy is also reduced by the proposed table. In conclusion, the proposed table can be used as a DNA sequence table in DNA cryptography to improve overall system security.

Keywords: DNA cryptography · DNA sequence table · Entropy · IoT application

1 Introduction

Novel encryption techniques tend to ensure system security more than traditional cryptographic methods either by combining two or more traditional techniques or by taking the advantage of biological characteristic of DNA encryption [1, 2]. With advancements in technology, more efforts are to ensure system security from attacker's perspective. A recent study sheds light on the fact that traditional cryptographic solutions either symmetric or asymmetric encryption are not secure any longer and cannot be used directly

as a standalone solution [3]. One of the presented solution is using DNA Cryptography only or in combination with traditional cryptography, where DNA computing, encoding, decoding, and biological simulation processes yield better security [4]. Recent advancements in DNA cryptography present the solutions based on symmetric and asymmetric cryptographic systems, which when deployed showed improved usability [5]. These advancements are primarily because of DNA computing being fast and secure than existing technique and indicating that in near future DNA chips will replace silicon chips in computer systems for highly fast processing. A single gram of DNA has 1021 bases which are equal to 108 of data, due to the compactness in the double-helical structure of DNA [6].

DNA computing was started by L. Adleman in 1994 to solve complex computational problems primarily [7]. Recent development shows that it uses DNA cryptography to ensure system security for the next generations. Specifically, proposed encryption algorithms apply DNA cryptography with a DNA table that serves as a foundation for encoding and decoding the ASCII characters using DNA bases. DNA has four bases which are known as Adenine (A), Guanine (G), Cytosine (C) and Thymine (T) for encoding. Initially, researchers in [8] have designed a base table using those bases and most techniques use the table either with or without iterations currently. This means that in one iteration code, “TTTT” sequence is representing A in the plaintext characters but in other iterations, this code may be used to represent Y. Only the code positions against characters changes but the sequence remains the same, and this is done to ensure that same plain text is retrieved at the other end after passing through this table. However, the analysis of the DNA encryption technique shows that frequency analysis reveals information about the key [9]. Thus, this paper proposes an enhanced DNA encoding table that is designed keeping in view the frequency count, randomness in codes and entropy of the table. Later, the proposed table is compared with the table in [8], followed by a complete encryption-decryption process and performance analysis.

The rest of the paper is organized as follows: Sect. 2 has a review of the work done so far. Section 3 gives a detailed insight into the proposed table. Section 4 has an experimental analysis, statistics followed by conditional entropy and its calculation and lastly Sect. 5 concludes this paper.

2 Related Work

DNA cryptography is an umbrella having technologies that are inspired by genetic entity DNA, ranging from Polymerize Chain Reaction (PCR) of DNA synthesis to digital coding using the same bonding and stimulation patterns as defined in DNA by nature [9]. This paper focuses on the study of techniques where DNA coding has been used to improve system security. DNA coding has four basic nitrogenous bases A,T,C,G and their representation using binary bases 0 and 1. Several DNA based cryptographic techniques have been proposed where DNA bases are used in combination with one-time pad (OTP) [10–12]. The technique in [10] uses microdots to save ciphertext while PCR is used at the decryption end. Whereas researchers in [11] use basic mathematical operations like DNA addition, DNA subtraction combined in a Feistel structure for system security. Meanwhile, the technique in [13] uses a lookup table to rearrange DNA bases,

where a dynamic table for 256 ASCII characters is created. Then, it applies iterations to change the positions of characters which followed by a mathematical series before it uses asymmetric encryption for the encryption-decryption process. The combination of the output from the asymmetric encryption process with chunks of dynamically encoded text produces the ciphertext.

Later, researchers in [14] introduce a dynamic ASCII table where random ASCII characters are assigned to DNA bases initially. The dynamic ASCII table brings a new insight into DNA cryptography that random table results in different ciphertexts with the same plaintext making it challenging for an attacker to get access to the table along with iteration. A mathematical series is used for iteration purposes in which every iteration changes the position of the characters dynamically. For example, plaintext "A" may be encoded as "AAAA" in one iteration but in the next iteration, it may be "ACCT" to provide the randomness to the ciphertext. The same plaintext is encoded over different iteration to yield different ciphertexts. This encoded text is modified using OTP and is followed by genomic conversion. The final ciphertext is a compressed form of the genomic conversion into an amino acid table.

A biological simulation-based technique is proposed in [15] whereby a unique DNA based encoding table has been introduced. In this technique, a random encoding table is introduced after every session resulting in session-based output. The approach encodes the same plaintext that has different outputs in every different session. On the other hand, authors in [16] remove limitations of OTP ciphertext using DNA and amino acid coding, followed by randomness evaluation using NIST tests. A biotic-DNA oriented secret key mechanism is introduced in [17] and they use genetic information gathered from biological systems. The technique in [18] is a combination of digital coding, traditional cryptography as well as PCR amplification. Digital coding and traditional encryption are used to encode the plaintext followed by PCR for key generation. The technique in [19] has the underlying foundation of the signature method and asymmetric encryption with DNA. Initially the plain text is converted into ASCII codes followed by binary code, and is transformed into a matrix. This data is transmitted physically as a biological molecule in DNA.

DNA cryptography is not limited to text encryption but is equally applicable in image encryption as well [20]. A technique proposed by Zhang et al. [21] is image-based encryption comprised of map lattices of linear, as well as non-linear coupled with spatiotemporal chaos. A technique proposed in [22] is a combination of the hyperchaotic system along with a genetic recombinant, for image encryption where the system proved good security for image encryption. Zhang et al. proposed an image encryption system [23] based on permutation algorithms. Mix chaotic mapping in addition to Josephus traversing is used in [24] for image encryption. A combination of the chaotic system has been proposed in [25] which gave good encryption results than the previous techniques. DNA cryptography has been implemented in cloud computing to improve system security by enabling socket programming [26, 27]. In [28] an architectural framework has been proposed where digital signatures have been used in combination with DNA. Robust DNA codes based on DNA sequence has been proposed in [29, 30].

A technique proposed in [31] has used the same base table as in [8] and it uses DNA computing for intrusion detection. In this technique, DNA encoding is used to convert

the network traffic data into DNA sequences. The idea behind selecting DNA computing is that it follows the same mechanism to detect diseases as an intrusion detection system does. The results of the technique show that DNA can be used for intrusion detection and can give better results using a better encoding method.

Analyzing the existing work proves that all these techniques use the same base table as in [8] with or without iterations. This base table has a unique code for all ASCII printable characters, alphabets, capital, small, numbers and special characters. This table has 96 codes for 96 different characters which are 26 capital, 26 small, 10 numbers and 34 special characters. The bases contain binary coding, A = 00, T = 01, G = 10, C = 11. Each character is coded using DNA bases such as “y” is coded as “AAAA”. These codes are unique, indicating no two characters can have the same code. Where primarily this table is used to introduce randomness in cipher text. This table being randomly generated can yield more or less randomness across iterations, so by carefully examining this fact a new table has been created, which will always yield better randomness. New table is static in nature but can improve security as it has more random encoding as compare to base table. Frequently occurring characters have high random codes than less frequently occurring. Where a detail description is provided in Sect. 3. Keeping these facts into consideration, a new encoding table has been designed and these two tables will be compared based on different parameters.

3 Proposed DNA Table

The proposed table has unique codes for all $48 * 2 = 96$ matrix as of base table as in Table 1. These codes are not assigned randomly but have the underlying foundation as follows:

- Frequency count of “71,013,156” characters from [32], and additional “5747” characters online, to rank these characters according to the number of occurrences.
- A character that has high-frequency count will have codes with all four bases without being repeated in a particular order, similarly going down to least frequency characters with repetition of DNA bases. High-frequency character “t” is coded as “ACTG”, having all four bases and when it is converted into its binary code is “00110110”. Meanwhile, the same character is coded as “TCCC”, 01111111 in [8] having more repeating bits in a sequence. This is due to the repetitions of English language basic characteristics. Thus, the code should have all four bases to reduce repetitions so that the bits are more random and less predictable for the attacker.
- Each DNA base is counted exactly 96 times as in Table 1. This has been carefully selected and designed having the probability $96/384$ for each base and the final entropy factor.

Table 1. The proposed DNA encoding table

Rank	Character	Frequency	DNA code	Rank	Character	Frequency	DNA code
1	e	7,741,972	ATCG	49	H	123,634	CGTC
2	t	5,507,785	ACTG	50	x	123,585	CTGC
3	a	5,263,861	ACGT	51	7	120,193	AGAC
4	o	4,729,276	ATGC	52	W	107,223	ACAG
5	n	4,535,686	AGTC	53	L	106,998	ATAG
6	i	4,527,428	AGCT	54	O	105,776	AGAT
7	s	4,186,244	GCTA	55	F	100,951	ACAT
8	r	4,137,989	TACG	56	Y	94,312	ATAC
9	h	2,955,955	TCAG	57	G	93,618	TATC
10	l	2,553,528	TGCA	58	J	78,794	TCCA
11	d	2,369,920	TGAC	59	z	66,509	TGTA
12	c	1,960,612	TAGC	60	j	65,894	TATG
13	u	1,613,333	TCGA	61	U	57,512	TCTG
14	m	1,467,476	GATC	62	q	54,288	TGTC
15	f	1,296,945	GCAT	63	:	54,102	GACC
16	p	1,255,599	GTCA	64)	53,753	GCGA
17	g	1,206,847	GTAC	65	(53,472	GTGA
18	y	1,062,140	AGCG	66	\$	51,586	GAGT
19	w	1,015,755	GACT	67	K	46,612	GCGT
20	,	985,065	CATG	68	;	36,839	GTGC
21	.	946,186	CTGA	69	V	31,104	CGCA
22	b	866,356	CTAG	70	*	20,772	CACG
23	v	653,397	CAGT	71	?	12,481	CTCG
24	0	546,333	CGTA	72	Q	11,872	CGCT
25	l	461,006	CGAT	73	/	8,198	CTCA
26	k	460,798	CATC	74	X	7,682	CACT
27	5	374,503	ACGC	75	&	6,539	AGAG
28	2	333,599	ATCA	76	Z	5,672	ACAC
29	T	325,562	ACTA	77	!	2,201	ATAT
30	S	304,999	AGTA	78	%	2,005	TGTG
31	“	284,771	ATGA	79	+	324	TATA
32	9	282,397	AGCA	80	>	89	TCTC

(continued)

Table 1. (continued)

Rank	Character	Frequency	DNA code	Rank	Character	Frequency	DNA code
33	A	280,987	ACGA	81	<	84	TCGC
34	M	259,574	TCGT	82	=	24	GAGA
35	–	252,382	TGCT	83	#	12	GTGT
36	C	229,383	TAGT	84	@	3	CACA
37	I	223,370	TGAT	85	{	2	CCCG
38	N	205,465	TCAT	86	}	2	CGCC
39	‘	204,593	TACT	87	[2	TGGT
40	4	192,545	GATG	88]	2	GTTG
41	3	187,640	GTAG	89	^	2	GAAG
42	8	182,681	GTCC	90	_	2	GTTG
43	B	169,490	GCTG	91		1	TAAT
44	6	153,881	GACG	92	~	1	TCCT
45	R	146,455	GCAG	93	\	1	CAAT
46	P	144,300	CTAC	94	‘	1	GAAT
47	E	138,459	CGAC	95	€	1	CGGT
48	D	129,645	CAGC	96	£	1	GTA A

The subsequent section gives a detailed insight into how this entropy is calculated and what is the ideal value for this entropy.

3.1 Entropy

Entropy is the measure of uncertainty in bits and this concept was introduced by Shannon in 1948 [32, 33]. The uncertainty of the cipher is the number of plaintext bits that must be recovered from scrambled ciphertext to get the message back, and this is measured via entropy. The entropy of a variable is the weighted average of optimal bit representation size such as the average size of an optically encoded message. Mathematically, entropy can be defined as in (1) [9, 34].

$$H(X) = - \sum_{x \in X} \Pr[X] \log_2(\Pr[X]) \tag{1}$$

Meaning, higher the probability of an event less the uncertainty. Here we are calculating the entropy of X with four bases as $X = \{A, C, T, G\}$. Now, the probability of each DNA base multiplied by log of its probability as modified for four values as in (2):

$$H(X) = -[P(A)\log_2(P(A)) + P(C)\log_2(P(C)) + P(G)\log_2(P(G)) + P(T)\log_2(P(T))] \tag{2}$$

The highest uncertainty is only achieved when the values are equally distributed. Figure 1 explains that the probability of an event that ranges from 0 to 1 and the entropy

can range from 0 to 1. This graph gives an insight into the entropy of an event where two outcomes are considered:

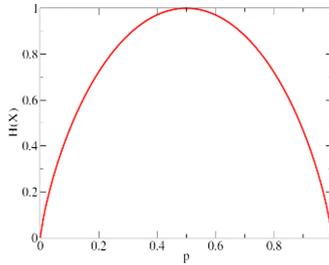


Fig. 1. Entropy and probability distribution

- The value of entropy is 0 for both the least and highest probability, which proves that if the probability of occurrence is 0, an event entropy will be 0, indicating that event will never happen. Similarly if the probability of an event is 1 means this event will always happen against the entropy is 0 because, in this scenario, there is no uncertainty about the information.
- The entropy of the system is maximum “1” when the probability is “ $1/2 = 0.5$ ”. This clearly indicates that all events have the same chance to occur. If the probability increases from “0.5” then entropy decreases and similarly if the probability decreases the entropy also decreases, because in a former event is less likely to occur whereas in later the event is more likely to occur.
- For a system where number of events increases, for example, it has four possible outcomes and the probability range from 0 to 1, its distribution differs. Each event has equal probability $1/4 = 0.25$ only then maximum entropy will be achieved. Here the entropy reaches a maximum value which is 2.

3.2 Entropy Calculation for DNA Table

This section calculates the entropy of the base table that originally described in [8] and followed by the entropy of the proposed table.

Entropy of Base Table

The entropy of the base table is calculated by examining the number of occurrences of each DNA base table in [8], whereby A = 83, C = 106, G = 96, and T = 99. And the total number of bases in the table is 384. By substituting these values in Eq. (2),

$$H(X) = - \left[\frac{83}{384} \log_2 \left(\frac{83}{384} \right) + \frac{106}{384} \log_2 \left(\frac{106}{384} \right) + \frac{96}{384} \log_2 \left(\frac{96}{384} \right) + \frac{99}{384} \log_2 \left(\frac{99}{384} \right) \right]$$

$$H(X) = 1.9164$$

Hence the minimum number of bits needed to encode all possible meaning of the table or number of bits of information per character is 1.92.

Entropy of the Proposed Table

The entropy of the proposed table is calculated by examining the number of occurrences of each DNA base in the table, where A = 96, C = 96, G = 96, and T = 96. And the total number of bases in the table is 384. By substituting these values in Eq. (2),

$$H(X) = -\left[\frac{96}{384}\log_2\left(\frac{96}{384}\right) + \frac{96}{384}\log_2\left(\frac{96}{384}\right) + \frac{96}{384}\log_2\left(\frac{96}{384}\right) + \frac{96}{384}\log_2\left(\frac{96}{384}\right)\right]$$

$$H(X) = 2$$

The minimum number of bits needed to encode all possible meaning in the proposed table or number of bits of information per character is 2. Higher entropy makes conducting frequency analysis harder in DNA cryptography. The DNA table with higher entropy introduces more uncertainty about the ciphertext when an attacker does not have any information about the plaintext. Meanwhile, a lack of good entropy can leave a cryptosystem vulnerable and unable to encrypt data securely. Later, Sect. 4 discusses the experimental result of the entropy for the proposed DNA table.

4 Experimental Results

To compare the performance of both DNA tables, this section describes the encryption algorithm [14] of DNA cryptography as follows..

1. Read input (plaintext)
2. Create a DNA sequence using dynamic DNA table
3. Convert sequence into 2 bit binary
4. XOR binary sequence with random binary key of equal length
5. Convert the sequence of step iv into DNA sequence
6. Use mRNA table to convert sequence from step 5
7. Transfer mRNA to tRNA
8. Divide tRNA into two and interchange their positions
9. Apply the reverse simulation (U to T)
10. Generate cipher text using amino acid table

This algorithm is implemented in MATLAB R2019b. The next subsections describe the performance measurements, where outputs of base and proposed DNA table are compared in terms of frequency count, time entropy, and conditional entropy of key given cipher.

4.1 Frequency Analysis of Encoded Text

After taking the input from the user, the DNA table is used to encode the plain text. Four DNA bases A, C, T, and G replace each plaintext character so an input string of

length, m becomes $m * 4$ in the encoded text. Given below are the graphs, indicating the frequency count of the input string of 3000 characters and 6000 characters as in Fig. 2 and Fig. 3 respectively. An input string of 3000 characters yields an encoded string $3000 * 4 = 12000$ DNA bases. Figure 2 gives the insight of DNA bases frequencies. From Subsect. 3.2 with the calculation of the proposed table entropy, the encoded count of all DNA bases should be equally distributed, thus making a string of 12000 bases has $12000/4 = 3000$ frequency per base. Figure 2 proves that the frequency count of the base table is not uniformly distributed, ranging from 1938–4830 when compared to the proposed DNA table with the range in between 2933 to 3036.

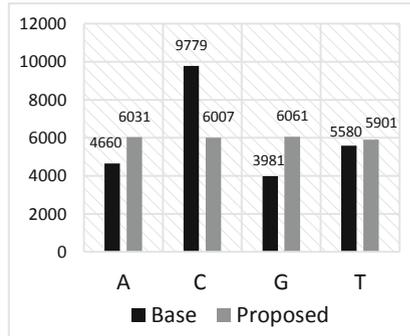
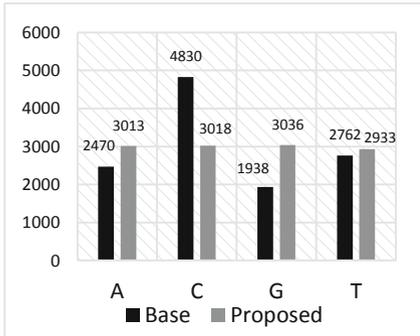


Fig. 2. Frequency count input 3000 characters **Fig. 3.** Frequency count input 6000 characters

Figure 3 shows the frequency count using another input of 6000 characters that creates $6000 * 4 = 24000$ DNA bases. Ideally, the frequency of each base should be 6000 or nearly equal to it, but the frequency count of the DNA base table ranges from 3981–9779 while the proposed DNA table creates the value near to 6000, ranging from 5901–6061.

4.2 Computational Time of Encoding Process

Figure 4 illustrates a time comparison to substitute the same plaintext of equal length using both DNA tables. The graph proves that time difference is negligible in case of small plaintext, but as the size of plaintext increases, the time required to encrypt the plaintext using the proposed DNA table has a significant difference as compared to the base table. Based on the graph, for an input of 50 characters, the time taken by the base table is 0.0539 ms while the proposed table is 0.0522 ms with the time difference is 0.0017 ms. But, as the size of plaintext increases to 1750 characters, the time difference is 0.0646 ms, and for 2500 characters, the time difference has increased to 0.9931 ms. This difference is mainly because of the frequency analysis performed prior to table creation, as most occurring characters are at the start so the loop does not iterate through the whole table. Instead, it immediately encodes the character and exits but in the base table, the characters are randomly placed so the time increases with the size of the input.

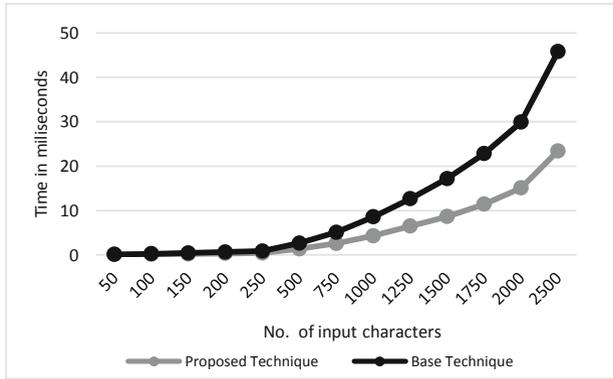


Fig. 4. Time comparison of encoding process using the DNA base and proposed table

4.3 Frequency Analysis of Ciphertext

At the end of this algorithm, the DNA cryptography generates the ciphertext. The text DNA sequences are mapped to the amino acid table. The final ciphertext is in the form of 26 English characters. Figure 5 implies the frequency analysis of the ciphertext with 300 characters plaintext as the input of the encryption algorithm. Based on both DNA tables, it is obvious that frequency ranges from 0–66 for the base technique whereas it ranges from 0–36 for the proposed technique. Meanwhile, Fig. 6 gives a frequency analysis of 2400 characters plaintext as the input. For the base technique, the frequency count ranges from 11–451, while for the proposed table, it ranges from 11–308. This frequency difference is important from an attacker’s perspective to conduct a frequency analysis attack due to the randomness of the ciphertext. This way, he cannot extract meaningful information about the plaintext from when he does not know the secret key [9, 35].

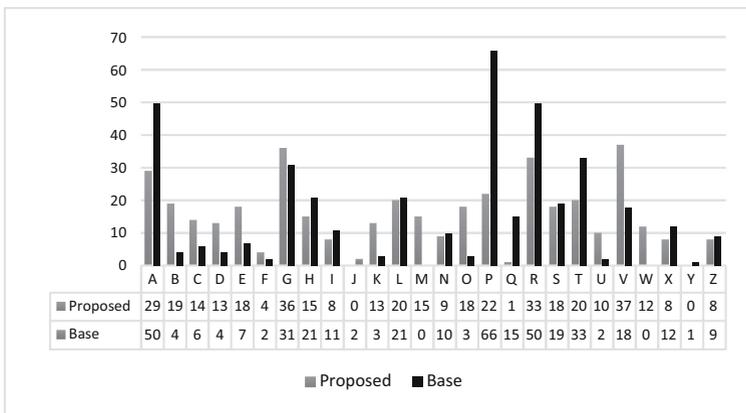


Fig. 5. Frequency analysis of ciphertext (with 300 characters)

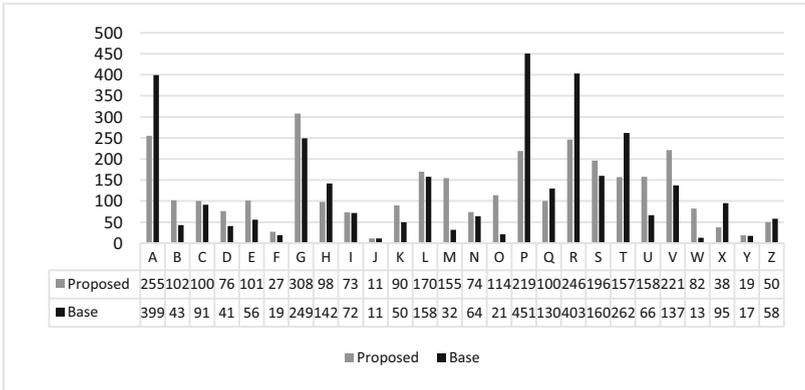


Fig. 6. Frequency analysis of ciphertext (with 2400 characters)

4.4 Computational Time of the DNA Encryption Technique

Time is an important factor when it comes to the computational complexity performance of the algorithm. Figure 7 shows the encryption time that is calculated for multiple inputs. The inputs are the number of characters in the plaintext and computational time is the total encryption time in seconds. The same input is provided to both experiments using the base and the proposed DNA tables. The graph indicates that the encryption time for the proposed table is significantly less than using the base table. Based on the graph, for an input of 600 characters, the time taken by the proposed table is 0.1624 s while and the base table is 0.248 s. One of the factors for this time difference is because the table is designed by taking into account that the characters with high rank or frequency must be in start so that loop traverse time reduces.

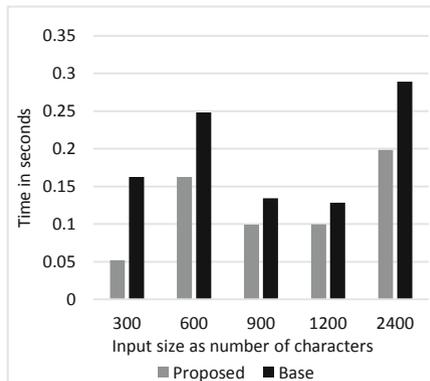


Fig. 7. Encryption time comparison

4.5 Entropy Analysis

The entropy of the cryptosystem varies as the number of inputs or sample space varies. Figure 8 gives the entropy of different ciphertexts, provided the same plaintext and the same random key. The X-axis is the input size or the number of characters, whereas the Y-axis is the entropy of ciphertext $H(C)$. For input size ranging from 300 to 2400, it is obvious that the entropy of the proposed technique is more than the entropy of the base technique. This difference is because at the time of encoding, at a very early stage of encryption algorithm the frequencies of encoded text are nearly equally distributed, and highly ranked characters have most random codes. As explained in [9], the entropy of English characters is $26 \log_2 26 = 4.7$ which is an ideal entropy value for the proposed cipher. In the case of 2400 characters as the input, the entropy of the proposed technique is approximately 4.4 which is not very less than ideal value, while for the base table, the entropy is in between 4.09 to 3.91 which is less than ideal.

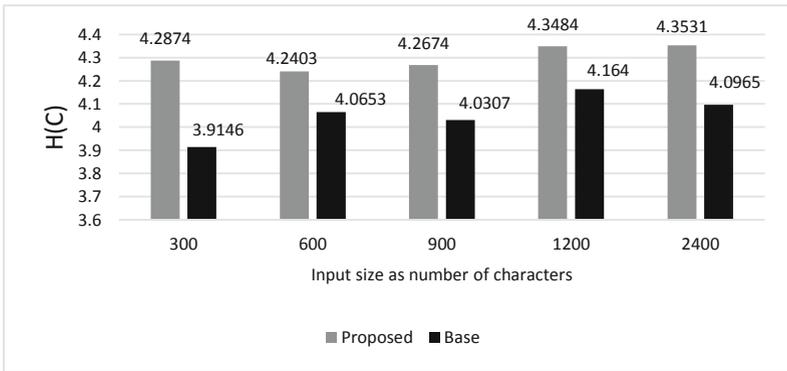


Fig. 8. Entropy of ciphertext

4.6 Conditional Entropy

According to Kerchoff’s Principle, the security of the cryptosystem depends primarily on having strong keys and keeping them secret, but not in the encryption-decryption algorithm because it can be accessed [35]. Thus, the system must ensure that there is no information leakage about key or plain text from the ciphertext. Conditional entropy is also called key equivocation when it comes to cryptography. Key equivocation of a cryptosystem can be described as in (3) with $M^N =$ plaintext with length N , $C^N =$ ciphertext with length N and $K =$ random key [24].

$$H(K/C^N) = H(K) + H(M^N) - H(C^N) \tag{3}$$

Key equivocation is a process in cryptanalysis where the attacker has only access to ciphertext and he tries to infer some information about the key. It is also known as a ciphertext-only attack. Given below is the calculation of conditional entropy in

both scenarios whereby the encryption uses the base and the proposed table. Key and plaintext are the same for both cases with the number of input characters of 1200. Using the equation in (3) to calculate the conditional entropy for base table, $H(K/C^N) = 0.9994 + 4.7615 - 4.164 = 1.5969$ and the conditional entropy for the proposed table, $H(K/C^N) = 0.9994 + 4.47615 - 4.3484 = 1.1272$.

This conditional entropy is the information being leaked or it can be stated as the amount of information of the key with the given cipher. The result shows that the proposed technique reveals less information about the key than the base technique. Hence, the proposed table technique can serve as a good substitute for the base technique for improvements in terms of frequency analysis, computational time, and entropy.

5 Conclusion

Traditional cryptographic techniques are designed based on substitution, and transposition operations. With advancements of technology however compromise the security of the cryptographic algorithms. Researchers propose new security solutions to overcome security issues including DNA cryptography which involves PCR, DNA synthesis and digital coding. DNA coding is mostly used for encryption-decryption techniques with a basic table of 96 ASCII characters to encode the plaintext into DNA bases. This research work has improved that table, by carefully constructing the table based on frequency analysis, randomness in code and entropy of the table. Later, the paper compares the performance of the proposed table with the base table which originally designed for DNA cryptography. The results indicate that the proposed table gives a balanced frequency of occurrence in the encoded text and ciphertext, reduces encoding time based on DNA bases and encryption time, better entropy of the ciphertext and finally the conditional entropy is less than the base technique. The entropy of the proposed table is slightly less than the base table but it has a huge impact on the output. Hence, this table can be used for better security and computational time of DNA encryption techniques.

Acknowledgement. The authors would like to thank all reviewers for their helpful comments. The authors would also like to thank the Ministry of Higher Education Malaysia for supporting the research under Fundamental Research Grant Scheme Vot No. FRGS/1/2019/ICT03/UTHM/03/1 and partially sponsored by Universiti Tun Hussein Onn Malaysia.

References

1. Lu, M.X., et al.: Symmetric-key cryptosystem with DNA technology. *Sci. China Ser. F: Inf. Sci.* **50**(3), 324–333 (2007)
2. Anam, B., Sakib, K., Hossain, M.A., Dahal, K.: Review on the advancements of DNA cryptography. *arXiv preprint arXiv:1010.0186* (2010)
3. Kabir, K.S., Chakraborty, T., Alim Al Islam, A.B.M.: SuperCrypt: a technique for quantum cryptography through simultaneously improving both security level and data rate. In: *Proceedings of the 2016 International Conference on Networking System and Security*, pp. 25–33 (2016)

4. Pelletier, O., Weimerskirch, A.: Algorithmic self-assembly of DNA tiles and its application to cryptanalysis. arXiv preprint [arXiv:cs/0110009](https://arxiv.org/abs/2001.01009) (2001)
5. Galbraith, S.D.: *Mathematics of Public Key Cryptography*. Cambridge University Press, New York (2012)
6. Zhang, Y., He, L., Fu, B.: Research on DNA cryptography. In: *Applied Cryptography and Network Security*. InTech, Rijeka (2012)
7. Adleman, L.M.: To combinatorial problems. *Science* **266**(5187), 1021–1024 (1994). <https://doi.org/10.1126/science.7973651>. Bibcode: 1994Sci...266.1021A. CiteSeerX 10.1.1.54.2565
8. Noorul Hussain, U., Chithralekha, T.: Inventors; assignee. A novel DNA encoding technique and system for DNA cryptography. India Patent 5107, CHE, 2012, 7 December 2012
9. Othman, H., Hassoun, Y., Owayjan, M.: Entropy model for symmetric key cryptography algorithms based on numerical methods (2015). <https://doi.org/10.1109/arcse.2015.7338142>
10. Borda, M., Tornea, O.: DNA secret writing techniques. In: *Proceedings of the 8th International Conference on Communications (COMM)*, pp. 451–456 (2010)
11. Zhang, X., Zhou, Z., Niu, Y.: An image encryption method based on the feistel network and dynamic DNA encoding. *IEEE Photonics J.* **10**(4), 3901014 (2018)
12. Zhang, Y., Liu, X., Sun, M.: DNA based random key generation and management for OTP encryption. *Biosystems* **159**, 51–63 (2017)
13. Biswas, R., Alam, K.M.R.: A technique for DNA cryptography based on dynamic mechanisms. *J. Inf. Secur. Appl.* **48**, 102363 (2019)
14. Hossain, E.M.S., Alam, K.M.R., Biswas, M.R.: A DNA cryptographic technique based on dynamic DNA sequence table. In: *Proceedings of the 19th International Conference on Computer and Information Technology*, 18–20 December 2016, North South University, Dhaka, Bangladesh. IEEE (2016)
15. Hussain, N., Rahman, U., Balamurugan, C., Mariappan, R.: A novel DNA computing based encryption and decryption algorithm. In: *International Conference on Information and Communication Technologies (ICICT 2014)*, *Procedia Computer Science*, vol. 46 pp. 463–475 (2015)
16. Ibrahim, F.E., Moussa, M.I., Abdalkader, H.M.: A symmetric encryption algorithm based on DNA computing. *Int. J. Comput. Appl.* **97**(16), 41–45 (2014)
17. Babu, E.S., Raju, C.N., Prasad, M.H.M.K.: Inspired pseudo biotic DNA based cryptographic mechanism against adaptive cryptographic attacks. *Int. J. Netw. Secur.* **18**(2), 291–303 (2016)
18. Cui, G., Qin, L., Wang, Y., Zhang, X.: An encryption scheme using DNA technology. In: *Proceedings of the 3rd International Conference on Bio-Inspired Computing: Theories and Applications, USA*, pp. 37–42. IEEE (2008)
19. Lai, X., Lu, M., Qin, L., Han, J., Fang, X.: Asymmetric encryption and signature method with DNA technology. *Sci. China Inf. Sci.* **53**(3), 506–514 (2010)
20. Kulsoom, A., Xiao, D., Aqeel-ur-Rehman, Abbas, S.A.: An efficient and noise resistive selective image encryption scheme for gray images based on chaotic maps and DNA complementary rules. *Multimed. Tools Appl.* **75**, 1–23 (2016). <https://doi.org/10.1007/s11042-014-2221-x>
21. Zhang, Y.Q., Wang, X.Y.: A symmetric image encryption algorithm based on mixed linear-nonlinear coupled map lattice. *Inf. Sci.* **273**, 329–351 (2014)
22. Wang, X.Y., Zhang, H.L.: A novel image encryption algorithm based on genetic recombination and hyper-chaotic systems. *Nonlinear Dyn.* **83**, 333–346 (2016)
23. Zhang, W., Yu, H., Zhao, Y.L., Zhu, Z.L.: Image encryption based on three-dimensional bit matrix permutation. *Signal Process.* **118**, 36–50 (2016)
24. Wang, X.Y., Zhu, X.Q., Zhang, Y.Q.: An image encryption algorithm based on Josephus traversing and mixed chaotic map. *IEEE Access* **6**, 23733–23746 (2018)

25. Parvaz, R., Zarebnia, M.: A combination chaotic system and application in color image encryption. *Opt. Laser Technol.* **101**, 30–41 (2018)
26. Prajapati Ashishkumar, B., Barkha, P.: Implementation of DNA cryptography in cloud computing and using socket programming. *IEEE* (2016)
27. Hammami, H., Brahmi, H., Yahia, S.B.: Secured outsourcing towards a cloud computing environment based on DNA cryptography, pp. 31–36. *IEEE* (2018)
28. Chouhan, D.S., Mahajan, R.P.: An architectural framework for encryption & generation of digital signature using DNA cryptography, pp. 743–748. *IEEE* (2014)
29. Keerthana Priya, S.V., Saritha, S.J.: A robust technique to generate unique code DNA sequence, pp. 3815–3820. *IEEE* (2017)
30. Sadeg, S., Gougache, M., Mansouri, N., Drias, H.: An encryption algorithm inspired from DNA. In: *Proceedings of the International Conference on Machine and Web Intelligence*, pp. 344–349. *IEEE* (2010)
31. Rashid, O.F., Othman, Z.A., Zainudin, S.: A novel DNA sequence approach for network intrusion detection system based on cryptography encoding method. *Int. J. Adv. Sci. Eng. Inf. Technol.* **7**(1), 183–189 (2017)
32. Jones, M.N., Mewhort, D.J.K.: Case-sensitive letter and bigram frequency counts from large-scale English corpora. *Behav. Res. Methods Instrum. Comput.* **36**, 388 (2004). <https://doi.org/10.3758/BF03195586>
33. Shannon, C.E.: A mathematical theory of communication. *Bell Syst. Tech. J.* **27**(3), 379–423 (1948). <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
34. Shannon, C.E.: A mathematical theory of communication. *Bell Syst. Tech. J.* **27**(4), 623–656 (1948). <https://doi.org/10.1002/j.1538-7305.1948.tb00917.x>
35. Mohan, M., Kavitha Devi, M.K., Jeevan Prakash, V.: Security analysis and modification of classical encryption scheme. *Indian J. Sci. Technol.* **8**(S8), 542–548 (2015). ISSN (Print): 0974-6846 ISSN (Online): 0974-564



Solving Generic Decision Problems by in-Message Computation in DNA-Based Molecular Nanonetworks

Florian-Lennert Adrian Lau^(✉), Regine Wendt, and Stefan Fischer

Institute of Telematics, Ratzeburger Allee 160, 23562 Lübeck, Germany
{lau,wendt,fischer}@itm.uni-luebeck.de

Abstract. One of the biggest unsolved problems in nanonetwork research is the actual construction of the components required for building such networks. Most existing ideas are limited to partial solutions of construction of nanodevices, computation within them, and communication between them. While many ideas are promising, the problem remains how to combine those various building blocks into operational and efficient nanonetworks.

