Enabling human physiological sensing by leveraging intelligent head-worn wearable systems.

Nhat Pham

Wolfson College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2022

To my beloved wife (Yen Duong), our child (David Pham), and my family.

I hope that the knowledge that we have discovered in this thesis in some ways could help us to better understand and design wearable computers to improve human lives.

Acknowledgements

Looking back at the last four years, it has been quite a memorable journey full of ups and downs, disappointing and exciting moments. At the end of that journey, I could see myself being another person, grown up and matured both professionally and personally. For that, I am deeply grateful to my supervisor, Professor Tam Vu, for all his guidance and support throughout the years. I have learnt from him not only the knowledge but also methodologies to lead research projects, find innovative ideas, efficiently manage time, mentor junior students and, most importantly, a strong desire for impactful research and to constantly challenge my limits. I am also very thankful to my supervisors, Professor Niki Trigoni and Professor Andrew Markham, for working with me in the last year of my DPhil study, spending a lot of effort to refine this thesis, and helping me to build up my academic career. I would like to thank Professor Phuc (VP) Nguyen for being my 'shadow' advisor and an 'older brother' to me, who I could always rely on in difficult times. I also want to thank Professor Ronald Clark and Professor Alastair Beresford for serving on my thesis committee, spending their time reviewing this thesis and providing valuable feedback.

I am thankful to all of my colleagues in the Mobile and Networked Systems Laboratory (MNS), including Hoang Truong, Nam Bui, Taeho Kim, and Anh Nguyen, for contributing to the development of this thesis. I am grateful for all of my research collaborators Professor Cecilia Mascolo, Professor Dong Ma, Professor Farnoush Banaei-Kashani, Professor Thang Dinh, Doctor Ann Halbower, Doctor Robin Deterding, Tuan Dinh, Minh Tran, Hong Jia, Zohreh Raghebi, Young Kwon, Carole Kline, and Tuan Nguyen for helping me with the development and evaluations of the proposed systems. Finally, I wish to express my gratitude to my beloved wife, Yen Duong, our child, David Pham, and my family, Tai Pham, My Ha, Hoa Thai, and Thi Pham, for their unconditional support which has been

my most powerful motivation throughout my DPhil program. To them, I dedicate this thesis.

Abstract

This thesis explores the challenges of enabling human physiological sensing by leveraging head-worn wearable computer systems. In particular, we want to answer a fundamental question, i.e., ***could we leverage head-worn wearables to enable accurate and socially-acceptable solutions to improve human healthcare and prevent life-threatening conditions in our daily lives?*** To that end, we will study the techniques that utilise the unique advantages of wearable computers to (1) facilitate new sensing capabilities to capture various biosignals from the brain, the eyes, facial muscles, sweat glands, and blood vessels, (2) address motion artefacts and environmental noise in real-time with signal processing algorithms and hardware design techniques, and (3) enable long-term, high-fidelity biosignal monitoring with efficient on-chip intelligence and pattern-driven compressive sensing algorithms.

We first demonstrate the ability to capture the activities of the user's brain, eyes, facial muscles, and sweat glands by proposing WAKE, a novel behind-the-ear biosignal sensing wearable. By studying the human anatomy in the ear area, we propose a wearable design to capture brain waves (EEG), eye movements (EOG), facial muscle contractions (EMG), and sweat gland activities (EDA) with a minimal number of sensors. Furthermore, we introduce a Three-fold Cascaded Amplifying (3CA) technique and signal processing algorithms to tame the motion artefacts and environmental noises for capturing high-fidelity signals in real time. We devise a machine-learning model based on the captured signals to detect microsleep with a high temporal resolution.

Second, we will discuss our work on developing an efficient Pattern-dRiven Compressive Sensing framework (PROS) to enable long-term biosignal monitoring on low-power wearables. The system introduces tiny on-chip pattern recognition primitives (TinyPR) and a novel pattern-driven compressive sensing technique (PDCS) that exploits the sparsity of biosignals.

They provide the ability to capture high-fidelity biosignals with an ultra-low power footprint. This development will unlock long-term healthcare applications on wearable computers, such as epileptic seizure monitoring, microsleep detection, etc. These applications were previously impractical on energy and resource-constrained wearable computers due to the limited battery lifetime, slow response rate, and inadequate biosignal quality.

Finally, we will further explore the possibility of capturing the activities of a blood vessel (i.e., superficial temporal artery) lying deep inside the user's ear using an ear-worn wearable computer. The captured optical pulse signals (PPG) are used to develop a frequent and comfortable blood pressure monitoring system called eBP. In contrast to existing devices, eBP introduces a novel in-ear wearable system design and algorithms to eliminate the need to block the blood flow inside the ear, alleviating the user's discomfort.

Statement of Originality

This thesis is submitted to the Department of Computer Science, University of Oxford, in fulfilment of the requirements for the degree of Doctor of Philosophy. This thesis is entirely my own work, except where otherwise stated.

Nhat Pham, Wolfson College.

Funding

The work described in this thesis was partially funded by the University of Oxford DPhil Scholarship, the Alfred P. Sloan Fellowship no. FG-2020-13110 (TV), ERC through Project 833296 (EAR), Nokia Bell Labs, NSF #2132112, NSF CNS/CSR #1846541, NSF SCH #1602428, Google Faculty Awards 2018, the Schramm Foundation and the Colorado Advanced Industries Accelerator (AIA).

Contents

1 Introduction **1**

1.1 Motivations 1
1.2 Challenges 3
1.3 Contributions 4
1.3.1 WAKE: Behind-the-ear biosignals sensing and microsleep detection. 4
1.3.2 PROS: Long-term head-based biosignal monitoring on low-power wearables. 5
1.3.3 eBP: In-ear blood vessels activity and pressure monitoring. . . 7
1.4 Limitations and future directions 8
1.5 Publications 8
1.6 Thesis Structure 12

2 Background **13**

2.1 Behind-the-ear biosignal sensing and microsleep detection 13
2.1.1 Microsleep detection application 13
2.1.2 Understanding microsleep 14
2.1.3 Electroencephalography (EEG) 16
2.1.4 Electrooculography (EOG) 16
2.1.5 Electromyography (EMG) 17
2.1.6 Electrodermal Activity (EDA) 17
2.1.7 Motion artefacts 17
2.1.8 Environmental noise 18
2.1.9 Related work 19
2.2 Long-term head-based biosignal monitoring on low-power wearables . 20
2.2.1 The importance of long-term biosignal monitoring. 20
2.2.2 Sparsity in biosignal monitoring 21
2.2.3 Low-power Microcontrollers 22

2.2.4 Pattern Recognition with Neural Networks 23
2.2.5 Tiny Machine Learning 23
2.2.6 Compressive Sensing 23
2.2.7 Related work 24
2.3 In-ear blood vessels activity and pressure monitoring. 25
2.3.1 Frequent blood pressure measurement 25
2.3.2 Fundamental of BP measurement 26
2.3.3 Photoplethysmography (PPG) 28
2.3.4 Related work 29
2.4 Summary 30

3 WAKE: Sensing head-based biopotentials and microsleep detection with a behind-the-ear wearable 31

3.1 Introduction 31
3.1.1 Goals and Challenges 31
3.1.2 Impact of environmental noise and motion artefacts 32
3.2 Exploring Microsleep biomarkers from behind the ears 33
3.3 System Overview 36
3.4 Mitigating Impact of Motions & Noise 38
3.5 Signal Processing 42
3.5.1 WAKE On-board processing. 42
3.5.2 WAKE Physiological Signals Extraction 45
3.6 Algorithms 46
3.6.1 Classification based on Feature Engineering 46
3.6.2 Deep Learning on Raw Data 48
3.7 Implementation 48
3.8 Performance Evaluation 51
3.8.1 BTE Signals Sensitivity Validation 51
3.8.2 Noise Suppression Performance 52
3.8.3 Microsleep Detection Performance 54
3.8.4 Usability Analysis 61
3.9 Summary 63