In this paper we use DNA as *the* basic building block for all components of nanonetworks. The inherent properties of this molecule are used to assemble complex nanostructures. DNA can be utilized to create both nanodevices and a communication mechanism. Properly designed DNA-molecules can even be utilized for computational purposes. In summary, DNA forms the base for an exhaustive nanonetwork concept.

This work specifically presents an approach how to solve arbitrary mathematical problems that can be modeled as boolean formulas using DNA-based nanonetworks by in-message computation. The computation itself is encoded in the assembly process of a message. This avoids often-stated space constraints for computations at the nanoscale, as the medium of transportation is commonly less constrained than the size of nanodevices dictates. This method thereby presents a constructive approach on how to actually create message molecules, rather than only proving the general possibility.

Keywords: Nanonetworks · Tile-based self-assembly · Nanostructures · Molecular communication · Decision problems

1 Introduction

After the concept of nanonetworks and their use in, e.g., medical applications have been adequately described in recent years, research groups worldwide are now looking into the question of how such nanonetworks can actually be constructed. A promising approach is the paradigm of self-assembly: components of nanonetworks such as sensors, actuators and even the messages exchanged

between them can grow like crystals [7] instead of being manually engineered by humans – a process hardly imaginable on the nanoscale [4].

Crystal formation is a natural instance of the principle of self-assembly. It is based on tiny and often simple components assembling themselves into larger and more complex structures according to local rules. In order to make this process usable in the context of building nanonetworks, the resulting complex structures should be nanodevices, nanosensors, message molecules, fully functional nanonetworks or even computers [2]. However, it is hardly possible to influence the process. A very fortunate exception are certain building blocks made from DNA. These also follow the paradigm of self-assembly and can be treated like crystals [16].

The basic entities we are looking at are DNA strands. It is possible to create arbitrary DNA strands in the laboratory [13]. One can actually create DNA building blocks out of DNA strands that behave like puzzle pieces [17]. These puzzle pieces of DNA are also called *tiles* or *DNA-tiles*. Tiles can be designed in such a way that they bind with other tiles only in a specific and predetermined way. It has been shown that this characteristic can be used to create almost any structure at the nanoscale [8]. Additionally, tiles can also perform computations through conditional binding processes [9, 17]. All these building processes execute – after an initial tile set has been provided – without the need for any further human intervention [7]. Also, all the materials required for this approach can be produced in a laboratory, so that the first rudimentary nanonetworks can really be built and operated in-vitro.

DNA-based nanonetworks, in their most primitive form, are unfortunately not programmable once the devices have self-assembled. Further, implementing conventional algorithms with just a small set of tiles is much less well researched than conventional programming languages. As a consequence, the purpose of a nanonetwork has to be known before its self-assembly, and the tiles have to be created and selected accordingly. While a proof of concept for this approach has already been presented in [7], it is still mostly unknown how nanonetworks for specific mathematical problems can be created, i.e., given a specific problem formulation, how many different tiles are needed and how they look like.

This paper presents a novel approach on how to create the necessary building blocks for specific, but arbitrary problems, namely those that *can be modeled as a boolean formula*. Previous findings that tile-based self-assembly systems are Turing-complete [17] are of mere theoretical interest, as the proof gives no algorithm on how to compute in a feasible manner [14]. This paper bridges the gap by offering a solution for the complexity classes AC^0 and NC^1 . AC^0 and NC^1 are classes of problems that describe strongly restricted circuits of limited size, among other [3, 15]. Important problems of these classes are basic arithmetic functions like addition and multiplication, as well as logical operations and the computation of thresholds. These form the basic necessary operations to create more complex communication protocols. Looking at possible applications, nanonetworks can be created that react, upon measuring environmental param-

eters, once a predefined threshold has been reached, thus increasing reliability. A logical AND can be used to establish a distributed consensus among nanodevices, as demonstrated in [7].

The remainder of the paper is structured as follows: Sect. 2 presents the necessary definitions to understand the mathematical models and the process of self-assembly in general. Section 3 introduces the definitions and general ideas behind DNA-based nanonetworks. Section 4 suggest an algorithms that creates message molecules for problems that can be modeled as boolean formulas. Section 5 explains a generic nanonetwork architecture. Section 6 summarizes by presenting a list of problems that can be solved by DNA-based nanonetworks via the developed procedure.

2 Preliminaries for Tile-Based Self-Assembly

This section gives a brief introduction and examples of tile-based self-assembly systems from Eric Winfree [17]. The notation used in the preliminaries is based on [5] and has already been presented in [7]. For a detailed definition of self-assembly systems and an overview of the most important results regarding self-assembly systems, please consult [11].

Tiles are the basic components of self-assembly systems. A 2D-tile is a square in \mathbb{Z}^2 and a 3D-tile is a cube in \mathbb{Z}^3 . Figure 1(a)–(d) show examples of two-dimensional and three-dimensional tiles. Models with focus on the mathematical functionality (a, b) and with focus on the biological components (c, d) are presented.

From here on, the dimension is largely omitted. Unless otherwise specified, two-dimensional tiles are implied by the symbol t .

Each side of a tile can have any number of *glues*, with a corresponding *binding strength*. The glues and their strengths are depicted by a number of black boxes on each side. In Fig. 1 all glues are of strength $s = 1$. The glue color is indicated by a label, here N, E, W and S.

Additionally, each tile has a *marker* in the middle. The marker is also represented by a label and can be implemented by a florescence marker biologically. It is used to encode semantic properties of a tile. An example would be the representation of an encoded truth value.

In the DNA-computing research community, tiles are made from DNA [11]. Glues are implemented by open DNA strands with freely selectable base sequences. This enables the implementation of both color, by the use of a specific base sequence, and strength, by the corresponding length of the open strands.

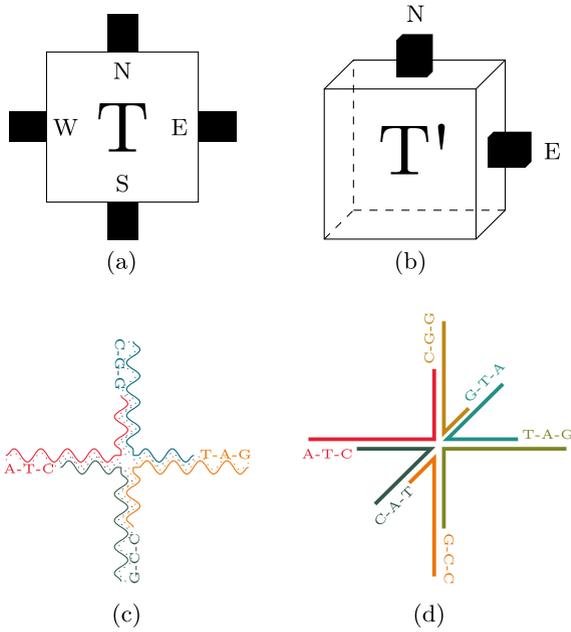


Fig. 1. Examples of tiletypes with mathematical and biological representation in 2D and 3D. The modeled glues are displayed as black cubes, their type as a label. The biological glues are illustrated by base sequences. [1]

Examples of DNA-tiles are shown in Fig. 1(c, d). The open strands with the displayed bases adenine (A), thymine (T), cytosine (C) and guanine (G) can only bind with a matching complementary sequence.

A tile can have no or exactly one neighbor on each side. Tiles t and t' only interact with each other when they are neighbors. The interaction rules are determined by the glues.

If glues are not explicitly shown or mentioned, the empty glue $\mathcal{L}_g = \emptyset$ with the label VOID and the strength 0 is assumed.

A tile t , with a glue strength 0 or no glue at a side can't interact with other tiles t' at that side.

Two tiles are of the same *tilettype*, iff they have identical glue color, glue strength and markers.

The *temperature* τ of a self-assembly system describes the minimum glue strength s that is required for a tile to form a stable bond with other tiles.

Two tiles bind with each other if they are neighbors and have glues of a suitable color and if the binding strength with all neighbors is at least equal to the temperature. Meaning that two tiles can only bind at temperature 2 if they share a glue with the same color and 2 black boxes.

The temperature τ models the physical temperature. If the temperature τ is increased, molecules move faster. The additional energy makes it more likely that bonds are destroyed or cannot form in a stable manner.

In this paper, τ is assumed as 2 or 3 and omitted in the notation. Temperatures 2 and 3 are sufficient to implement all functionality relevant to the scenarios discussed in this paper.

In order for tiles to bind together, they must be rotated to match each other. In biological systems based on DNA this happens automatically. Mathematical models require that tiles are not rotatable.

Due to the intrinsic properties of DNA, errors are unavoidable [17]. However, the number of errors during the self-assembly process can be reduced if special tiletypes are used. In [11], procedures are described that greatly reduce the probability of errors in two dimensions. In [7] this reduction is further explained in the context of nanonetworks and in [1] a procedure for three dimensions is introduced.

The *binding strength* s of tiles t and t' is equal to the number of matching glues with all neighbors of a tile.

The *total binding strength* of a tile t is the sum of binding strengths between t and all of its correct neighbors. Incorrect neighbors do not contribute to the total binding strength. An *assembly* is the result of two or more tiles forming a stable binding with each other.

The *border* of an assembly α is a subset α' of α . The border only contains tiles with at least one unoccupied neighboring position. A tile t correctly binds with an assembly α if the total binding strength of t with neighboring tiles of assembly α is at least equal to the temperature τ . Only border-tiles of assemblies may interact with free tiles.

The *growth front* of n -dimensional assemblies is a subset of positions of \mathbb{Z}^n . A position is part of the growth front iff it is unoccupied and neighbor to a border-tile that has a positive glue strength at that side.

The positions of the growth front change when tiles are added to or removed from the assembly. It is assumed that exactly one tile is added or removed at any discrete time.

The initial assembly α_0 at time 0 is called the *seed-assembly* or *seed-tile* σ .

Starting with σ , tiles are added non-deterministically to the assembly at the growth front. Due to the inherent non-determinism of self-assembly systems, it can be difficult to create tilesets T that grow into desired structures.

An *assembly sequence* of a self-assembly system is a sequence $S = \langle \alpha_0, \alpha_1, \dots \rangle$, while α_{i+1} is created from α_i by adding a tile to α_i . The last element of a finite sequence is called the *result* or *terminal*.

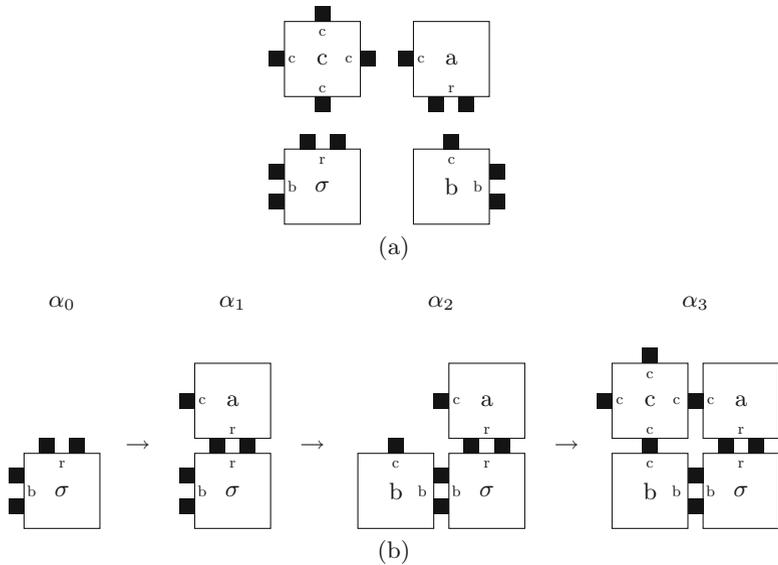


Fig. 2. (a) Example for a tileset of a 2D-self-assembly systems. (b) The corresponding assembly sequence of the tileset from (a). The seed-tile is called σ and the temperature is 2. $S = \langle \alpha_0, \dots, \alpha_3 \rangle$ shows the three steps until the assembly is terminal. [7]

Figure 2(a) shows an example tileset of a 2D-TAM with seed-tile σ . The assembly sequence of the TAM \mathcal{T}_τ at temperature $\tau = 2$ is depicted in Fig. 2(b). α_0 describes the assembly at time 0. α_1 , α_2 and α_3 show the assembly after the addition of one tile at a time. In α_3 , the required correct glues originate from two different neighboring tiles.

3 DNA-Based Nanonetworks

In the following sections different components are defined, which are necessary for the creation of DNA-based nanonetworks. The idea of using DNA-tiles as construction material for nanonetworks, their computations and communication mechanisms goes back to [7]. For a detailed description of DNA-based nanonetworks see [7].

First, a communication mechanism based on tiles is described. This is followed by an explanation of how receptors and ligands can be realized using DNA. The individual components are then combined and formally defined as DNA-based nanonetworks.

3.1 Message Molecules

Since self-assembly systems can also be used as computational models, the presented approach has the advantage that computations can be integrated into the assembly process of a message molecule.

A message molecule can be designed in such a way that specific tiles are necessary to achieve a fully assembled message molecule. By making sure that these tiles are only distributed under certain conditions – e.g. in case of an event – they can be interpreted as input for a computation. Other tiles, also required for the computation, can be kept in the medium as required [7].

This method ensures that, for example, a ligand necessary for a binding reaction is only formed on a message molecule once a computation has been finalized. A ligand is the part of a molecule that can form a bond with a receptor.

Tiles and assemblies are subject to Brownian motion. Brownian motion can be used as a distribution mechanism to transport tiles to the required positions. The process is largely random, which is why a large number of message molecules is required.

Any decision problem [14] can be solved by a self-assembly system at temperature 2 [17]. However, it is unclear whether nanobots will perform computations in the same way as macroscopic computers and whether there is enough space for complex computations at the nanoscale [6]. Further, it is mostly unknown how to create a space efficient self-assembly system that solves a given decision problem using message molecules. In the following sections a generalized method that creates message molecules for many decision problems is explained.

Since self-assembly systems solve a large number of challenges at once, while they can already be created in the laboratory, it is compelling to use this previously proven technology.

Definition 1. *A message molecule \mathcal{M}_Φ is a tiling set T , which computes the boolean formula Φ and creates a ligand in the case of a successful computation.*

3.2 Forming Ligands

The ligands are binding sites on message molecules, which enable a correct binding with nanobots. These are modeled by the glues of tiles.

The variable size and shape of message molecules require additional tiletypes to always provide a uniform starting point for the composition of a ligand. Furthermore, it should only form if a computation has been successfully completed.

Figure 3 shows an example of a ligand. Since it is a temperature 2 system, the two tiles with the label “R” can only bind to the assembly once the middle row of the adjacent message molecule is complete. This in turn depends on the successful assembly of the remaining assembly [7]. After the whole ligand is finalized, the ligand can bind to the receptor from Fig. 4.

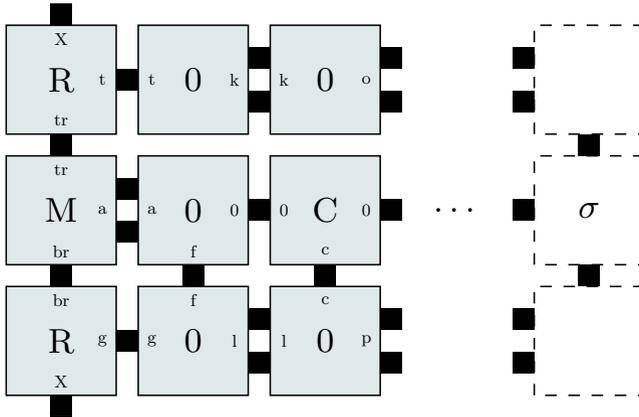


Fig. 3. Generic ligand of a message molecule at temperature 2. The tiles with marker “R” can only attach to the assembly if tile C has a right neighbor. The white tiles represent an adjacent message molecule that has to be finalized before the ligand can form. [7]

3.3 Message Receptors

Receptors are the parts of nanobots that can bind message molecules. They are modeled by tiles with appropriate glues. These can be generated by simple assemblies. Unlike ligands, they are not tied to the successful computation of a formula Φ .

Receptors can be of any shape as long as they bind their corresponding ligand without overlapping. In this case the temperature restriction must be observed. Furthermore, the externally available glues of a receptor must be at least one tile apart from each other to prevent premature binding of parts of the ligand to the receptor.

Figure 4 shows an example of a possible receptor for a message molecule. The gray squares represent any part of a ligand message. The black squares represent individual glues of strength 1. The labels of the receptor’s glues encode the binding condition. If the receptor is correctly designed, it can “recognize” the outermost three tiles of the message molecule at once.

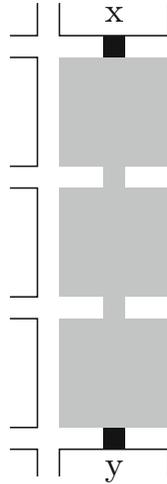


Fig. 4. Generic receptor, which can stably bind message molecules at temperature 2. The grey tiles represent the ligand of a message molecule with unbound glues x and x that are spaced one tile apart from each other. The white tiles are part of a nanobot with unbound glues x and y that allow message molecules to bind. [7]

4 Creating Generic Message Molecules

Although it is known that self-assembly systems are Turing-complete at temperature 2, this finding cannot be transferred directly to message molecules. In this section a procedure is presented, with which many mathematical problems can be compiled to message molecules. This demonstrates the potential versatility of DNA-based nanonetworks.

Theoretical computer science has proven that different types of problems can be transformed into each other. Since decision problems can easily be processed at the nanoscale, these are of particular interest. Theorem 1 proves a procedure that creates a tiling set that assembles into message molecules that solve a given boolean formula Φ .

Theorem 1. *For each decision problem modeled as a boolean formula Φ , a corresponding message molecule can be created that executes the same computation and only forms a ligand if the result of the computation is “1”.*

Proof. Let Φ be a boolean formula. We consider every truth assignment that satisfies the formula Φ .

Every boolean formula may be transformed into disjunctive normal form [12]. This is done by setting up the truth table for said formula. Table 1 shows an example.

Table 1. Example truth table for the formula $A \wedge B$.

A	B	$A \wedge B$
1	1	1
1	0	0
0	1	0
0	0	0

For each line of the truth table that contains “1” in the last column a clause is created. A clause is a formula ϕ_i that only contains literals, negative literals and the logical \wedge . The process is repeated for every row of the truth table. All resulting formulas are then combined with the logical \vee operator. The result is a canonical formula that performs an equivalent computation as the original boolean formula Φ . The produced formula is called the disjunctive normal form of a formula Φ . The size of the result can be further reduced by procedures like McCluskey or Karnaugh maps [10].

For every formula in disjunctive normal form a tileset can be generated as follows:

For every sub formula a message molecule is created. The number of rows of each message molecule is equal to the number of literals times 3. Figure 5 shows an example for the formula from Table 1.

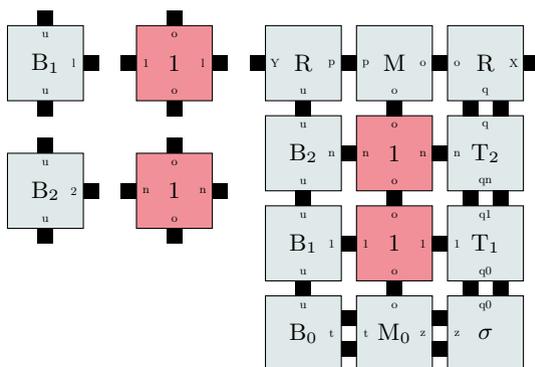


Fig. 5. Example tilesets for the DNF formula from truth Table 1. For every row i of the truth table that evaluates to “1”, two tiletypes for each literal are created (left). These assemble into the message molecule (right).

The truth value of each literal is represented by the label of the corresponding tile as well as by its glues. Only if all literals are part of the message molecule a ligand is formed. Each ascending row of the message molecule can only form once the previous row assembled fully.

Since there is a message molecule for each sub formula and all the sub formulas are combined by a logical \vee , the completion of a single message molecule is sufficient to communicate a successful computation. \square

5 Arbitrary Nanonetworks for Boolean Formulas

In [6] a list of relevant mathematical operations for nanonetworks is presented. Most of them can be solved by representing them as a boolean formula Φ that can be transcribed into a message molecule by applying the methodology from Sect. 4.

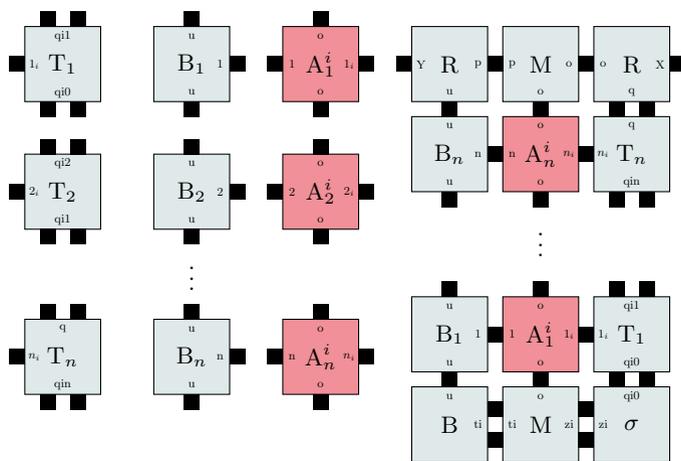


Fig. 6. Tileset (left) and fully assembled message molecules (right) for a boolean formula Φ . The tileset to the left contains three tiletypes for every literal in line i of the corresponding truth table and only contains tiles that are used more than once. To the right, the message molecules for the clauses of the disjunctive normal form (DNF) of the formula Φ are displayed.

Figure 6 shows an example for the general architecture with n input bits. i represents the index of the row of a truth Table 2 that evaluates to “1”. The truth table represents an arbitrary boolean formula Φ . The disjunctive normal form of the formula Φ can directly be derived from the table.

The assembly process of the message molecules starts with the seed-tile σ at the bottom right. Since the temperature is 2, The bottom row of M and B can form, as well as the right stack of tiles $T_1 \dots T_n$. The remaining assembly process can only proceed if the red tiles A_n^i , that represent the literals of each clause, are present. Once all required tiles A_n^i are bound to the assembly the tiles $B_1 \dots B_n$ can attach, followed by the left tile R , which completes the ligand.

Table 2. Generic truth table for a boolean formula Φ . The columns represent all occurring sub formulas. Each row i that evaluates to “1” represents a truth assignment that satisfies the formula Φ and serves as input for a message molecule. The red line shows an example.

ϕ_1	ϕ_2	...	Φ
1	1		1
1	0		1
		⋮	
0	1		0
0	0		1

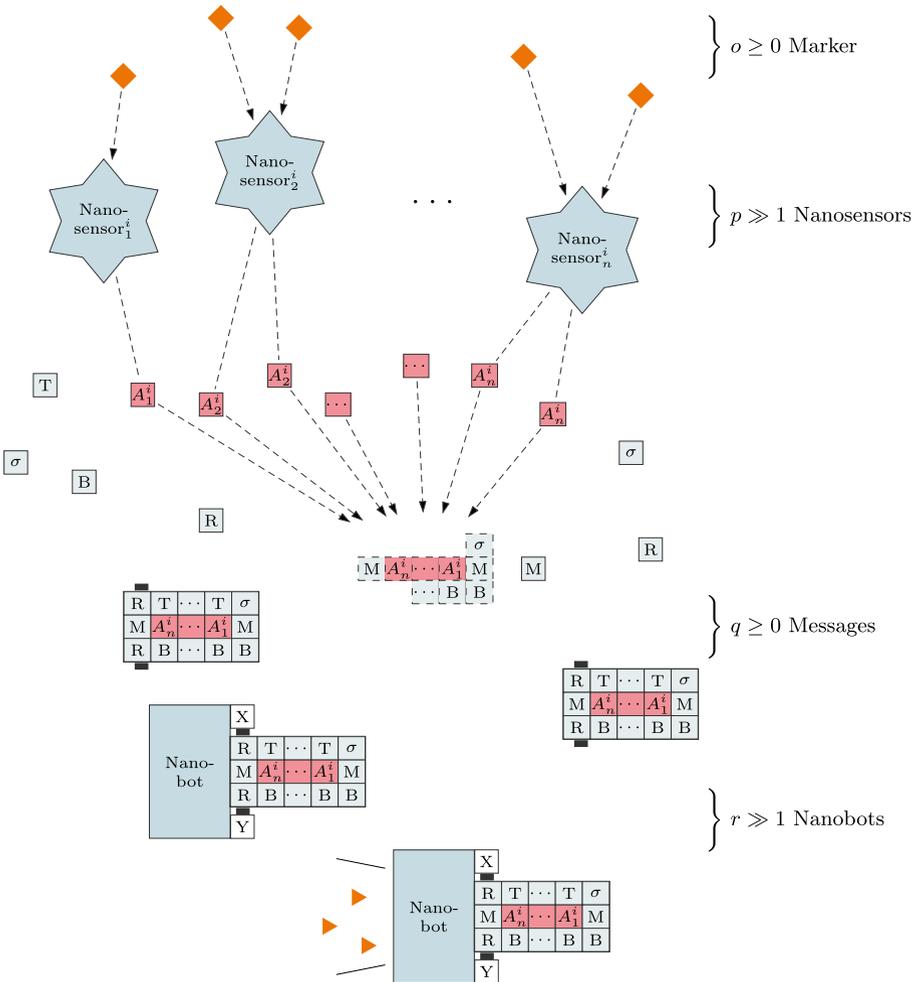


Fig. 7. The general reference architecture for a DNA-based nanonetwork. It consists of $o \geq 0$ markers (orange rhomboids), $p \gg 1$ nanosensors, $q \geq 0$ message molecules and $r \gg 1$ Nanobots. (Color figure online)

The binding of one of the created message molecules to a nanobot suffices to implement the required logical \vee . Combined, the result is a behavior that maps the original problem. The corresponding nanonetwork is shown in Fig. 7. It depicts a set of nanosensors i_n that measure specific markers (orange rhomboids) and release tiles A^i_n into a medium once a successful measurement has been conducted. Together with tiles that are always present in the shared medium, the tiles assemble into message molecules that compute the clauses of the given boolean formula Φ .

Only in the case of a successful computation a ligand is formed and the resulting message molecule can bind to a nanobot which then can release medical payload or communicate the measurement (orange triangles).

Apart from message molecules, receptors and ligands, nanosensors and nanobots can also be created from tiles. Consequently, we define a nanonetwork as a tileset. Nanobots and nanosensors are modeled as hollow cubes, which may be opened once a specific message molecules or markers binding to them.

Definition 2. *A tile-based nanonetwork \mathcal{N}_Φ is a tileset $\mathcal{N}_{S_e} \cup \mathcal{N}_R \cup \mathcal{M}_\Phi$. \mathcal{N}_{S_e} is a tileset for nanosensors. \mathcal{N}_R is a tileset for nanobots and \mathcal{M}_Φ a tileset for message molecules, which compute the function Φ .*

6 Conclusion and Future Work

This work advances the modeling framework for DNA-based nanonetworks presented in [7]. While [7] showed that it is in-fact possible to create nanonetworks that solve arbitrary decision problems, it was not elaborated how to achieve this goal. The viability of the approach was exemplified by the implementation of a DNA-based nanonetwork for the mathematical operation AND with four bits.

In this work we closed the gap by presenting and proving a procedure that creates message molecules for any decision problem that can be modeled by a boolean formula. This increases the number of known, feasible computations by DNA-based nanonetworks tremendously. In [6] a list of mathematical operations that are of interest to nanonetworks has been assembled. Table 3 shows a subset of the entries. The left column shows problems that were derived from medical use-case scenarios, while the right column displays additional problems from the same complexity class. For all of the depicted problems a corresponding decision problem can be defined and solved by our methodology. This can easily be deduced since all of the problems can be solved by circuits, which directly implies that they can be expressed as a boolean formula.

As predicted by [6], the computations of AC^0 -messages and NC^1 -messages are relatively simple to implement at the nanoscale.

The complexity classes AC^0 and NC^1 describe circuits with specific limitations. Boolean circuits characterized by AC^0 are of polynomial size and constant depth in regard to the number of input bits. The gates of NC^1 have two inputs per gate at most, but can be of logarithmic depth in regard to the number of input bits. For a detailed description consult [6]. Both complexity classes represent comparably easy problems.

Table 3. List of problems sorted by complexity classes. The class NC^1 includes the class AC^0 and the problems in AC^0 are therefore considered *easier*. Both AC^0 and NC^1 describe problems that require very little space to be computed. Table adapted from [6].

	Medical problems	Additional problems
AC^0 -message:	ADD	ODD/EVEN
	SUB	DIV ₂
	SIGN	MOD ₂
	INC	INV
	AND/OR	LOG ₂
NC^1 -message:	MULT	MIN/MAX
	DIV	PARITY
	EXP	IT-MULT
	MAJOR	MOD
	THRES	
	IT-ADD	
	AVG	

However, the size of a truth table grows exponentially with the number of different literals in the boolean formula that serves as an input. Therefore it isn't always possible to find a small set of message molecules that solves a given problem. However, for many of the presented scenarios our methodology suffices.

Tile-based self-assembly systems are indeed Turing-complete and a universal Turing-machine can be created from tiles. However, the number of required tiletypes for a tileset that encodes a Turing-machine is huge and often infeasible. The presented methodology provides a feasible approach for many common input-sizes.

For mathematical systems, more efficient solutions can often be constructed for specific problems. It is possible and likely that smaller message molecules exist for individual operations. The general approach presented in this paper works for all boolean formulas, specific, fixed operations however, can often be solved in a more efficient manner.

The next logical step is the construction of tilesets for the most important problems from Table 3. The problems ADD, MULT and THRES are of particular interest as they are among the most commonly used operations in modern programming. Space-efficient message molecules for those operations could be combined to encode the behavior of primitive communication protocols.

Furthermore, the feasibility of the presented approach could be tested in a wet-lab experiment. Successfully simulating the assembly process of the message molecules would illustrate the practical usefulness and generality of tile-based nanonetworks, as the presented procedure is capable of solving numerous problems.

References

1. Bende, P., Lau, F., Fischer, S.: Error-resistant scaling of three-dimensional nanoscale shapes on the basis of DNA-tiles. In: Proceedings of the 6th ACM International Conference on Nanoscale Computing and Communication (ACM NanoCom 2019), Dublin, Ireland (2019)
2. Büther, F., Lau, F.-L., Stelzner, M., Ebers, S.: A formal definition for nanorobots and nanonetworks. In: Galinina, O., Andreev, S., Balandin, S., Koucheryavy, Y. (eds.) NEW2AN/ruSMART/NsCC -2017. LNCS, vol. 10531, pp. 214–226. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67380-6_20
3. Clote, P., Kranakis, E.: Boolean Functions and Computation Models. Texts in Theoretical Computer Science. An EATCS Series, 1st edn. Springer, Heidelberg (2002)
4. Kallenbach, N., Ma, R.I., Seeman, N.C.: An immobile nucleic acid junction constructed from oligonucleotides. *Nature* **305**(5937), 829 (1983)
5. Lathrop, J.I., Lutz, J.H., Summers, S.M.: Strict self-assembly of discrete Sierpinski triangles. In: Cooper, S.B., Löwe, B., Sorbi, A. (eds.) CiE 2007. LNCS, vol. 4497, pp. 455–464. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-73001-9_47
6. Lau, F., Büther, F., Gerlach, B.: Computational requirements for nano-machines: there is limited space at the bottom. In: Proceedings of the 4th ACM International Conference on Nanoscale Computing and Communication, ACM NanoCom 2017, Washington DC, USA (2017)
7. Lau, F., Büther, F., Geyer, R., Fischer, S.: Computation of decision problems within messages in DNA-tile-based molecular nanonetworks. *Nano Commun. Netw.* **21**, 100245 (2019). <https://doi.org/10.1016/j.nancom.2019.05.002>
8. Lau, F., Stahl, K., Fischer, S.: Techniques for the generation of arbitrary three-dimensional shapes in tile-based self-assembly systems. *Open J. Internet of Things* **4**(1), 126–134 (2018). https://www.ronpub.com/ojiot/OJIOT_2018v4i1n10.Lau.html. Special Issue: Proceedings of the International Workshop on Very Large Internet of Things (VLIoT 2018) in Conjunction with the VLDB 2018 Conference in Rio de Janeiro, Brazil
9. Lee, J.Y., Shin, S.Y., Park, T.H., Zhang, B.T.: Solving traveling salesman problems with DNA molecules encoding numerical values. *Biosystems* **78**(1), 39–47 (2004). <https://doi.org/10.1016/j.biosystems.2004.06.00>. <http://www.sciencedirect.com/science/article/pii/S0303264704001157>
10. McCluskey Jr., E.: Minimization of Boolean functions. *Bell Syst. Tech. J.* **35**(6), 1417–1444 (1956)
11. Patitz, M.J.: An introduction to tile-based self-assembly and a survey of recent results. *Natural Computing* **13**(2), 195–224 (2013). <https://doi.org/10.1007/s11047-013-9379-4>
12. Quine, W.: The problem of simplifying truth functions. *Am. Math. Monthly* **59**(8), 521–531 (1952)
13. Rothemund, P.W.K.: Folding DNA to create nanoscale shapes and patterns. *Nature* **440**(7082), 297–302 (2006). <https://doi.org/10.1038/nature04586>
14. Turing, A.M.: On computable numbers, with an application to the entscheidungsproblem. *Proc. London Math. Soc.* **2**(1), 230–265 (1937)
15. Vollmer, H.: The NC hierarch. In: Introduction to Circuit Complexity. Texts in Theoretical Computer Science An EATCS Series, pp. 107–171. Springer, Heidelberg (1999). https://doi.org/10.1007/978-3-662-03927-4_5

16. Watson, J.D., Crick, F.H.: A structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953). <http://www.nature.com/nature/dna50/watsoncrick.pdf>
17. Winfree, E., Liu, F., Wenzler, L., Seeman, N.C.: Design and self-assembly of two-dimensional DNA crystals. *Nature* **394**(6693), 539–544 (1998)



A Model for Electro-Chemical Neural Communication

Maryam Hosseini¹, Reza Ghazizadeh¹, and Hamed Farhadi²(✉)

¹ Department of Electrical and Computer Engineering, Birjand University, Birjand, Iran

{maryam.hosseini,rghazizade}@birjand.ac.ir

² Department of Information Science and Engineering, KTH Royal Institute of Technology, Stockholm, Sweden
farhadih@kth.se

Abstract. The neuro-spike communication is conducted using electro-chemical nervous signal transmissions between neurons and synapses. The nervous signal is composed of a sequence of electrically charged ions exchange in the neurons. It passes to other from one neuron to another one through the process of release and a combination of chemical substances in synapses. The neuro-spike communication is subject to disruptions due to different biological factors that impact the permanence of neural communications. In this paper, we investigate the performance of a neuro-spike communication between two neighboring neurons. We first present a mathematical model to capture the inherent biological characteristics of the nervous system. Next, the error probability of signal detection as a function of biological parameters has been characterized. Finally, we study the impacts of some specific medicines on the parameters of neuro-spike communication in the diseases of Multiple Sclerosis and Alzheimer's.

Keywords: Neural communication · Axonal noise · Cooperative communication synapses · Synaptic channel · Neuro-spike communication

1 Introduction

Neuro-spike communication in biological nervous systems is a promising research field that is expected to have impacts on brain-machine communication system design and medical science. The neuro-spike communication is a heterogeneous communication process comprising electrical and chemical communications. In the nervous transmission, a signal may be blocked, may be changed into several spikes, or maybe added to other spikes and makes complex or wrong patterns of spikes. In addition, because of certain types of nervous system diseases or using specific types of medicines, a fault signal may be generated while there has been

no nervous signal to transmit. It is a challenging task to model the biological nervous system as a mathematical model.