4 PROS: efficient Pattern-dRiven cOmpressive Sensing for Low-Power Biosignal-based Wearables 65
4.1 Introduction . . . 65
4.1.1 Goals and Challenges . . . 65
4.2 PROS System Overview . . . 66
4.3 Tiny pattern recognition primitives . . . 68
4.3.1 Key challenges and designs . . . 68
4.3.2 Pattern Recognition as the Rare Event Detection Problem . . . 69
4.3.3 Informative Feature Extraction . . . 70
4.3.4 Efficient Design for Recognition Model . . . 71
4.4 Pattern-Driven Compressive Sensing . . . 73
4.4.1 PDCS framework design . . . 74
4.4.2 Sparsity variations among patterns . . . 75
4.4.3 Optimal wavelet domains search . . . 76
4.4.4 Recovery with Pattern Information and Block Sparse Bayesian Learning . . . 78
4.5 Hardware and OS Optimisations . . . 79
4.6 Implementation . . . 83
4.7 Evaluations . . . 84
4.7.1 Datasets Preparation. . . . 84
4.7.2 TinyPR Primitives. . . . 85
4.7.3 Pattern-driven Compressive Sensing. . . . 87
4.7.4 Hardware runtime performance. . . . 90
4.7.5 Epileptic seizures detection use case. . . . 92
4.8 Summary . . . 94

5 eBP: Sensing in-ear blood vessels to monitor blood pressure with a wearable computer. 95
5.1 Introduction . . . 95
5.1.1 Goals and Challenges . . . 95
5.1.2 Applications and Impacts . . . 96
5.2 System Overview . . . 97
5.2.1 Motivation and challenges for in-ear BP monitoring: . . . 97
5.2.2 Proposed solutions: . . . 99
5.3 System Hardware Design . . . 100
5.3.1 Design considerations . . . 100

5.3.2 Implementation 102
5.4 In-ear Blood Pressure Estimation Algorithms 106
5.4.1 Systolic BP Measurement 107
5.4.2 Mean Arterial Pressure Detection 108
5.4.3 Diastolic BP Measurement 110
5.5 PPG Qualification and Signal Processing 111
5.6 Evaluation 114
5.6.1 Experimental Methodology 115
5.6.2 System Performance 115
5.6.3 Power consumption 117
5.6.4 Prediction Stability 117
5.6.5 Optimal sensor location 118
5.6.6 User Experience Survey 118
5.7 Summary 119

6 Conclusions **121**
6.1 Concluding Remarks 121
6.2 Future Work 121
6.3 Long-term Vision 125

Bibliography **126**

List of Figures

1.1 a) A standard biosignals (EEG/EOG/EMG) monitoring system, b) an wearable biosignal sensing system developed by Emotiv Inc. 2

1.2 WAKE is a novel behind-the-ear wearable device for head-based biosignals sensing. By monitoring biosignals from the brain, eye movements, facial muscle contractions, and sweat gland activities from behind the user's ears, WAKE can detect microsleep with a high temporal resolution. We introduce a Three-fold Cascaded Amplifying (3CA) technique to tame the motion artefacts and environmental noises for capturing high-fidelity signals. 5

1.3 PROS is an efficient pattern-driven compressive sensing framework for low-power wearables. PROS eliminates the conventional trade-off between signal quality, response time, and power consumption by introducing tiny pattern recognition primitives and a novel pattern-driven compressive sensing technique that exploits the sparsity of biosignals. 6

1.4 eBP is a wearable system that can measure blood pressure frequently and comfortably from inside a user's ear. eBP has three key components: (1) a light-based pulse sensor attached to an inflatable pipe that goes inside the ear, (2) a digital air pump with a fine controller, and (3) a BP estimation algorithm. In contrast to existing devices, eBP introduces a novel technique that eliminates the need to block the blood flow inside the ear, alleviating the user's discomfort. 7

2.1 Wakefulness neural pathway. 14

2.2 Standard EEG, EOG, EMG, and EDA electrode locations. 15

2.3 The response of blood artery to outer pressure that causes the measurement of BP. 27

3.1 BTE vs. standard locations. 33

3.2 BTE anatomy. 34

3.3 EEG, EOG, EMG and EDA signals captured from behind the ears. . 35

3.4 WAKE system overview. 36
3.5 WAKE hardware module. 37
3.6 3CA model. 39
3.7 EMG saturation w/o AGC. 43
3.8 $\Sigma - \Delta$ modulation. 44
3.9 Signals captured by WAKE during the transition between awake and microsleep. 45
3.10 3D model. 49
3.11 BTE silicon earpieces. 50
3.12 WAKE circuit. 50
3.13 BTE EEG, EOG, EMG and EDA signals vs. ground-truth. 52
3.14 Motion noise power reduction. 53
3.15 Walking motion noise suppression. 53
3.16 Driving motion noise suppression. 54
3.17 3CA noise reduction (In an Office). 54
3.18 3CA noise reduction (At Home). 55
3.19 3CA noise reduction (Inside a Car). 55
3.20 Experiment setup. 56
3.21 Learning curves of training SorNet on the test-set setting. 60
3.22 Active power usage. 61
3.23 CPU core thermal profile. 62
3.24 User study. 63

4.1 PROS system overview. 67
4.2 Pattern-driven Compressing Sensing framework. 74
4.3 Sparsity variations among patterns in a wavelet (db2) domain. 76
4.4 Efficiency Curve. 80
4.5 Tickless Sleep. 81
4.6 PROS hardware platform. 83
4.7 SPSW TinyPRs (tiny configuration) performance in 10-fold cross-validations. 85
4.8 TinyPRs (tiny configuration) performance in 10-fold cross-validations. 86
4.9 SPSW TinyPRs (tiny configuration) performance in unknown subjects evaluations. 87
4.10 TinyPRs (tiny configuration) performance in unknown subjects evaluations. 87
4.11 ABSZ optimal domain search. 88

4.12 ABSZ recovery quality in *bior*4.4 domain. 89
4.13 Runtime experiment setup. 90
4.14 Open-source hardware and biosensing platforms. 92

5.1 eBP system. 97
5.2 Ear canal anatomy. 99
5.3 In-ear BP module design. 101
5.4 In-ear PPG sensor and balloon design. 103
5.5 The pressure vs. diameter curve of the balloon. 104
5.6 eBP hardware. 106
5.7 eBP prototype. 106
5.8 Amplitude vs. Pressure. 107
5.9 Systolic fraction β detection. 109
5.10 Illustration of the inconsistency of conventional peak-to-peak computation. 110
5.11 In-ear PPG signal (a) with corresponding amplitude and pressure (b). 110
5.12 First derivative of PPG amplitude discloses diastolic BP. 111
5.13 Peak Interval Variance versus Entropy variance. 112
5.14 Results of median and Gaussian smoothing filter on amplitude (a) and pressure trend (b), and the balloon structure (c). 113
5.15 Experiment setup to compare eBP with the KonQuest device. 114
5.16 Bland-Altman plot comparing eBPś measurements and groundtruth . 116
5.17 Pearson correlation coefficients of eBPś estimation and groundtruth. . 116
5.18 Mean and SD error in cross-validation. 117
5.19 Intraclass correlation coefficient of eBP and ground truth. 118
5.20 The SNR of PPG signals at different sensor locations. 119
5.21 User experience questionnaire results. 120

Glossary

3CA Three-fold Cascaded Amplifying.

AAMI Association for the advancement of medical instrumentation.

AASM American Academy of Sleep Medicine.

ADC Analog to digital converter.

AE Active electrode.

AGC Adaptive gain control.

ANS The autonomic nervous system.

BP Blood pressure.

BSBL Block sparse bayesian learning.

BSS Blind source separation.

BTE Behind-the-ear.

CF Compression factor.

CMRR Common-mode rejection ratio.

CNN Convolutional neural networks.

COTS Commercial-off-the-shelf.

CS Compressive Sensing.

DBP Diastolic blood pressure.

DNN Deep neural networks.

DSP Digital signal processor.

DVFS Dynamic voltage frequency scaling.

DWT Discrete wavelet transform.

eBP Ear-worn blood pressure measurement wearable system.

ECG Electrocardiogram.

EDA Electrodermal activity.

EDS Excessive daytime sleepines.

EEG Electroencephalogram.

EMG Electromyograph.

EOG Ectrooculography.

F2DP Feed Forward differential preamplifying.

FDA The United States food and drug administration.

FIR Finite impulse response.

ICA Independent components analysis.

ICC Intraclass correlation coefficient.

IMU Inertial measurement unit.

LOSOCV Leave-One-Subject-Out Cross Validation.

MCU Low-power microcontroller.

MFCC Mel-frequency cepstral coefficients.

ML Machine learning.

MPA Maximum Pulse Amplitude.

MSI Maximum sparsity index.

MWT Maintenance of wakefulness test.

PCA Principal component analysis.

PDCS Pattern-driven compressive sensing.

PED Peak envelope detector.

PIV Peak interval variability.

PoI Pattern of interest.

PPG Photoplethysmography.

PROS An pattern-driven compressive sensing framework for low-power wearables.

PSG Polysomnography.

SBP Systolic blood pressure.

SCL Skin conductance level.

SCR Skin conductance response.

SLD Square law detector.

SNR Signal to noise ratio.

SSIM Structural similarity index.

SVM Support vector machine.

TFLM TensorFlow Lite for Microcontrollers.

TinyPR Tiny pattern recognition primitive.

WAKE A behind-the-ear wearable system for microsleep detection.

Chapter 1

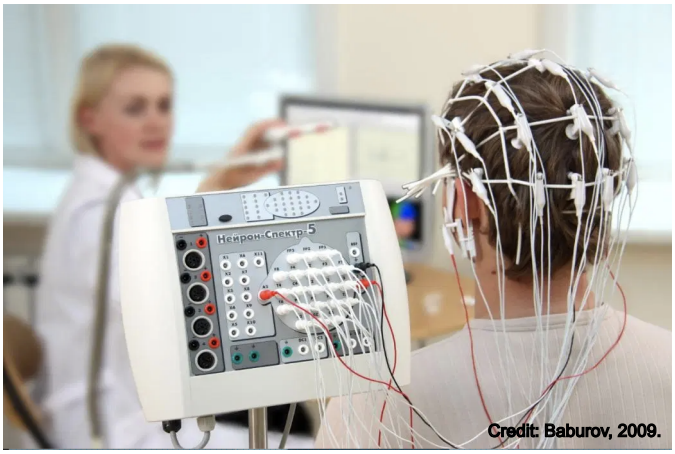
Introduction

1.1 Motivations

The wearable healthcare market is experiencing significant growth in recent years, reaching over $121 billions globally in 2021 and is predicted to surpass $390 billions by 2030 [1, 2, 3]. With more than 500 million units shipped annually, head-worn wearables such as headphones, earphones, headbands, hats, etc., hold an enormous potential to be the platforms of new innovations [4]. Furthermore, they are already socially accepted to be a part of our daily life. As a result, introducing more sensing modalities into these wearables is less likely to obstruct the user's everyday activities, resulting in easy adoption for newly developed research and technology in this field [5].

The human head is a prime location for wearable computers on the human body. It houses essential organs and tissues such as the brain, the eyes, facial muscles, blood vessels, and sweat glands. By listening to their generated biosignals and interactions, we can unlock numerous novel healthcare applications. For example, using just facial muscle signals (i.e., electromyography (EMG)) alone, one can infer the stress level of a user [6, 7] and the eating habit and the type of food consumed by the user [8, 9, 10]. When combining these EMG signals with brain signals (i.e., electroencephalogram (EEG)), one can further understand the user's emotional states [11, 12], their pain and suffering level [13], or sedation level during surgery [14]. Even the current gold standard for sleep quality monitoring in hospitals is based on brain signals (EEG), chin muscle signals (EMG), and eye movement signals (i.e., electrooculography (EOG)) [15, 16] captured from the head. The signals can also be used for measuring attention deficit hyperactivity disorder (ADHD) [17, 18], detecting autism onset [19], or assessing meditation efficacy [20].

However, most existing sensing methods are cumbersome, intrusive, and expensive, primarily suitable only for stationary and short-term usages such as in clinics or



a) Standard Biosignals Sensing

b) Emotiv wearable system

Figure 1.1: a) A standard biosignals (EEG/EOG/EMG) monitoring system, b) an wearable biosignal sensing system developed by Emotiv Inc.

hospitals. For example, existing solutions, such as EEG sensor caps [21, 22], provide very high resolution and accuracy but require users to carry a bulky data collection hub and to wear tens to hundreds of electrodes on their heads (Fig. 1.1a). Other existing wearable devices such as Emotiv (Fig. 1.1b) [23], NeuroSky MindWave [24], BrainLink Pro [25], Muse [26], Kokoon [27], Versus [28], Neuroon Open [29], Naptime [30], etc., which were designed to capture a small number of these signals, are required to be worn on the user's face or head. They are inconvenient, intrusive, and less likely to be adopted as a long-term and everyday wearable solution despite the potential benefit they can bring.

This thesis explores the challenges of enabling human physiological sensing by leveraging head-worn wearable computers. In particular, we want to answer a fundamental question, i.e., ***could we leverage head-worn wearables to enable accurate and socially-acceptable solutions to improve human healthcare and prevent life-threatening conditions in our daily lives?*** To that end, we will study the techniques that utilise the unique advantages of wearable computers to (1) facilitate new sensing capabilities to capture various biosignals from the brain, the eyes, facial muscles, sweat glands, and blood vessels, (2) address motion artefacts and environmental noise in real-time at the electrical level, and (3) enable long-term, high-fidelity biosignal monitoring with efficient on-chip intelligence and pattern-driven compressive sensing algorithms. While proposed sensing techniques focus on specific applications such as blood pressure monitoring, microsleep detection, and epileptic seizure monitoring, the developed software and hardware platforms could be used for a wide range of applications in the mobile healthcare and human-computer interaction

domains where long-term and reliable physiological and cognitive sensing is essential. The potential examples include sleep improvement, focus monitoring, worker safety management, eating habit monitoring, closed-loop pain coaching, etc.

1.2 Challenges

We have identified several challenges that prevent us from enabling physiological sensing with head-worn systems. They are clarified below.

- **Underexplored sensing locations.** In this thesis, we explore new sensing locations (i.e., inside and on the back of the ears) on the human head that could achieve sensing sensitivity, wearability, and social acceptability. Since these sensing locations are underexplored in the literature, the existing sensing techniques cannot be applied. In the case of in-ear blood pressure sensing, the feasibility of the method has yet to be previously confirmed in the literature.

- **Heavy noise and motion artefacts.** Heavy artefacts created by motion and noise coupled from the environment in daily use are the long-standing challenge limiting the practical uses of wearable biosignal sensing systems. It is difficult to ensure high-fidelity signals with limited sensing capabilities and computing resources on wearables. In this thesis, we will tackle motion artefacts generated by human activities such as walking or driving. Furthermore, we will also look at how to mitigate noise generated from electromagnetic radiation in the surrounding environment.

- **Weak and overlapping biosignals.** To achieve wearability and social acceptability, the chosen sensing locations are located further away from the signal sources. Thus, the captured signals are often weak, buried under the noise floor, and overlap with each other in multiple orders of magnitude. This thesis will explore the properties of the captured signals and propose techniques to ensure signal integrity.

- **Long-term, high-fidelity biosignal monitoring.** Many important healthcare applications, such as epileptic seizure or microsleep detection, require high-fidelity biosignal to be monitored over a long period of time (e.g., days or weeks). This results in a significant amount of wireless data communication and drains the battery quickly. In this thesis, we will discuss the proposed techniques to

achieve both high-fidelity signals and long battery life for long-term monitoring applications.

- **Energy and resource constraints.** Finally, wearable devices typically have small batteries and extremely constrained computing resources (e.g., MHz CPU, KB of memory). Thus, all of the developed algorithms and techniques need to be energy and computationally efficient so that they can be deployed on wearable devices.

In chapter 3, we will tackle the challenges of underexplored sensing locations, heavy noise and motion artefacts, and weak biosignals to enable the ability to sense activities of the brain, eyes, facial muscle, and sweat glands with wearable devices. In chapter 4, we will address the challenges of long-term, high-fidelity biosignal monitoring and resource constraints to make our proposed solutions more practical in the real world. Finally, in chapter 5, we will further explore a new possibility of capturing blood vessels' activities inside the ear canal to measure blood pressure.