Several statistical approaches have been investigated in [1–3] to model nervous systems. Signal estimation and signal detection in a nervous communication system subjected to noise and several random disturbances have been investigated in [1]. In this work, a mathematical model for a neuro-spike communication link has been developed for the cases of a synapse and multiple cooperative synapses. In [2], the binary stochastic channels are used to model nervous processes. Also, the detection error at the receiver is computed. In [4] the axonal-synaptic channel is modeled as a multiple-input-single-output (MISO) communication link and the error probability of an optimum detector for the axonal-synaptic channel is computed. In [3], several sources of randomness have considered in the model of a neuro-spike communication link, and the bit error rate for signal detection is computed. Also, an optimum receiver is designed in [3] to enhance the bit error rate. The authors have derived the closed form equation for the signal detection threshold and the optimum input spike rate. The results depicted a high efficiency in achievable bit rate with the proposed system design. In [5], a model is presented for signal propagation in nanomachine to neuron communications based on molecular communications, where the behavior of such a system as a function of the frequency is characterized. It is shown that in a frequencies range of about 3–84 Hz, a nanomachine is able to successfully communicate with a biological neuron with an acceptable time delay of about 13.5–43 ms.

In this work, we consider a neuro-spike communication link between two successive neurons with several synapses between them. We describe the transmission of action potentials along the axon as an additive white Gaussian noise (AWGN) and consider the axon in the presynaptic neuron as an AWGN channel. Next, we use a stochastic binary Z-channel to model the release of neurotransmitters. Because the communication in the synaptic cleft has a molecular nature, we use a binary stochastic X-channel to model this process. Also, we model the aggregation of synaptic channels effects as a binary Z-channel with an aggregation crossover probability. Finally, We compute the error probability of signal detection using the developed model. The model and the underlying analysis can be used to investigate the impact of different biological parameters on the performance of the neuro-spike communication link. This study could serve as an initial step in the analysis of the impacts of specific medicines or experimental treatments on special nervous system diseases such as Multiple Sclerosis and Alzheimer. Since the access to real data regarding the communication in the neocortex is difficult, similar to most related works in this field, our assumption and findings are based on the insights from the physiology of the brain.

The remainder of this paper is organized as follows. In Sect. 2, a physiological background of the central nervous system is presented. A mathematical system model for a neuro-spike communication link is described in Sect. 3. In Sect. 4, the error probability of nervous signal detection using the developed model is evaluated. The nervous communication performance subject to some nervous

system diseases using the developed model is discussed in Sect. 5. Simulation and numerical results are presented in Sect. 6. Finally, concluding remarks are drawn in Sect. 7.

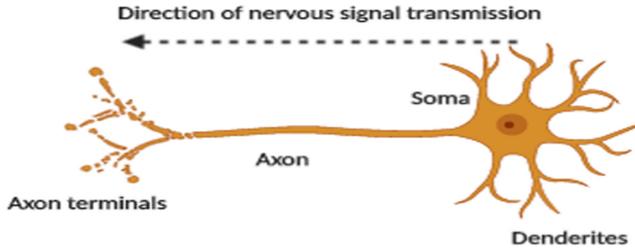


Fig. 1. Structure of a neuron in the central nervous system representing its main functional parts.

2 Biological Background of Neuro-Spike Communication

The human central nervous system consists of billions of interconnected neurons that are connected successively together by synaptic clefts that are named synapse [6]. A typical neuron, shown depicted in Fig. 1, mainly is made of several segments that are named as dendrites, cell body (soma), and axon. The outer layer of neuron is called a membrane. The axon is covered by myelin sheath with periodic gaps that nodes of Ranvier are located on them. The myelin sheath with insulating the axon speeds up signal propagation along the axon. A nervous signal first, enters the neuron through dendrites which are located on the top of soma, next, passes through the axon pathway and then, leaves neuron by axon terminals. The synapse is a small gap that passes a nervous signal only in one direction, from the presynaptic neuron to the postsynaptic neuron, see Fig. 2.

When there is no signal to transmit via the nervous system, the neuron is in a resting manner and is polarized with an intracellular potential about -95 to -65 mV. Electrically charged ion flows of potassium (K^+), sodium (Na^+), chloride (Cl^-), cause the transmitting signal throughout the nervous system. These ions enter the neuron or exit from that via the ion channels located on the soma and dendrites (cation and anion channels) and, on the nodes of Ranvier (Sodium and potassium channels). The ion exchanges between inside and outside of the neuron, change the membrane potential of the neuron to either a higher level or lower than the resting manner potential. With a potential increasing high enough about 20 mV to reaches a firing threshold level, the neuron will be excited, the membrane will be depolarized, and the firing will happen. When a neuron fires, an action potential (spike) about 90 mV at a time period of 1 ms will be generated in the neuron. The potential increase is called the excitatory-postsynaptic-potential (EPSP). Conversely, a potential decrease to a lower value than the resting manner potential causes the hyperpolarization of

the neuron membrane that is called inhibitory-postsynaptic-potential (IPSP). As firing happens, a spike passes along the axon. The spike jumps along the axon, from a node to the next node, and reaches the axon terminals. It requires to pass through the synapses to excite the next neuron [7]. When a spike reaches an axon terminal, the depolarization leads to opening the calcium channels and causes an influx of calcium ions (Ca^{++}) into the presynaptic neuron [8]. An increase in calcium ions causes the release of chemical substances called neurotransmitters into the synapse cleft. Neurotransmitters, in turn, bind to the receptor of the postsynaptic neuron, and by changing of permeability features of the neuron, making the cation or anion channel open. Opening cation channels conduct positively charged ions into the neuron, and thus, increases its potential to a value larger than the threshold and leads to a spike firing. Inversely, with opening anion channels, negatively charged ions conduct into the neuron, and due to a decrease of potential to a smaller value of resting potential, the neuron will be inhibited or in another point of view, its sensitivity to the next nervous signal will be reduced.

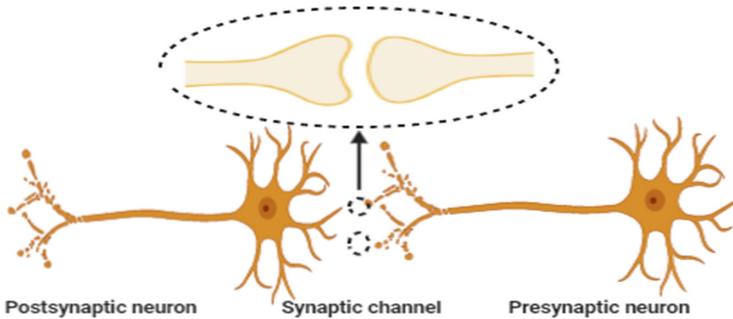


Fig. 2. Two successive neurons which are connected by chemical synaptic clefts; the nervous signal passes from presynaptic neuron to the postsynaptic neuron via multiple synapses.

3 Neuro-Spike Communication Model

Figure 3 depicts the mathematical model for a neuro-spike communication link consisting of two successive neurons and multiple synapses between them, which is shown in Fig. 2. This model is complex and heterogeneous and thus, it is split to several blocks which are investigated separately in the following.

3.1 Transmitter (Presynaptic Neuron)

A presynaptic neuron as a transmitter should pass the nervous signal through the axon and then, releases neurotransmitter into synapses. The nervous signal, as the

input of the neuro-spike communication model, could be modeled as a sequence of delta functions that is so-called spike train and is defined as follows [1]:

$$x(t) = \sum_i \delta(t - t_i), \quad (1)$$

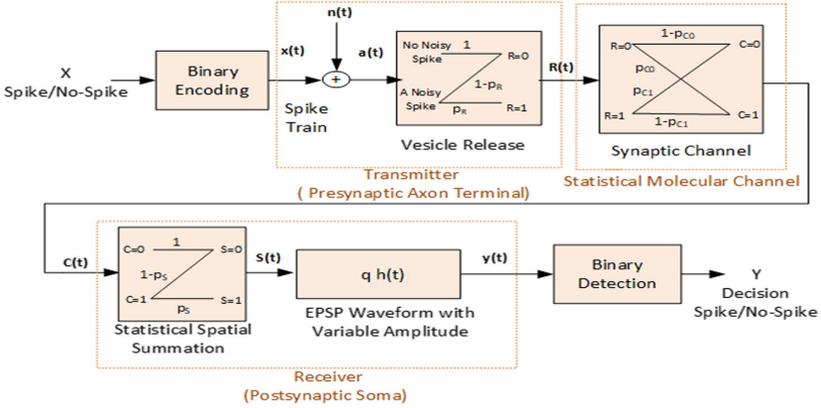


Fig. 3. Representation of the mathematical model for a neuro-spike communication link.

where $\delta(t)$ expresses the delta function, and t_i is the time duration in which the i -th spike occurs. The achieved signal at the axon terminal is obtained as

$$a(t) = x(t) + n(t), \quad (2)$$

where $n(t)$ implies the axonal noise that is assumed to have Gaussian distribution over a bandwidth BW_n with the variance σ_n^2 . The power-spectral-density of $n(t)$ is [3]

$$S_n(f) = \begin{cases} \frac{\sigma_n^2}{2BW_n}, & -BW_n \leq f \leq BW_n \\ 0, & \text{o.w.} \end{cases} \quad (3)$$

therefore, the axon signal to noise ratio (SNR) can be obtained as follows:

$$SNR_{ax} = \frac{1}{S_n(f)} \int_0^\infty x^2(t) dt = \frac{2BW_n}{\sigma_n^2} \int_0^\infty x^2(t) dt. \quad (4)$$

The axon SNR depends on the characteristics of the axon. Therefore, the higher value of SNR implies the enhancement of the axon health and leads to the more smoothing pathway of the nervous signal transmission throughout the axon. The main reason for some nervous system diseases such as multiple sclerosis (MS) [9] is the weak passing of signals through the axon. Therefore, the value of axon SNR in these nervous diseases is small. It is expected that

specific nervous medicines and treatments be effective to smooth the pathway of transmitting nervous spikes in the axon and thereby increase the value of SNR_{ax} .

The process of releasing neurotransmitters into the synapse can be modeled as a stochastic binary Z-channel, with a crossover probability of p_R . The binary input of this channel is equal to 1 when there is ‘a spike’ or is equal to 0 when there is ‘no spike’. Also, the binary output $R(t)$ is 1 or 0, respectively when neurotransmitters release happens or not. The probabilities p_R and $1 - p_R$ imply the cases of a noisy spike leads to the release and a noisy spike is failed to leading to the release. The proper synthesis of neurotransmitters, the on-time opening of calcium channels, and the influx of sufficient calcium ions into the neuron, and the perfect chemical combination in the axon terminals increase the value of p_R [10].

3.2 Synaptic Channel

Communication in the synaptic channel is mainly due to the activity of released chemical substances by the presynaptic neuron, and the opening and closing of ion channels on the postsynaptic neuron. Therefore, the synaptic channel has a molecular nature and we model it by a Z-binary channel with synaptic error probabilities p_{c0} and p_{c1} . The binary input of this channel is $R(t)$, that represents the release or not release of neurotransmitter into the synapse. Besides that, the binary output is $C(t)$ which is equal to 1 and deals to the opening of cation channels, otherwise is 0. Probabilities $1 - p_{c1}$ and $1 - p_{c0}$, respectively are equivalent to cases that neurotransmitters are released and cation channels open, and there is no release and cation channels remain closed. Somewhere in this paper, all these stochastic parameters are called synaptic parameters.

The values of synaptic parameters mainly dependent on the synaptic channel characteristics. Therefore, in some special nervous system diseases due to synaptic disruptions, the values of p_{c0} and p_{c1} are considerably high, while with the appropriate performance of synaptic channel the values of these parameters are negligible. To facilitate the describing of the synaptic channel performance we use a new concept as synapse operation probability which is obtained as

$$p_{Ch} = 1 - p_{Ch_e}, \quad (5)$$

where p_{Ch_e} expresses the error probability of synaptic channel and is defined as follows:

$$p_{Ch_e} = p_{C0pr} \{R = 0\} + p_{C1pr} \{R = 1\}. \quad (6)$$

3.3 Receiver (Postsynaptic Neuron)

There are hundreds to thousands of synapses between adjacent every pair of successive neurons [6]. While a spike transmission, each synapse has a basic role in the decreasing or increasing the membrane potential of the postsynaptic neuron, respectively. The excitation or inhibition impacts of all synapses, aggregate

in the soma of the postsynaptic neuron. Synapses that open cation channels on the postsynaptic neuron and cause to increase of membrane potential are called cooperative synapses. Also, the operation which leads to the aggregation of cooperative synapses impacts is called spatial summation [6]. The process of spatial summation between two neurons i and j can be modeled as a stochastic binary Z-channel with the crossover probability $p_{S_{ij}}$. Considering the physiology background presented in Sect. 2, this probability increases by an increase in the number of cooperative synapses between two neurons [11], and it will be decreased by increasing of firing threshold of the postsynaptic neuron. We can model the spatial summation as

$$p_{S_{ij}} = \frac{1}{|V_{th_j} - V_{rest_j}|} \left(\frac{1}{|V_{th_j} - V_{rest_j}|} + \exp(-N_{ij} p_{Ch_{ij}}^l) \right)^{-1}, \quad (7)$$

where N_{ij} is the number of cooperative synapses between presynaptic neuron i and postsynaptic neuron j . V_{rest_j} and V_{th_j} state the resting potential and firing threshold of the postsynaptic neuron. Also, $p_{Ch_{ij}}^l$ represents the operation probability of the l -th synapse between two neurons i and j , which is obtained by Eq. (5).

The output signal of the binary spatial summation Z-channel is $S(t)$, which in a short time slot is indicated by the binary variable S . Therefore, $S = 1$ implies a spike firing that means cooperative synapses successfully excite the postsynaptic neuron. Besides, $S = 0$ express that excitation of the postsynaptic neuron is failed and no spike is generated. Also, the binary input variable of this channel is $C = 1$, equivalent to the opening of cation channels and entering positive ions into the postsynaptic neuron or $C = 0$ which means cation channels are not opened. Notice that indices i and j can be removed because a general neuron link has been considered in this model.

On the other side, after a neurotransmitter release, the membrane potential changing of the postsynaptic neuron lasts about 15 ms. Therefore, another neurotransmitter that opens the same channel still could increase the membrane potential, and thereby, the excitation rate increases. Thus, the results of successive releases of neurotransmitters aggregate together. This process that states the postsynaptic neuron response to the successive releases is called temporal summation [6] and can be modeled as $q.h(t)$. In this modeling, $h(t)$ corresponds to the EPSP waveform and deals to the postsynaptic neuron response to a single neurotransmitter release, and q deals to the variable amplitude of the temporal summation. The value of q changes with the number of neurotransmitters releases. Also, $h(t)$ is modeled as an alpha function as follows [1]:

$$h(t) = \frac{h_p}{t_p} \exp\left(1 - \frac{t}{t_p}\right) u(t), \quad (8)$$

where h_p and t_p are the peak EPSP magnitude and the corresponding time, respectively, and $u(t)$ indicates the unit step function whose value is one for $t > 0$, and zero otherwise. Also, the probability density function (PDF) of q can be represented as the k -th order Gamma-distribution [12]

$$p(q) = \frac{\beta^k}{(k-1)!} q^{(k-1)} \exp(-\beta), \quad (9)$$

both β and k determine the distribution spread. The parameter k modify the variability of q therefore, the case with $k = 1$ refers to an exponential distribution with the highest variability and $k = \infty$ refers to a delta-function is independent to the variability in q .

4 Nervous Spike Detection

In this section, we evaluate the bit error rate (BER) of nervous spike detection in a neuro-spike communication link. As shown in Fig. 3, the binary variable X implies the existence or absence of a nervous spike in the presynaptic neuron, respectively by $X = 1$ and 0. Also, Y states the binary decision of spike existence or absence in the postsynaptic neuron, respectively by $Y = 1$ and 0. The output signal in the receiver measured over the period $0 \leq t \leq T$ is:

$$y(t) = h(t) * \sum_i q_i S_i \delta(t - t_i), \quad (10)$$

where q_i is the variable amplitude of EPSP waveform in response to the i -th nervous spike. S_i is a binary variable stating the spike fire in the soma of the postsynaptic neuron. Also, the symbol $*$ indicates the convolution operation. We consider the period T is divided into several time slots, and each time slot is small enough in which only one spike may occur. We can express the output signal in a single time slot as follows:

$$y(t) = S \cdot q \cdot h(t). \quad (11)$$

Thus, the following rules (Y_0, Y_1) relates the output signal to its binary equivalent:

$$\begin{aligned} Y_0; Y = 0 &\longrightarrow y(t) = 0, \\ Y_1; Y = 1 &\longrightarrow y(t) = S \cdot q \cdot h(t). \end{aligned} \quad (12)$$

Considering initial probabilities as $p_0 = pr \{X = 0\}$ and $p_1 = 1 - p_0 = pr \{X = 1\}$, we can formulate the likelihood ratio for the model as follows [2]:

$$L_x(y) = \frac{f \{Y|X = 1\}}{f \{Y|X = 0\}}, \quad (13)$$

in this equation, $f \{Y|X = 1\}$ implies the probability distribution function of the binary output in the postsynaptic soma conditioned on the spike existence in the presynaptic neuron. Thereby, we can write decision rules according to the model, base on (10) as [1]

$$\begin{cases} \text{if } L_x(y) \geq L_0 & \text{then } Y_1 \text{ is true} \\ \text{if } L_x(y) \leq L_0 & \text{then } Y_0 \text{ is true} \end{cases} \quad (14)$$

where L_0 states the decision threshold which can be defined as follows:

$$L_0 = \frac{p_0 + \frac{1}{SNR_{ax}}}{p_1}. \quad (15)$$

We can also represent the likelihood ratio as a function of other stochastic parameters of the model. Therefore, we rewrite it as

$$L_S(y) = \frac{f\{Y|S=1\}}{f\{Y|S=0\}}, \quad (16)$$

and thereby, the decision rules change to

$$\begin{cases} \text{if } L_S(y) \geq L_1 & \text{then } Y_1 \text{ is true} \\ \text{if } L_S(y) \leq L_1 & \text{then } Y_0 \text{ is true} \end{cases} \quad (17)$$

where L_1 is the new decision threshold and is defined as

$$L_1 = \frac{L_0 A_1 - A_3}{-L_0 A_2 + A_4}, \quad (18)$$

where

$$\begin{aligned} A_1 &= pr\{S=0|X=0\} = 1 - p_{C0}p_S \\ A_2 &= pr\{S=1|X=0\} = p_{C0}p_S \\ A_3 &= pr\{S=0|X=1\} = 1 - p_{C0} + p_R(p_{C0} - p_S + p_S p_{C1}) \\ A_4 &= pr\{S=1|X=1\} = p_R p_S (1 - p_{C1}) + (1 - p_R)p_S p_{C0}. \end{aligned} \quad (19)$$

Now, according to [13], the $L_S(y)$ can be represented as follows:

$$L_S(y) = \int_0^\infty p(q) \frac{pr\{Y|q : S=1\}}{pr\{Y|S=0\}} dq \quad (20)$$

where $pr\{Y|q : S=1\}$ represents the binary output probability conditioned on the variable amplitude of the EPSP waveform. Then, supposing the AWGN bandwidth is large enough to satisfy $BW_n t_P > 1$, we can simplify $L_S(y)$ as follows:

$$L_S(y) = \int_0^\infty \frac{\beta^k q^{(k-1)}}{(k-1)!} \exp(-\beta q + 2q \cdot r(y) - q^2 E_h) dq, \quad (21)$$

where $r(y) = \int_0^T h(t)y(t)dt$, and $E_h = \frac{\exp(2)T_p h_p^2}{4}$.

Finally, the average error probability of spike detection in the receiving neuron can be represented as

$$p_{error} = p_0 p_{false} + p_1 p_{miss} \quad (22)$$

where

$$\begin{aligned} p_{false} &= pr\{Y=1 | X=0\} = pr\{L_S(y) \geq L_1 | S=0\} (1 - p_S p_{C0}) \\ &\quad + pr\{L_S(y) \geq L_1 | S=1\} (p_S p_{C0}), \end{aligned} \quad (23)$$

and

$$\begin{aligned}
p_{miss} = & pr \{Y = 0 \mid X = 1\} = pr \{L_S(y) < L_1 \mid S = 0\} (1 + p_S p_R (-1 + p_{C1}) \\
& + (1 - p_R)(-p_S p_{C0})) + pr \{L_S(y) < L_1 \mid S = 1\} (p_R(1 - p_{C1}) p_S \\
& + (1 - p_R)(p_S p_{C0})).
\end{aligned} \tag{24}$$

5 Discussion on Nervous System Diseases and the Impact of Medicines on Neuro-Spike Communication Performance

In this section, the relation between random parameters of the developed model and some nervous system diseases is investigated. Considering the communication background of the nervous system presented in Sect. 2, a nervous signal may be blocked either while passing the axon, or transmitting in the synaptic channel.

For example in Multiple Sclerosis (MS) which is a mobility disability nervous disease, due to the demyelination of axon, the spike may be disrupted or blocked while passing the axon [9, 14]. In MS disease, while passing a spike in the axon, a large number of positive ions flow out of the neuron and the nervous spike will be blocked. Therefore, the value of (SNR_{ax}) in such diseases is low. Fampridine (Fampyra $\text{\textcircled{R}}$) is known as an efficient medicine for MS and walking disability diseases [15]. Fampridine by blocking the potassium channels prevents excessive efflux of positive electrical charges. As a result, SNR_{ax} increases, and the transmission of the nervous signal through the axon facilitates.

In some other nervous system diseases, the release probability p_R is too small. In such cases, special medicines such as clonidine act on the presynaptic neuron and through a prolonged inhibition of this neuron increase the release probability of neurotransmitters [10].

Also, synaptic channel disruptions cause some nervous diseases such as, Parkinson, Schizophrenia, and Alzheimer [16–19]. In these diseases in which the nervous spike passing is failed at the synaptic channel, the value of synaptic parameters p_{C0} and p_{C1} are high. Exercise, enough and good quality sleep as well as the hormone Leptin act on the nervous system to facilitate nervous signal passing through the synaptic channel. Thus, values of p_{C0} and p_{C1} decrease and abilities of learning and memorizing [20] will be improved.

6 Numerical Results

In this section, the bit error rate (BER) of a neuro-spike communication link based on the developed model is investigated. Simulations are carried out in the environment of MATLAB [21]. According to [1], we set the predefined parameters as $h_p = 2$ mV and $T_p = 1$ ms for the EPSP waveform. The variable amplitude of EPSP waveform, q , is considered as a Gamma distribution with parameters $\beta = 1$ and $k = 1$. We also considered 10^7 transmitted spikes and computed

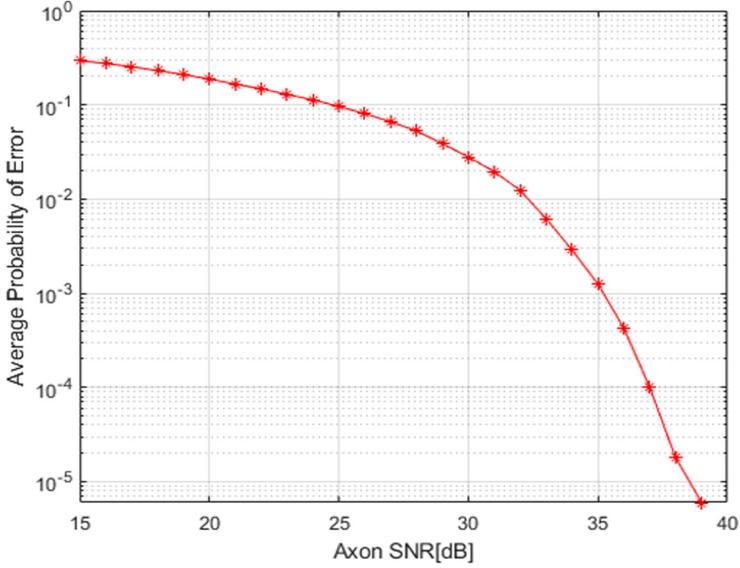


Fig. 4. The average probability of error versus the axon SNR in a neuro-spike communication link with $p_R = 0.8$, $p_{C0} = 0.01$, $p_{C1} = 0.01$, $p_S = 0.4$.

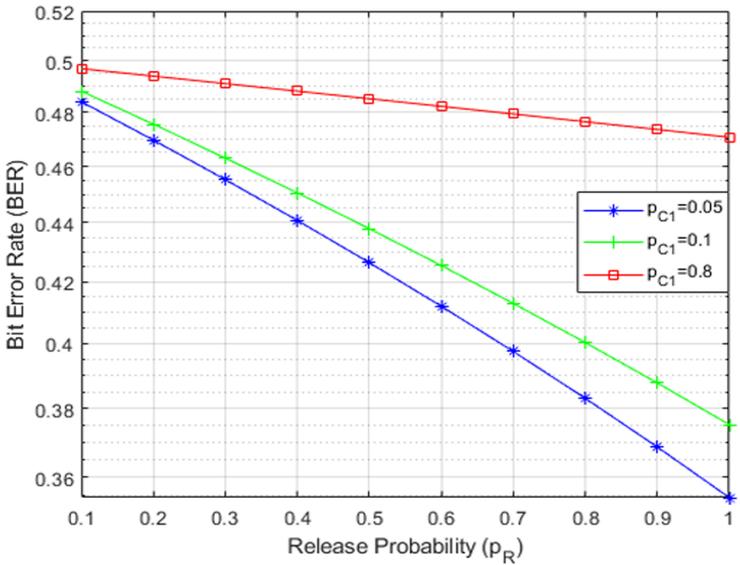


Fig. 5. The bit error probability curves versus the release probability for different parameters of the synaptic channel in a neuro-spike communication link with $p_{C0} = 0.009$, $p_S = 0.6$, $SNR_{ax} = 15$ dB.

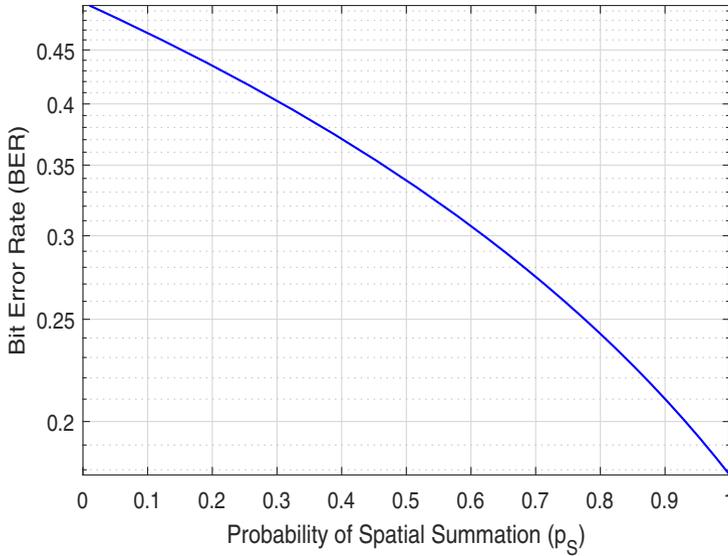


Fig. 6. The bit error rate versus the spatial summation probability in a neuro-spike communication link with $p_R = 0.8$, $p_{C0} = 0.01$, $p_{C1} = 0.09$, $SNR_{ax} = 15$ dB.

the average performance. In the following, the results of the error probability analysis versus stochastic parameters are depicted.

Figure 4 shows the error probability of spike detection in the postsynaptic neuron versus the axon SNR changes. It is observed that by an increase in SNR_{ax} , the spike transmission in the neuron facilitates and as a result, the error probability improves. In cases of axonal diseases such as MS who suffer from weak passing of signal in the axon, SNR_{ax} is low and thus, as we can observe in Fig. 4, the correct spike detection in the receiver with a high probability will be failed.

In Fig. 5, the average error probability versus the releases probability of neurotransmitters p_R for different cases of $p_{C1} = 0.05$, 0.1, and 0.8 is depicted. It can be seen that with an increase of p_R , the error probability of spike detection decreases. Also, it is observed that for the worst case of $p_{C1} = 0.8$ increasing p_R could not change the error probability. The reason is that in such a case, the synaptic channel is approximately disrupted and thus, the error probability is not sensitive to other parameters.

In Fig. 6, the average error probability versus the spatial summation probability is shown. With an increase of p_S , the average error probability decreases. The reason is that spatial summation probability directly is dependent on the number of cooperative synapses between two neurons. Therefore, more synapses cooperate to excite the postsynaptic neuron the detection of the signal will be more successful.

7 Conclusion

In this paper, the performance of a neuro-spike communication link has been studied. First, we developed a model of neuro-spike communication which consists of two successive neurons that are connected via multiple synapse cleft. Next, we evaluated the error probability of signal detection in this system using the proposed model. The simulation results reveal the strong dependence of signal detection to the disruption factors such as axonal noise, release probability of neurotransmitters, synaptic channel parameters, and spatial and temporal summation. We also have studied the impact of different nervous system medicines on these stochastic parameters. As part of our future work, we model the cooperation of synapses as a biological concept of synaptic plasticity and develop the neuro-spike communication model by considering the medicine effect.

References

1. Manwani, A., Koch, C.: Detecting and estimating signals over noisy and unreliable synapses: information-theoretic analysis. *Neural Comput.* **13**(1), 1–33 (2001)
2. Maham, B.: A communication theoretic analysis of synaptic channels under axonal noise. *IEEE Commun. Lett.* **19**(11), 1901–1904 (2015)
3. Aghababaiyan, K., Shah-Mansouri, V., Maham, B.: Joint optimization of input spike rate and receiver decision threshold to maximize achievable bit rate of neuro-spike communication channel. *IEEE Trans. Nanobiosci.* **18**(2), 117–127 (2018)
4. Maham, B.: Spike detection in axonal-synaptic channels with multiple synapses. In: *ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1130–1134. IEEE (2019)
5. Galluccio, L., Palazzo, S., Santagati, G.E.: Characterization of signal propagation in neuronal systems for nanomachine-to-neurons communications. In: *2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 437–442. IEEE (2011)
6. Hall, J.E.: *Guyton and Hall Textbook of Medical Physiology e-Book*. Elsevier Health Sciences, Amsterdam (2010)
7. Ramezani, H., Akan, O.B.: Impacts of spike shape variations on synaptic communication. *IEEE Trans. Nanobiosci.* **17**(3), 260–271 (2018)
8. Lai, K.-O., Ip, N.Y.: Central synapse and neuromuscular junction: same players, different roles. *Trends Genet.* **19**(7), 395–402 (2003)
9. Pagnini, F., Bosma, C.M., Phillips, D., Langer, E.: Symptom changes in multiple sclerosis following psychological interventions: a systematic review. *BMC Neurol.* **14**(1), 222 (2014)
10. Knight, D., Cunnane, T.C., Lavidis, N.A.: Effect of chronic clonidine treatment on transmitter release from sympathetic varicosities of the guinea-pig vas deferens. *Br. J. Pharmacol.* **134**(7), 1480–1486 (2001)
11. Fares, T., Stepanyants, A.: Cooperative synapse formation in the neocortex. *Proc. Natl. Acad. Sci.* **106**(38), 16:463–16:468 (2009)
12. Bekkers, J.M., Clements, J.D.: Quantal amplitude and quantal variance of strontium-induced asynchronous EPSCs in rat dentate granule neurons. *J. Physiol.* **516**(1), 227–248 (1999)

13. Maham, B., Kizilirmak, C.: Neuro-spike communications with multiple synapses under inter-neuron interference. *IEEE Access* **6**, 962–968 (2018)
14. Shi, J., Wu, X., Chen, Y.: Study on dalfampridine in the treatment of multiple sclerosis mobility disability: a meta-analysis. *PloS ONE* **14**(9) (2019)
15. Fragoso, Y.D., Adoni, T., Alves-Leon, S.V., Apostolos-Pereira, S.L., Barreira, A.A., Brooks, J.B.B., Claudino, R., Correa, E.C., Ferreira, M.L.B., Finkelsztejn, A., et al.: Real-life experience with fampridine (fampyra®) for patients with multiple sclerosis and gait disorders. *Neuro Rehabil.* **39**(2), 301–304 (2016)
16. Gao, R., Penzes, P.: Common mechanisms of excitatory and inhibitory imbalance in schizophrenia and autism spectrum disorders. *Curr. Mol. Med.* **15**(2), 146–167 (2015)
17. Osimo, E.F., Beck, K., Marques, T.R., Howes, O.D.: Synaptic loss in schizophrenia: a meta-analysis and systematic review of synaptic protein and mRNA measures. *Mol. Psychiatry* **24**(4), 549–561 (2019)
18. Picconi, B., Piccoli, G., Calabresi, P.: Synaptic dysfunction in Parkinson's disease. In: Kreutz, M., Sala, C. (eds.) *Synaptic Plasticity*, vol. 970, pp. 553–572. Springer, Vienna (2012). https://doi.org/10.1007/978-3-7091-0932-8_24
19. Laurén, J., Gimbel, D.A., Nygaard, H.B., Gilbert, J.W., Strittmatter, S.M.: Cellular prion protein mediates impairment of synaptic plasticity by amyloid- β oligomers. *Nature* **457**(7233), 1128–1132 (2009)
20. Harvey, J., Solovyova, N., Irving, A.: Leptin and its role in hippocampal synaptic plasticity. *Prog. Lipid Res.* **45**(5), 369–378 (2006)
21. Release, M.: *The mathworks*, Inc., Natick, Massachusetts, United States, vol. 488 (2013)

Connected Wearables Sensors for Healthcare Applications



Activity Monitoring Using Smart Glasses: Exploring the Feasibility of Pedometry on Head Mounted Displays

Zhiquan You¹, Farnaz Mohammadi², Emily Pascua¹, Daniel Kale¹, Abraham Vega¹,
Gian Tolentino¹, Pedro Angeles¹, and Navid Amini¹(✉)

¹ California State University, Los Angeles, Los Angeles, CA 90032, USA
namini@calstatela.com

² University of California, Los Angeles, Los Angeles, CA 90095, USA

Abstract. Fitness tracking, fall detection, indoor navigation, and visual aid applications for smart glasses are rapidly emerging. The performance of these applications heavily relies on the accuracy of step detection, which has rarely been studied for smart glasses. In this paper, we develop an accelerometer-based algorithm for step calculation on smart glasses. Designed based on a salience-analysis approach, the algorithm provides a highly accurate step calculation. An activity monitoring application for a commercial Android-based smart glasses (Vuzix M100) is designed and realized for algorithm evaluation. Experimental results from 10 participants wearing the smart glasses running our application achieved average step detection error of 2.6% demonstrating the feasibility of our salience-based algorithm for performing pedometry on smart glasses.

Keywords: Smart glasses · Accelerometer · Activity monitoring · Salience · Peak-to-Peak

1 Introduction

In recent years, interest in smart glasses has been substantially growing, as has the number of influential companies, such as Amazon and Facebook, that are announcing their entry into the smart glasses market. It is expected that the combined market size for smart glasses will grow at an exponential rate of 76% annually and will reach \$16B by 2025 [1]. The use of commercial smart glasses for fitness tracking [2, 3], indoor navigation [4, 5], fall detection [6, 7], vision enhancement in the visually impaired [8–10] has already gained momentum and it is expected that such applications will comprise a considerable sector of the smart glasses market. The performance of these applications as well as many other emerging services is dependent on the accuracy of pedometry (step detection and counting) on smart glasses.

Z. You and F. Mohammadi – Contributed equally.

Due to the proliferation of accelerometers in consumer products and by virtue of their low power consumption, accelerometer-based step calculation has emerged as the most popular approach to perform pedometry. Many existing step counting solutions have shown that when a user is walking, the measurements of the accelerometer (on the user's body) will regularly change and such changes can be used to calculate the number of steps taken by the user.