1.3 Contributions

By utilising the ubiquity and unique advantages of head-worn wearables, this thesis focuses on the challenges of capturing head-based biosignals on energy and resource-constrained wearables to enable unconventional healthcare applications. Below, we will highlight our research contributions through each system's development.

1.3.1 WAKE: Behind-the-ear biosignals sensing and microsleep detection.

Every year, the U.K. and U.S. economy lose more than $411 billion because of work performance reduction, injuries, and traffic accidents caused by microsleep. To mitigate microsleep's consequences, a discreet, reliable, and socially acceptable microsleep detection solution throughout the day, every day, is required. Unfortunately, existing solutions do not meet these requirements.

In this project, we propose WAKE (Fig. 1.2), a novel behind-the-ear wearable device for microsleep detection. By monitoring biosignals from the brain, eye movements, facial muscle contractions, and sweat gland activities from behind the user's ears, WAKE can detect microsleep with a high temporal resolution. Finally, we introduce a Three-fold Cascaded Amplifying (3CA) technique to tame the motion artefacts and environmental noises for capturing high-fidelity signals.

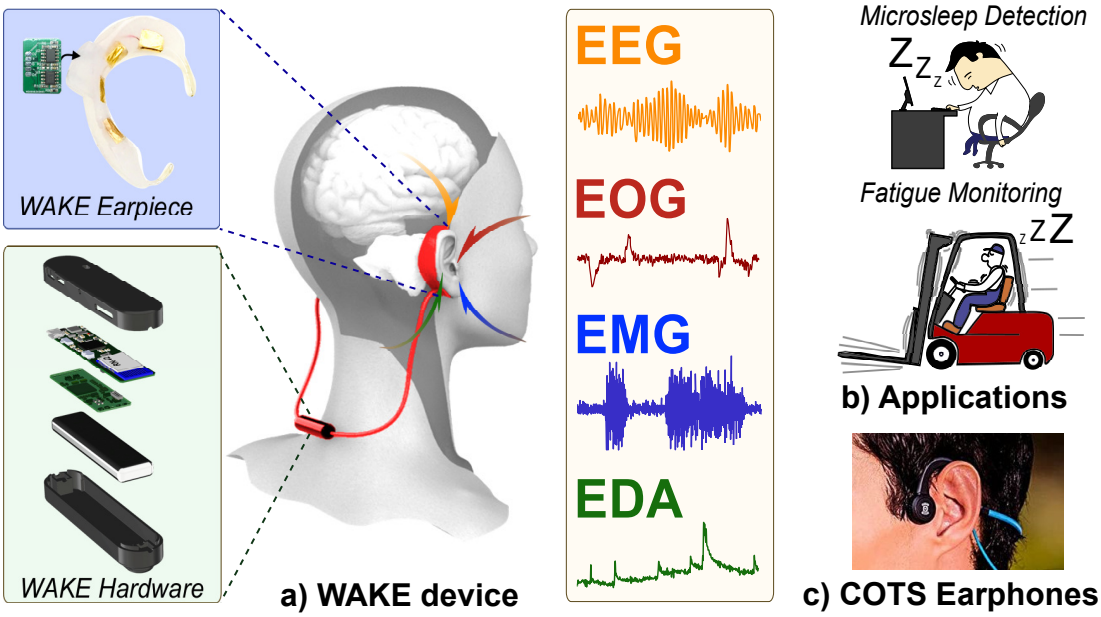


Figure 1.2: WAKE is a novel behind-the-ear wearable device for head-based biosignals sensing. By monitoring biosignals from the brain, eye movements, facial muscle contractions, and sweat gland activities from behind the user's ears, WAKE can detect microsleep with a high temporal resolution. We introduce a Three-fold Cascaded Amplifying (3CA) technique to tame the motion artefacts and environmental noises for capturing high-fidelity signals.

Through our prototyping, we show that WAKE can suppress motion and environmental noise in real-time by 9.74-19.47 dB while walking, driving, or staying in different environments, ensuring that the biosignals are captured reliably. We evaluated WAKE using gold-standard devices on 19 sleep-deprived and narcoleptic subjects. The Leave-One-Subject-Out Cross-Validation results show the feasibility of WAKE in microsleep detection on an unseen subject with average precision and recall of 76% and 85%, respectively.

1.3.2 PROS: Long-term head-based biosignal monitoring on low-power wearables.

While the global healthcare market of wearable devices has been growing significantly in recent years and is predicted to reach $60 billion by 2028, many important healthcare applications, such as seizure monitoring, drowsiness detection, etc., have not been deployed due to the limited battery lifetime, slow response rate, and inadequate biosignal quality.

This study proposes PROS (Fig. 1.3, an efficient pattern-driven compressive sens-

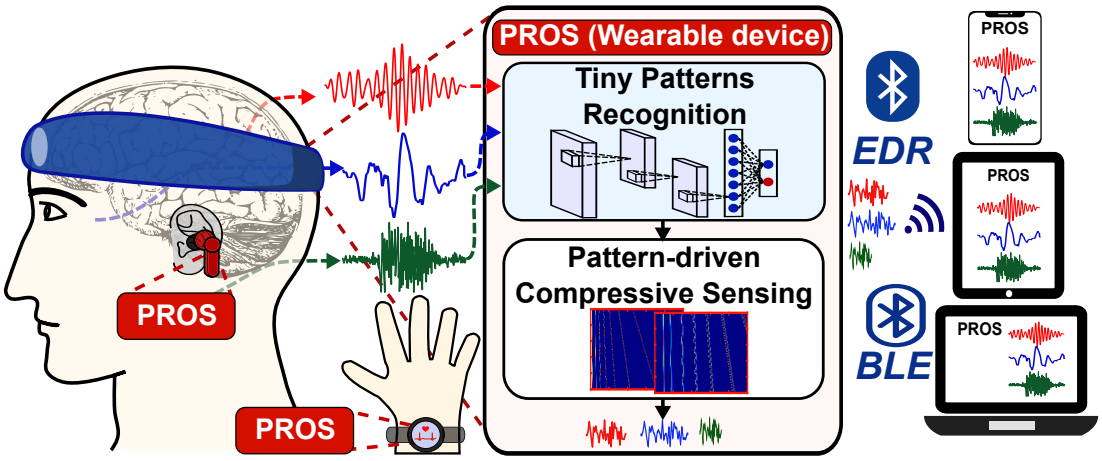


Figure 1.3: PROS is an efficient pattern-driven compressive sensing framework for low-power wearables. PROS eliminates the conventional trade-off between signal quality, response time, and power consumption by introducing tiny pattern recognition primitives and a novel pattern-driven compressive sensing technique that exploits the sparsity of biosignals.

ing framework for low-power wearables. PROS eliminates the conventional trade-off between signal quality, response time, and power consumption by introducing tiny pattern recognition primitives and a pattern-driven compressive sensing technique that exploits the sparsity of biosignals. Specifically, we (i) develop tiny embedded machine learning models to eliminate irrelevant biosignal patterns, (ii) efficiently perform compressive sampling of relevant biosignals with appropriate sparse wavelet domains, and (iii) optimise hardware and OS operations to push the processing efficiency. PROS also provides an abstraction layer, so the application only needs to care about detected relevant biosignal patterns without knowing the optimisations underneath.

We have implemented and evaluated PROS on two open biosignal datasets with 120 subjects and six biosignal patterns. The experimental results on unknown subjects of a practical use case, such as epileptic seizure monitoring, are very encouraging. PROS can reduce the streaming data rate by 24X while maintaining high fidelity signal. It boosts the power efficiency of the wearable device by more than 1200% and enables the ability to react to critical events immediately on the device. The memory and runtime overheads of PROS are minimal, with a few KBs and 10s of milliseconds for each biosignal pattern, respectively.

1.3.3 eBP: In-ear blood vessels activity and pressure monitoring.

Frequent blood pressure (BP) assessment is key to the diagnosis and treatment of many severe diseases, such as heart failure, kidney failure, hypertension, and hemodialysis. Current "gold-standard" BP measurement techniques require the complete blockage of blood flow, which causes discomfort and disruption to normal activity when the assessment is done repetitively and frequently. Unfortunately, patients with hypertension or hemodialysis often have to get their BP measured every 15 minutes for a duration of 4-5 hours or more. The discomfort of wearing a cumbersome and limited mobility device affects their everyday activities.

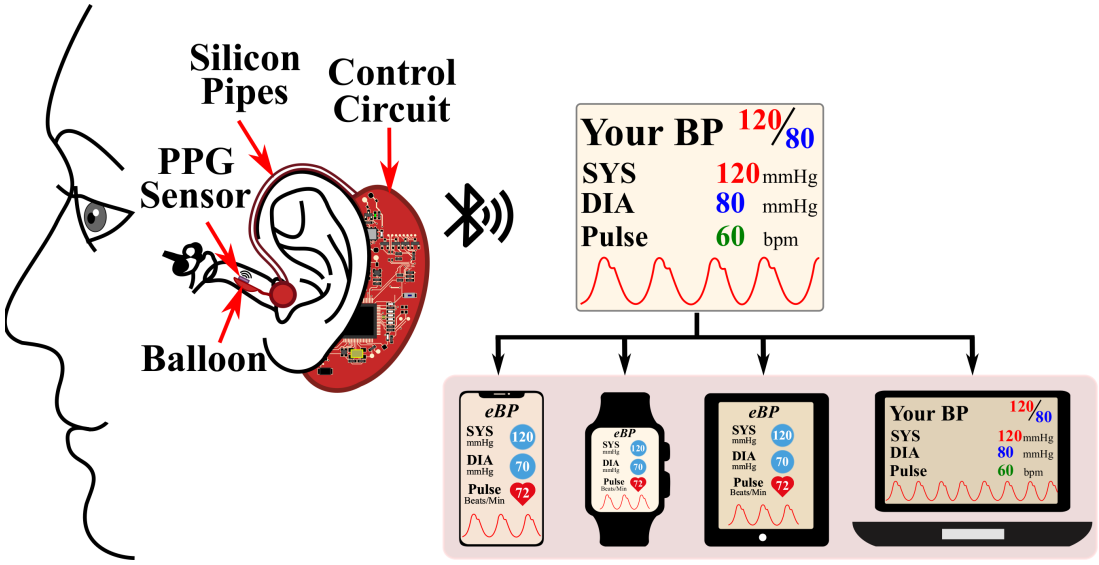


Figure 1.4: eBP is a wearable system that can measure blood pressure frequently and comfortably from inside a user's ear. eBP has three key components: (1) a light-based pulse sensor attached to an inflatable pipe that goes inside the ear, (2) a digital air pump with a fine controller, and (3) a BP estimation algorithm. In contrast to existing devices, eBP introduces a novel technique that eliminates the need to block the blood flow inside the ear, alleviating the user's discomfort.

In this work, we propose a device called eBP (Fig. 1.4) to measure BP from inside the user's ear, aiming to minimize the measurement's impact on users' normal activities while maximizing its comfort level. eBP has three key components: (1) a light-based pulse sensor attached to an inflatable pipe that goes inside the ear, (2) a digital air pump with a fine controller, and (3) a BP estimation algorithm. In contrast to existing devices, eBP introduces a novel technique that eliminates the need to block the blood flow inside the ear, which alleviates the user's discomfort.

We prototyped eBP custom hardware and software and evaluated the system through a comparative study on 35 subjects. The study shows that eBP obtains the average error of 1.8 mmHg and -3.1 mmHg and a standard deviation error of 7.2 mmHg and 7.9 mmHg for systolic (high-pressure value) and diastolic (low-pressure value), respectively. These errors are around the acceptable margins regulated by the FDA's AAMI protocol, which allows mean errors of up to 5 mmHg and a standard deviation of up to 8 mmHg.

1.4 Limitations and future directions

In this thesis, we focus on addressing the fundamental challenges to enable long-term and efficient biosignal monitoring from important sources on the human head and demonstrate the potential applications that could be enabled. However, there are also several important topics that we have yet to be able to cover, such as security, privacy, device optimisation, user comfort and form factor, etc. As the developed systems are early prototyping versions, our user study is limited to evaluating the performance with the medical gold standards and partially investigating users' acceptability. Furthermore, we could integrate feedback components into the proposed sensing system to prevent life-threatening conditions. A straightforward example is to warn the users or caregivers when negative events, e.g., microsleeps, seizures, or abnormal blood pressure measurements, are detected. The mobile application or wearable device could produce warnings by playing sounds or vibrations. A more sophisticated direction is stimulating the nervous systems with electrical pulses or optical signals to promote wakefulness [31], relieve seizure activities [32], or treat hypertension [33]. We reserve a more comprehensive study on these exciting topics in future research.

1.5 Publications

The main contributions of this thesis have been published at the following conferences and journals. I have highlighted my personal and my collaborators' contributions below.

1. "Detection of Microsleep Events with a Behind-the-ear Wearable System." **Nhat Pham**, Tuan Dinh, Taeho Kim, Zohreh Raghebi, Nam Bui, Hoang Truong, Tuan Nguyen, Farnoush Banaei-Kashani, Ann Halbower, Thang Dinh, Phuc Nguyen, and Tam Vu.

IEEE TMC 2021 - IEEE Transactions on Mobile Computing *(IF: 5.577)*.
This paper presents a comprehensive study on our proposed ear-worn biosignal sensing wearable systems, the ability to robustly capture head-based biosignal, and deep learning algorithms to detect microsleep events. I am the lead author and the main contributor to the ideas, sensing hardware development, signal processing implementation, experiments, and writing of the paper. My collaborators help with machine learning development, data collection, and microsleep labelling. This work will be presented in Chapter 3 of the thesis.

2. "WAKE: A Behind-the-ear Wearable System for Microsleep Detection."
Nhat Pham, Tuan Dinh, Zohreh Raghebi, Taeho Kim, Nam Bui, Phuc Nguyen, Hoang Truong, Farnoush Banaei-Kashani, Ann Halbower, Thang Dinh, and Tam Vu.
ACM MobiSys 2020 - The 18th ACM International Conference on Mobile Systems, Applications, and Services. *(Acceptance Ratio: 19.4%)*
This paper proposes a novel ear-worn biosignal sensing wearable system by identifying the behind-the-ear locations and devising a three-fold cascaded amplifying technique on the hardware to capture multiple head-based biosignals with minimal number of sensors reliably. Similarly, I am the lead author and the main contributor to the ideas and experiments of the paper. My collaborators help with machine learning algorithms, data collection and labelling. This work will be presented in Chapter 3 of the thesis.

3. "Demo: Earable - An Ear-Worn Biosignal Sensing Platform for Cognitive State Monitoring and Human-Computer Interaction."
Nhat Pham, Taeho Kim, Frederick M Thayer, Anh Nguyen, and Tam Vu.
ACM MobiSys 2019 - The 17th ACM International Conference on Mobile Systems, Applications, and Services.
This paper demonstrates the feasibility of our developed ear-worn biosignal sensing in cognitive state monitoring and human-computer interaction use cases such as relaxation improvement and hand-free drone control by using head-based biosignal. I am the lead author and the main contributor to the ideas and experiments of the paper. This work will be discussed in Chapter 3 of the thesis.

4. "PROS: an Efficient Pattern-Driven Compressive Sensing Framework for Low-Power Biopotential-based Wearables with On-chip Intelligence."

Nhat Pham, Hong Jia, Minh Tran, Tuan Dinh, Nam Bui, Young Kwon, Dong Ma, Phuc Nguyen, Cecilia Mascolo, and Tam Vu.
ACM MobiCom 2022 - The 28th ACM International Conference on Mobile Computing and Networking. *(Acceptance Ratio: 17.8%)*
This paper proposes an efficient compressive sensing framework to enable long-term biosignal monitoring on low-power wearable devices by introducing tiny on-chip pattern recognition models, novel pattern-driven compressive sensing techniques and employing state-of-the-art hardware and OS optimisations. I am the lead author and the main contributor to the ideas, algorithm design and implementation, experiments, and writing of the paper. My collaborators contribute partially to the efficient machine learning design, data processing, and paper writing. This work will be presented in Chapter 4 of the thesis.