Plenty of research studies on activity monitoring, especially step detection have been conducted [11]. A number of such studies aim at detecting steps for handheld devices such as smartphones [12]. Other researchers have transformed the problem of step detection into a case in pattern recognition and employed a machine learning technique for pedometry [13, 14]. Another group of studies devised algorithms that target step detection in ankle-worn, shin-worn, or waist-worn devices [15, 16]. To the best of our knowledge, step detection techniques suited for smart glasses have not been adequately explored. Smart glasses cannot offer similar activity tracking utility levels compared to the typical handheld, waist-worn, or ankle-worn wearable devices [17]. Head-worn devices typically contain a limited number of sensors and they are vulnerable to external sources of error. Moreover, it is typical for smart glasses to require low computing power by virtue of their size, heat, and battery issues [18, 19].

In order to motivate the problem, we conducted an experiment, where we attached 6 inertial measurement units (IMUs) produced by Shimmer Research to different regions on a subject body and asked him to take 220 steps. We use the windowed peak detection technique, proposed in [20] as an optimal step detection method for wearable low-power devices, to derive the number of steps for each IMU. Table 1 summarizes the number of steps counted by each IMU; we observe that lower parts of the body and especially, the shin, thigh, and waist areas performed well; this can be attributed to the higher impact of each step (foot striking the ground) on those areas comparing to other locations.

Table 1. Counted steps vs. the location of IMU

Sensor location	Shin	Thigh	Waist	Forearm	Upper arm	Head
Steps counted	228	199	230	188	179	114
Error($\frac{real-est.}{real}$)(%)	4%	10%	5%	15%	19%	48%

Based on the above experiment, in this paper, we propose a step detection algorithm for smart glasses, which can identify steps accurately and in real-time. The algorithm is composed of a *signal preprocessing* phase, *axes combination* phase, *saliency calculation and analysis* phase, and a *peak detection* phase. The proposed algorithm utilizes only the accelerometer data for step calculation since smart glasses are commonly designed with an accelerometer, while other sensors such as the GPS and gyroscope are not always available. Furthermore, the proposed algorithm requires a low level of computing power from smart glasses.

The contributions of this paper are threefold. First, we propose an algorithm for step detection and calculation on smart glasses based on accelerometer data. Second,

in two experiments, we investigate the step detection accuracy of our algorithm on commercial smart glasses as well as on IMUs attached to multiple on-body locations. Third, we incorporated the proposed algorithm into an activity monitoring application for Android-based smart glasses for real-time step calculation. The rest of this paper is organized as follows. Section 2 introduces some related studies. Sections 3 and 4 describe the designed data collection and data analysis phases, respectively. Section 5 covers our experimental results and Sect. 6 discusses constraints and future directions of our investigation. Finally, Sect. 7 concludes this study.

2 Related Work

There are numerous algorithms and research studies conducted on step detection based on smartphones and waist-mounted devices, with high accuracies [20]. Some algorithms explore other ways to detect steps beyond the basic peak-valley relationship, such as the method proposed by Kumar et al., [21] discussing the use of the linear relationship between the amplitude of acceleration and the frequency of steps, effectively detecting steps through their pilot trials. Most of the research in the field of step detection focuses on 5 body parts to mount the sensor: wrist, pocket (waist level), thigh, ankle, and foot. The most reliable step count algorithms come from insole pressure sensors that have the ability to detect the pressure applied to the sensor once a subject takes a step [22, 23]. The impulse (force) generated from each foot striking the ground in walking is more pronounced in those areas. In general, lower parts of the body such as the ankle and foot absorb most of the impact of stepping and therefore, devices placed on these body parts leads to the most accurate results [24].

Smartwatches and smartphones are among the electronic equipment that is capable of recording and analyzing motion signals [25–27]. However, smartphones are not necessarily always taken in the same or relative location. In addition, these devices are extremely sensitive to non-ambulatory activities [28]. Moreover, hand and arm movements when the device is being carried, do not necessarily correlate with stepping and may create artifacts on the signal. On the other hand, smartwatches provide a tremendous amount of information by continuously monitoring of the user and measurement of their physiological parameters. Although, the data acquired from smartwatches are under the question of reliability, due to subtle and often wrist movements, compared with other parts of the body where the sensors are usually placed [26]. One inevitable shortcoming of data based on wearable devices is that they are not generalizable and unstable across different brands [29], which is the case in smartphones as well.

The most popular step detection algorithm uses peak-valley extraction to detect steps from the accelerometer data. The three-axis accelerometer data is combined into a single acceleration vector, which is then graphed where each peak is considered to be a step [30, 31]. A more advanced algorithm would be a filtering system which places thresholds for each peak to be considered a step, as there will be multiple peaks from the accelerometer that is not considerable enough to be detected as a peak.

Most step counting algorithms are created for devices inside the pocket [32]. For instance, Brajdic and Harle [20] provided evaluations considering various algorithms for step detection on smartphones. They evaluated several algorithms such as windowed

peak detection (WPD), hidden Markov model, and continuous wavelet transform, which were the most promising approaches. They found that the simplest algorithm with the best accuracy is the windowed peak detection algorithm.

A number of studies have employed a modified version of the peak-to-peak algorithm to detect steps from an accelerometer signal. For instance, the peak detection algorithm proposed in [33] produces a refined method that enables the algorithm to detect peaks in periodic and quasi-periodic signals. This method is noted to have high efficiency in peak detection within high and low-frequency signals. Another innovation based on step detection for head-mounted sensors is being used for detecting user movements when using virtual reality systems. The idea of these virtual reality systems is to immerse the user into a virtual world in which their movement in real life is synchronized to their virtual avatar [34].

3 Sensor Platforms and Pilot Experiments

We run a pilot study to investigate the step detection accuracy of our algorithm on commercial smart glasses as well as on the IMUs attached to multiple body parts. The first experiment is through activity monitoring via a smart glass device for the head signal, and another is using the IMUs for pocket, head, and foot-worn IMUs. Experiments are conducted in an outdoor environment at the California State University, Los Angeles campus. The participants include 10 students (average age 26.3, $SD = 5.2$, average height 170 cm, $SD = 11.3$ cm). For both experiments, each test consisted of walking 100 steps at their usual pace repeated 10 times per individual. For this pilot study, none of the participants reported any limitations in their mobility or vision. In the following, we explain both experiments.



Fig. 1. Vuzix M100 device on test subject for data collection.

3.1 Step Detection on the Smart Glass

A popular Android-based smart glass in the research and industry communities, namely the Vuzix M100, with a sampling frequency of 100 Hz is used in the first experiment. Figure 1 shows a subject wearing the Vuzix M100 with the activity monitoring Android application running. The application utilizes our salience-based algorithm to detect the steps in real time. Figure 2 is a screenshot of the Android application.

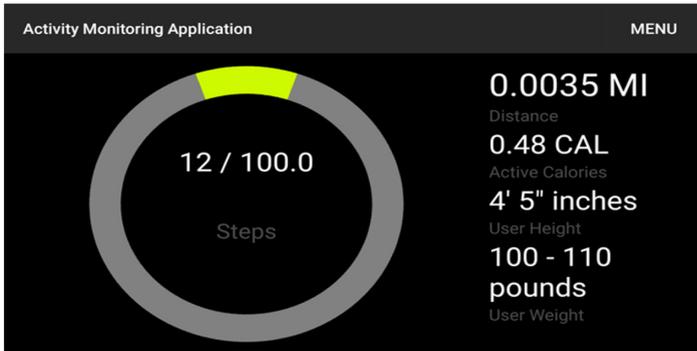


Fig. 2. Graphical user interface of application on Vuzix M100.

3.2 Step Detection on IMUs

In experiment 2, we employ the Shimmer3 [35] IMU unit to collect the acceleration signal from the head, foot, and pocket locations. The step detection is performed off-line on a desktop computer. Figure 3 shows the Shimmer worn by the same subject on the head.

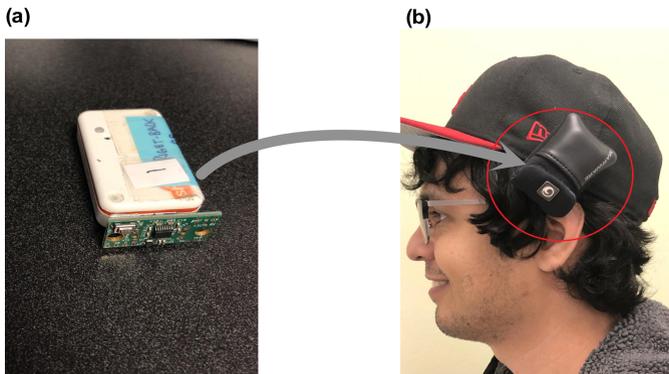


Fig. 3. (a) The Shimmer3 IMU, (b) the Shimmer sensor is worn by a subject on the head.

4 Methods

We propose a novel algorithm to identify the steps using the accelerometer signal from a smart glass device and compare the accuracy of our results against the WPD method. The WPD uses a moving average to smooth the accelerometer signal, and it detects single peaks via a fixed size sliding window; it finds the maximum value in the window, shifts the window, and discards the chosen maximum if it persists for two windows in a row. This algorithm has been identified as an optimal step detection approach for wearable devices owing to its computationally inexpensive nature [20, 36]. Figure 4 presents the block diagram of our step detection approach; the description of each step will follow in the preceding subsections.

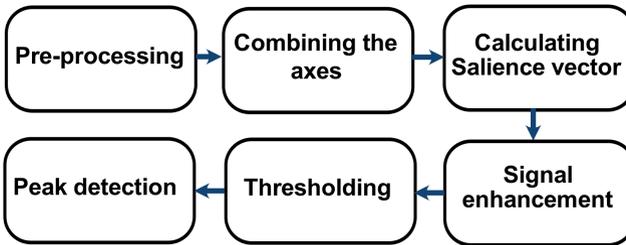


Fig. 4. The block diagram representing the flow of the approach.

4.1 Pre-processing

The signal preprocessing consists of calibration and low-pass filtering. First, the calibration process is utilized to reduce the drift errors and offsets from the raw acceleration signals. Second, a moving average filter is used to suppress the high-frequency noise of the calibrated inertial signals. Details of the preprocessing phase can be found in [29].

4.2 Combining the Axes

The three axes (x , y , z) are combined to a single magnitude measure. We use (1) to calculate the magnitude of combined acceleration in all directions using the Euclidean distance method:

$$r = \sqrt{A_x^2 + A_y^2 + A_z^2} \quad (1)$$

4.3 Peak Detection with Saliency Algorithm

We calculate the saliency of each acceleration sample point. Considering all possible intervals in a given signal of size N , saliency is the length of the longest interval for which a sample is a maximum [37, 38]. The key property of the saliency values algorithm is

that the starting point of each step has a large salience value. Hence, steps can be found by locating these prominent points.

As an example, salience can be visualized as shown in Fig. 5. Considering the fourth sample in the signal shown in Fig. 5; despite the fact that it has a higher magnitude than the seventh sample, it has a smaller salience comparing to the seventh sample, $s(4) = 4$ as opposed to $s(7) = 8$. The term salience vector represents the resulting signal containing the salience of each sample in the original signal. The list of salience vector of the example signal shown in Fig. 5 can be found in (2). We use salience signals to find each stride that has the largest salience samples over a time threshold of 30 s. Once the time threshold is met, the salience vector is then processed through a peak detection algorithm to determine the steps of the filtered data.

$$s(k) = \{14, 1, 2, 4, 1, 13, 8, 1, 4, 2, 1, 5, 2, 1, 15\}, \quad 1 \leq k \leq 15 \quad (2)$$

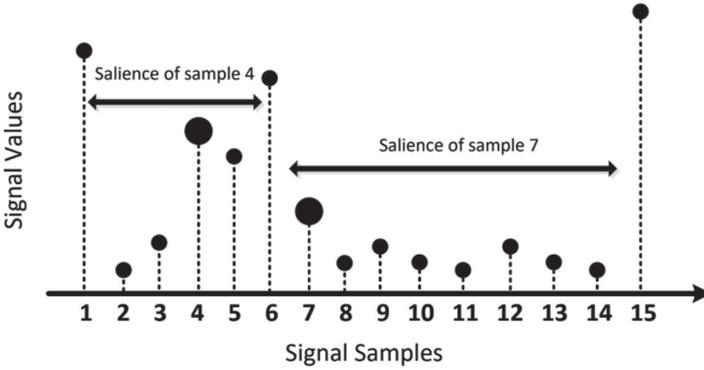


Fig. 5. Salience of an example data sequence.

We have implemented all three algorithms found in [38] for computing the salience vector, s , of a signal. These algorithms include: 1) basic salience computation, 2) partial salience allocation, and 3) sliding window analysis. We found the sliding window approach to be the most suitable one for the purpose of step detection, as it offers faster execution time and it addresses the issues pertaining to the signal's origin. The following briefly explains three algorithms with which we can calculate salience that is done after combining the three axes of the signal and obtaining the r vector [37].

Basic Scheme. To find the length of the longest interval for which a sample is maximum in a signal, the following steps are required:

1. Initialization: Since each sample is a maximum with respect to itself, all salience values are set equal to 1. Thus, the length of the analysis interval (n) is set as $n = 2$.
2. Subdividing the input signal of length N into analysis intervals of length n . The length of the last interval will be between 1 and n .
3. Finding the maximum point within each interval and assigning a salience of n to the corresponding sample.

4. Increasing the size of the analysis interval by one and going back to step 2 until n reaches N .

The last calculated salience vector will be reported.

Partial Salience Allocation Scheme. As an alternative to the basic scheme, partial salience allocation is used in the case of computationally expensive functions. This algorithm utilizes tabular representation of the samples as instead of finding maxima over intervals with increasing lengths, one can determine the maximum over frames of decreasing lengths. With j being the position of the global maxima in the signal, the value in this position is a maximum for the previous $j-1$ first samples. Accordingly, assuming that j is known, one can compute the maximum over $j-1$ first samples only and assign it a salience of $j-1$ and so on, until a salience has been assigned to the first sample.

Sliding Window Salience Computation Scheme. Partial salience allocation is faster than the basic scheme, however, this method cannot find salience values for all the samples in the input signal. Furthermore, the problems related to the samples near the boundaries are not addressed in this scheme. The intuition behind the sliding window scheme is to apply a window to the input signal and move the window sample by sample in order to address the boundary problem. In this method, a sliding window $w_M(i)$, of length M , where $M < N$, and origin i , is placed at the beginning of the signal and is shifted sample by sample towards the endpoint of the signal. Saliences $S(k, i)$, $S_{left}(k, i)$, $S_{right}(k, i)$ are then computed with respect to the sliding window. If the proper window size is chosen, the running saliencies obtained will suffice for most applications. We define local saliencies with reference to the position of a given window. Moreover, running saliencies $S^*(k)$, $S_{left}^*(k)$, $S_{right}^*(k)$, are defined as the maximum of the local saliencies of sample k for all window positions (previous and current). Thus, running saliencies are non-decreasing. Here we outline the sliding window analysis scheme:

1. Initialization: all saliencies S^* are set equal to 1 and all saliencies S_{left}^* and S_{right}^* are set equal to zero.
2. Applying the partial salience allocation scheme to the samples within the sliding window with origin i and computing the local $S(k, i)$, $S_{left}(k, i)$, and $S_{right}(k, i)$ saliencies, where $k = i, i+1, \dots, i+M-1$.
3. Updating running saliencies as follows: for each sample k , $S^* = S(k, i)$, $S_{left}^*(k, i) = S_{left}(k, i)$, $S_{right}^*(k) = S_{right}(k, i)$ if their values will be increased, otherwise, we skip this step.
4. Shifting the sliding window to the right by one sample, i.e., ($i \rightarrow i+1$) and going back to step 2.

Once the sliding window reaches the right boundary of the input signal, the final salience vector, $S_{final}(k)$ is given by (3):

$$S_{final}(k) = S_{left}^*(k) + S_{right}^*(k) + 1, \quad k = 1, 2, \dots, M \quad (3)$$

We discard the first and last $M - 1$ element of the salience vector since these values are in the boundary. Due to this problem, it is important to properly choose the size of the sliding window.

4.4 Signal Enhancement

In this step, we compute a vector u , which is defined as:

$$u = \frac{(r \cdot s)}{\max(s)} \quad (4)$$

where s is the salience vector and dot (\cdot) represents an element-wise multiplication. The idea behind deriving u is to make peaks of r more pronounced and to diminish the rest of the samples.

4.5 Thresholding and Peak Detection

We extract all the elements of u that exceed a certain threshold, that is given by (5) as potential cycle indices [38].

$$threshold = \frac{2}{3} \max(u) \quad (5)$$

Then the difference vector, d , between adjacent extracted indices is calculated. We normalize d around its mean and extract one of the two indices of points that fall within this range, i.e., $|d - \text{mean}(d)| < \text{mean}(d)$. This elimination phase helps to increase the accuracy of our stride detection algorithm by counting two close peaks only once. The number of such points is indicative of the number of steps taken and in fact, these points are the start/endpoints of each step. Finally, we return a list of indices that these points correspond to.

Among the three algorithms to calculate the salience vector, the sliding window scheme is more efficient than the basic and partial salience allocation. Hence, we will base our results on the calculations from the sliding window algorithm. We choose the window size of 94 [38] and calculate the salience vector.

Figure 6a shows the combined accelerometer signal, r , and the red lines represent the occurrence of a step. Figure 6b exhibits the salience vector of the combined acceleration signal. Figure 6c depicts the enhanced salience vector where the peaks (associated with steps) become pronounced, and that this subfigure corresponds to the vector, u , where we enhance the signal. As shown, the peaks are more significant and distinguishable from the rest of the signal.

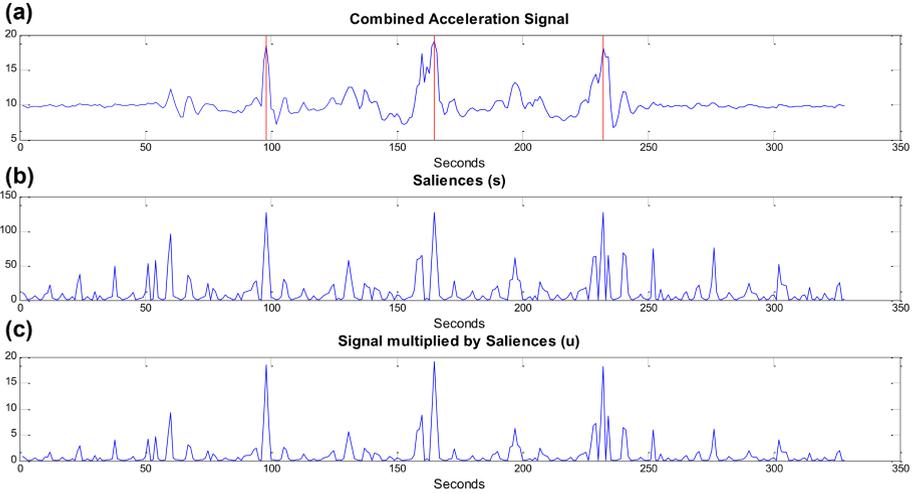


Fig. 6. Original signal (a), the signal’s salience vector (b), and the salience vector after normalization (c).

4.6 Error Determination

To quantify the differences between the signals from different parts in the body namely foot, head, and pocket, we calculate the error for each trial and each person as written in (6) and (7). The error in our experiments is defined as the absolute difference of the number of steps calculated by the algorithm from the ground truth [36]:

$$e_j(k) = \frac{|steps_j(k) - \widehat{steps}_j(k)|}{steps_j(k)} \quad (6)$$

with j being the subject number, and k being the trial number. The term $steps_j(k)$ is the true number of steps that subject j took in trial k , and $\widehat{steps}_j(k)$ is the step calculated by the algorithm. Each subject performs each trial 10 times, thus, the averaged error for subject j and overall repeats (e_j) is given by (7):

$$e_j = \frac{\sum_{k=1}^n e_j(k)}{n} \quad (7)$$

5 Results

We evaluate the step detection performance of our algorithm on commercial smart glasses as well as IMUs attached to the head and other parts of the body. In the pilot trial, we collected gait data from 10 subjects and applied our proposed algorithm to these signals. In this section, we compare the performance of the algorithm for the foot, pocket, and head signals, and provide error analysis for the results.

5.1 Accuracy Comparison

Figure 7 graphs the changes in step detection error from the WPD method to our salience-based approach. The average step detection error of our approach (2.6% with SD of 2.8) was significantly lower than that of the WPD method (21.8% with SD of 9.7).

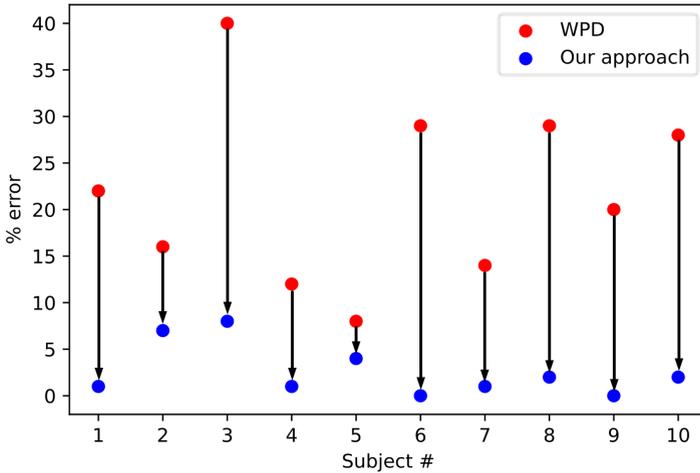


Fig. 7. Comparing the error for step counting from the head signal between the WPD method and our approach.

5.2 Error Analysis

Table 2 lists the average step detection for three IMU locations in experiment 2. For both the WPD method and our approach, the foot location performs remarkably well and achieves similar results with an average error of 2.9% (SD = 1.6) and 2.7% (SD = 2.4), respectively.

We note moderate improvements in step detection for the pocket location with our approach (6% with SD of 3.5) comparing to the WPD (7.2% with SD of 6.5). The head location benefitted the most from our salience-based approach, where the step detection error declined from 25.6% (SD = 9.4) to 3.9% (SD = 2.4). The error range for the head location was from 13% to 39% in the WPD method; however, with our approach, the range was narrowed down to 0% to 8%.

Table 2. Average step detection error in experiment 2.

Subject	WPD			Our approach		
	Head	Pocket	Foot	Head	Pocket	Foot
1	31%	10%	1%	8%	1%	0%
2	23%	5%	5%	3%	13%	2%
3	35%	2%	2%	5%	2%	2%
4	21%	6%	0%	4%	8%	6%
5	15%	1%	2%	0%	6%	7%
6	38%	9%	4%	5%	4%	4%
7	20%	23%	5%	2%	7%	1%
8	13%	10%	3%	3%	5%	0%
9	21%	1%	3%	2%	9%	4%
10	39%	5%	4%	7%	5%	1%
Average	25.6%	7.2%	2.9%	3.9%	6.0%	2.7%
St. Dev.	9.4	6.5	1.6	2.4	3.5	2.4

6 Discussion and Future Work

During the process of this research, we are attempting to develop an algorithm that further improves on the current step detection algorithms by collecting additional data for testing and applying the salience algorithm to filter the noisy data received from accelerometer sensors. We focused on analyzing smart glass data received from the sensors. There are several factors that make it difficult to achieve the same accuracy from head data than it is from the foot data. One of the most challenging issues with collecting data from the head is all of the movement the head experiences throughout walking. Every slight acceleration from the head can be falsely interpreted as a step, which is where the salience algorithm aids in fixing. Moreover, in our trials, the subjects are asked to walk at their normal pace that is relatively slow which is more challenging. In the future, we intend to investigate the performance of our algorithm for various speeds. As people age, the angular displacement of the head changes and the body will have a decreased ability to attenuate accelerations from trunk to head [39]. As another future direction, we will investigate our algorithm's performance on older adults.

Providing an improvement in accuracy on current step detection algorithms can also be applied on a myriad of other fields ranging from physical health measurements, medical studies, and personal localization. Although step counters are not commonly used in clinical research, steps per days can now be considered as vital signs for crucial medical information in the future [24].

Before activity monitors can be adequately used for clinical research, they must first prove there is a connection between steps taken and future occurrence of diseases. Having accurate measurements of physical activity is crucial during medical studies

as any slight mismeasurement could result in obscuring actual links between physical activity and certain ailments that affect physical movements [40].

Another future application of this research can be implemented in personal localized positioning systems based on steps detection and the calculated distance the user has travelled [41]. A group of researchers from Seoul National University worked on a method to determine the user's location in indoor settings. Due to the lack of GPS measurements for indoor environments, these researchers resorted to step detection in combination with pedestrian dead reckoning to capture the user's current location. They are able to achieve step detection through several conditions such as handheld texting, in pocket of pants front and back [14]. For indoor position to be accurate the direction in which the user is walking and their heading is crucial to be determined, which is accomplished by using the magnetometers to find their absolute direction [42]. Furthermore, the algorithm developed in this paper can be used on other data sets from other studies. The data set from [43] is of accelerometer data collected from two groups of individuals who are sighted and blind using the aid of walking cane or guide dog. The paper [43] discusses the interesting differences between the sighted and blind walkers.

Here, we perform the analysis assuming the data we are analyzing is walking signal, in other words, walking is granted, however, the salience algorithm can be employed to detect whether the subject is actually walking in the first place. The proposed method is computationally inexpensive, performs well on slow walking data, and is a low battery consumer. These qualities make this approach efficient and highly practical.

7 Conclusions

A biomechanical process (called Pronation [44]) occurs during ambulation that allows the body to naturally absorb shock as each foot strikes the ground. Since the head is the farthest body location from the ground, the accelerations become significantly dampened [45], making it an unsuitable signal to detect steps. On the other hand, applications in fitness tracking, indoor navigation, user authentication, vision enhancement for the visually impaired will comprise a considerable portion of the smart glasses market and they are all dependent on accurate step counts. With this motivation, we implement and evaluate an accurate and reliable method for step detection using head acceleration signals. Furthermore, our algorithm achieves superior step counting performance when applied to the acceleration signal collected from other body parts. Our results demonstrate the feasibility of our salience-based algorithm for performing pedometry on smart glasses.

References

1. Park, Y., Yun, S., Kim, K.H.: When IoT met augmented reality: Visualizing the source of the wireless signal in AR view. In: *MobiSys 2019 - Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, pp. 117–129 (2019)
2. Level Smartglasses by VSP Global. <https://appadvice.com/app/level-smart-glasses/1336625786>. Accessed 01 Aug 2020
3. Vue Glasses by Vigo Technologies Inc. <https://www.vueglasses.com/>. Accessed 01 Aug 2020
4. GlassUp F4 AR Smartglasses. <https://www.glassup.com/en/f4/>. Accessed 01 Oct 2020

5. MAD Gaze Smartglasses. <https://www.madgaze.com/>. Accessed 01 Aug 2020
6. abEye Smart Eyewear. <https://www.abeye.tech/>. Accessed 01 Aug 2020
7. Ellcie's Smartglasses. <https://ellcie-healthy.com/en/home/>. Accessed 01 Aug 2020
8. eSight by eSight Corp. <https://esighteyewear.com/>. Accessed 01 July 2020
9. OrCam MyEye 2.0. <https://www.orcam.com/en/myeye2/>. Accessed 01 Aug 2020
10. Amini, N., et al.: Design and evaluation of a wearable assistive technology for hemianopic stroke patients. In: Proceedings of the 2020 International Symposium on Wearable Computers, pp. 7–11 (2020)
11. Rhudy, M.B., Mahoney, J.M.: A comprehensive comparison of simple step counting techniques using wrist- and ankle-mounted accelerometer and gyroscope signals. *J. Med. Eng. Technol.* **42**(3), 236–243 (2018)
12. Pan, M.S., Lin, H.W.: A step counting algorithm for smartphone users: design and implementation. *IEEE Sens. J.* **15**(4), 2296–2305 (2014)
13. Zhang, H., Yuan, W., Shen, Q., Li, T., Chang, H.: A handheld inertial pedestrian navigation system with accurate step modes and device poses recognition. *IEEE Sens. J.* **15**(3), 1421–1429 (2015)
14. Park, S.Y., Heo, S.J., Park, C.G.: Accelerometer-based smartphone step detection using machine learning technique. In: International Electrical Engineering Congress, iEECON 2017, pp. 1–4 (2017)
15. Kim, J.W., Jang, H.J., Hwang, D.H., Park, C.: A step, stride and heading determination for the pedestrian navigation system. *J. Glob. Position. Syst.* **3**(1–2), 273–279 (2004)
16. Rantalainen, T., Karavirta, L., Pirkola, H., Rantanen, T., Linnamo, V.: Gait variability using waist-and ankle-worn inertial measurement units in healthy older adults. *Sensors* **20**(10), 2858 (2020)
17. Saeedi, R., Amini, N., Ghasemzadeh, H.: Patient-centric on-body sensor localization in smart health systems. In: 2014 48th Asilomar Conference on Signals, Systems and Computers, pp. 2081–2085 (2015)
18. Amini, N., Vahdatpour, A., Dabiri, F., Noshadi, H., Sarrafzadeh, M.: Joint consideration of energy-efficiency and coverage-preservation in microsensors networks. *Wirel. Commun. Mob. Comput.* **11**(6), 707–722 (2011)
19. Amini, N., Miremadi, S.G., Fazeli, M.: A hierarchical routing protocol for energy load balancing in wireless sensor networks. In: Canadian Conference on Electrical and Computer Engineering, pp. 1086–1089 (2007)
20. Brajdic, A., Harle, R.: Walk detection and step counting on unconstrained smartphones. In: UbiComp 2013 - Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 225–234 (2013)
21. Naqvib, N.Z., Kumar, A., Chauhan, A., Sahni, K.: Step counting using smartphone-based accelerometer. *Int. J. Comput. Sci. Eng.* **4**(5), 675 (2012)
22. Crea, S., Cipriani, C., Donati, M., Carrozza, M.C., Vitiello, N.: Providing time-discrete gait information by wearable feedback apparatus for lower-limb amputees: usability and functional validation. *IEEE Trans. Neural Syst. Rehabil. Eng.* **23**(2), 250–257 (2015)
23. Ngueleu, A.M., et al.: Design and accuracy of an instrumented insole using pressure sensors for step count. *Sensors* **19**(5), 984 (2019)
24. Bassett, D.R., Toth, L.P., LaMunion, S.R., Crouter, S.E.: Step counting: a review of measurement considerations and health-related applications. *Sports Med.* **47**(7), 1303–1315 (2016)
25. Pham, V.T., Nguyen, D.A., Dang, N.D., Pham, H.H., Tran, V.A., Sandrasegaran, K., Tran, D.T.: Highly accurate step counting at various walking states using low-cost inertial measurement unit support indoor positioning system. *Sensors* **18**(10), 3186 (2018)
26. Genovese, V., Mannini, A., Sabatini, A.M.: A smartwatch step counter for slow and intermittent ambulation. *IEEE Access* **5**, 13028–13037 (2017)

27. Zhang, X., Xu, W., Huang, M.C., Amini, N., Ren, F.: See UV on your skin: an ultraviolet sensing and visualization system. In: Proceedings of the 8th International Conference on Body Area Networks, pp. 22–28 (2013)
28. Case, M.A., Burwick, H.A., Volpp, K.G., Patel, M.S.: Accuracy of smartphone applications and wearable devices for tracking physical activity data. *JAMA – J. Am. Med. Assoc.* **313**(6), 625–626 (2015)
29. Wang, L., Liu, T., Wang, Y., Li, Q., Yi, J., Inoue, Y.: Evaluation on step counting performance of wristband activity monitors in daily living environment. *IEEE Access* **5**, 13020–13027 (2017)
30. Lee, H.H., Choi, S., Lee, M.J.: Step detection robust against the dynamics of smartphones. *Sensors* **15**(10), 27230–27250 (2015)
31. Abadleh, A., Al-Hawari, E., Alkafaween, E.A., Al-Sawalqah, H.: Step detection algorithm for accurate distance estimation using dynamic step length. In: Proceedings - 18th IEEE International Conference on Mobile Data Management, MDM, pp. 324–327 (2017)
32. Kang, X., Huang, B., Qi, G.: A novel walking detection and step counting algorithm using unconstrained smartphones. *Sensors* **18**(1), 297 (2018)
33. Scholkmann, F., Boss, J., Wolf, M.: An efficient algorithm for automatic peak detection in noisy periodic and quasi-periodic signals. *Algorithms* **5**(4), 588–603 (2012)
34. Caserman, P., Krabbe, P., Wojtusich, J., von Stryk, O.: Real-time step detection using the integrated sensors of a head-mounted display. In: 2016 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2016 - Conference Proceedings, pp. 003510–003515 (2017)
35. Burns, A., et al.: SHIMMERTM - a wireless sensor platform for noninvasive biomedical research. *IEEE Sens. J.* **10**(9), 1527–1534 (2010)
36. Apostolopoulos, I., Coming, D.S., Folmer, E.: Accuracy of pedometry on a head-mounted display. In: Conference on Human Factors in Computing Systems - Proceedings, pp. 2153–2156 (2015)
37. Amini, N.: Transmission Power Management for Wireless Health Applications. UCLA Electron. Theses Diss. (2012)
38. Mertens, C., Grenez, F., Schoentgen, J.: Speech sample salience analysis for speech cycle detection. In: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH (2009)
39. Miraftebi, A., et al.: Macular SD-OCT outcome measures: comparison of local structure-function relationships and dynamic range. *Investig. Ophthalmol. Vis. Sci.* **57**(11), 4815–4823 (2016)
40. Ewald, B., McEvoy, M., Attia, J.: Pedometer counts superior to physical activity scale for identifying health markers in older adults. *Br. J. Sports Med.* **44**(10), 756–761 (2010)
41. Vahdatpour, A., Amini, N., Sarrafzadeh, M.: Toward unsupervised activity discovery using multi-dimensional motif detection in time series. In: IJCAI International Joint Conference on Artificial Intelligence (2009)
42. Ilkovičová, Ľ., Kajánek, P., Kopáčik, A.: Pedestrian indoor positioning and tracking using smartphone sensors, step detection and map matching algorithm. In: International Symposium on Engineering Geodesy pp. 20–22 (2016)
43. Flores, G.H., Manduchi, R.: WeAllWalk: an annotated dataset of inertial sensor time series from blind walkers. In: Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS 2016, vol. 11, no. 1, pp. 1–28 (2016)
44. Chan, C.W., Rudins, A.: Foot biomechanics during walking and running. In: Mayo Clinic Proceedings, vol. 69, no. 5, pp. 448–461. Elsevier (1994)
45. Maslivec, A., Bampouras, T.M., Dewhurst, S., Vannozzi, G., Macaluso, A., Laudani, L.: Mechanisms of head stability during gait initiation in young and older women: a neuro-mechanical analysis. *J. Electromyogr. Kinesiol.* **38**, 103–110 (2018)



Real-Time Human Activity Recognition Using Textile-Based Sensors

Uğur Ayvaz¹ , Hend Elmoughni² , Asli Atalay³ , Özgür Atalay² ,
and Gökhan Ince¹ 

¹ Computer Engineering Department, Istanbul Technical University, Istanbul, Turkey
{ayvaz18,gokhan.ince}@itu.edu.tr

² Faculty of Textile Technologies and Design, Istanbul Technical University,
Istanbul, Turkey
{elmoughni,atalayoz}@itu.edu.tr

³ Textile Engineering Department, Marmara University, Istanbul, Turkey
asli.atalay@marmara.edu.tr

Abstract. Real-time human activity recognition is a popular and challenging topic in sensor systems. Inertial measurement units, vision-based systems, and wearable sensor systems are mostly used for gathering motion data. However, each system has drawbacks such as drift error, illumination, occlusion, etc. Therefore, under certain circumstances, they are not efficient alone in activity estimation. To overcome this, hybrid sensor systems were used as an alternative approach in the last decade. In this study, a human activity recognition system is proposed using textile-based capacitive sensors. The aim of the system is to recognize the basic human actions in real-time such as walking, running, squatting, and standing. The sensor system proposed in this study is used to collect human activity data from the participants with different anthropometrics and create an activity recognition system. The performance of the machine learning models is tested on unseen activity data. The obtained results showed the effectiveness of our approach by achieving high accuracy up to 83.1% on selected human activities in real-time.