5. "eBP: Frequent and comfortable blood pressure monitoring from inside human's ears."
   Nam Bui, **Nhat Pham**, Jessica Barnitz, Phuc Nguyen, Hoang Truong, Taeho Kim, Anh Nguyen, Zhanan Zou, Nicholas Farrow, J. Xiao, Robin Deterding, Thang Dinh and Tam Vu.
   ACM CACM 2021 - Communications of the ACM Journal, 8/2021. *(Research Highlights article.)*
   This paper presents an ear-worn wearable system to monitor blood pressure from in-ear arteries. I contribute to the sensing hardware and signal processing development, experiments, and writing of the manuscript. This work is discussed in Chapter 5 of the thesis.

6. "eBP: A Wearable System For Frequent and Comfortable Blood Pressure Monitoring."
   Nam Bui, **Nhat Pham**, Jessica Barnitz, Phuc Nguyen, Hoang Truong, Taeho Kim, Anh Nguyen, Zhanan Zou, Nicholas Farrow, J. Xiao, Robin Deterding, Thang Dinh and Tam Vu.
   ACM MobiCom 2019 - The 25th ACM Intl' Conf. on Mobile Computing and Networking. *(Acceptance Ratio: 16.1%)*
   Best Paper Award, Communication of the ACM, ACM SIGMOBILE, ACM GetMobile Research Highlights 2021, 2020, 2019.
   This paper proposes a wearable system that can sense blood vessel activities inside the ear and use the captured signals to enable comfortable and frequent

blood pressure monitoring. In this project, I am the sensing hardware development lead and the second author of the paper. I contribute to idea formulation, wearable and sensing hardware design to sense in-ear PPG signal and pressure, implementation, evaluations, and partially to the paper writing. This work is discussed in Chapter 5 of the thesis.

7. "eBP: Frequent and comfortable blood pressure monitoring from inside human's ears."
   Nam Bui, **Nhat Pham**, Hoang Truong, Phuc Nguyen, J. Xiao, Robin Deterding, and Tam Vu.
   ACM GetMobile 2019 - Mobile Computing and Communications, 12/2019.
   This magazine article discusses the concept of blood pressure measurement with in-ear arteries. I contribute to the sensing hardware and signal processing development, experiments, and writing of the manuscript. This work is discussed in Chapter 5 of the thesis.

The work in this thesis also contributed to the following papers and submitted manuscripts:

8. "Painometry: Wearable and Objective Quantification System for Acute Postoperative Pain."
   H. Truong, N. Bui, Z. Raghebi, M. Ceko, **N. Pham**, P. Nguyen, A. Nguyen, T. Kim, K. Siegfried, E. Stene, T. Tvrdy, L. Weinman, T. Payne, D. Burke, T. Dinh, S. D'Mello, F. Banaei-Kashani, T. Wager, P. Goldstein, and T. Vu.
   ACM MobiSys 2020 - The 18th ACM International Conference on Mobile Systems, Applications, and Services. *(Acceptance Ratio: 19.4%)*
   A head-worn wearable system to quantify pain is proposed in this paper. The work of the thesis partly inspires the sensing hardware and techniques to capture head-based biosignals reliably.

9. "Epileptic Seizure Detection and Experimental Treatment: A Review."
   Taeho Kim, Phuc Nguyen, **Nhat Pham**, Nam Bui, Hoang Truong, Sangtae Ha, Tam Vu
   Frontiers in Neurology 2020, Jul. 2020.
   This paper presents a survey for epileptic seizure detection and treatments. The development of head-worn biosignal wearables of the thesis also contributes to the study as a potential solution for long-term seizure monitoring applications.

10. "A Large-Scale Study of a Sleep Tracking and Improving Device with Closed-loop and Personalized Real-time Acoustic Stimulation."
    A. Nguyen, G. Pogoncheff, B. Dong, N. Bui, H. Truong, **N. Pham**, L. Nguyen, S. Ha, T. Vu.
    Submitted to **Science Translational Medicine**.
    This paper presents a large-scale study of a biosignal wearable system to monitor and improve sleep. The work of this thesis inspires the biosignal sensing hardware and the technique to capture head-worn biosignals robustly.

11. "UR2M: Uncertainty and Resource-aware Wearable Event Detection on Microcontrollers."
    Hong Jia, Young Kwon, Dong Ma, **Nhat Pham**, Lorena Qendro, Tam Vu, Cecilia Mascolo.
    Submitted to **MLSys 2023** - Sixth Conference on Machine Learning and Systems.
    In this paper, an uncertainty and resource-aware event detection framework for low-power microcontrollers is proposed through the ideas of cascaded and shared layers, which are partly inspired by the work of this thesis.

1.6 Thesis Structure

The rest of the thesis is organised as follows. Chapter 2 will provide an overview of background materials and related literature. Chapter 3 will present a wearable system called WAKE, techniques to reliably capture multiple head-based biosignals originating from the brain, eye movements, facial muscle contractions, and sweat gland activities, and its application in microsleep detection. Chapter 4 proposes a pattern-driven compressive sensing framework (PROS) that enables highly efficient biosignal sensing wearables for long-term monitoring applications such as microsleep and seizure detection. In Chapter 5, we will we will further explore the feasibility of capturing the activities of a blood vessel (i.e., superficial temporal artery) lying deep inside the user's ear using an ear-worn wearable computer enable comfortable and frequent blood pressure monitoring. Finally, Chapter 6 will conclude the thesis and discuss the future work for this research direction.

Chapter 2

Background

This chapter will introduce the background and the importance of microsleep detection, long-term biosignal monitoring, and blood pressure measurement, which motivates our work. We will also discuss the related methods and the gaps in current literature that this thesis could contribute to.

2.1 Behind-the-ear biosignal sensing and microsleep detection

2.1.1 Microsleep detection application

More than 65 million people in the U.K. and the U.S. suffer from Excessive Daytime Sleepiness (EDS) due to sleep deprivation, obstructive sleep apnea, and narcolepsy [34, 35]. EDS often results in frequent lapses in awareness of the environment (i.e. microsleeps). Healthy people with sleep deprivation usually experience microsleep [34]. Shift workers, nighttime security guards, and navy sailors with sleep problems have a 1.6x higher risk of being injured, causing 13% of all work injuries [36]. Sleepy drivers are at a 3x higher risk of an accident, causing one in five fatal car crashes [37]. People with sleep apnea also suffer from microsleep. The microsleep issue due to sleep apnea alone leads to a loss of nearly $150 million every year due to daily work performance reduction and vehicle accidents [38]. Additionally, more than half of Narcoleptic people are unemployed because of uncontrollable microsleep [39]. They often use Amphetamines to keep themselves awake, resulting in many drug overdose cases [40]. Combined, the sleepiness problem of drivers and the workforce costs the U.K. and U.S. economies more than $411 billion annually [41, 35]. As a result, it is an urgent need for an unobtrusive, reliable, and socially acceptable microsleep detection solution throughout the day.

2.1.2 Understanding microsleep

The Orexin system is a wakefulness network throughout the whole central nervous system, as illustrated in Fig. 2.1. It promotes neuron activity in the mid-brain, the cerebrum, and the visual cortex. These neuronal activities are represented through brain waves, such as fast Beta (β) and Alpha (α) waves while the brain is wakeful and conscious and the slow Theta (θ) waves when the brain experiences sleepiness. Furthermore, studies on animals [42] have shown that Orexin neurons modulate pupil size, eyelid position, and possibly convergence and eye alignment via motoneurons of multiple muscle fibres. As a result, the wakefulness state is also represented by the movements and activities of the eyes. Additionally, several studies [43] have shown that Orexin regulates wakefulness in the autonomic nervous system (ANS) by activating the ANS through projections to the ventrolateral medulla (VLM) and spinal cord, causing the inhibition of sleep. The changes in sympathetic tone are, in turn, represented by changes in facial muscles and sweat gland activity.

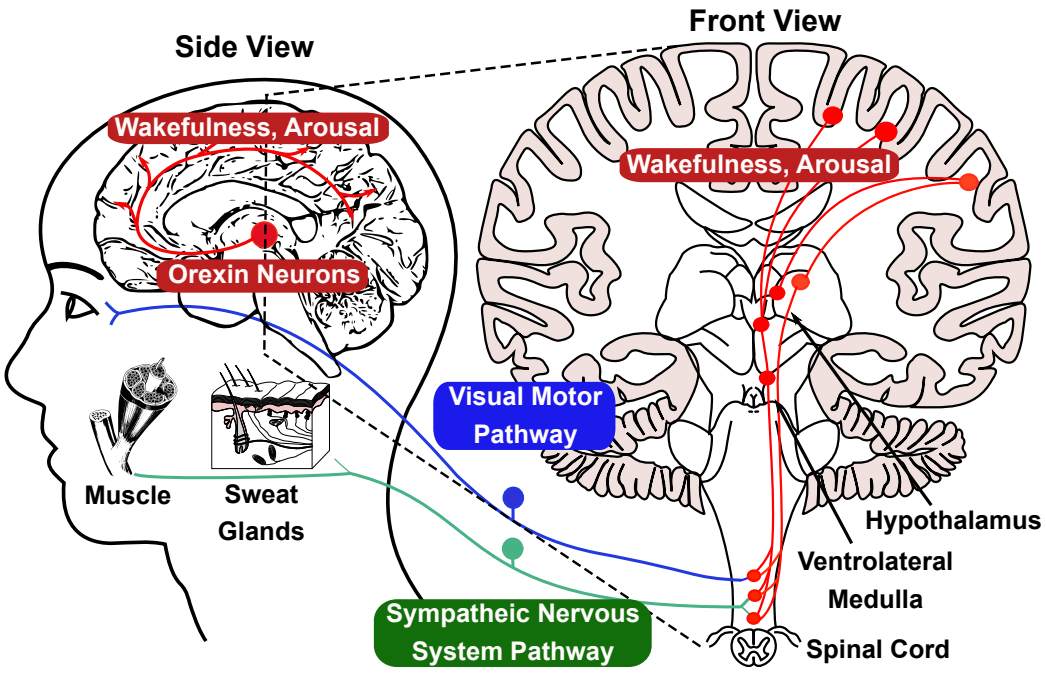


Figure 2.1: Wakefulness neural pathway.

Microsleep is the temporary episode of losing consciousness and is the key to capturing the transition from wakefulness to sleep. A microsleep episode can last from a few to 30 seconds, and people can still wake up after an episode. Microsleep manifests itself both behaviorally (slow-rolling eyes, gradual eye-lid closure, head nods [44]) and electrically (shift in electroencephalography (EEG) from fast α and

β waves to slower θ activities [45]). These manifestations link to the inhibition of the Orexin system. Microsleep is extremely dangerous for tasks requiring constant awareness since people who experience MS are usually unaware of them and still believe that they are awake the whole time [46]. This often happens with people who have excessive daytime sleepiness (EDS).

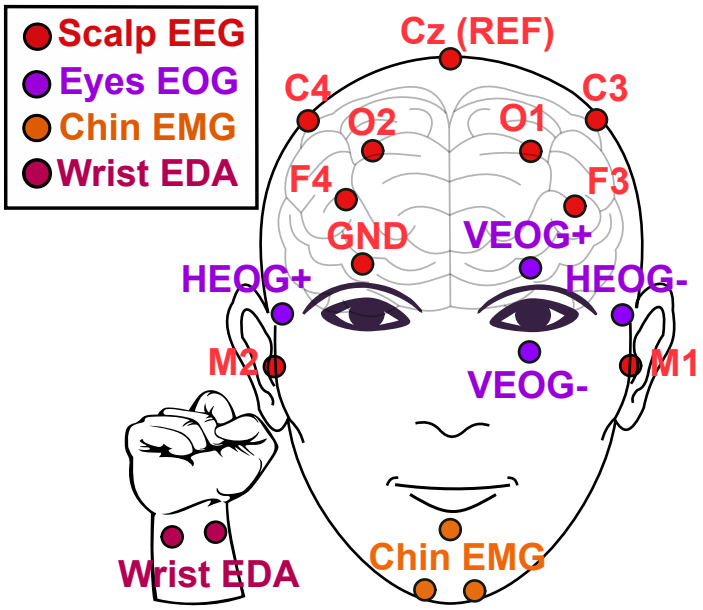


Figure 2.2: Standard EEG, EOG, EMG, and EDA electrode locations.

Conventionally, the need for placing multiple sensors on the user's head to capture different biomarkers for accurate microsleep detection makes it challenging to build a wearable and socially acceptable system. As illustrated in Fig. 2.2, several electrodes (e.g. at least 9 in the standard polysomnography (PSG) system [47]) are usually needed to be placed on the user's scalp to capture brain waves. A wearable camera or 2-4 biopotential electrodes can be placed on the user's eyes to capture eye movements. To capture facial muscle contractions, electrodes are placed on the user's chin. Lastly, sweat gland activity is often captured by electrodes on the wrist or the fingers. With this amount of sensors at different locations on the user's head and face, achieving wearability and social acceptability for microsleep detection is not a trivial task. These studies confirm that there are four key bio-markers that we need to capture for microsleep detection. The remaining questions are *(1) where to place the sensors, (2) how many sensors are sufficient, and (3) how the sensors can be made to capture this information?* We will discuss these questions in Chapter 3 of the thesis.

2.1.3 Electroencephalography (EEG)

Discovered by two Nobel-winning physiologists, Hodgkin and Huxley, in 1952, biologically generated electricity is directly linked to the movements of potassium and sodium ions of nerve impulses [48]. Electroencephalography (EEG) is the method to measure the voltage difference between two points on the human scalp. These voltage differences are created by the flow of ions within neurons and glial cells inside the brain. The amplitude of measured voltages could be from a few nV to 100s of uV, depending on location and brain activity. The spectrum of EEG has a range from 0.1 - 100Hz [48, 49]. It is divided into several notable frequency bands, such as Alpha (α), Beta (β), Theta (θ), Delta (δ), and Gamma (γ), representing different cognitive states of the brain. These brain waves are rhythmic in the time domain. In addition to the rhythmic activities, there are several transient EEG patterns, such as spike and sharp waves, vertex waves, and sleep spindles. They can be used to identify epileptic seizure activity or classify sleep stages.

α, β, and θ waves are important brain waves linked to microsleep activities. β waves are fast brain rhythms ranging from 13 to 35Hz. They are often observed in the frontal cortex and when a subject is highly alert. α waves have a frequency range from 8 to 12 Hz. They are dominant in the cerebral cortex's posterior regions (e.g., occipital and parietal lobes). They appear when an adult subject is relaxed but still conscious. θ waves have a relatively lower frequency band than α and β, i.e., 4 - 8Hz. θ rhythms appear when the subject falls asleep and loses consciousness. Strong oscillations of θ waves could be observed in the hippocampus. The suppression of α wave and increase in θ activity marks the transition from awake to sleep, where the subject loses awareness of the environment and control over behaviour.

2.1.4 Electrooculography (EOG)

Similar to the brain, the eyes also generate electricity with their activities. The measurement for this electricity is called EOG. At rest, there is a relatively stable voltage potential, called corneo-retinal potential, between the cornea (front of the eye) and the Bruch's membrane on the retina (back of the eye) [50]. The corneo-retinal potential is positive at the cornea and negative at the retina. By placing electrodes either above and below the eye (vertical EOG) or to the left and right of the eye (horizontal EOG), activities of the eye, such as blinks, saccades, opening and closure, lead to changes in the voltage potential across two measurement electrodes. EOG signal typically has an amplitude between 400uV and 1000uV and a frequency range

from 0.1 to 10Hz [51]. EOG plays an essential role in detecting microsleeps. When the sleep pressure increases, the eyeblinks become irregular, and the eyeballs start to roll slowly [52].