Keywords: Wearable capacitive sensors · Human activity recognition · Onset-offset detection

1 Introduction

The field of Human Activity Recognition (HAR) aims to monitor and model human behaviors and body kinematics. In daily life, humans perform different body motions to sustain life routines, to carry out duties or to meet their needs. A combination of the motions of different body parts represents a specific activity such as walking, running, jumping, sitting, standing, lying on the sofa, etc. where each body motion has some distinctive characteristics on the signal level. Last decade, the progress in wearable sensor technologies made it possible to

collect useful data regarding body kinematics, muscle response, physical and neural changes that occur during the action. The estimation of these distinctive features is challenging, however capturing and modeling human movements provide utmost beneficial information for autonomous computer-aided proactive and personalized services. Therefore, in the last two decades, scientists have been intensively working on monitoring and modeling the body movements and utilizing the power of computers in terms of data processing, high computational capacities, and re-usability of activity data.

In HAR, vision-based sensors [1–3], embedded sensors such as accelerometer, gyroscope, magnetometer, etc., [4–6] and various body-mounted sensors [7–11] are popular technologies that are used to collect motion data. One of the most widely used sensors in HAR are the Inertial Measurement Unit sensors (IMUs). IMUs are capable of tracking kinematics of the user when they are attached to body joints. Although IMUs are rather low-cost, the localization of the IMUs is still a challenging task and IMUs are mostly prone to the drift problem that leads to cumulative error in sensing.

Currently, each sensor system has its limitations in terms of reliability, precision, power consumption and cost. There is no silver bullet for HAR yet and it has been a progressive field of research for the past two decades. With the development of the wearable textile-based sensors, motion data collection became more effective since they are manufactured from flexible and body conformal textile material. The body-fitted wearable sensors are good at capturing the activity pattern as they are worn directly on body joints. Textile based sensors are widely used in a number of applications including sports/recreation [12, 13], elderly care [14, 15], rehabilitation [16, 17], gaming [18–21], and robotics [22, 23] etc. Herein, low-cost textile-based sensors manufactured [24] for this study were used to establish a wearable HAR system. Stretchable wearable braces are used for sensor attachment and placement on knees. It is feasible to measure different body kinematics by attaching them to belts, wristbands, or elbows.

The textile-based sensors have embedded capacitive, resistive, optical and piezoelectric properties that are able to sense strain, pressure, touch, temperature, humidity [25, 26]. In this study, the textile-based capacitive strain sensors are used that measure the capacitance variation depending on the elongation during lower limb motion for HAR system. We tested our prototype system on four different lower limb motion activities; walking, running, squatting, and standing. We initially chose these basic activities before moving on to more complex activities, since these activities have distinctive signal patterns. We collected activity data from multiple participants. Each participant wore these knee braces and performed 4 different activities. Finally, we evaluated the performance of our HAR system by using signal segmentation and machine learning methods.

2 Related Studies

Although IMUs are one of the most commonly used sensors, they are not sufficient alone in terms of the performance in HAR, and scientists use some other

auxiliary sensors. Wu et al. [27], used a combination of wearable flexible sensor and accelerometer to recognize activities of elderly people during the rehabilitation period. Hu et al. [28] manufactured a sensor system using flexible fabrics and conductive yarns that is attachable to the knee joints. An electrogoniometer was used as a reference sensor to calibrate their system. They compared the motion capturing performance of their system with the VICON [29] motion capturing system and stated that their system is able to accurately detect knee joint movements both indoor and outdoor without hysteresis.

Another example for the sensor fusion in HAR is reported by Leier et al. [30] who developed fall detection and activity recognition system. They manufactured a smart wearable sensor for people working in challenging conditions to detect accidental situations and give feedback.

Beside the classical machine learning algorithms such as those Support Vector Machine (SVM) [31], Decision Tree (DT) [32], Random Forest (RF) [33], k-Nearest Neighbor (kNN) [34], etc., Nguyen et al. [35] investigated a new machine learning approach for human activity recognition. They used an ensemble algorithm based on voting technique. Their model was trained on two static datasets called Mobile Health (MHEALTH) [36] and University of Southern California-Human Activities Dataset (USC-HAD) [37].

Vu et al. [38] produced a flexible sensor by padding conductive ink that contains conductive carbon nanotubes to a spandex fabric to detect basic human motions. They attached their sensor to a commercial muscle pants to collect data from the upper thigh. They obtained best results in terms of recognition accuracy with RF algorithm.

Apart from offline learning approaches, Bhat et al. [39] developed a framework that performs online learning and inference in HAR. They used a combination of strain sensor and accelerometer. They also used a low-cost Internet of Things (IoT) device to test their system. The policy gradient algorithm achieved great performance in HAR.

The novelty of our proposed system is to provide a new textile-based sensor system in HAR. The capacitive properties of the sensors are feasible to mimic human motions when they are attached to body joints. In this study, we tested the performance of our HAR system in terms of the accuracy and speed of activity classification in real-time.

3 Activity Recognition System Using Textile-Based Sensors

In this section, the infrastructure of the proposed human activity recognition system is introduced as illustrated in the flow diagram of Fig. 1: 1) Data acquisition of basic human activities, 2) Preprocessing, 3) Feature extraction, and 4) Classification.

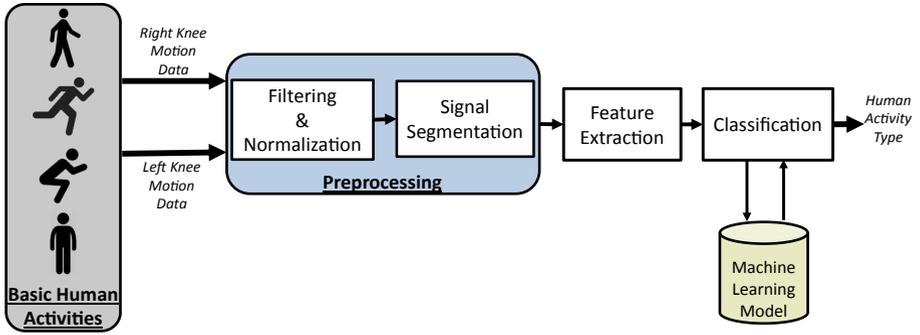


Fig. 1. Flow diagram of the proposed HAR system

3.1 Data Acquisition System

Textile-Based Capacitive Sensor. The sensor system used in this study is manufactured using low-cost textile materials with capacitive properties [24, 40]. The sensors can be mounted to apparel pieces such as knee braces to sense the motion kinematics. The capacitive strain sensors consist of two pieces of knitted conductive fabrics and a silicone insulator between these fabrics. The capacitance value changes depending on the strain value of the fabric. The capacitance variation provides information about the movement when the sensors are attached to body joints such as knees. The structure of the capacitive sensor is illustrated in Fig. 2.

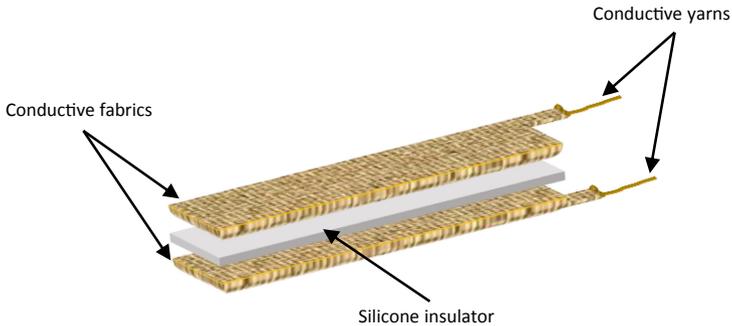


Fig. 2. Structure of the textile-based strain sensor

Design of the Knee Braces with Sensors. As legs are the most active body parts involved in many movements, we decided to track the knee joints. The knee joint movements yield information about the patterns of sports activities like walking, running, squatting, etc. One of the novelties of this study is producing a new wearable sensor to estimate human activities in real-time. Therefore, the

sensors are affixed to braces worn on both knees throughout the data collection in the training and testing procedures (Fig. 3).

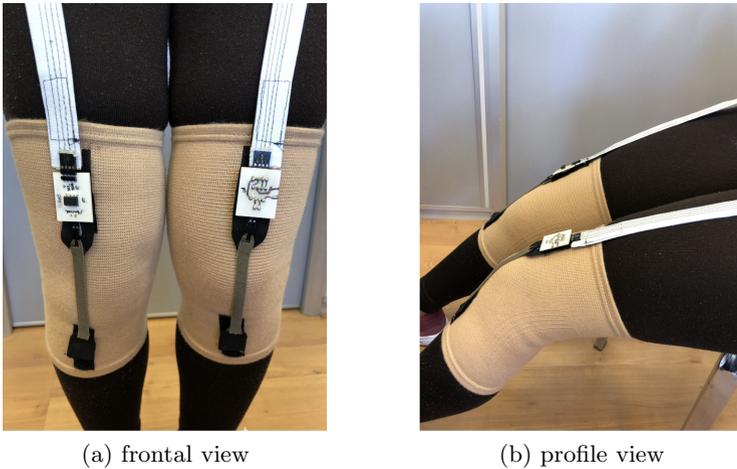


Fig. 3. Knee brace sensors

These sensors can easily be worn and taken off by the participants. In addition, they can be attached to fabrics on different body joints such as wristbands and elbows.

Design of the Hardware. Figure 4 shows a diagram of the data transmission process. The data transmission lines of the two knee sensors as shown in Fig. 3 are connected to a Transmitter Bluno (TB) module performing Bluetooth-based communication via a microcontroller.

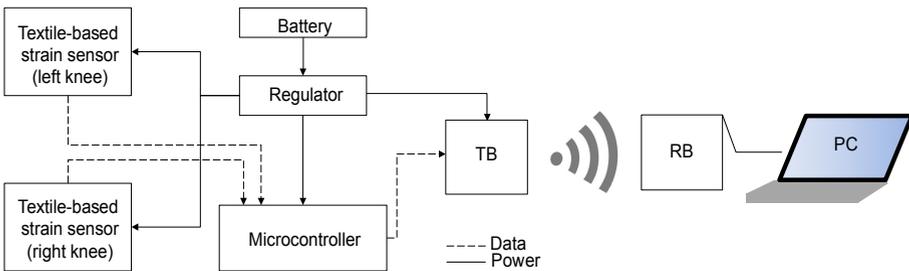


Fig. 4. Circuit diagram for data collection and transmission

A low-3.7V lipo battery is sufficient to power the TB module as well as the microcontroller and is attached to a small pocket case. The microcontroller

connected to capacitive sensor measures the capacitance difference between the two ends of each sensor and transmits the values through conductive yarns to TB. These measurements are transferred to a Receiver Bluno (RB) module via Bluetooth simultaneously. The RB is attached to the USB port of the computer in our setup. The data collection process starts as soon as the Bluno connection is established. The data transfer rate is set to 50 Hz.

3.2 Preprocessing

Signal Filtering and Normalization. The TB transfers the following sensory data to the RB: 1) timestamp value indicating the exact time when data samples are read, and capacitance values that are read from 2) the Left Knee (LK) and 3) Right Knee (RK) at that time. A smoothing filter called Savitzky-Golay [41] is applied to the incoming data. Figure 5 displays the raw data belonging to different activity classes on the left column and, the filtered data on the right column.

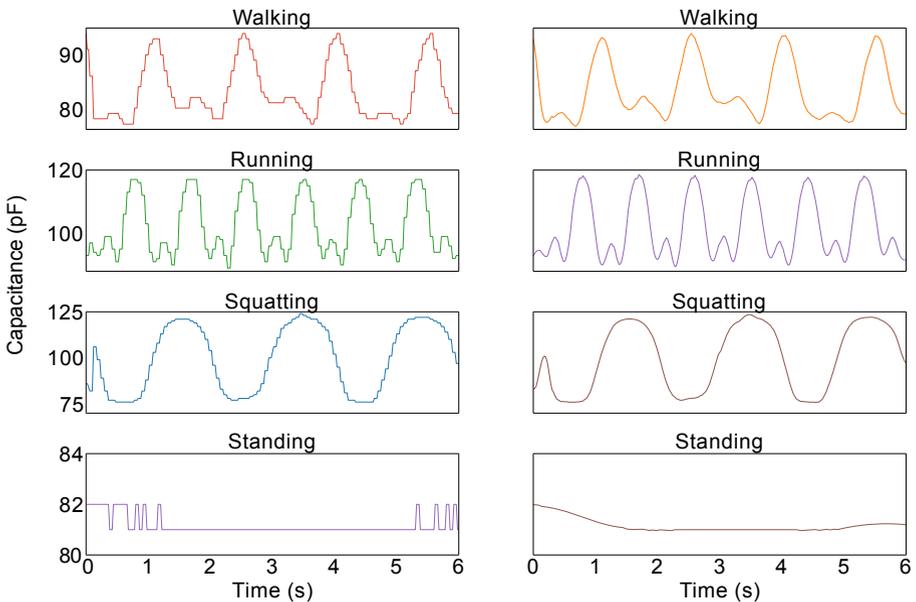


Fig. 5. Recorded signals of human activities before smoothing (left column) and after smoothing (right column)

In this study, we create different feature sets for obtaining the optimal HAR performance of our HAR system. We applied two different data normalization approaches by taking the ratio and the difference of raw sensor data retrieved from the right and left knees simultaneously. This approach is intended to

improve the data integrity and make the features more robust, especially when one (or more) of the following changes: 1) the user of the braces, 2) the initial positioning of the braces at the knee of the user, and 3) the location of the braces while performing the activity.

Signal Segmentation. We applied two different types of signal segmentation methods based on 1) Sliding Window, 2) Onset-Offset Detection to investigate the effect of these two segmentation approaches on the performance of activity classification.

Sliding Window (SW) [42] is one of the most commonly used signal segmentation methods. We heuristically fixed the window and shifting size to 128 samples (2.56 s) and 64 samples (1.28 s), respectively. Therefore, two consecutive frames had 50% of data samples in common. We applied the First In First Out (FIFO) inventory valuation method.

Onset-Offset Detection (OOD) based segmentation approach is an adaptive segmentation technique. We applied an algorithm to extract each onset and offset from walking, running and squatting signals. A single onset-offset tuple is created using the samples starting with a left local minimum that is followed by a peak and then ending with the right local minimum.

3.3 Feature Extraction

In this section, we will explain the features extracted from the frames processed using the SW and OOD techniques, separately.

The SW technique provides a fixed-sized data frame with 128 samples in each time slot [43]. Figure 6 shows one instantaneous frame obtained from each activity. It is noteworthy to mention that during running, the participant takes almost twice as many steps as he/she takes during walking. During squat motion, the signals obtained from both knees are mostly amplitude-shifted versions of each other. During standing, the signals are more or less stationary. After normalizing the capacitance values retrieved from RK and LK sensors (i.e., the ratio and difference of RK and LK signals), we extracted the 8 statistical features given in Table 1.

Since OOD approach is an adaptive method for extracting frames, frame size is equal to the number of samples between the detected onset and offset and varies depending on the speed and step size of the activity, if any. A single frame captured arbitrarily between the detected onset and offset for each activity is illustrated in Fig. 7.

It is observed that walking and running activities have similar patterns; however, both the maximum and minimum values are larger in running, whereas the duration of running action is generally shorter. In the squat motion, the dynamic range between the maxima and minima regarding the capacitance change is larger due to the extended range of the knee activity. A frame for the standing action is not explicitly shown here, since it has no significant onset-offset difference. The statistical features in Table 1 are extracted from each consecutive frame obtained by applying OOD.

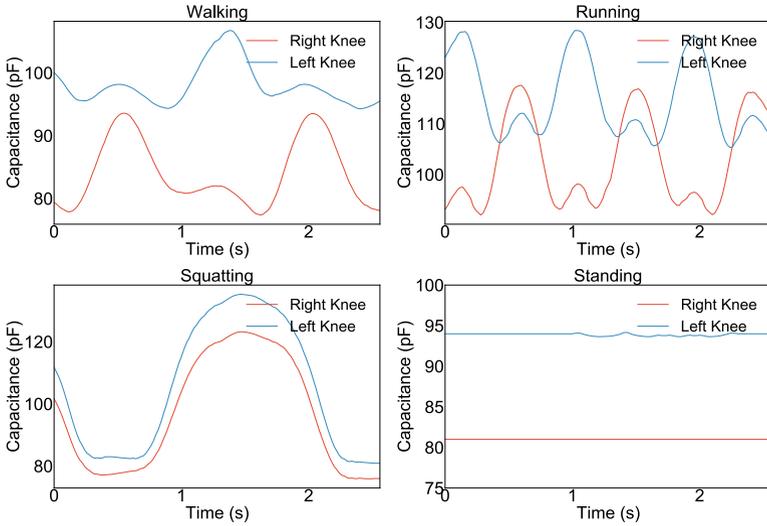


Fig. 6. Signals in one frame for each activity using shifting window method

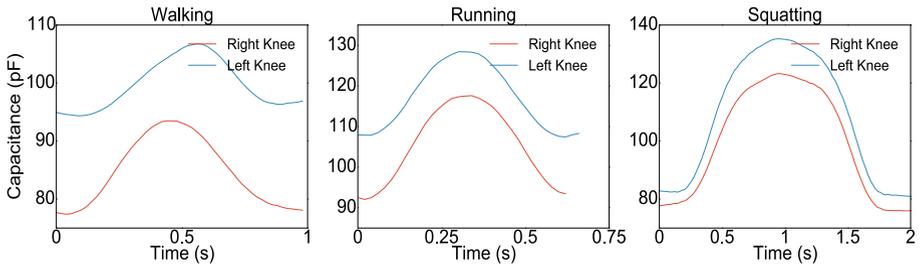


Fig. 7. Signals in one frame for each activity using onset offset detection method

As a consequence, each feature instance is labeled with action names and recorded in the feature set. The statistical features are used to model the distribution and tendency of data, because they provide distinctive properties for each activity class. Herein, the statistical features are the same for both segmentation methods.

3.4 Classification

Four different machine learning algorithms are proposed to be used in the framework of this paper. These algorithms are SVM, kNN, RF and DT. Each algorithm is trained using different parameters to create the corresponding models as given in Table 2. It is noteworthy to mention that the Radial Basis Function (RBF) [44] for SVM as a kernel achieved higher performance compared to other kernel functions.

Table 1. Features extracted from each frame.

Features	Definition
Mean	Central tendency of the data in one frame
Median	Value which divides the data in 2 equal parts
Min	Minimum value in a frame data
Max	Maximum value in a frame data
Standard deviation	Average distance between each quantity and mean
Kurtosis	Measure of whether the data has profusion of outliers or lack of outliers relative to a normal distribution
Skewness	Measure of the asymmetry of the probability distribution of a real-valued random variable
Quantiles	Points in a distribution that relate to the rank order of values in that distribution

Table 2. Hyperparameters of ML models

Model	Hyperparameter	Value
SVM-1	Kernel	RBF
	Gamma	0.01
	C	10
SVM-2	Kernel	RBF
	Gamma	“scale”
	C	1
DT	Criterion	“gini”
	Number of estimators	200
	Random state	0
kNN	Number of neighbors	5
	Leaf size	20
	Weights	“uniform”
RF	Criterion	“gini”
	Number of estimators	100
	Random state	0

4 Experiments

4.1 Experimental Framework

In this study, we used a specific sports activities dataset to compare the effect of two signal segmentation methods in classification. We investigated the performances of five different machine learning models and finally perform a real-time HAR. Firstly, we collected activity data from the participants using the knee brace sensors presented in the Sect. 3.1. Participants between the age of 21 to

30 and were assessed. Body weight and height of participants ranged from 62 to 75 kg and 168 to 178 cm, respectively. Each participant wears the braces and performs four different sports activities in different time slots. An experimenter checks the connection between the sensor and the computer to ensure that there is no interruption or distortion during the recording.

A PC with the following features; Intel Core i5 2.90 GHz with 8 GB RAM was used. On the PC side, all implementations involving signal processing, feature extraction, and classification were carried out using Python 3.7 platform.

4.2 Dataset

Our training and test data were collected from 3 participants with different anthropometrics in different sessions. The training data set include 36 min long data (12 min for each participant consisting of three minutes for each activity). The total number of data samples collected for the training process is 108.000 ($36 \times 60 \text{ s} \times 50 \text{ Hz}$). The testing data include 12 min long data (4 min for each participant consisting of one minute for each activity). The size of testing data is about 36.000 ($12 \times 60 \text{ s} \times 50 \text{ Hz}$).

4.3 Evaluation Criteria

Each machine learning model (i.e., SVM-1, SVM-2, DT, kNN, RF) is trained with the training set and tested with the test data. One of the main purposes of this study is to inspect the effects of two signal segmentation methods not only on the classification accuracy, but also on the response time considering the real-time constraints. Therefore, we evaluated the performance of each algorithm by using the 10-fold cross-validation method in terms of classification accuracy (Acc) and execution time (ET [s]) in an offline manner. In the real-time HAR system, however, we selected the system with the model showing the optimal performance.

5 Results

To demonstrate the importance of the features extraction in the offline tests, we compared the performance of the proposed system using feature vectors with the selected eight Statistical Features (w. SF) and without the statistical features (wo. SF), i.e. all samples in the frame (e.g., 128 samples) were used in the feature vector. In each experiment, we used the same classifiers with the same parameters.

Table 3 shows the classification results obtained using SW-based segmentation. This table is used to investigate the performance of different combinations of the signal processing pipeline including 1) Single sensor on one knee, 2) Two sensors without normalization, 3) Two sensors normalized by their signal ratio and, 4) Two sensors normalized by their signal difference.

Table 3. Results of classification using SW-based segmentation

Feature sets	MODELS									
	SVM-1		SVM-2		DT		kNN		RF	
	Acc	ET[s]	Acc	ET[s]	Acc	ET[s]	Acc	ET[s]	Acc	ET[s]
Single sensor (wo. SF)	0.399	15.00	0.661	5.80	0.717	4.08	0.803	1.01	0.762	28.89
Single sensor (w. SF)	0.771	1.95	0.515	1.68	0.796	0.26	0.786	0.16	0.785	8.21
Two sensors without normalization (wo. SF)	0.367	30.79	0.643	11.80	0.73	8.69	0.767	2.52	0.772	41.89
Two sensors without normalization (w. SF)	0.782	3.51	0.648	2.24	0.831	0.52	0.773	0.28	0.827	12.24
Two sensors normalized by ratio (wo. SF)	0.464	9.82	0.615	8.52	0.703	6.90	0.745	1.94	0.754	32.46
Two sensors normalized by ratio (w. SF)	0.495	1.45	0.611	0.98	0.751	0.14	0.682	0.13	0.762	7.16
Two sensors normalized by difference (wo. SF)	0.376	14.42	0.585	5.95	0.683	3.77	0.735	1.11	0.752	26.14
Two sensors normalized by difference (w. SF)	0.745	1.33	0.683	1.11	0.742	0.21	0.749	0.14	0.76	6.97

Overall results show that RF model achieved the highest accuracy on most of the feature sets. However, the highest accuracy is obtained by DT model with a value of 83.1% using two sensors without normalization and with statistical features. Although RF has the best overall performance in terms of classification accuracy, it is the laziest model in terms of ET. Herein, kNN proves to be the fastest model. However, DT can be chosen as the ideal model for our real-time HAR system with an acceptable processing latency. In addition, the results indicate that the SF have characteristic properties helping in the improvement of the classification accuracies in most of the classification models. It is also observed that the proposed normalization scheme by using neither ratio nor difference of the two sensor signals have contributed to an improvement in terms of accuracy.

In Table 4 the classification results of the onset-offset detection based segmentation approach is given. The best ET (0.20 s) is observed using kNN model. Herein, we omit the onset-offset dataset (wo. SF) since each detected onset-offset frame has a different number of samples and doesn't have a regular shape.

The detailed classification performance in terms of accuracy, precision, recall and F1-score for the DT model using SW-based segmentation and for the RF

Table 4. Results of classification using OOD-based segmentation

Feature sets	SVM-1		SVM-2		DT		kNN		RF	
	Acc	ET [s]	Acc	ET [s]						
Single sensor (w. SF)	0.782	3.65	0.473	2.94	0.801	0.34	0.758	0.20	0.812	9.48

model using OOD-based segmentation results are examined in Tables 5 and 6, respectively.

Inspecting the precision and recall values of the RF and DT algorithms for both segmentation approaches, we obtained similar results. Although these results show that RF is well-suited at classifying different types of sports activities, it is not an adequate model for real-time systems. Furthermore, the highest F1-Score is obtained by using the DT model using two sensors without normalization and SW-based segmentation.

The Confusion Matrices (CM) for the DT model using SW-based segmentation and for the RF model using OOD-based segmentation showing highest accuracy results are examined in Figs. 8a and b, respectively.

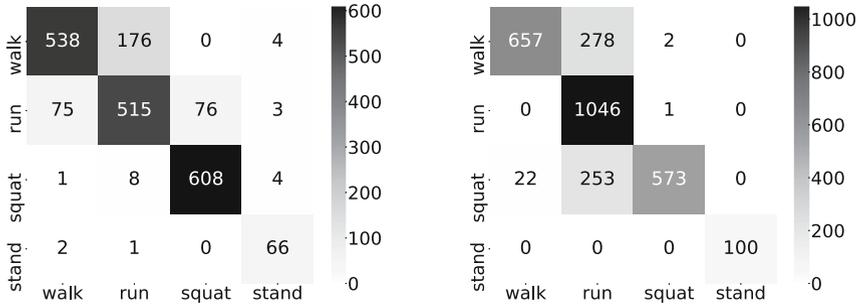
As can be clearly seen, the HAR models are more prone to misclassification between the human activities of walking and running. Besides, in OOD based classification, the models mostly confuse the activity pairs of walking and running, and running and squatting. Another important aspect to emphasize is that the number of instances in the standing activity are relatively less than other activities. Considering the performance of the recognition system is quite excellent for standing behaviour, a balanced data set after the segmentation having equal number of instances in each class would have drastically improved classification performance compared to what we have demonstrated here.

Table 5. Classification performance of different models using two sensors without normalization and SW-based segmentation

Model	Accuracy	Precision	Recall	F1-score
SVM-1	0.78	0.86	0.78	0.79
SVM-2	0.642	0.72	0.64	0.61
DT	0.831	0.83	0.83	0.83
kNN	0.773	0.84	0.77	0.78
RF	0.827	0.83	0.83	0.82

Table 6. Classification performance of different models using single sensor with OOD-based segmentation

Model	Accuracy	Precision	Recall	F1-score
SVM-1	0.782	0.84	0.75	0.75
SVM-2	0.473	0.53	0.48	0.40
DT	0.801	0.86	0.81	0.80
kNN	0.758	0.80	0.73	0.73
RF	0.812	0.87	0.81	0.81



(a) DT Model using SW segmentation (b) RF Model using OOD segmentation

Fig. 8. Confusion matrices of different HAR models

6 Conclusion

In this study, a novel approach is applied using textile-based knee sensors to recognize specific human activities, such as walking, running, squatting, and standing. These sensors have capacitive properties and are able to measure capacitance variation during the movement. We attached these sensors to knee braces to accurately measure the variations. We implemented two different signal segmentation algorithms, i.e. the sliding-window method and onset-offset detection method before statistical features were extracted. Using different classifiers, such as SVM, kNN, RF, DT, we evaluated the proposed system using performance criteria, such as classification accuracy and execution time. The overall classification results show that although RF attained the highest accuracy, it is the slowest model, thus not sufficient for our real-time HAR system. Therefore, DT model showing similar accuracy, but drastically improved execution speed is ideally recommended for HAR.

Due to Covid-19, we conducted this study with a limited size dataset. The classification accuracy also depends on the size of data. We believe that a higher accuracy will be obtained in a more balanced data set. In the future, we plan to improve our experimental setup and collect more activity data from multiple participants. Besides, only knee bracer sensors are used in this study. We aim to improve our sensor framework by attaching new sensors to full pants instead of knee braces. Our ultimate goal is to design a textile-based sensor that can be worn on the whole body. Moreover, in the next research, we will investigate the contribution of auxiliary sensors such as smartphone IMUs on the performance of the HAR system.

Acknowledgement. This research is partially funded by Marie Skłodowska-Curie Individual Fellowships (IF) as part of the project “Textile based soft sensing actuators for soft robotic applications - TexRobots”, (Grant No: 842786).

References

1. Desai, C., Ramanan, D.: Detecting actions, poses, and objects with relational phraselets. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7575, pp. 158–172. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33765-9_12
2. Donahue, J., et al.: Long-term recurrent convolutional networks for visual recognition and description. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2625–2634 (2015)
3. Nazir, S., Yousaf, M.H., Velastin, S.A.: Evaluating a bag-of-visual features approach using spatio-temporal features for action recognition. *Comput. Electr. Eng.* **72**, 660–669 (2018)
4. Ronao, C.A., Cho, S.-B.: Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Syst. Appl.* **59**, 235–244 (2016)
5. Hassan, M.M., Uddin, Md.Z., Mohamed, A., Almogren, A.: A robust human activity recognition system using smartphone sensors and deep learning. *Future Gener. Comput. Syst.* **81**, 307–313 (2018)
6. Lara, O.D., Labrador, M.A.: A survey on human activity recognition using wearable sensors. *IEEE Commun. Surv. Tutor.* **15**(3), 1192–1209 (2012)
7. Labrador, M.A., Lara Yejas, O.D.: *Human Activity Recognition: Using Wearable Sensors and Smartphones*. CRC Press, Boca Raton (2013)
8. Li, F., Shirahama, K., Nisar, M.A., Köping, L., Grzegorzec, M.: Comparison of feature learning methods for human activity recognition using wearable sensors. *Sensors* **18**(2), 679 (2018)
9. De Leonardis, G., et al.: Human activity recognition by wearable sensors: comparison of different classifiers for real-time applications. In: 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA), pp. 1–6. IEEE (2018)
10. Davila, J.C., Cretu, A.-M., Zaremba, M.: Wearable sensor data classification for human activity recognition based on an iterative learning framework. *Sensors* **17**(6), 1287 (2017)
11. Jiang, W., Yin, Z.: Human activity recognition using wearable sensors by deep convolutional neural networks. In: Proceedings of the 23rd ACM International Conference on Multimedia, pp. 1307–1310 (2015)
12. Stetter, B.J., Ringhof, S., Krafft, F.C., Sell, S., Stein, T.: Estimation of knee joint forces in sport movements using wearable sensors and machine learning. *Sensors* **19**(17), 3690 (2019)
13. Ray, T., et al.: Soft, skin-interfaced wearable systems for sports science and analytics. *Curr. Opin. Biomed. Eng.* (2019)
14. Wang, Z., Yang, Z., Dong, T.: A review of wearable technologies for elderly care that can accurately track indoor position, recognize physical activities and monitor vital signs in real time. *Sensors* **17**(2), 341 (2017)
15. Awais, M., et al.: An internet of things based bed-egress alerting paradigm using wearable sensors in elderly care environment. *Sensors* **19**(11), 2498 (2019)
16. Porciuncula, F., et al.: Wearable movement sensors for rehabilitation: a focused review of technological and clinical advances. *PM&R* **10**(9), 220–232 (2018)
17. Dorsch, A.K., King, C.E., Dobkin, B.H.: Wearable wireless sensors for rehabilitation. In: Reinkensmeyer, D.J., Dietz, V. (eds.) *Neurorehabilitation Technology*, pp. 605–615. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-28603-7_29

18. Ahram, T., Falcão, C.: *Advances in Human Factors in Wearable Technologies and Game Design*. Springer, Heidelberg (2020). <https://doi.org/10.1007/978-3-319-60639-2>
19. Cao, M., Xie, T., Chen, Z.: Wearable sensors and equipment in VR games: a review. In: Pan, Z., Cheok, A.D., Müller, W., Zhang, M., El Rhalibi, A., Kifayat, K. (eds.) *Transactions on Edutainment XV*. LNCS, vol. 11345, pp. 3–12. Springer, Heidelberg (2019). https://doi.org/10.1007/978-3-662-59351-6_1
20. Wei, S.-Y., et al.: Runplay: action recognition using wearable device apply on Parkour game. In: *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pp. 133–135 (2016)
21. Nam, S.-H., Kim, J.-Y.: Dance exergame system for health using wearable devices. *IEEE Access* **6**, 48224–48230 (2018)
22. Moschetti, A., Cavallo, F., Esposito, D., Penders, J., Di Nuovo, A.: Wearable sensors for human-robot walking together. *Robotics* **8**(2), 38 (2019)
23. Özcan, B., Caligiore, D., Sperati, V., Moretta, T., Baldassarre, G.: Transitional wearable companions: a novel concept of soft interactive social robots to improve social skills in children with autism spectrum disorder. *Int. J. Soc. Robot.* **8**(4), 471–481 (2016)
24. Atalay, A., et al.: Batch fabrication of customizable silicone-textile composite capacitive strain sensors for human motion tracking. *Adv. Mater. Technol.* **2**(9), 1700136 (2017)
25. Gonçalves, C., da Silva, A.F., Gomes, J., Simoes, R.: Wearable e-textile technologies: a review on sensors, actuators and control elements. *Inventions* **3**(1), 14 (2018)
26. Wang, J., Chunhong, L., Zhang, K.: Textile-based strain sensor for human motion detection. *Energy Environ. Mater.* **3**(1), 80–100 (2020)
27. Wu, Y., Qi, S., Hu, F., Ma, S., Mao, W., Li, W.: Recognizing activities of the elderly using wearable sensors: a comparison of ensemble algorithms based on boosting. *Sens. Rev.* (2019)
28. Sufeng, H., Dai, M., Dong, T., Liu, T.: A textile sensor for long durations of human motion capture. *Sensors* **19**(10), 2369 (2019)
29. Merriaux, P., Dupuis, Y., Boutteau, R., Vasseur, P., Savatier, X.: A study of vicon system positioning performance. *Sensors* **17**(7), 1591 (2017)
30. Leier, M., Jervan, G., Allik, A., Pilt, K., Karai, D., Fridolin, I.: Fall detection and activity recognition system for usage in smart work-wear. In: *2018 16th Biennial Baltic Electronics Conference (BEC)*, pp. 1–4. IEEE (2018)
31. Amari, S., Si, W.: Improving support vector machine classifiers by modifying kernel functions. *Neural Netw.* **12**(6), 783–789 (1999)
32. Freund, Y., Mason, L.: The alternating decision tree learning algorithm. In: *ICML 1999*, pp. 124–133 (1999)
33. Liaw, A., Wiener, M., et al.: Classification and regression by randomforest. *R News* **2**(3), 18–22 (2002)
34. Keller, J.M., Gray, M.R., Givens, J.A.: A fuzzy k-nearest neighbor algorithm. *IEEE Trans. Syst. Man Cybern.* **4**, 580–585 (1985)
35. Nguyen, H.D., Tran, K.P., Zeng, X., Koehl, L., Tartare, G.: Wearable sensor data based human activity recognition using machine learning: a new approach. *arXiv preprint arXiv:1905.03809* (2019)
36. Banos, O., et al.: mHealthDroid: a novel framework for agile development of mobile health applications. In: Pecchia, L., Chen, L.L., Nugent, C., Bravo, J. (eds.) *IWAAL 2014*. LNCS, vol. 8868, pp. 91–98. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-13105-4_14

37. Zhang, M., Sawchuk, A.A.: USC-HAD: a daily activity dataset for ubiquitous activity recognition using wearable sensors. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing, pp. 1036–1043 (2012)
38. Chi, V., Kim, J.: Human motion recognition by textile sensors based on machine learning algorithms. *Sensors* **18**(9), 3109 (2018)
39. Bhat, G., Deb, R., Chaurasia, V.V., Shill, H., Ogras, U.Y.: Online human activity recognition using low-power wearable devices. In: Proceedings of the International Conference on Computer-Aided Design, p. 72. ACM (2018)
40. Ozlem, K., Atalay, O., Atalay, A., Ince, G.: Textile based sensing system for lower limb motion monitoring. In: Masia, L., Micera, S., Akay, M., Pons, J.L. (eds.) ICNR 2018. BB, vol. 21, pp. 395–399. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-01845-0_79
41. Luo, J., Ying, K., Bai, J.: Savitzky-Golay smoothing and differentiation filter for even number data. *Sig. Process.* **85**(7), 1429–1434 (2005)
42. Kozina, S., Lustrek, M., Gams, M.: Dynamic signal segmentation for activity recognition. In: Proceedings of International Joint Conference on Artificial Intelligence, Barcelona, Spain, vol. 1622, p. 1522 (2011)
43. Anguita, D., Ghio, A., Oneto, L., Parra, X., Reyes-Ortiz, J.L.: A public domain dataset for human activity recognition using smartphones. In: Esann (2013)
44. Shi, H., Xiao, H., Zhou, J., Li, N., Zhou, H.: Radial basis function kernel parameter optimization algorithm in support vector machine based on segmented dichotomy. In: 2018 5th International Conference on Systems and Informatics (ICSAI), pp. 383–388. IEEE (2018)



Extraction of Respiratory Signals and Respiratory Rates from the Photoplethysmogram

Shenglang Xiao¹, Pengfei Yang¹(✉), Luyao Liu², Zhiqiang Zhang³, and Jiankang Wu⁴

¹ Xidian University, Xi'an 710071, Shaanxi, China

slxiao@stu.xidian.edu.cn, pfyang@xidian.edu.cn

² University of Science and Technology Beijing, Beijing 100083, China

17888841360@163.com

³ University of Leeds, Leeds LS2 9JT, UK

Z.Zhang3@leeds.ac.uk

⁴ University of Chinese Academy of Sciences, Beijing 100049, China

jkwu@ucas.ac.cn

Abstract. Respiration rate (RR) is an important indicator of human health assessment which can be estimated by extracting respiratory signals from the photoplethysmogram (PPG). The goal of this study is to propose an alternative method, for obtaining accurate estimation of respiratory rate (RR) from the PPG signal. The proposed algorithm is based on the multiple autoregressive models and autocorrelation analysis (AC-AR). In AC-AR, the autoregressive model (AR) is applied to determining the dominant respiratory rate from the PPG, and autocorrelation is applied to reduce the effect of clutter in the three respiratory-induced variations. Meanwhile, this paper introduced signal quality indices (SQI) to improve reliability of results. This algorithm is tested using an open source database: The CapnoBase benchmark dataset, which comprising 42 eight-minute PPG recording and respiratory signal acquired from both children and adults in different clinical setting. Compared with that of existing method in the literature, the average absolute error percentage (AAEP) of the proposed algorithm is less than 3.72%, which demonstrated that our presented AC-AR bring a significant improvement in accuracy.

Keywords: Respiratory rate (RR) · Photoplethysmography (PPG) · AR model · Data fusion

1 Introduction

Respiratory rate (RR) is one of the indicators used by hospitals to monitor patients for abnormal conditions, such as cardiac, respiratory arrest, systemic inflammatory response syndrome (SIRS), and renal failure [1]. Adults have a normal respiratory rate of 8–20 breaths per minute (bpm) [2]. In a study of respiratory abnormalities, 54% of cardiac

arrest patients had at least one $RR > 27$ bpm three days before cardiac arrest [3]. Therefore, it is essential to monitor patients' respiratory. However, although pulse oximetry can be used to continuously measure heart rate (HR) and peripheral oxygen saturation (SpO₂), continuous estimation of RR requires additional equipment, such as measurement of gas flow. Therefore, there is a need to improve the accuracy of RR estimates from the electrocardiogram (ECG), the photoplethysmogram (PPG) obtained from pulse oximeters [4, 5]. This paper focuses on extracting RR from PPG signals. Pulse oximeters estimate blood oxygen saturation (SpO₂) based on Beer-Lambert's law, which indicates that the light intensity decays exponentially as it passes through the medium and the degree of attenuation is related to wavelength [6]. Therefore, we can use PPG to show the change in blood volume in the finger over time. The PPG signal includes a pulse component and a constant component, and the respiratory signal and the heartbeat signal are included in the pulse component [7].

The modulation of PPG signals by respiratory cycle includes a variety of ways, including amplitude modulation (AM), frequency modulation (FM), and baseline wander (BW) [8]. To extract the respiratory modulation signal from the PPG, the most common method is to detect the peak and trough of the PPG signal and obtain the respiratory modulation signal by calculation. In peak-trough detection in the time-domain, we define the time-series of peaks in the PPG to be a set of pairs $\{t_{pk,i}, y_{pk,i}\}_{i=1\dots N_{pk}}$, and the time-series of troughs in the PPG to be a set of pairs $\{t_{tr,i}, y_{tr,i}\}_{i=1\dots N_{tr}}$, where N_{pk} and N_{tr} are the number of peaks and troughs, respectively. $N_{pk} \neq N_{tr}$ will be caused by noise in the signal or misdetection of the detection algorithm [6, 9, 10]. The time-series of peak and trough will be used to derive three different respiration-modulated signals, representing three different kinds of information about respiration [6]. (1) Respiration leads to change in cardiac output, causing respiratory-induced amplitude variation (RIAV), that is, change in peripheral pulse intensity. RIAV is defined as the height difference between two adjacent peaks and troughs. Therefore, $y_{RIAV} = \{t_i, y_{pk,i} - y_{tr,i}\}_{i=1\dots N_{tr}}$. (2) Respiration causes periodic changes in heart rate, namely respiratory-sinus arrhythmia (RSA). It appears that the heart rate increases during inhalation and decreases during exhalation, thereby causing respiratory-induced frequency variation (RIFV), which is defined as the time interval between successive PPG peaks. Therefore, $y_{RIFV} = \{t_i, t_{i+1} - t_i\}_{i=1\dots N_{pk}}$. (3) Respiration causes change in the pressure in the chest, causing blood exchange between the pulmonary and systemic circulation. Leading to a change in the baseline of perfusion, called respiratory induced intensity variation (RIIV). RIIV appears as the change in the amplitude of the peak of PPG waveform. Therefore, $y_{RIIV} = \{t_i, y_{pk,i}\}_{i=1\dots N_{pk}}$. There are also other respiratory modulation signals, such as pulse width variability [11], which can be used to estimate the RR (see Fig. 1).

Respiration modulates the PPG in different ways. Different methods of modulation signal extraction have been proposed in a number of literatures, which are discussed in Sect. 2. In Sect. 3, the improved Incremental-Merge Segmentation (IMS) algorithm and peak detection algorithm are introduced, and a combined algorithm for spectral analysis is proposed to improve the accuracy of RR estimation. Databases and evaluation methods are described in Sect. 4. Section 5 shows the results of RR estimation using the proposed algorithm. The significance and results of this study are discussed in Sect. 6.

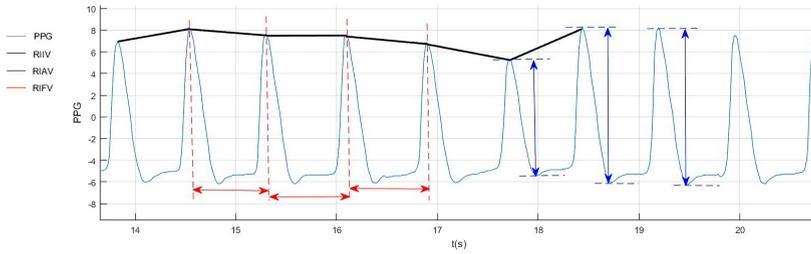


Fig. 1. PPG waveform and three respiration modulation signals. RIIV is the change in the baseline of perfusion; RIAV is the change in peripheral pulse intensity; RIFV is periodic changes in heart rate.

2 Related Work

Different algorithms have been proposed to estimate RR from PPG, such as digital filters [12], fast Fourier transforms (FFT) [6], wavelet decomposition [13] and hidden semi-Markov model [14]. Autoregressive model (AR) [9, 15], principal component analysis (PCA) [16] and artificial neural network (NN) [17], have all been successfully applied to various PPG databases with good estimation results. Some studies use neural networks to analyze the three modulation signals to select the best waveform for the algorithm design. There are also studies that use data fusion to combine estimates of multiple modulation signals [6, 18]. However, these methods have higher requirements for time domain waveforms. This problem can be solved by autocorrelation analysis. Autocorrelation analysis is a mathematical tool for finding repetitive patterns, such as periodic signals masked by noise. Since the respiratory signal can be viewed as a noisy periodic signal, the autocorrelation analysis can be used to calculate the respiratory rate [19]. In the autocorrelation signal, each peak (except the first) represents a period of strong autocorrelation, and the period with the greatest correlation can be regarded as the RR.

In order to extract effective information from chaotic PPG signal, researchers have proposed various methods. Byung S. Kim et al. used independent component analysis (ICA) to reduce motion artifact [16]. Despite so many advances, the use of pulse oximeters to measure respiratory rate has only recently been used commercially, because there are more reliable methods of RR estimation in clinical settings, such as spirometry or capacitance. Therefore, it is important to come up with a reliable method for PPG. A common method now is to introduce the signal quality index (SQI) to evaluate the signal quality [20]. If the PPG signal does not carry meaningful physiological information, it will not be algorithmically estimated. The lack of quality indicators may lead to serious clinical errors, and the introduction of evaluation indicators can improve accuracy and reliability.

To overcome this limitation, we designed an algorithm that uses all available respiratory-induced waveform to achieve significant accuracy. In this study, we propose an algorithm that combines the results of the three respiratory-induced variations described above, and use the AC-AR algorithm to estimate respiratory rate.

3 Proposed Algorithm

As for RR estimation, the most essential is the extraction of respiratory modulation signals. The main methods for extracting respiratory modulation signals are peak detection and signal quality assessment. Before the peak detection, the pre-processing procedure should be carried out first. A high-pass filter is applied to remove the dc component of the PPG signal. Then the PPG is segmented into pulses using IMS algorithm and artifacts are detected which are used to calculate signal quality (see Fig. 2). If the assessed quality is low, the RR estimation is not provided. This paper proposes an improved method for peak detection and signal evaluation, and then uses spectrum analysis and data fusion to estimate respiratory rate. In the following sections, we will describe RR estimation in more detail.

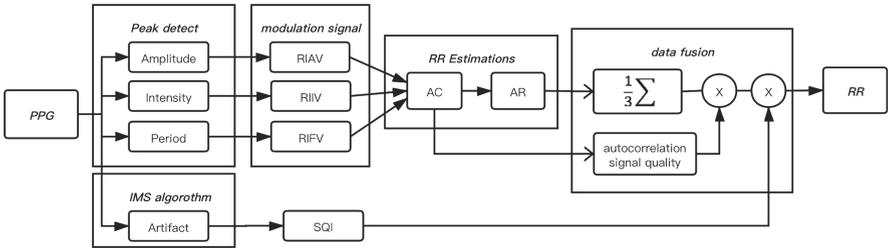


Fig. 2. The AC-AR algorithm flowchart.

3.1 Peak Detection

The general principle of peak detection is that any singular point of a differentiable signal corresponds to a zero-crossing point or two inflection points in its derivative signal. This paper proposes a new method for peak detection (see Fig. 3). This method does not need to solve the second derivative, also does not need to solve the inflection point of the first derivative. Therefore, the computational efficiency can be improved to facilitate real-time processing. The specific method is shown in the *Algorithm 1* below. In order to avoid the impact of PPG signal amplitude changes on peak detection and verification, a 10 s sliding window is used for PPG waveform with an overlap time of 5 s.

In order to improve detection accuracy, peak verification is needed which mainly considers two factors, the amplitude threshold and the time interval threshold. The method of setting the amplitude threshold is shown below. Let thresh1 be the ninth decile and thresh2 be the first decile. Then,

$$\text{thresh3} = \text{thresh2} + 0.7 * (\text{thresh1} - \text{thresh2}) \quad (1)$$

and then,

$$\text{highdiff} = \text{abs}(\text{peaks} - \text{thresh1}) \quad (2)$$

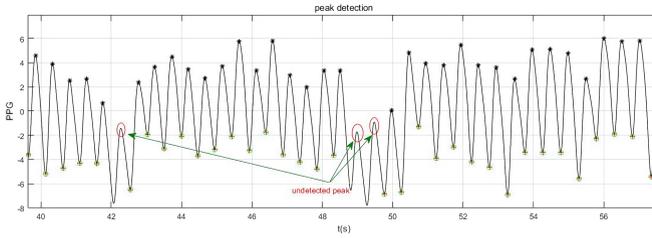


Fig. 3. The result of peak detection algorithm

$$\text{middlediff} = \text{abs}(\text{peaks} - \text{thresh3}) \quad (3)$$

$$\text{lowdiff} = \text{abs}(\text{peaks} - \text{thresh2}) \quad (4)$$

where *peaks* is the time-series of peaks. If condition

$$(\text{highdiff} < \text{middlediff}) \& (\text{highdiff} < \text{lowdiff}) \quad (5)$$

is met, the peak point is recorded.

Algorithm 1 Peak detection algorithm

```

1: data ← PPG;
2: diff ← diff(data);
3: left_diff ← diff [1: end-1];
4: logical_left ← logical(left_diff>0);
5: right_diff ← diff [2: end];
6: logical_right ← logical(right_diff>0);
7: peaks ← find (logical_left & logical_right = 1) + 1;

```

where *diff(X)* calculates the difference between *X* adjacent elements along the first array dimension whose size is not equal to 1; *logical(A)* converts *A* to an array of logical values. Any non-zero element in *A* will be converted to the logical value 1 (true), and zero to the logical value 0 (false); *find(X)* returns a vector containing a linear index of each non-zero element in the array *X*.

In terms of time interval, because the pulse wave is mainly regulated by the heartbeat, and the normal person's resting heart rate is 60–100 bpm, the peak of the time interval corresponding to this range will be detected. For trough detection, the PPG waveform shows that the minimum value between two peaks is the trough. The time-series of peak and trough will be used to derive three different respiration-modulated signals, representing three different kinds of information about respiration.

3.2 Signal Quality Index

Since there are motion artifacts and noise that cannot be filtered out in the PPG, the quality of the PPG needs to be evaluated. The signal quality evaluation method used in this paper is analyzed for consistency. First, the PPG pulse is divided into line segments

using Incremental-Merge Segmentation (IMS) algorithm. According to the shape of the line segment, it is distinguished into effective signals and noise. This paper calculates the ratio of artifact and clip in the signal as the SQI for the signal quality.

The IMS algorithm can be used for real-time processing with a sliding window structure [21]. The algorithm only needs to set a parameter m (the number of points moved each time, mainly related to the sampling rate). The principle is to divide the PPG signal into n m -length segments, calculate the slopes of these segments, and merge them with the same slope, and the different slopes are divided into new Line segments. After the IMS algorithm, each PPG pulse is represented as a straight line from the beginning of the pulse to the end of the primary peak of the pulse (see Fig. 4).

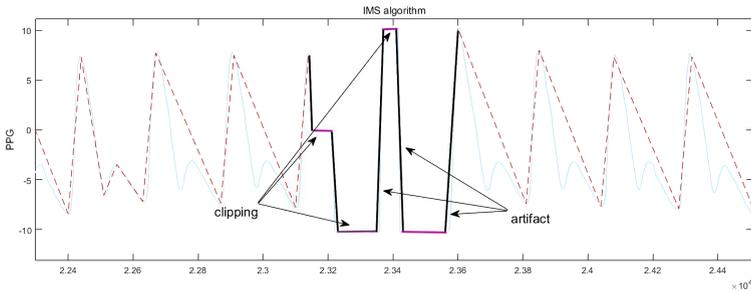


Fig. 4. The result of IMS algorithm and Artifact detection

Since the upslope and downslope line segments have a one-to-one correspondence, the upslope line segments are analyzed separately. If the amplitude and slope of the upslope line segment both exceed the threshold, it is regarded as artifact; if the slope is zero, it is regarded as clip; the line segment immediately after the clip is also artifact. SQI is calculated according to the ratio of artifact and clip in the PPG for signal quality. If the SQI is less than the threshold, then the data window is labeled as low RR estimation quality (see Fig. 5).

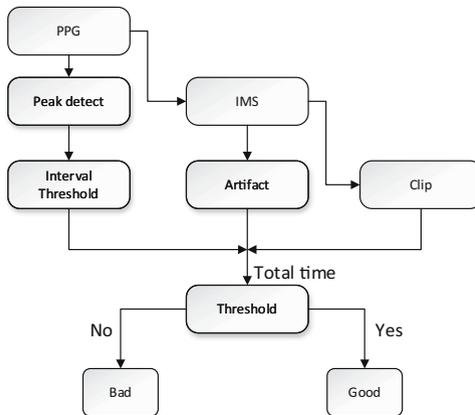


Fig. 5. SQI algorithm flowchart

3.3 Estimation of Respiratory Rate

After the time-series of peak and trough are obtained from the peak detection, the above method is used to calculate three kinds of respiratory modulation signals: RIIV, RIFV, and RIAV. Because these modulation signals are unevenly-sampled time-series, they are resampled at $f_s = 4$ Hz, using linear interpolation. Each resampled time-series is normalized using a zero-mean unit-variance transformation, so that the amplitudes of the three modulation signals are unified to the same range for subsequent spectral analysis. Then use a high-pass filter to remove the low-frequency signal, and a moving average filter to smooth the signal. The next step is to extract the respiratory rate.

Autocorrelation Analysis. Autocorrelation analysis is a mathematical tool for finding repetitive patterns, such as periodic signals masked by noise. Since the respiratory signal can be viewed as a noisy periodic signal, the autocorrelation analysis can be used to calculate the respiratory rate. The autocorrelation formula is as follows,

$$\rho_x(\tau) = \frac{E[(x_i - \mu)(x_{i+\tau} - \mu)]}{\sigma^2} \tag{6}$$

where x_i is the time-series of signal, $x_{i+\tau}$ is the time-series translated by τ units, μ is the mean, and σ^2 is the variance. An autocorrelation sequence $C[\tau]$ can be combined by the value of formula from $\tau = 0$ to $\tau = n - 1$. In the autocorrelation signal, each peak (except the first) represents a period of strong autocorrelation, and the period with the greatest correlation can be regarded as the RR. Therefore, we can use autocorrelation analysis to obtain the periodicity of the respiratory signal.

We apply the autocorrelation to analyze the signal as Fig. 6 shows. As the result, the autocorrelation coefficient waveform contains the breath rate signal, and it overcomes the effect of noise and clutter. At last, we can analyze these coefficients by AR model to acquire the more accurate RR.

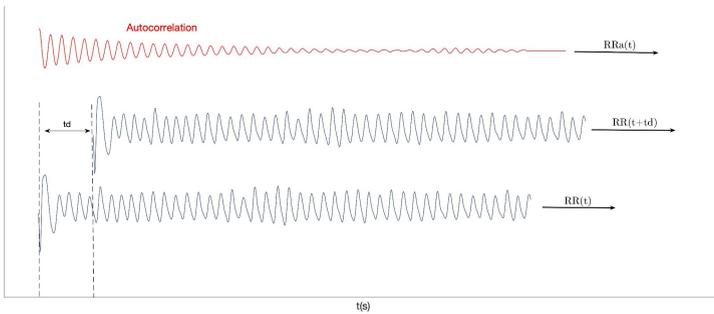


Fig. 6. Autocorrelation analysis diagram

AR Model. AR model is an alternative to the discrete Fourier transform (DFT) and one of the methods for high-resolution spectral estimation of short-term sequences. In

biomedical engineering, AR models are widely used for spectrum analysis of heart rate variability and electroencephalography analysis. In AR model, each point in the time-series is a regression of its past points. The number M of past points used is called the order of AR model. AR model can be regarded as a filter, which divides the time-series into predictable time series and prediction error series. Compared with the DFT, it provides a smoother and more intuitive power spectrum, and yet is more complicated. The AR model is defined as,

$$x[n] = \sum_{i=1}^M a_i x[n-i] + \varepsilon[n] \quad (7)$$

where M is the model order, a_i is the weight, and $\varepsilon[n]$ is the prediction error and follows $\varepsilon \sim N(0, \sigma)$. The least squares method is used to minimize the prediction error $\varepsilon[n]$ to obtain the optimal parameter a_{opt} . Matrix the above formula,

$$x = Xa + \varepsilon \quad (8)$$

when the prediction error $\varepsilon[n]$ reaches the minimum, the parameter a_{opt} is optimal, that is,

$$\varepsilon = x - Xa_{opt} = 0 \quad (9)$$

$$X^T \varepsilon = X^T (x - Xa_{opt}) = 0 \quad (10)$$

$$X^T x = X^T X a_{opt} \quad (11)$$

$$\left(X^T X\right)^{-1} \left(X^T X\right) a_{opt} = a_{opt} = \left(X^T X\right)^{-1} X^T x \quad (12)$$

Another point of AR model is the choice of model order M . Different orders have different effects in AR model. In practice, by fitting the sequence to multiple orders, the order with the best effect is selected. The most common selection criterion is Akaike's Information Criterion (AIC),

$$AIC(M) = N \cdot \ln\left(\sigma_p^2\right) + 2M \quad (13)$$

where σ_p^2 is the variance of the prediction error $\varepsilon[n]$. The best model order is M that minimizes AIC.

Then, the time-series spectrum $R(e^{j\omega})$ can be obtained by multiplying the square of the transfer function and the variance of the prediction error,

$$R(e^{j\omega}) = \left|H(e^{j\omega})\right|^2 \sigma_p^2 \quad (14)$$

where $H(e^{j\omega})$ is the transfer function of AR model,

$$H(e^{j\omega}) = \frac{1}{1 - a_1 e^{-j\omega} - \dots - a_M e^{-jM\omega}} \quad (15)$$

The autocorrelation method can remove the noise in the periodic signal with the characteristics described above. Each peak of the autocorrelation sequence represents a period of strong autocorrelation, so the period of the autocorrelation corresponds to the period of the original signal. Therefore, the autocorrelation signal of the respiratory modulation signal used as the input signal of AR spectrum analysis to estimate the RR.

3.4 Data Fusion

In order to improve the accuracy and reliability of RR estimation, data fusion can be performed on three kinds of respiratory modulation signals. A common fusion method is to average the spectrum of the three kinds of modulation signals, and the maximum value is selected as RR. This paper proposes an improvement method.

Due to the autocorrelation signal of the modulation signal can be analyzed as a breathing signal, and the waveform of the autocorrelation signal is more regular. First use the IMS algorithm to segment the autocorrelation signal. Since the autocorrelation signal of a normal breathing signal is approximately a sine wave, its variance value is small, and its mean value is close to 1, that is, the waveform is relatively stable (see Fig. 7). An analysis of variance is performed on the autocorrelation signal, and autocorrelation signal quality (ASQ) is used as an indicator,

$$\text{ASQ} = \text{var}/\text{mean} \quad (16)$$

where var is the variance of these line segments, and mean is the mean of these line segment. The spectrums of the modulation signal with ASQ less than a certain range are processed by average.

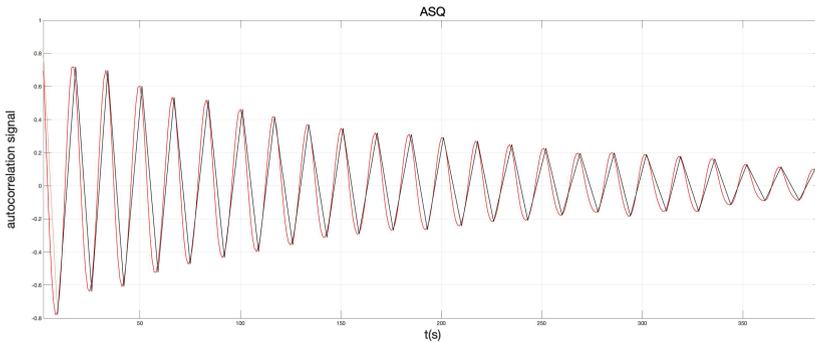


Fig. 7. Autocorrelation signal quality

4 Materials and Methods

The method was implemented and tested using the MATLAB software framework, v.R2018a (Mathworks, Natick, MA, USA), and it was designed to be used on each single window (independently of the others). An open-source dataset, the CapnoBase benchmark dataset (available at www.capnobase.org) was used for the analysis described in this paper. The database which contained PPG signals, ECG signals, and respiratory signals was collected by Karlen et al. The sampling frequency was 300 Hz. These data were collected from 59 children (mean age 8.7) and 35 adults (mean age 52.4). The author of the database randomly selects a part of it, and then combines it into a new data set, which contains 42 data segments with a duration of 8 min. Each recorded CO₂

tracing waveform was used as a reference “gold standard” record for RR. Each breath on the carbon dioxide map in the database has been manually marked by the research assistant, and the reference RR value is derived based on the average time between two consecutive breaths using annotations [6].

Before data analysis, preprocessing is performed. We remove the linear trend of the signal to avoid errors caused by data offset and then use low-pass filtering to remove high-frequency noise. The three methods mentioned above are then used to extract the respiratory modulation signal. The signals are resampled since they are unevenly-sampled time-series. Each resampled time-series is normalized using a zero-mean unit-variance transformation. In order to increase the reliability of the signal, a signal evaluation index (SQI) is introduced to evaluate the quality of the modulated signal. In this study, RR was estimated to be within a reasonable range of breathing frequencies set at 4 to 65 breaths per minute.

To extract the RR from the PPG, a common method is to use a sliding window to segment the PPG time-series, and each window obtains an RR. This experiment uses two windows size of 30 s and 60 s, and estimates the RR every 3 s and 6 s, respectively. Based on the estimated value and the reference value, performance was assessed by calculating the mean absolute error (MAE) and average absolute error percentage (AAEP) in breaths per minute for each record, defined as,

$$\text{MAE} = \frac{1}{w} \sum_{i=1}^w |y_i - y_{ref,i}| \quad (17)$$

$$\text{AAEP} = \frac{1}{w} \sum_{i=1}^w \frac{|y_i - y_{ref,i}|}{y_{ref,i}} \quad (18)$$

where w is the number of reference value, y_i is estimate value, and $y_{ref,i}$ is reference value. The observation value of each algorithm is compared with the reference observation value, and the measurement error of the observation value of each algorithm is calculated. The first 64 s are not used for performance measurements because they are used to initialize high-pass filters and sliding window. All RR estimation methods, including the single modulation methods, ignore the measurement errors of the windows containing artifacts automatically detected by the algorithm.

5 Results

According to the experiment, for the signal with a sampling frequency of 300 Hz, the IMS algorithm can obtain a better result when $m = 10$. It provides a good tradeoff between calculating load and time resolution for pulse peak detection. Different time windows have no significant effect on RR measurement errors, but larger windows can slow down the real-time response of the algorithm. But when the time window is too small, the lower respiratory rate cannot be detected. Therefore, we eclectically selected the time window of 60 s for analysis. Firstly, the PPG signal is analyzed from the time-frequency domain. Figure 8(a) is the PPG signal, and the spectrum analysis is shown in Fig. 8(b). From the frequency spectrum, we can find that the energy of the breath rate signal is weaker compared with the heart signal and its harmonics. Therefore, it is necessary to

extract the respiratory modulation signal from PPG signal to avoid the interference of heartbeat signal. As can be seen from Fig. 9, after using the AC-AR algorithm proposed in this paper, the spectrum is concentrated near RR.

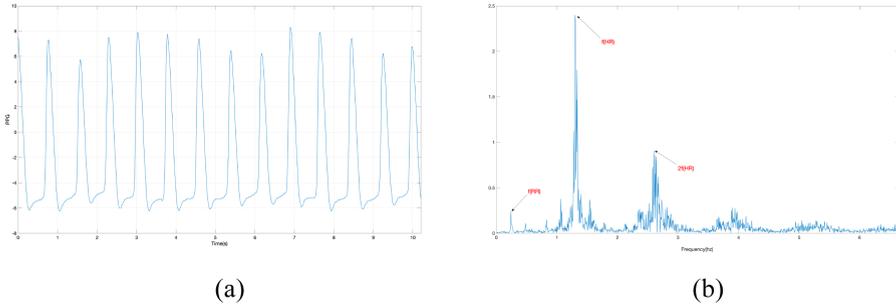


Fig. 8. PPG signal (a) and frequency spectrum (b)

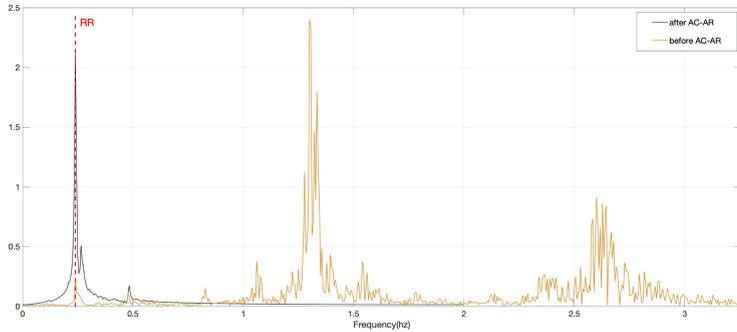


Fig. 9. The frequency spectrum of 180 s respiratory signal before AC-AR (yellow) and after AC-AR (black). The red dotted line is the respiratory rate. (Color figure online)

After signal preprocessing, we get the respiratory modulation signal from the PPG signal. Figure 10 shows the comparison between the reference respiratory signal and the respiratory modulation signal extracted using the peak detection algorithm proposed in this paper. The extracted respiratory signal is basically similar to the reference respiratory signal, which is of great help to the subsequent analysis. Following the signal processing method previously mentioned, respiratory rates are acquired through the AC-AR algorithm. By the above formula and reference respiratory rate, we can calculate the MAE and AAEP. From Table 1, we can see that the result is much improved after using SQI. In the CapnoBase database, some signals are chaotic, and the results with large errors will be obtained by using these signals to analyze. Using SQI can avoid these errors, which is beneficial to the reliability of clinical results. Meanwhile, compared with AR model, the AAEP of the AC-AR algorithm decreased by about 1%. From Table 2, the accuracy is improved to some extent after data fusion with ASQ, which proves that the feasibility

of data fusion using this method. Compared with averaging the spectrum directly, using ASQ can dynamically select a better spectrum as the result according to the quality of autocorrelation signal.

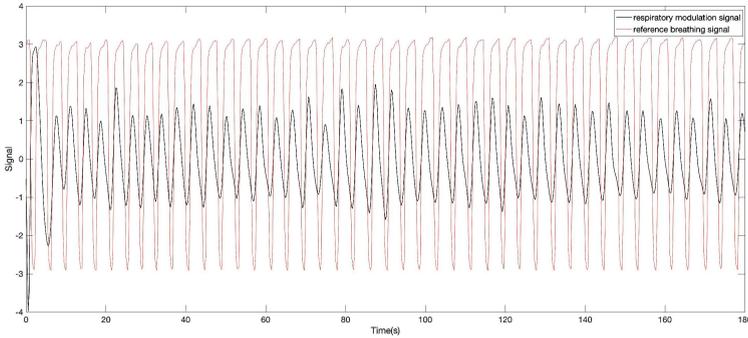


Fig. 10. Comparison of respiratory modulation signals with reference breathing signals. After processing, the respiratory signals can be extracted normally.

Table 1. Comparison of results

	AAEP	
	AC	AC-AR
Before using SQI	8.37%	7.56%
After using SQI	4.27%	3.72%

Table 2. The result of ASQ

	RIIV	RIFV	RIAV	ASQ
AAEP	6.61%	5.61%	8.55%	3.72%

As can be seen from the boxplot (see Fig. 11), the MAEs are quite different when using one of the modulation signals alone. The results of RIIV are obviously better than the other two, indicating that RIIV has the strongest modulation of PPG signal. After using data fusion and SQI, the experimental results are obviously better. Among them, the results of data fusion using ASQ were better than SQI analysis of single respiratory modulated signal, indicating that it is necessary to conduct quality analysis of modulated signal. Moreover, the result of data fusion using autocorrelation signal quality is better than that of spectrum averaging. Without considering the outliers, the error of the AC-AR algorithm is 0.12 ± 0.36 bpm.

To further evaluate the algorithm, a scatter plot is drawn for analysis (see Fig. 12). Best Fit stands for PPG respiratory rate equal to the reference respiratory rate, and the

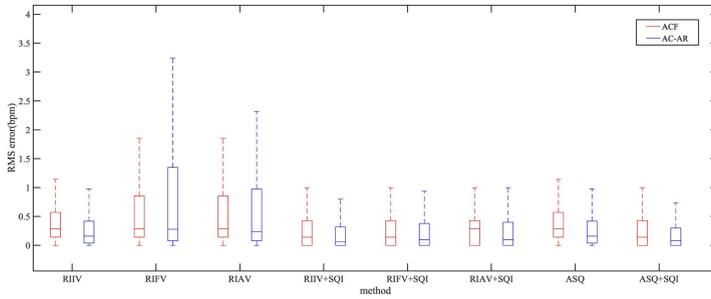


Fig. 11. Results for the CapnoBase benchmark dataset using 60 s windows. The boxplots give the RMS Error for the different RR estimation methods.

closer the vertical distance to the line, the better the result. As can be seen from the figure, results are concentrated near the Best Fit, indicating that the AC-AR algorithm identified and eliminated a majority of high error estimations. This also shows that the algorithm proposed in this paper is feasible and accurate.

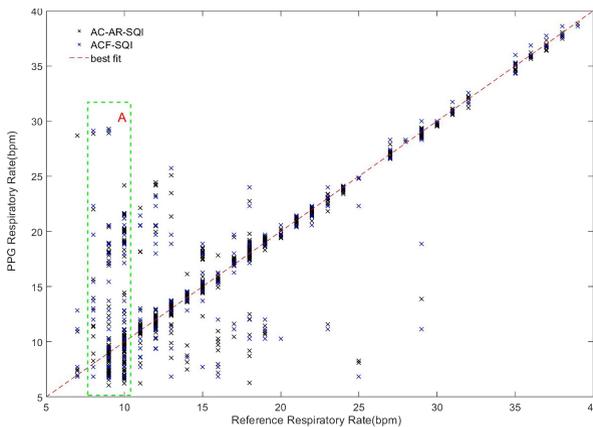


Fig. 12. Scatter plot comparing the reference RR obtained from capnometry with the PPG RR obtained from the ACF and AC-AR algorithm. The AC-AR eliminates the estimations with large error (distance from Best fit). Box A: The signal has been badly distorted.

6 Conclusion

In this paper, we improved the method of obtaining respiratory modulation signals and proposed a new analysis method that can be used in combination with other respiratory frequency analysis methods to improve the accuracy and robustness of respiratory frequency estimation. The autocorrelation method can remove noise in a periodic signal having the above characteristics. Each peak of the autocorrelation sequence represents

a period of strong autocorrelation, so the period of the autocorrelation corresponds to the period of the original signal. Autoregression (AR) model uses the time history of the signal to extract the important information hidden in the signal. Therefore, the autocorrelation signal of the respiratory modulation signal can be used as the input signal of AR spectrum analysis to estimate the RR. Several experiments have been performed on different datasets with different methods. The experimental results show that the average absolute error percentage (AAEP) is less than 3.72%. It is proved that the method of autocorrelation combined with autoregressive model used to extract respiratory rate from PPG is feasible and reliable. Finally, it can be seen from the boxplot that the result of each algorithm has a lot of outliers. The problem is that when the waveform of the PPG signal is relatively chaotic, the respiratory modulation signal extracted from it is not reliable. If the baseline drift of the PPG signal is severe, there will be errors in the peak detection results, which will cause some peaks to be missed. To solve this problem, our next goal is to better remove motion noise so that the respiratory rate can be extracted from people in motion.

References

1. Van Leuvan, C.H., Mitchell, I.: Missed opportunities? An observational study of vital sign measurements. *Crit. Care Resuscitation* **10**(2), 111–115 (2008)
2. Hooker, E.A., et al.: Respiratory rates in emergency department patients. *J. Emerg. Med.* **7**(2), 129–132 (1989)
3. Birrenkott, D.A., Pimentel, M.A., Watkinson, P.J., Clifton, D.A.: A robust fusion model for estimating respiratory rate from photoplethysmography and electrocardiography. *IEEE Trans. Biomed. Eng.* **65**(9), 2033–2041 (2018)
4. Walsh, J.A., et al.: Novel wireless devices for cardiac monitoring. *Circulation* **130**(7), 573–581 (2014)
5. Hubner, P., et al.: Surveillance of patients in the waiting area of the department of emergency medicine. *Medicine* **94**(51), Article no. e2322 (2015)
6. Karlen, W., Raman, S., Ansermino, J.M., Dumont, G.A.: Multiparameter respiratory rate estimation from the photoplethysmogram. *IEEE Trans. Biomed. Eng.* **60**(7), 1946–1953 (2013)
7. Charlton, P.H., Birrenkott, D.A., Bonnici, T., Pimentel, M.A.F.: Breathing rate estimation from the electrocardiogram and photoplethysmogram: a review. *IEEE Rev. Biomed. Eng.* **11**, 2–20 (2018)
8. Birrenkott, D.: Respiratory quality indices for ECG- and PPG-derived respiratory data. D.Phil transfer of status report, Department of Engineering and Science, University of Oxford, Oxford, U.K (2015)
9. Pimentel, M.A.F., et al.: Toward a robust estimation of respiratory rate from pulse oximeters. *IEEE Trans. Biomed. Eng.* **64**(8), 1914–1923 (2017)
10. Meredith, D., et al.: Photoplethysmographic derivation of respiratory rate: A review of relevant respiratory and circulatory physiology. *J. Med. Eng. Technol.* **36**(1), 60–66 (2012)
11. Lázaro, J., Gil, E., Bailón, R., Mincholé, A., Laguna, P.: Deriving respiration from photoplethysmographic pulse width. *Med. Biol. Eng. Comput.* **51**, 233–242 (2013)
12. Nilsson, L., et al.: Monitoring of respiratory rate in postoperative care using a new photoplethysmographic technique. *J. Clin. Monit. Comput.* **16**(4), 309–315 (2000)
13. Leonard, P., et al.: An algorithm for the detection of individual breaths from the pulse oximeter waveform. *J. Clin. Monit. Comput.* **18**(5/6), 309–312 (2004)

14. Pimentel, M.A.F., Santos, M.D., Springer, D.B., Clifford, G.D.: Heart beat detection in multimodal physiological data using a hidden semi-Markov model and signal quality indices. *Physiol. Meas.* **36**, 1717–1727 (2015)
15. Takalo, R., Hytti, H., Ihalainen, H.: Tutorial on univariate autoregressive spectral analysis. *J. Clin. Monit. Comput.* **19**, 401–410 (2005)
16. Kim, B.S., Yoo, S.K.: Motion artifact reduction in photoplethysmography using independent component analysis. *IEEE Trans. Biomed. Eng.* **53**(3), 566–568 (2006)
17. Johansson, A.: Neural network for photoplethysmographic respiratory rate monitoring. *Med. Biol. Eng. Comput.* **41**, 242–248 (2003)
18. Khreis, S., Ge, D., Rahman, H. A., Carrault, G.: Breathing rate estimation using kalman smoother with electrocardiogram and photoplethysmogram. *IEEE Trans. Biomed. Eng.* **67**, 893–904 (2019)
19. Nguyen, N.T.P., Lyu, P., Lin, M.H., Chang, C., Chang, S.: A short-time autocorrelation method for noncontact detection of heart rate variability using CW doppler radar. In: 2019 IEEE MTT-S International Microwave Biomedical Conference (IMBioC), China, pp. 1–4 (2019)
20. Orphanidou, C., Bonnici, T., Charlton, P., Clifton, D., Vallance, D., Tarassenko, L.: Signal-quality indices for the electrocardiogram and photoplethysmogram: derivation and applications to wireless monitoring. *IEEE J. Biomed. Health Inf.* **19**(3), 832–838 (2015)
21. Karlen, W., Ansermino, J.M., Dumont, G.: Adaptive pulse segmentation and artifact detection in photoplethysmography for mobile applications. In: 34th Annual International Conference of the IEEE EMBS (2012)



An Ultra-low-Power Integrated Heartbeat Detector for Wearable Sensors

Antoine Gautier¹, Marine Dael¹, Robin Benarrouch^{2,3,4}, Benoit Larras² ,
and Antoine Frappé² 

¹ Yncréa Hauts-de-France, Lille, France

² Univ. Lille, CNRS, Centrale Lille, Univ. Polytechnique Hauts-de-France, Yncréa Hauts-de-France, UMR 8520 - IEMN, 59000 Lille, France
benoit.larras@yncrea.fr

³ STMicroelectronics, Technology and Design Platforms, 38920 Crolles, France

⁴ University of California Berkeley, Berkeley, CA 94704, USA

Abstract. To optimize energy consumption in wearable sensor networks, an efficient scheme is to set the sensors in sleep mode and wake them up to engage communication. However, synchronicity between the sensors needs to be assured by always-on local oscillators. This work proposes a different topology that takes advantage of the heart beat to wake-up wearable sensors. The electrocardiogram (ECG) is detected by two probes and then converted into a pulse signal. Using 28-nm FD-SOI CMOS technology, this solution is implemented on a circuit consuming 19 nW at a 900 mV supply voltage, hence suitable for long term and wearable applications.

Keywords: Heartbeat detection · Integrated circuit design · 2-electrode sensing · FD-SOI technology

1 Introduction

In the context of wearable systems, every integrated sensor needs a long battery life in order to avoid frequent replacements. Because communication is the most energy-hungry part of each sensor, data exchanges between the sensors (or nodes) and aggregators have to be reduced to a minimum, while maintaining the synchronization between all the nodes.

A conventional solution to address this problem is putting the nodes to sleep most of the time, and regularly wake up the nodes at defined time steps. During the wake-up phase, the aggregator can establish specific communication with certain nodes, which remain active, whereas the others return in sleep mode. While reducing the average power consumption of the nodes, it requires an always-on local oscillator in each node. Moreover, all the oscillators have to be in sync so that the communication between the aggregator and the nodes can be established.

On the human body, the heartbeats can be interpreted as a clock signal with a low duty cycle, and act as a synchronized wake-up signal [1,9]. An example of a heart-rate paced communication is depicted in Fig. 1. A typical application for this slow-paced scheme is posture or pose recognition [11,12]. Once a heartbeat occurs, all the nodes and the aggregator are activated. The aggregator sets up a communication with some nodes, *e.g.* Node1, and the other nodes go back in sleep mode. Compared to the previous scheme, the local oscillators are replaced by independent heartbeat detectors in each node. This method has the advantage to ensure synchronicity for the wake-up phase between all the nodes distributed on the entire body.

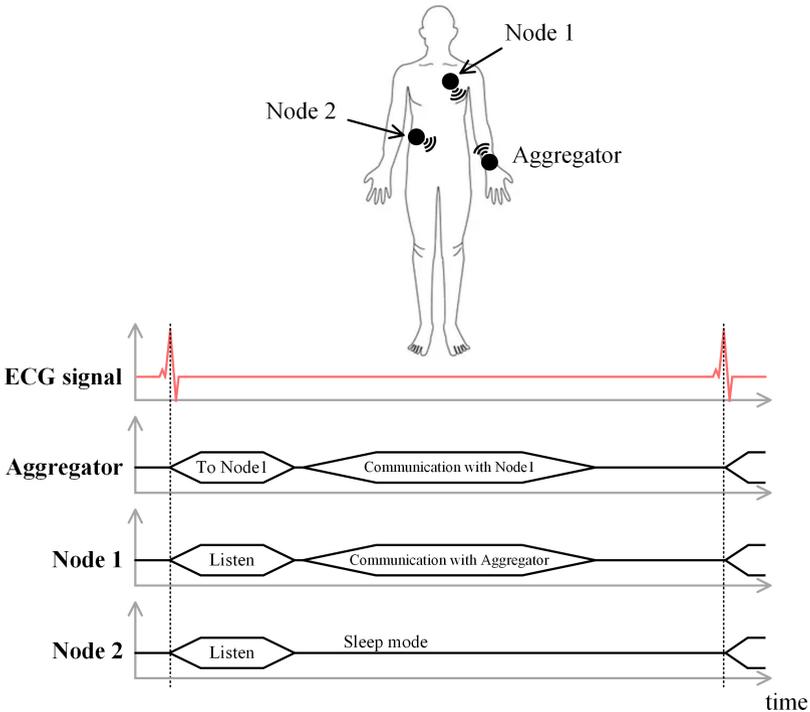


Fig. 1. Example of heart-rate scheduled communication between wearable sensors.

For an always-on system, using Components Off-The-Shelf (COTS) is not an option, as the power of such a system is in the range of tens of milliwatts, like in [14]. In the state-of-the-art, integrated heartbeat detection is done by digitizing the ECG signal with an Analog-to-Digital Converter (ADC), and then processing the digital data with a Digital Signal Processor (DSP) [4,10]. However, this scheme is truly efficient when the objective is to reconstruct the ECG signal with high fidelity [5,7,15,17]. In the proposed context, only the fact that a heartbeat occurred is important, not the waveform in itself. Therefore, an

energy-hungry structure ADC + DSP is not relevant to be embedded on each node. Moreover, in all the systems, a reference electrode located far from the sensing electrodes is needed to get rid of parasitics, 50 Hz interference. A 2-electrode heartbeat detector circuit focusing on timing and not on signal reconstruction has been integrated in [8], consuming 58 nW. An external microcontroller (not taken into account in the energy consumption figure) is nonetheless still needed to adjust the comparator threshold. In [3], Bose *et al.* present a 2-electrode heartbeat detection system with self-adaptation of the comparator threshold consuming 504 nW in the analog front-end. However, in these 2 reference works, the comparator threshold is adapted compared to an absolute reference, and needs to be changed dynamically depending on the baseline drift of the ECG signal. The calibration circuit is therefore active all the time and increases the energy consumption of the whole system. Moreover, the last work relies on the fact that the 2 electrodes are far from each other (tens of centimeters) to sense a significant difference between the ECG signals, which limits the embedded character and comfort of the system.

This paper proposes a standalone heartbeat detector circuit functioning without an external microcontroller, making use of the 28-nm FD-SOI CMOS process for self-calibration through body biasing. The detection threshold is set using the common-mode of the ECG signal as a reference. Therefore, once calibrated, the decision threshold does not need to change. The 28-nm FD-SOI CMOS technology process also offers efficient co-integration with advanced-pitch digital circuits for embedded processing using the heart-rate as a clock signal. The proposed circuit was designed to function with 2 cm-spaced electrode signals lower than 1 mV, without a need for an additional reference electrode on the body, and thus completely integrable on a wearable device. The proposed circuit has been simulated with a 900 mV supply V_{DD} and consumes 19 nW.

This paper is organized into the following sections. Section 2 provides an overview of the system, Sect. 3 explains the circuit blocks in detail, Sect. 4 presents the results from both system-level and transistor-level simulations and Sect. 5 concludes the paper.

2 System Overview

The proposed structure is based on the differential ECG heart rate measured between 2 input electrodes, depicted in Fig. 2. The objective is to detect the high amplitude of the R-wave in the QRS complex for each heartbeat. The 2 electrodes are capacitively coupled to the inputs of the differential amplifiers for DC offset suppression. The input biasing voltage is set to $V_{DD}/2$ through large resistors in the T Ω range allowing input impedances orders of magnitude larger than the electrodes impedances $1/G_1$ and $1/G_2$. These resistors are implemented using transistor-based pseudo-resistors, yielding a silicon occupation of 10 μm^2 , and thus compatible with on-chip integration. The resulting high-pass filters are called **HPF1** and **HPF2**.

Since the electrodes are very close from each other, they sense the same ECG signal with very few differences. The closer the 2 ECG signals are, the

more the difference needs to be amplified to be sensed. However, the energy consumption of the analog front-end increases with the amplifier gain. To amplify the difference between the ECG signals without increasing the amplifier gain, the values of the capacitances in the filters **HPF1** and **HPF2** are voluntarily mismatched so that the 2 cut-off frequencies are not matched anymore, and as a result the 2 filtered ECG signals are different. The filters original cut-off frequency is set 15 Hz to attenuate baseline drift and motion artifact. The influence of the voluntary frequency mismatch is studied further in Sect. 4.1.

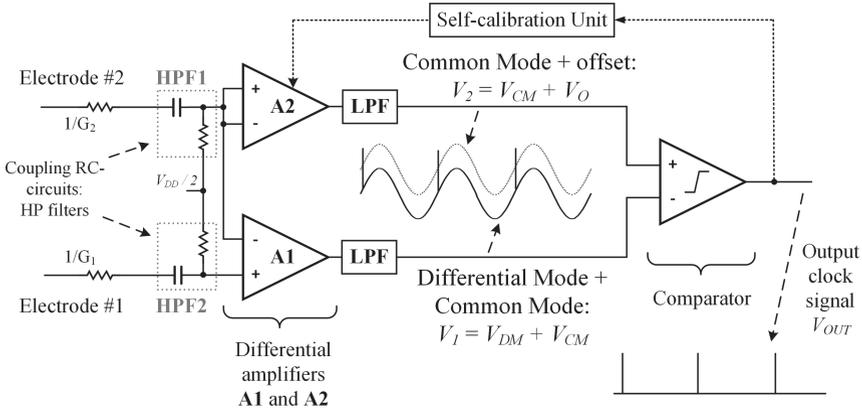


Fig. 2. Block diagram of the proposed circuit.

The differential ECG signal is first amplified by amplifier **A1** with a differential gain G_D . Amplifier **A1** is used to amplify both the differential signal of the probes corresponding to the heart rate (typically in the range of 100 μV to 1 mV from [16]), and also the common mode of the input signals with a lower gain G_{CM} . It also includes a Common-Mode Feedback (CMFB) loop for internal stability. Amplifier **A2** is designed the same way as **A1**, especially with a similar G_{CM} , and permits amplifying the common mode reference only, by putting the same signal at both inputs of the amplifier. The output of **A1**, *i.e.* V_1 , is thus composed of the amplified differential signal V_{DM} , and the amplified common mode V_{CM} . The output of **A2**, *i.e.* V_2 , is composed of the amplified common mode V_{CM} and an intentional offset V_O of at least 10 mV. The amplifiers have to be designed so that noise is negligible compared to V_O , thus making the circuit noise tolerant. Then, by generating similar common-mode components in V_1 and V_2 , they can be neglected if only the difference of V_1 and V_2 is taken into account. Besides, the proposed structure avoids using filters with different bandwidths by extracting separately differential and common modes, hence avoiding large capacitors. This is a significant advantage for scalability and integration purposes.

Identical low-pass filters (**LPF** in Fig. 2) are inserted after each amplifier to reduce 50 Hz residual frequency component, due to the lack of the reference elec-

trode. A threshold comparator is then used to compare V_1 and V_2 and generate the desired clock signal V_{OUT} . V_1 and V_2 having the same common mode, the differential signal V_{DM} , representing the presence of a heartbeat, is eventually compared to the offset V_O . V_O has to be set so that it is lower to the amplified R-wave:

$$|V_O| \leq |V_{DM}|. \quad (1)$$

When a heartbeat occurs, V_{DM} crosses V_O , the threshold comparator toggles and generates the desired pulse.

Finally, a self-calibration unit uses the system output signal V_{OUT} to adjust the offset V_O generated by amplifier **A2**. The obtained V_O has to respect the condition expressed in (1), so that the low-duty-cycled targeted clock signal is output at V_{OUT} . The calibration process is done once, off-line, at device start-up, since there is no need to change the threshold value while the device stays in the same measurement conditions.

The main advantage of the proposed structure is that the reference signal of the comparator is based on the common-mode voltage output of the amplifiers, so that only the differential voltage component between the 2 electrodes can be considered. As a result, in terms of amplifier design, the Common-Mode Rejection Ratio (CMRR) requirement is lowered compared to a single-amplifier structure where the same design is used for both differential signal amplification and common-mode attenuation.

However, since the proposed scheme relies on an identical common-mode gain in 2 distinct amplifiers, the local mismatch between similar transistors in **A1** and **A2** is to be addressed. For that purpose, it is mitigated at the layout level by placing matched transistors next to one another using interdigitated gates for connection. The self-configuration unit feedback loop ensures a correct behavior of the system despite the remaining transistor mismatch.

3 Circuit Design

3.1 Amplifier A1

The structure of the amplifiers used in this approach is a conventional 3-stage amplifier structure. It is composed of a differential pair and a gain stage. The electrical schematic of amplifier **A1** is shown in Fig. 3. For the differential pair, composed of M_{11} and M_{12} , the inputs are connected to PMOS transistors instead of the NMOS transistors to mitigate the $1/f$ noise. The gain stage, composed of 2 transistors in the conventional amplifier structure (M_{15} and M_{16} in Fig. 3), is doubled with the purpose of increasing the gain (addition of M_{17} and M_{18}). The body voltage of M_{17} is fixed and set to $V_{DD}/2$. For this application, there is no need for a high CMRR because the system does not need to be linear and detect all the components of the heartbeats. Thus, the design effort can be set on lowering the current consumption while still acquiring the R-wave with high precision. A CMFB circuit is added to the structure for stability and ensures operation at a fixed common-mode voltage for the first 2 stages. It presents the particularity to directly drive the body voltage of transistors M_{13} and M_{14} .

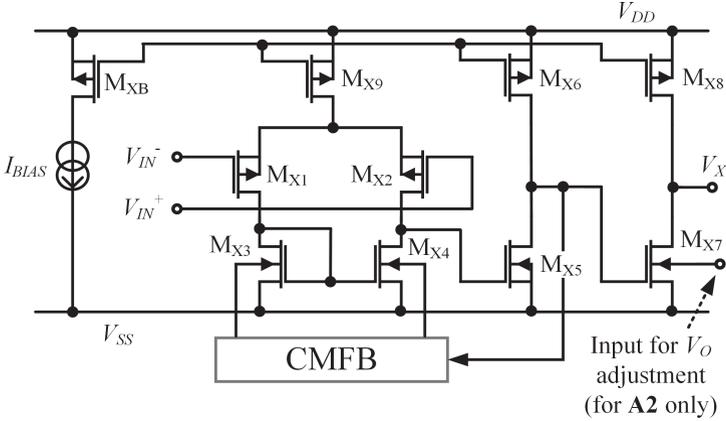


Fig. 3. Electrical schematic of the designed amplifiers **A1** and **A2**, including each a Common-Mode Feedback (CMFB) loop. The naming convention is $X = 1$ for **A1**, $X = 2$ for **A2**. In **A2**, the bulk of M_{27} is connected to the self-calibration unit for V_O adjustment.

3.2 Amplifier A2

Amplifier **A2** is designed using the same structure as in amplifier **A1**, shown in Fig. 3. Both CMFB circuits in **A1** and **A2** share the same voltage reference. However, in the last stage, the transistor M_{27} is modified so that the bulk voltage can be modified by the self-calibration unit, Fig. 3. For simplicity reasons, the possible range of M_{27} bulk voltage is from 0 V to V_{DD} . Increasing or decreasing this voltage will decrease or increase the output offset, respectively. This feature is therefore used to generate the additional offset V_O at the output of the amplifier.

3.3 Threshold Comparator

The designed threshold comparator is a conventional operational amplifier used as a comparator, as shown in Fig. 4. The differential gain of the amplifier is maximized so that the slightest difference between V_{IN}^+ and V_{IN}^- saturates the output V_{COMP} between V_{SS} and V_{DD} . Besides, the transistors' dimensions are designed to optimize the slew-rate, and thus the response rapidity of the comparator.

3.4 Self-calibration Unit

The self-calibration unit acts as the feedback loop for adjusting the offset voltage of **A2**, Fig. 5. It takes V_{OUT} as an input and charges the capacitor C_C with a constant current I_{CHARGE} , depending on the duty cycle of V_{OUT} . To calibrate the system, the resulting voltage V_C must be between 2 reference voltages V_{refL} and V_{refH} , defined externally depending on the user. If V_C is over V_{refH} , it means

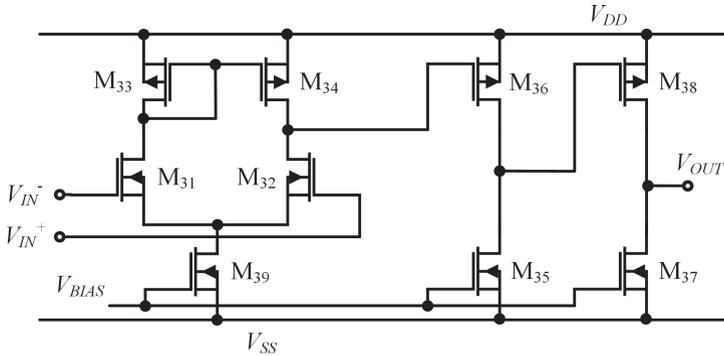


Fig. 4. Electrical schematic of the designed threshold comparator.

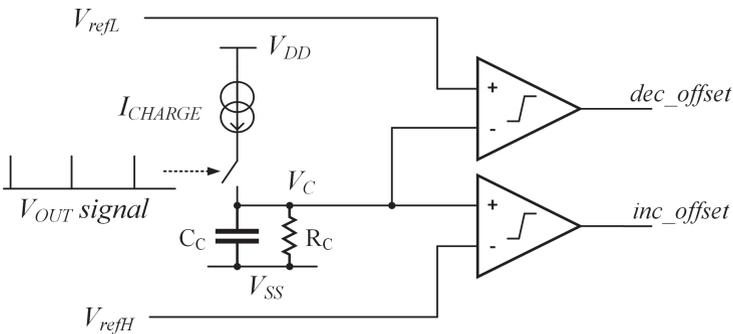


Fig. 5. Block diagram of the proposed scheme for the self-calibration unit.

that the offset component V_O is too low, and the signal inc_offset is set. If V_C is below V_{refL} , the signal dec_offset is set instead. The digital signals inc_offset and dec_offset are then used to command a charge pump circuit allowing to increase and decrease the body biasing of M_{27} , respectively. In addition, a resistor R_C is used to discharge the capacitor C_C . The passive components R_C and C_C can be implemented off-chip so that the resulting time constant can be adapted externally to each person’s heartbeat. Once the value of V_O is calibrated, it does not need to change while the system is on-line, therefore the self-calibration unit can be put in sleep mode.

4 Results

4.1 System-Level Model

The proposed structure has been simulated using *Matlab*[®], and the intermediate signals are shown in Fig. 6. The ECG signal is generated by the *ECGSYN* application from Physionet [6, 13], 50 Hz coupling due to the absence of a reference electrode, Fig. 6-(a). Its amplitude and shape are similar to that of a

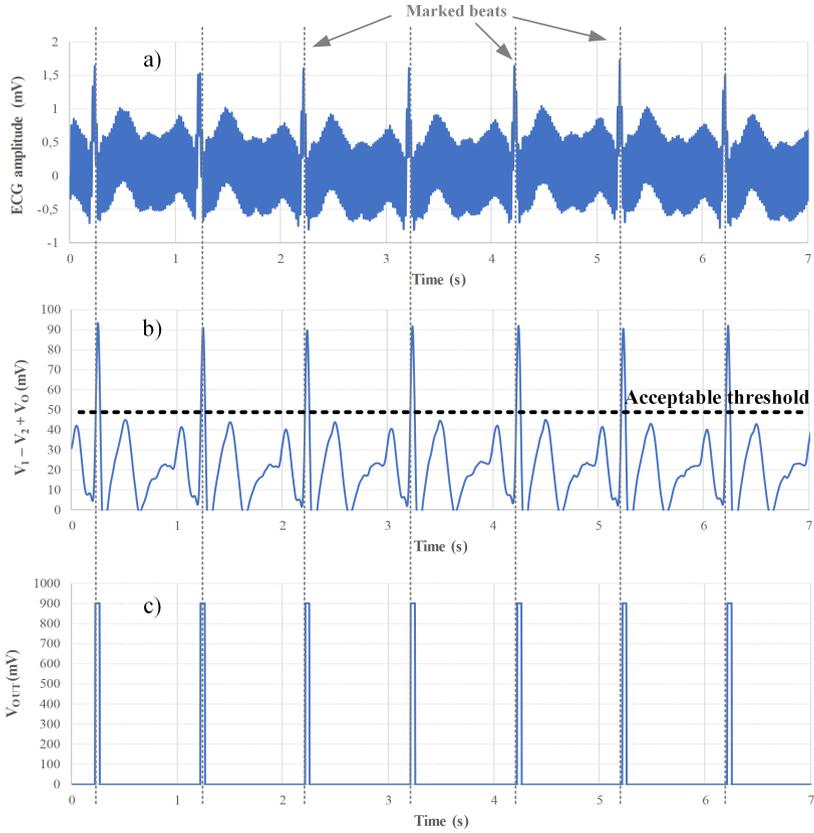


Fig. 6. Transient response of different voltages in the simulated system using *Matlab*[®]. a) Input signal at electrode #1, 50Hz coupling. b) Difference between the signals V_1 and V_2 at the output of the amplifiers **A1** and **A2**, without taking the threshold V_O into account. An acceptable value of V_O is indicated on the graph. c) Output signal V_{OUT} .

measured ECG signal [13], sensed by both electrodes as they are next to each other. A random electrode conductivity is modeled, depending on the electrode material itself and its placement on the skin. The highpass filters **HPF1** and **HPF2** are designed to have a cut-off frequency 15 Hz, allowing the attenuation of motion artifacts and baseline wandering caused by the respiration [2]. The model includes a random variation of the passive elements values, extracted from transistor-level Monte-Carlo simulations. Moreover, a voluntary mismatch between the highpass filters capacitances is added to enhance the differences between the 2 input signals. Given simulation results, a 40% capacitance mismatch can produce a detectable differential voltage corresponding to the R-wave at the output of **A1**. However, a capacitance mismatch superior to 150% bring the filter cut-off frequency closer 50 Hz, and produces a differential voltage

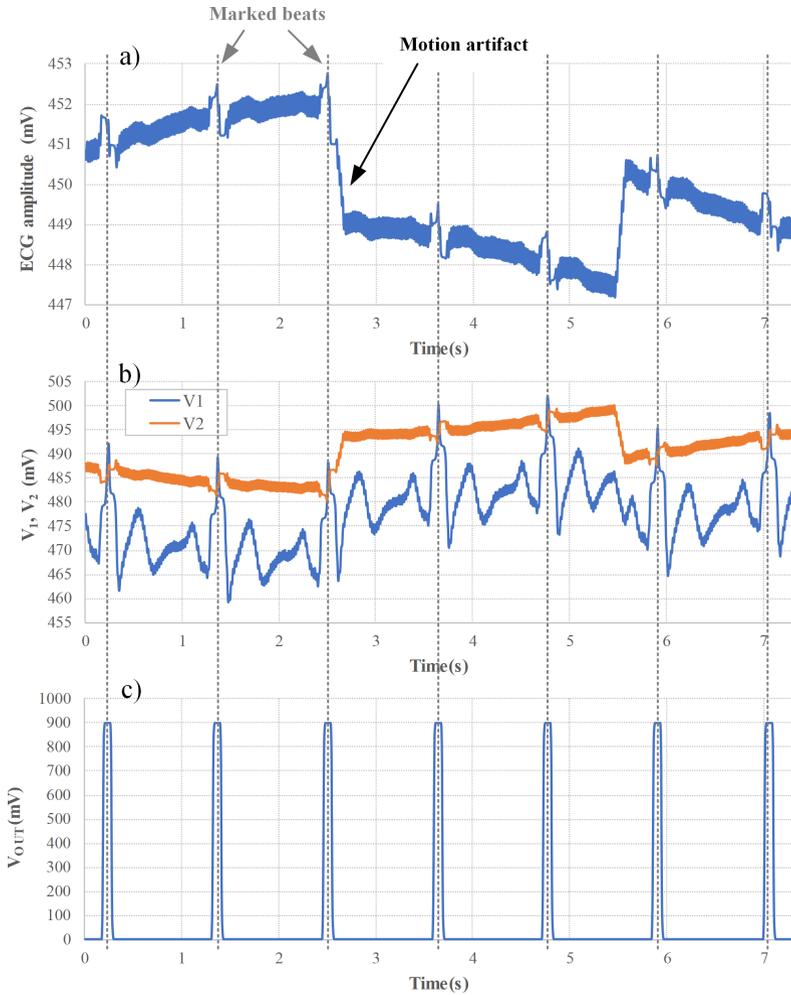


Fig. 7. Transient response of different voltages in the system simulated using *Spectre*[®]. a) Input signal at electrode #1, 50 Hz coupling, baseline drift and motion artifact. b) Intermediate signals at the output of the amplifiers **A1** and **A2**. c) Output signal V_{OUT} .

including also amplified environmental artifacts and not only the R-wave. The recovery of the heartbeat is not possible anymore. A capacitance mismatch of 100% is thus set for the behavioral simulations.

Amplifiers **A1** and **A2** are then modeled, with a differential gain G_D and a common-mode gain G_{CM} . The CMRR value of the amplifiers is derived from those 2 gain values. For a defined value of the CMRR, the model generates 100 ECG signals for statistical purposes, and tests if an acceptable value of the threshold V_O exists after the processing of each signal, as in Fig. 6-(b). If this value exists, a test is

considered a success and the signal V_{OUT} is output, Fig. 6-(c). Otherwise, the test is considered a failure. From the simulation, a minimum CMRR of 41 dB is needed to ensure a 95% success rate. This requirement is thus used to refine the amplifier design described in Sect. 3.1.

4.2 Transistor-Level Simulations

The circuit is designed using STMicroelectronics 28-nm FD-SOI CMOS. This technology process allows a fine tuning of the transistors back biasing, needed for previously described scheme. Besides, the 28-nm FD-SOI CMOS process allows efficient co-integration with advanced-pitch digital circuits for embedded processing using the heart-rate as a clock signal.

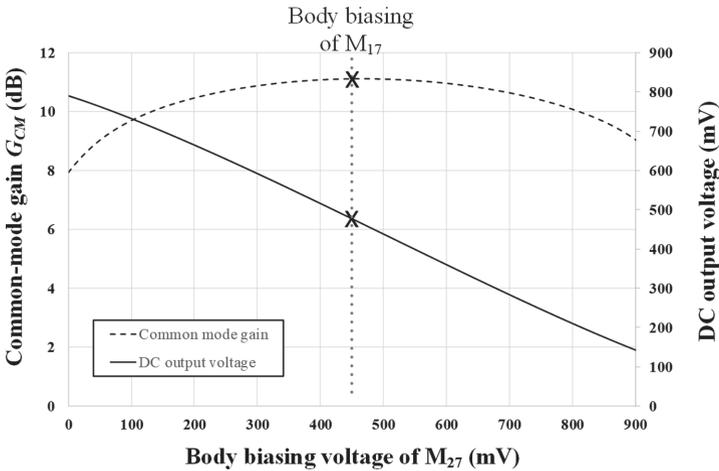


Fig. 8. DC response of the DC output voltage of **A2** depending on the body biasing voltage of M_{27} .

The circuit behavior has been simulated using *Spectre*[®]. The voltage supply V_{DD} is set to 900 mV. The current bias of the amplifiers I_{BIAS} is set to 2 nA, while the bias of the comparator is set to 1 nA. For transistor mismatch mitigation, the transistor lengths have been set to 3 times the minimal allowed length. The ECG signal is also generated by the *ECGSYN* application, and 50 Hz coupling, baseline drift and a motion artifact, using frequency components and amplitudes described in [2]. The obtained input wave is displayed in Fig. 7-(a). The outputs of both amplifiers **A1** and **A2**, V_1 and V_2 , respectively, are shown in Fig. 7-(b). While the amplified common mode is still present in V_1 and V_2 , the high differential gain G_D allows to extract the expected R-wave in V_1 . As in Fig. 7-(b), the common-mode voltage reference allows reacting to the environmental perturbations. In the self-calibration unit, V_{refL} and V_{refH} are set so that the offset voltage V_O of V_2 is 16 mV. The output voltage of the heartbeat

detector V_C is shown in Fig. 7-(c). Considering the offset calibration, Fig. 8 shows that the DC output voltage of **A2** can be tuned in that range with a variation of the body biasing voltage of M_{27} that does not modify the common-mode gain G_{CM} of **A2**, thus enabling self-calibration without risking a behavioral change in the amplifiers.

The amplifiers consume 7.5 nW each, and the comparator consumes only 4.2 nW at 60 bpm. The system power consumption is 19 nW without taking into account the automatic offset calibration unit, since it is in sleep mode while on-line.

Since they are the most power-hungry part of the system, the characteristics of the amplifiers are compared with state-of-the-art ECG amplifiers, Table 1. In this work, even though the CMRR is far below that of conventional instrumentation amplifiers used for ECG detection [3, 4, 8, 17], it is still sufficient for the application since signal reconstruction is not needed. Moreover, since the CMRR constraint is lowered, an ultra-low-power design can be envisioned for the amplifiers, yielding to a total power consumption reduction of more than 50% compared to an amplifier used for a similar application [8]. This system is thus suitable for near-sensor integration.

Table 1. Characteristics of state-of-the-art ECG amplifiers

	[4]	[17]	[8]	[3]	This work
Application	ECG recording	ECG recording	Single heartbeat detection	Single heartbeat detection	Single heartbeat detection
Technology process	65-nm bulk	0.35- μ m bulk	0.18- μ m bulk	0.18- μ m bulk	28-nm FD-SOI
CMRR	>80 dB	>65 dB	68 dB	>50 dB	48.9 dB
Amplifier power consumption	64 nW	320 nW	50.4 nW	504 nW	7.5 nW (x2)
Self-adaptative detection threshold ?	No	No	No	Yes (92 nW)	Yes (4.2 nW on-line)

5 Conclusion

This work presents a low power circuit for wearable systems detecting the ECG signal and converting the heart rate into a digital signal. This solution permits waking up the sensors from time to time (e.g. once or twice per second), and consumes only 19 nW at a 900 mV supply voltage. The applications for this integrated circuit are in medical monitoring and more generally in high autonomy

wearable systems. Compared to state-of-the-art on-chip solutions, the structure of this circuit is simpler and self-adjustable, since there is no need to reproduce the exact shape of the signal. Thereby, a high CMRR is not required unlike in conventional instrumentation amplifiers, and the global consumption is lowered.

Acknowledgment. This work has been initiated in the frame of the Human Body Intranet collaborative project between STMicroelectronics, BWRC at UC Berkeley and IEMN. The authors wish to acknowledge Andreia Cathelin, Jan Rabaey and Andreas Kaiser for their support. This work was supported in part by the European CHIST-ERA JEDAI program under number ANR-19-CHR3-0005-01.

References

1. Benarrouch, R., et al.: Heartbeat-based synchronization scheme for the human intranet: modeling and analysis (2020). <http://arxiv.org/abs/2005.05915>
2. Berwal, D., Vandana, C., Dewan, S., Jiji, C., Baghini, M.: Motion artifact removal in ambulatory ECG signal for heart rate variability analysis. *IEEE Sens. J.* **19**(24), 12432–12442 (2019). <https://doi.org/10.1109/JSEN.2019.2939391>
3. Bose, S., Shen, B., Johnston, M.L.: A 20 μ W heartbeat detection system-on-chip powered by human body heat for self-sustaining wearable healthcare. In: 2020 IEEE International Solid-State Circuits Conference - (ISSCC), pp. 408–410 (2020). <https://doi.org/10.1109/ISSCC19947.2020.9063071>
4. Chen, Y., et al.: An injectable 64 nW ECG mixed-signal SoC in 65 nm for arrhythmia monitoring. *IEEE J. Solid State Circuits* **50**(1), 375–390 (2015). <https://doi.org/10.1109/JSSC.2014.2364036>
5. Dong, Y., et al.: A 1.8 μ W 32 nV/ $\sqrt{\text{Hz}}$ current-reuse capacitively-coupled instrumentation amplifier for EEG detection. In: 2017 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1–4, May 2017. <https://doi.org/10.1109/ISCAS.2017.8050494>
6. Goldberger, A., et al.: PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* **23**(101), e215–e220 (2003). <https://doi.org/10.1161/01.CIR.101.23.e215>
7. Ha, H., et al.: A bio-impedance readout IC with digital-assisted baseline cancellation for two-electrode measurement. *IEEE J. Solid State Circuits* **54**(11), 2969–2979 (2019). <https://doi.org/10.1109/JSSC.2019.2939077>
8. He, D.D., Sodini, C.G.: A 58 nW ECG ASIC with motion-tolerant heartbeat timing extraction for wearable cardiovascular monitoring. *IEEE Trans. Biomed. Circuits Syst.* **9**(3), 370–376 (2015). <https://doi.org/10.1109/TBCAS.2014.2346761>
9. Li, H., Tan, J.: Heartbeat-driven medium-access control for body sensor networks. *IEEE Trans. Inf. Technol. Biomed.* **14**(1), 44–51 (2010). <https://doi.org/10.1109/TITB.2009.2028136>
10. Izumi, S., et al.: Normally off ECG SoC with non-volatile MCU and noise tolerant heartbeat detector. *IEEE Trans. Biomed. Circuits Syst.* **9**(5), 641–651 (2015). <https://doi.org/10.1109/TBCAS.2015.2452906>
11. Larras, B., Frappé, A.: On the distribution of clique-based neural networks for edge AI. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **1** (2020). <https://doi.org/10.1109/JETCAS.2020.3023481>
12. Laurijssen, D., Saeys, W., Truijten, S., Daems, W., Steckel, J.: Synchronous wireless body sensor network enabling human body pose estimation. *IEEE Access* **7**, 49341–49351 (2019). <https://doi.org/10.1109/ACCESS.2019.2910636>

13. McSharry, P., Clifford, A., Tarassenko, L., Smith, L.: A dynamical model for generating synthetic electrocardiogram signals. *IEEE Trans. Biomed. Eng.* **3**(50), 289–294 (2003). <https://doi.org/10.1109/TBME.2003.808805>
14. Xiang, J., Dong, Y., Xue, X., Xiong, H.: Electronics of a wearable ECG with level crossing sampling and human body communication. *IEEE Trans. Biomed. Circuits Syst.* **13**(1), 68–79 (2019). <https://doi.org/10.1109/TBCAS.2018.2879818>
15. Yan, L., et al.: A 680nA ECG acquisition IC for leadless pacemaker applications. *IEEE Trans. Biomed. Circuits Syst.* **8**(6), 779–786 (2014). <https://doi.org/10.1109/TBCAS.2014.2377073>
16. Yoo, J., Yan, L., Lee, S., Kim, Y., Yoo, H.: A 5.2 mW self-configured wearable body sensor network controller and a 12 μ W wirelessly powered sensor for a continuous health monitoring system. *IEEE J. Solid State Circuits* **45**(1), 178–188 (2010). <https://doi.org/10.1109/JSSC.2009.2034440>
17. Zhang, J., Zhang, H., Sun, Q., Zhang, R.: A low-noise, low-power amplifier with current-reused OTA for ECG recordings. *IEEE Trans. Biomed. Circuits Syst.* **12**(3), 700–708 (2018). <https://doi.org/10.1109/TBCAS.2018.2819207>



Anxiety Detection Leveraging Mobile Passive Sensing

Lionel M. Levine¹, Migyeong Gwak¹(✉) , Kimmo Kärkkäinen¹,
Shayan Fazeli¹ , Bita Zadeh², Tara Peris¹, Alexander S. Young¹,
and Majid Sarrafzadeh¹

¹ University of California, Los Angeles, Los Angeles, CA 90095, USA
{lionel,mgwak,kimmo,shayan,majid}@cs.ucla.edu, tperis@mednet.ucla.edu,
ayoung@ucla.edu

² Chapman University, 1 University Drive, Orange, CA 92866, USA
bzadeh@chapman.edu

Abstract. Anxiety disorders are the most common class of psychiatric problems affecting both children and adults. However, tools to effectively monitor and manage anxiety are lacking, and comparatively limited research has been applied to addressing the unique challenges around anxiety. Leveraging passive and unobtrusive data collection from smartphones could be a viable alternative to classical methods, allowing for real-time mental health surveillance and disease management. This paper presents eWellness, an experimental mobile application designed to track a full-suite of sensor and user-log data off an individual's device in a continuous and passive manner. We report on an initial pilot study tracking ten people over the course of a month that showed a nearly 76% success rate at predicting daily anxiety and depression levels based solely on the passively monitored features.

Keywords: Mobile application · Anxiety · Remote mental health monitoring · Passive sensing · Machine learning

1 Background and Introduction

Within the spectrum of mental health disorders, Anxiety disorders are the most common class of psychiatric problems affecting both children and adults [7, 9, 17], with up to one in three people in the US meeting full diagnostic criteria by early adulthood [13, 25]. This manifests in the form of roughly 7 to 9% of the population in the US suffering from a specific phobia, 7% from social anxiety disorder, and 2 to 3% each from panic disorder, agoraphobia, generalized anxiety disorder, and separation anxiety disorder [4]. Individuals with anxiety disorders contend with substantial distress and impairment. They are at heightened risk for a host of negative long-term outcomes including depression, substance abuse, educational underachievement, and poor physical health [5, 19, 27].

The optimal method for the prevention or care of mental illness is early identification, diagnosis, and proactive treatment [26]. Time-sensitive intervention is

therefore crucial for preventing conditions from becoming chronic and debilitating. However, traditional methods of psychiatric assessment, including clinical interviews and self-reports, are limited in their ability to provide just-in-time interventions as well as early identification. They depend heavily on retrospective summaries collected in clinical settings, conditions that often result in reporting biases, inaccurate recall, or late and ineffectual treatment.

Additionally, anxiety disorders are, for the most part, vastly overlooked and under-treated in the community; only 15–30% of anxious individuals in the community receive treatment of any kind. Recent research has found strikingly high levels of anxiety among college-age youth. Indeed, 58.4% of college-aged youth report feeling “overwhelmed by anxiety” [3]. Several other recent studies document the high proportion of college students meeting full diagnostic criteria for an anxiety disorder [8]. At the same time, young adults are particularly overlooked within the health care system, with rates of screening, identification, and referral falling below those of either children or adults [27]. Given this landscape, there remains a pressing need for tools that improve early identification of anxiety symptoms, provide users with the platforms to monitor their activities, raise awareness of factors impacting on their wellbeing, and provide a mechanism for intervention should an anxiety episode escalate.

The growing ubiquity of consumer devices, among them smartphones, smartwatches, and in-home sensors, all equipped with an array of sensors and user-logs, have resulted in an unprecedented opportunity to catalog and quantify the daily aspects of an individual’s life, creating repositories of personalized information [23].

While much has been noted about the insidious aspects of such surveillance capabilities, there is also significant potential for such monitoring, if harnessed and utilized by the individuals themselves, to dramatically improve their healthcare outcomes. Such tools could potentially allow the user to accurately track their behaviors and habits, compare personal activities with population-level baselines, establish outlier behaviors with their peers, and even motivate behavioral change and the promotion of healthy habits.

There is significant potential for such monitoring, if harnessed and utilized by the individuals themselves, to improve their healthcare outcomes dramatically. This potential has long been recognized with physical behavior and physiological health, as both are extensively tracked. In contrast, mental health is largely overlooked.

The notable exception to this trend has been the success in remote stress monitoring that has been achieved with physiological stress monitoring of features like heart-rate variability and Galvanic Skin Response that is accomplished by wearable sensing devices like smartwatches to determine stress level [10]. While such approaches have demonstrated efficacy, they are limited in their potential applicability by requiring the wearing of a physical sensing device, and provide little contextual awareness as to the causes of stress that are encountered. More recent advances have attempted to compensate for the restrictions

in activity detection leveraging novel sensing modalities including wireless signal fluctuations around the body, but such efforts are still in their infancy [24].

Specifically, The capability to track behavioral metrics and associate them to mental health, although intimately linked, has not been definitively established. This owes to the significant difficulty in correlating monitorable behaviors and corresponding mental health. Behavioral patterns both within (e.g., the transition from weekday to weekend) and across individuals (e.g., simple differences in how many men and women carry their phones) are simply too diverse and too subject to confounding factors beyond mental health to allow for easy correlations. Nevertheless, the growing challenges around mental health, necessitate exploring the possibility further.

Recent efforts have explored whether pervasive mental health monitoring could be feasible through a smartphone and the embedded sensors, such as motion sensors, ambient light, microphone, camera, Global Positioning System (GPS), proximity, and touch screen [6, 10, 18, 20]. These efforts have shown the promise of this approach in successfully tying behavioral monitoring to mental health; however, such approaches have not translated into fully mature frameworks, and have focused almost exclusively on depression-related conditions, which while often spoken in conjunction with anxiety, manifest in distinct ways [12].

The advantages of leveraging a smartphone-based platform are that the continuous collection of quantitative data potentially provides a more reliable indicator of an individual's risk at any given time, as well as offering a mechanism for just-in-time intervention should a mental health episode occur [6]. Conversely, smartphone-derived data present several challenges, some of which have already been noted, which can result in limited accuracy owing to differences in behavioral patterns across users, and the indirect manner of detection [12].

We present a system for the remote monitoring of mental health symptoms, their fluctuation, and their attendant disruption to personal functioning, called eWellness. The eWellness framework is designed to capture a broad spectrum of remote monitoring, survey data acquisition, secure data transmission and management, data analytics, and visualization.

The primary component of eWellness is a mobile application that facilitates data collection and transmission harvested from an array of sensors and usage logs from a user's smartphone. The data is collected passively, pre-processed, and transmitted through a secure gateway to the cloud, where it is securely stored, and indexed using a scalable database.

Concurrently the eWellness application includes an active querying component where users can be prompted with Ecological Momentary Assessments (EMA) of their mental health status. This architecture is complemented by a back-end analytic engine, capable of mapping observed metrics and exogenous data sources to a user's mental health state, based on adaptive statistical models, and advanced machine learning algorithms. The system is designed to monitor overall mental health as well as acute crisis events in both a retrospective and predictive capacity.

2 Framework

2.1 Server

Data from the study, both sensor feeds and usage-logs, along with user-generated EMA responses, are first encrypted, cached locally on the user's device, and then transmitted to a secure remote server, where it is stored in an encrypted scalable MySQL database.

2.2 eWellness Data Collection

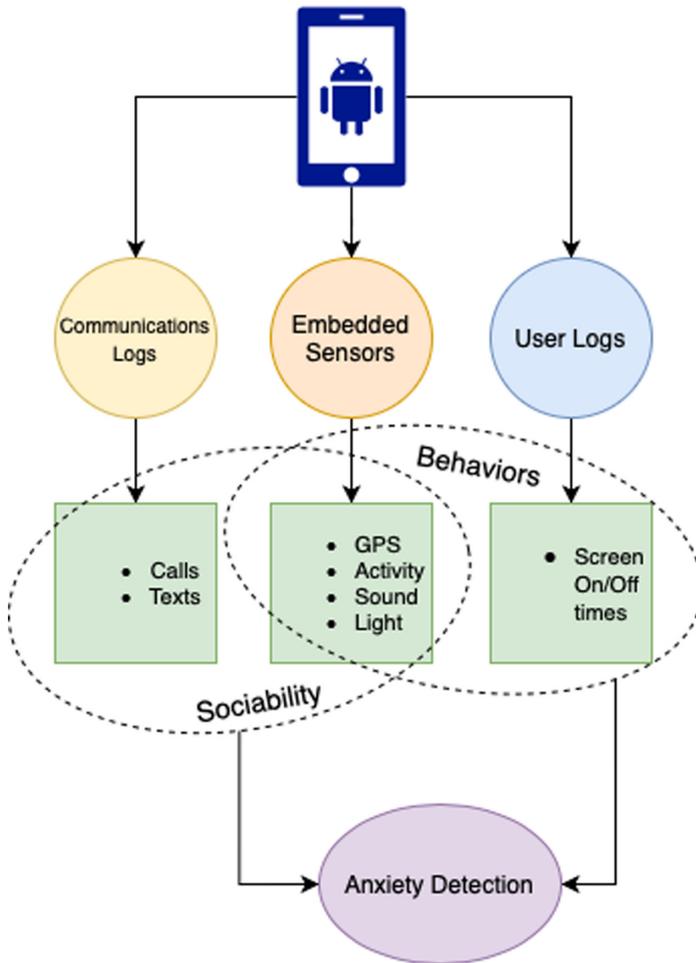


Fig. 1. eWellness data collection hierarchy

The eWellness mobile application, developed for android devices, collects passive behavioral data derived from communications logs, embedded sensors, and user-logs capturing (Fig. 1) the following metrics:

- **Communication:** monitors incoming and outgoing phone calls and text messages, including the duration of phone calls, the number of texts and phone calls, and unique individuals contacted. This does not assess the content of communications or the recipient of the communication, beyond establishing a unique contact.
- **Location:** is periodically sampled using GPS, network, and Wi-Fi detection. Prompts for a new location after moving 5 m, up to once a minute. This metric leverages the Google Fused Location API. The application does not track specific locations; instead, it keeps a total distance traveled using the vectorized haversine distance function.
- **Ambient Sound:** is a numeric measure, designed to detect speech and communication above 50 decibels using the phone’s microphone. It samples every 5 min for 5 s. This metric does not capture the audio files of communications and merely documents the sound frequency and decibel level as numeric values.
- **Activity and Movements:** leverage the device’s accelerometer, gyroscope, and GPS tracking. Activity is sampled every 60 s. In order to determine stationary and moving activity-type, the application leverages Google’s Activity Recognition API.
- **Light:** detects light level associated with possibly being in an outdoor or indoor location. This sensor is sampled every 6 s.
- **Phone Use:** is user-log monitoring the device’s screen on-time.

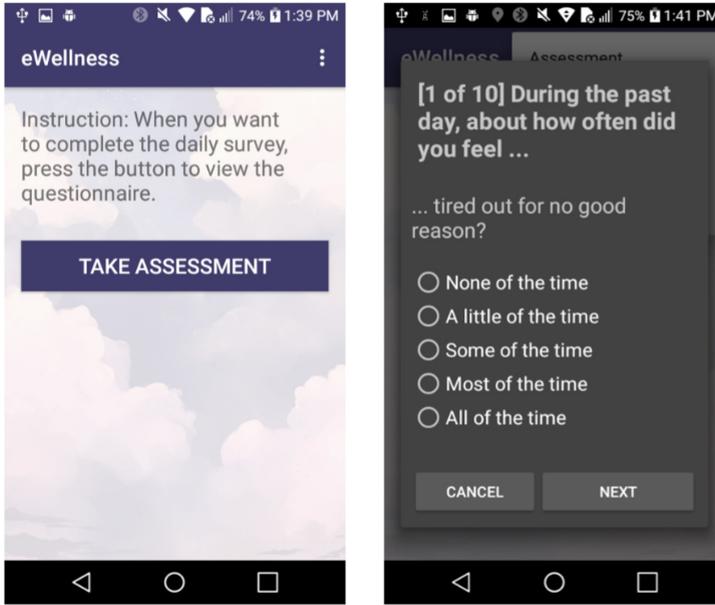
From these raw values, we derived daily aggregated features from these metrics to infer both a user’s sociability, and behavioral patterns. These were then used to learn a model for prediction of anxiety symptom severity. We obtained statistical characteristics, such as minimum, maximum, mean, standard deviation, the 25th, 50th, and 75th percentiles, of the numeric values of noise exposure and the ambient luminescence. The number of activity transitions and duration of each physical activity per day also became a significant metric of identifying mentally distressed days.

2.3 Limiting Personally Identifiable Data Collection

Recognizing the potentially invasive nature of applications like this, data collection was carefully scoped to avoid the collection of Personally Identifiable Information (PII) that could link a particular user to a particular dataset. For example, when attempting to gauge sociability, the application logs the total number of phone calls made, total time on the phone, and the number of unique contacts called; the identities of specific callers were not tracked. This has the consequence of introducing a degree of obscurity into an observed finding (e.g., as the application is unable to differentiate between calls to friends and calls to

a customer-service hotline). At the same time, in the interest of both respecting privacy and ensuring the acceptability of the app, these efforts were felt to be necessary constraints on data collection.

3 Pilot Study Methodology



(a) Landing page.

(b) Daily EMA questionnaire.

Fig. 2. Screenshots of eWellness

An IRB-approved pilot study was conducted on a dozen individuals who use smartphone devices with Android version 5.0 and above. Participants were recruited from the university community, and included both students and staff. Study participants did not have a reported history of mental illness. Participants were asked to download and install the eWellness application (Fig. 2), and then run it on their phone for a month. Passive data was collected continuously by the application throughout the month. Participants were asked to answer EMA daily through the eWellness app, but did not provide any other personal information, such as name, gender, age, during participation.

The Kessler Psychological Distress Scale (K10) [2] is a validated measure of psychological distress over the past 30 d, which is used for clinical and epidemiological purposes. It has a notable success in measuring feelings of anxiety along with depression. For this pilot, the K10 was modified to assess criteria over the

previous 24 h period. The modified K10 prompted the users as daily EMA to measure their feelings of anxiety and depression. The K10 is composed of ten questions, structured on the following standardized template, “Over the past 24 h, how often have you...”, to which users can provide one of five standardized responses: All of the time, Most of the time, Some of the time, A little of the time, and None of the time). These responses are scored on a range from five (All of the time) through one (None of the time). The minimum possible score of K10 is 10, and the maximum possible score is 50. K10 results are categorized into four levels of psychological distress: low distress, moderate distress, high distress, and very high distress. Table 1 details the stress threshold scores. These results were leveraged as a label for the classification of supervised learning.

Table 1. Categorization of K10 Scores [1].

K10 Score	Level	Samples (N = 146)
10–15	Low distress	91
16–21	Moderate distress	29
22–29	High distress	21
30–50	Very high distress	5

4 Results

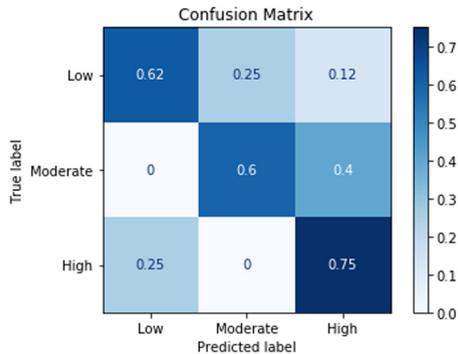
Only 10 participants answered at least seven days of EMAs and provided successful passive sensing data throughout the month. Our analysis focused on a fully supervised learning approach, and only labeled samples were included. For this pilot study, we used 146 daily samples to identify daily anxiety and depression levels. The Z-Score normalization was applied to the features to reach normalized values from different participants.

We selected 25 features that have a relatively higher correlation with the raw K10 score. Table 2 provides a detailed list of feature labels and associated descriptions.

For the 4-class classification, we used 5-fold Cross-Validation (CV) with four models: K-Nearest Neighbors (KNN), Extra-Trees (ET), Support Vector Machine (SVM), and Multilayer Perceptron (MLP). The class weight was automatically applied to the models inversely proportional to the class frequencies to train the imbalanced dataset. The highest classification accuracy achieved was around 76% with the extra-trees model. We also applied the under-sampling technique to improve the performance of an imbalanced dataset. Samples from the low distress class were removed randomly to make uniformly distributed class labels. Samples from the very high distress class were also ignored. A confusion-matrix (Fig. 3) demonstrates that the average score of classifying three classes is 0.65.

Table 2. 25 features most highly correlated to K10 scores

Feature name	Description
total-messages	Total # of text messages-received
is-silent-count	Number of instances no noise was detected
freq-std	Standard deviation for noise frequency
freq-25%	Noise frequency 25th percentile value
deci-std	Standard deviation for noise decibel
deci-50%	Noise decibel 50th percentile value
deci-75%	Noise decibel 75th percentile value
rms-max	Root mean squared measure of audio over time
act-transition	Activity tracking count when in transition
still-cnt	Total count of time user was still
tilting-cnt	Total count of instances the user was tilting
on-foot-cnt	Total count of the instances the user was still
on-bicycle-cnt	Total count of time user was riding a bike
on-foot-dur	Duration of time on foot
on-bicycle-dur	The duration the user spent on a bike
elapsed-device-on	Count of time the phone was active
elapsed-device-off	Count of time the phone was inactive
light-std	Standard deviation for luminescence value
light-25%	Luminescence 25th percentile value
light-50%	Luminescence 50th percentile value
loc-speed-mean	Average speed traveled in a day
loc-alt-mean	Mean Altitude Location
loc-alt-std	Standard deviation of the altitude
loc-alt-75%	Altitude 75th percentile value
loc-alt-max	Altitude max value

**Fig. 3.** 3-class (Low, Moderate, and High distress) classification confusion matrix.

5 Discussion

5.1 Relevant Features

There are some notable and counter-intuitive findings regarding what data elements proved to be most-highly correlated to mental health. It is not surprising to note the presence of features closely related to physical activities (e.g., Duration of time spent biking or walking) as such activities have been definitively linked to mental health [22].

What is somewhat less intuitive is the presence of multiple audio and light sensing features. Audio sensing was included in the protocol under the hypothesis that a moderate level of sound could be indicative of pro-social activities like being outdoors or in group settings. Conversely, overly loud or quiet noise profiles could be indicative of stressful environments or isolated conditions that could be deleterious to mental health. But while interesting in theory, there are many confounding causes of noise that, by limiting ourselves to solely capturing the frequency and decibel levels of the sound, we would fail to distinguish. (intuitively, someone watching TV at home alone could register the same noise profile as someone out to dinner with friends).

Similarly, it was hypothesized that light sensing could be indicative of an individual being outside, which has been shown to positively correlate to mental health [16], however here too, many confounding factors would impact light readings, foremost among them, that the user would actually have to have their phones out and exposed when outside for the light sensor to register it.

The authors note that Sound and Light sensing is notable in that both were the most frequently sampled of all features. It is possible that the high degree of granularity of readings afforded to these particular sensing modalities explains their relevance. Regardless, the authors suggest additional work is needed to understand whether or not these features are indeed more universally indicative of mental health, and explore why that is potentially the case.

5.2 Limitations of the Study

While 10 subjects completing one-months worth of continuous data represents a critical validation of the technology and its potential utility, the dataset is too small to achieve statistically significant results. Additionally, this pilot was scoped to only include individuals without a clinical diagnosis of Anxiety. Consequently, there were insufficient cases of user-reported mental distress, particularly moderate or severe cases, in order to classify them effectively. Additional studies are planned to enlarge our dataset and include a cohort of individuals with diagnosed mental health conditions.

5.3 Accuracy of Labeling

The authors feel there is significant concern about the veracity of user self-reported labeling of mental health that was leveraged in this study. When con-

structuring the experimental design, focus was placed on maximizing user participation in the study. At the time, the primary concern the authors had was that participants would fail to submit a sufficient number of survey responses. Therefore the protocol was designed to combat this, by prompting users to fill out a daily EMA in the application via push-notification, with manual outreach to users who failed to complete an EMA within 48 h, as well as designing the K10 to be a simple to complete multiple-choice assessment. This combination resulted in successfully encouraging active participation in the study; however, there was no mechanism designed to confirm or validate that the resulting inputs were an accurate reflection of a user's actual wellbeing.

It is highly likely, therefore, that at least some users were motivated to respond quickly, and not necessarily accurately. This would result in users simply selecting the default answer of no reported anxiety to each question.

Furthermore, there may have been a reluctance among users to accurately report out mental health issues given perceived embarrassment or stigma associated with poor mental health. Under-reporting of mental health issues is a persistent issue that plagues the domain more generally, and isn't limited to this study [11], however failing to account for under-reporting is a notable issue.

Finally, even well-meaning participants may have failed to accurately represent their mental health state due to their either overlooking, or mischaracterizing, stresses they encountered. This is particularly true when comparing responses across users, where baseline expectations of stress may vary wildly among participants, with prior work for instance demonstrating a clear association between gender and reported wellbeing [21], the result being that one participant's perception of a 'normal' day, might easily be classified as a low or moderately-stressed day by another.

Solving this challenge is essential for ultimately achieving the intended goal of accurate classification of mental illness, for unlike alternative labeling exercises, where quantifiable metrics are possible, here the labeling of an objective state, mental illness, particularly when physiological monitoring is not available, is entirely reliant on subjective inputs, ones that are difficult to accurately capture, and even more difficult to standardize across users.

The authors recommend that future studies will have to address these concerns by better anticipating and correcting for challenges with accurate labeling of mental health.

There are a number of possible remedies to this. In the questionnaire itself, careful structuring of the questions can engage users to provide more thought-out results [14]. Cross-validating questions designed to ensure internal consistency are also an effective means of ensuring user accuracy [15].

Consideration should also be given to alternative methods for collecting labels. Interviewing subjects to determine their mental health, for instance, would likely produce more accurate results, although would have attendant tradeoffs of its own, such as reducing the number of labels that could effectively be captured.

Educating participants on the presentations of Anxiety could also be key towards a more accurate and consistent recall of symptoms. Finally, developing the trust of participants through engagement and transparency, could help to solicit more honest engagement.

5.4 Subject Heterogeneity

The activities tracked by the eWellness app showcase significant heterogeneity across subjects in-terms of usage-patterns. Variables like distance-traveled, number of texts and calls, and physical activity levels, are all far more likely to be impacted by the individual's lifestyle, than their mental health on any given day.

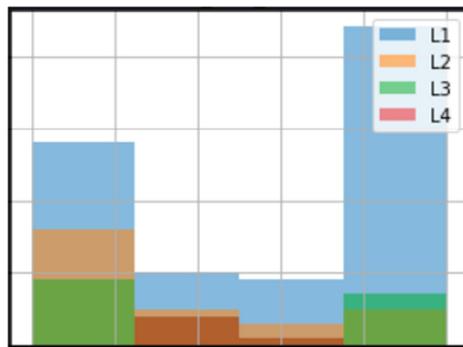


Fig. 4. Histogram of normalized values of Duration on Foot for the 4 labels of stress

Figure 4 showcases a fairly typical distribution, in this case the duration spent on foot in a given day, bucketed into quartiles, with the 4 labels of interest (with L1 or Level-1 corresponding to Low-distress, L2 to Moderate Distress, L3 to High-Distress, and L4 to Very high Distress), in this classification superimposed. While intuitively more time spent on foot may be associated with better mental health, here we observe no clear pattern.

It was therefore assumed that primary-success would be achieved by classifying mental health within users across time, once their baselines for normative behavior were established, rather than across users. The limitations of this initial dataset did not allow for adequate classifying by individual; however, the fact that classification success was achieved by bundling samples across all subjects is remarkable in its indication that cross-subject learning in this domain could be possible. The authors suspect that part of this result likely stems from normalization performed on the data to account for habitual differences in subject usage. By normalizing the data in this manner, the absolute number is rendered largely moot, and instead variances in user patterns are highlighted, as it is likely the day-to-day variations that are more reflective of shifts in mental health. Additional data collection is necessary to validate this finding.

5.5 Usability

Attempting to gauge the viability of the concept, participants in the pilot were asked to submit a voluntary anonymized post-study questionnaire regarding their perceptions about the application and its data collection practices. All participants responded. A significant majority described the application as somewhat (40%) or mostly (40%) useful. Likewise, all users endorsed feeling comfortable with the application, and only one user expressed reservations about the data being collected.

All participants obtained detailed accounting of the data that was collected as part of their onboarding process to the study. No individual declined to participate after learning the precise nature of what was being tracked. This sampling suggests that, particularly among the young adults who are more accustomed to digitized lives, there is less concern about data collection through their mobile devices. Limiting the collection of PII could be sufficient to assuage most privacy concerns.

The primary issue users had with the application was its battery consumption resulting from heavy over-sampling of the sensors. Future iterations of the application will seek to optimize battery usage by minimizing the sampling frequency.

6 Conclusion

Remote health monitoring of mental health, when done so leveraging passive and unobtrusive data collection, could be a useful alternative for conducting real-time mental health surveillance. This paper presents eWellness, an experimental mobile application designed to track a full-suite of sensor and log data off a user's device continuously and passively. An initial pilot study tracking ten people over a month showed a nearly 76% success rate at predicting daily anxiety levels based solely on the passively monitored features. Our current approach may prove useful at tracking longitudinal trends in an individual's mental health, as well as providing a platform for just-in-time interventions to mental health crises. Additional work is needed to refine both the technology and analytics.

References

1. Chapter - k10 scoring. <https://www.abs.gov.au/ausstats/abs@.nsf/lookup/4817.0.55.001Chapter92007-08>
2. Andrews, G., Slade, T.: Interpreting scores on the kessler psychological distress scale (k10). *Aust. New Zealand J. Public Health* **25**(6), 494–497 (2001)
3. American College Health Association: American college health association-national health assessment II: Reference group executive summary spring 2016 (2016)
4. American Psychiatric Association: what are anxiety disorders? (2019). <https://www.psychiatry.org/patients-families/anxiety-disorders/what-are-anxiety-disordersg>

5. Bardone, A.M., Moffitt, T.E., Caspi, A., Dickson, N., Stanton, W.R., Silva, P.A.: Adult physical health outcomes of adolescent girls with conduct disorder, depression, and anxiety. *J. Am. Acad. Child Adolesc. Psychiatry* **37**(6), 594–601 (1998)
6. Ben-Zeev, D., Scherer, E.A., Wang, R., Xie, H., Campbell, A.T.: Next-generation psychiatric assessment: using smartphone sensors to monitor behavior and mental health. *Psychiatric Rehabil. J.* **38**(3), 218 (2015)
7. Bitsko, R.H., et al.: Epidemiology and impact of health care provider-diagnosed anxiety and depression among us children. *J. Dev. Behav. Pediatr. JDBP* **39**(5), 395 (2018)
8. Bruffaerts, R., et al.: Mental health problems in college freshmen: prevalence and academic functioning. *J. Affect. Disord.* **225**, 97–103 (2018)
9. Cartwright-Hatton, S.: *Anxiety of childhood and adolescence: challenges and opportunities* (2006)
10. Ciabattoni, L., Ferracuti, F., Longhi, S., Pepa, L., Romeo, L., Verdini, F.: Real-time mental stress detection based on smartwatch. In: 2017 IEEE International Conference on Consumer Electronics (ICCE), pp. 110–111. IEEE (2017)
11. Corrigan, P.W., Watson, A.C.: The paradox of self-stigma and mental illness. *Clin. Psychol. Sci. Pract.* **9**(1), 35–53 (2002). <https://doi.org/10.1093/clipsy.9.1.35>. <https://onlinelibrary.wiley.com/doi/abs/10.1093/clipsy.9.1.35>
12. Fukazawa, Y., Ito, T., Okimura, T., Yamashita, Y., Maeda, T., Ota, J.: Predicting anxiety state using smartphone-based passive sensing. *J. Biomed. Inf.* **93**, 103151 (2019)
13. Hammerness, P., Harpold, T., Petty, C., Menard, C., Zar-Kessler, C., Biederman, J.: Characterizing non-OCD anxiety disorders in psychiatrically referred children and adolescents. *J. Affect. Disord.* **105**(1–3), 213–219 (2008)
14. Jenn, N.C.: Designing a questionnaire. *Malays. Fam. Phys. Off. J. Acad. Fam. Phys. Malays.* **1**(1), 32–35 (2006). note =PMID: 26998209
15. Li, F., Harmer, P., Duncan, T.E., Duncan, S.C., Acock, A., Yamamoto, T.: Confirmatory factor analysis of the task and ego orientation in sport questionnaire with cross-validation. *Res. Quart. Exerc. Sport* **69**(3), 276–283 (1998). <https://doi.org/10.1080/02701367.1998.10607694>. pMID: 9777664
16. Triguero-Mas, M., et al.: Natural outdoor environments and mental and physical health: relationships and mechanisms. *Environ. Int.* **77**, 35–41 (2015). <https://doi.org/10.1016/j.envint.2015.01.012>
17. Merikangas, K.R., et al.: Lifetime prevalence of mental disorders in us adolescents: results from the national comorbidity survey replication-adolescent supplement (NCS-A). *J. Am. Acad. Child Adolesc. Psychiat.* **49**(10), 980–989 (2010)
18. Narziev, N., Goh, H., Toshnazarov, K., Lee, S.A., Chung, K.M., Noh, Y.: STDD: short-term depression detection with passive sensing. *Sensors* **20**(5), 1396 (2020)
19. Newby, J.M., Mewton, L., Williams, A.D., Andrews, G.: Effectiveness of transdiagnostic internet cognitive behavioural treatment for mixed anxiety and depression in primary care. *J. Affect. Disord.* **165**, 45–52 (2014)
20. Osmá, J., Plaza, I., Crespo, E., Medrano, C., Serrano, R.: Proposal of use of smart-phones to evaluate and diagnose depression and anxiety symptoms during pregnancy and after birth. In: IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), pp. 547–550. IEEE (2014)
21. Pinquart, M., Sörensen, S.: Gender differences in self-concept and psychological well-being in old age: a meta-analysis. *J. Gerontol. Ser. B* **56**(4), P195–P213 (2001). <https://doi.org/10.1093/geronb/56.4.P195>

22. Stevens, T.: Physical activity and mental health in the United States and Canada: evidence from four population surveys. *Prev. Med.* **17**, 35–47 (1988). [https://doi.org/10.1016/0091-7435\(88\)90070-9](https://doi.org/10.1016/0091-7435(88)90070-9)
23. Swan, M.: Health 2050: the realization of personalized medicine through crowd-sourcing, the quantified self, and the participatory biocitizen. *J. Pers. Med.* **2**(3), 93–118 (2012)
24. Taylor, W., Shah, S.A., Dashtipour, K., Zahid, A., Abbasi, Q.H., Imran, M.A.: An intelligent non-invasive real-time human activity recognition system for next-generation healthcare. *Sensors* **20**(9), 2653 (2020). <https://doi.org/10.3390/s20092653>
25. Topper, M., Emmelkamp, P.M., Watkins, E., Ehring, T.: Prevention of anxiety disorders and depression by targeting excessive worry and rumination in adolescents and young adults: a randomized controlled trial. *Behav. Res. Therapy* **90**, 123–136 (2017)
26. Wagner, S., et al.: Mental health interventions in the workplace and work outcomes: a best-evidence synthesis of systematic reviews. *Int. J. Occup. Environ. Med. (The IJOEM)* **7** 1–14 (2016)
27. Woodward, L.J., Fergusson, D.M.: Life course outcomes of young people with anxiety disorders in adolescence. *J. Am. Acad. Child Adolesc. Psychiatr.* **40**(9), 1086–1093 (2001)

Author Index

- Ahmed, Iqrar 3
Amini, Navid 153
Angeles, Pedro 153
Atalay, Asli 168
Atalay, Özgür 168
Ayvaz, Uğur 168
- Bahsi, Hayretdin 77
Bartunik, Max 92
Benarrouch, Robin 199
- Dael, Marine 199
Dudley-McEvoy, Sandra 46
- Elmoughni, Hend 168
Enkoji, Ai 59
- Farhadi, Hamed 137
Fazeli, Shayan 212
Fischer, Georg 92
Fischer, Stefan 121
Frappé, Antoine 199
- Gautier, Antoine 199
Ghavami, Mohammad 46
Ghazizadeh, Reza 137
Gwak, Migyeong 212
- Hakala, Jaakko 34
Halder, Senjuti 3
Hämäläinen, Matti 34
Hosseini, Maryam 137
- Inatti, Jari 18, 34
Imdad, Maria 106
Ince, Gökhan 168
- Kale, Daniel 153
Kärkkäinen, Kimmo 212
Katz, Marcos 3
Key, Matthew 59
- Khalesi, Banafsheh 46
Kilpijärvi, Joni 34
Kirchner, Jens 92
Kissi, Chaïmaâ 18
Kuusik, Alar 77
- Larras, Benoit 199
Lau, Florian-Lennert Adrian 121
Levine, Lionel M. 212
Li, Ming 59
Liu, Luyao 184
- Mahdin, Hairulnizam 106
Marroquin, Aaron 59
Mohammadi, Farnaz 153
Mouni, Boppana Udaya 106
Myllylä, Teemu 34
Myllymäki, Sami 34
- Pascua, Emily 153
Peris, Tara 212
Pomalaza-Raez, Carlos 18, 34
Prabhakaran, B. 59
Puttock, James 46
- Ramli, Sofia Najwa 106
Richter, Martin 92
- Sahar, Shakira 106
Särestöniemi, Mariella 3, 18, 34
Sarrazfzadeh, Majid 212
Sohani, Behnaz 46
- Thalhofer, Thomas 92
Tiberi, Gianluigi 46
Tolentino, Gian 153
Tsiopoulos, Leonidas 77
- Vain, Jüri 77
Vega, Abraham 153

Wald, Christian 92

Wendt, Regine 121

Wu, Jiankang 184

Xiao, Shenglang 184

Yang, Pengfei 184

You, Zhiquan 153

Young, Alexander S. 212

Zadeh, Bitá 212

Zhang, Zhiqiang 184