2.1.5 Electromyography (EMG)

EMG is a technique to measure the electricity generated by muscle motor units' activities. A motor unit consists of multiple muscle fibres connected to a motor neuron. When the motor unit is activated, a neuron impulse is transmitted to the muscle fibres. A corresponding action potential is evoked in all the connected fibres to contract or relax the muscles [53]. By placing electrodes on the skin surface, we could capture the activities of surrounding muscle groups. EMG signal has a typical amplitude of several millivolts and a frequency spectrum ranging from 0.1Hz to 500Hz [54]. Chin EMG, which measures the facial muscle activities around the chin, is an important biomarker for microsleep. During the transition from wakefulness to sleep, chin EMG power will start to drop, representing the loss of muscle control when a subject becomes unconscious [55].

2.1.6 Electrodermal Activity (EDA)

EDA is the measurement of skin conductance by applying a small, constant DC voltage across two electrodes. EDA signal varies with the activity of sweat glands where electrodes are placed. As the sympathetic nervous system controls sweating, EDA could be used as an indicator of emotional and psychological arousal. When a subject is highly aroused, the increased sweat gland activity will lead to an increase in the EDA signal. EDA could be captured at locations with a high density of sweat glands, such as on the palms of hands, fingertips, and soles of the feet. The back of the ear is also a potential location for EDA measurement due to a high level of sweating activity [56]. EDA has two components, i.e., phasic and tonic. The phasic components are fast-changing activities that respond to external stimuli. On the other hand, the tonic component is the slow-changing electrodermal level that represents general arousal and alertness level [57].

2.1.7 Motion artefacts

Motion artefact is a significant roadblock that prevents practical usage of biosignal sensing systems. EEG and EOG signals are heavily affected by motion due to their

low amplitude (i.e., microvolts). There are two main sources of motion, i.e., (1) micro-movements of the sensing electrodes and (2) triboelectric effects on the measurement cables. The electrode-skin contact and the connection between sensing electrodes to the readout circuit could be modelled as a series of capacitors and resistors [58]. During motion, fluctuations in this electrical pathway result in unwanted voltage potentials that superimpose with the biosignals. After analogue-to-digital conversion, motion artefacts are not distinguishable from meaningful signals. Thus, various methods are needed on multiple components of the sensing systems, including electrodes, readout circuits, and signal processing, to be effective [59]. Software methods such as Independent Components Analysis (ICA) and Blind Source Separation (BSS) are often employed to mitigate motion artefacts. However, the requirements for real-time processing and a minimal number of sensors make it difficult to employ these methods on wearable devices [60].

2.1.8 Environmental noise

Noise coupling from the environment is another critical issue for biosignal sensing systems. Environmental noise could come from various sources, such as main power lines, power transformers, motors, etc. Almost all electrical equipment connected to the main power line has a power transformer. These transformers convert high-voltage alternative current (120 or 240 VAC) to low-voltage current that could be used by the equipment. During operation, the alternative magnetic field generated by transformers could easily interfere with biosignal sensing circuits creating unpredictable noise. Similarly, electric motors also generate strong magnetic fields that could induce tiny noise currents inside a sensing circuit [53]. The main power line is another significant source of environmental noise. Depending on the country, these power lines are typically 120/240 VAC at 50/60Hz. This AC couples into the measurement through capacitive coupling where the human body acts as a coupling medium [61]. Since the main power line is not perfectly sinusoidal, we often observe strong power of harmonics, i.e., up to 5^{th}-order from the fundamental frequency [53]. Since environmental noise could be multiple orders of magnitude stronger than biosignals, differential measurement is employed to eliminate the common noise and extract small differential biosignals. Common-mode rejection ratio (CMRR) is the metric to quantify the ability to suppress common noise of a measurement system [62].

2.1.9 Related work

Existing microsleep detection systems mainly use scalp EEG, eye tracking with EOG or cameras, IMU, and infrared light.

Video-Polysomnography. The Maintenance of Wakefulness Test (MWT) using Video-Polysomnography (EEG, EOG, EMG, and facial videos) [63] is the medical gold standard to quantify microsleep based on the electrical signals from the human head, such as brain waves, eyes ball movements, chin muscle tone, and behaviours including eyelid closure, eye blinks, and head nods. This method requires a complicated setup performed by trained technicians in a controlled clinical environment. Thus, it is not applicable for daily usage.

Camera-based solutions. Using cameras is another solution to detect microsleep. This approach is the most affordable and common method to detect microsleep for drivers [64, 65]. The camera-based approach only captures the outer reflection of sleepiness, such as eyelid closure and head nods and ignores the other physiological signatures of sleepiness (e.g., brain and muscle activities) [66]. Thus, it cannot capture microsleep reliably if the episode happens while the subject's eyes still open, which often occurs [67]. Furthermore, using cameras raises strong privacy concerns [68]. Besides, cameras are also limited by environmental light conditions. While wearable cameras can address this issue, wearing a camera on the face is not socially acceptable in daily use. For example, the mixed criticisms of Google Glass on its privacy [69], and form-factor [70] have shown that wearable cameras are not easily accepted by the public.

Sensing-based solutions. EEG and EOG signals have been widely used to detect microsleep [71, 72]. However, the conventional devices used to capture those signals can only be used in a controlled environment and are not socially acceptable. IMU sensors can be used to approximate body motion (e.g., head nods) corresponding to microsleep (smartwatch [73], hairband [65]). It is challenging to distinguish between normal body motion like agree-nodding and body motion due to microsleep. In addition, infrared light reflection methods monitor the eyelid movement of the subject, such as Vigo [74] and BlinQ [75]. These devices cannot recognise the inner physiological state, and their reliability has not been thoroughly evaluated.

Drowsiness vs. Microsleep detection. In literature, there are several drowsiness detection and monitoring works such as [76, 77, 78]. Drowsiness and microsleep detection works, however, should not be treated equally. In particular, drowsiness is a physiological state defined when there exists sleep pressure, which may cause slower

reaction time or compromise vision but does not mean fatal as our brain is still conscious of the surrounding environment [66]. Microsleep, on the other hand, is the brief and often fatal duration in which the brain loses consciousness [46]. Additionally, drowsiness detection works (especially in driving scenarios) use various methods to quantify the existence of drowsiness, such as steering pattern monitoring, vehicle position in lane monitoring, driver eye/face monitoring, etc. [76]. They are different from the ones (i.e. brain activity represents microsleep periods) used for microsleep. Particularly, in [77], the authors detect drowsiness by quantifying wake/sleep epochs; however, they also state that their system is not sensitive enough to detect microsleep events. In [78], lane deviation in a driving simulation, which is pointed out in [76] that it is not a reliable metric, is used as the indicator of drowsiness.

Ear-EEG Systems. In-ear sensing system for EEG signals is particularly useful for sleep monitoring like LIBS [79], Hearables [80], Ear-EEG [81]. However, wearing the device inside the ear might impact hearing capability and the user's daily activities. cEEGrids [82] and Gu et al. [83] compared scalp EEG with behind-the-ear (BTE) EEG and demonstrated that EEG data could be recorded behind the ear. However, the ability to detect microsleep from BTE has yet to be evaluated. Vital signs monitoring using wearable and mobile sensors has also been investigated in various studies such as breathing measurement [84], tongue-teeth localization (EEG, EMG, and skin surface deformation) [85], stress estimation (heart rate variability (HRV), galvanic skin response, and EMG) [86], mental health management (EEG, hemoencephalography, and HRV) [87], and eating detection (IMU, microphone, and proximity sensor) [88, 89]. However, to the best of our knowledge, there are no existing works to detect microsleep from wearable BTE sensors accurately and reliably.

2.2 Long-term head-based biosignal monitoring on low-power wearables

2.2.1 The importance of long-term biosignal monitoring.

The wearable healthcare market has been experiencing significant growth in recent years, and it is predicted that healthcare wearable devices will be the next generation of personal telemedicine practice. This is especially important for patients with chronic diseases and after surgery, where constant monitoring is essential to prevent fatalities [90]. However, many wearable-enabled healthcare applications have not been

deployed due to limited battery lifetime, slow response rate, and inadequate biosignal quality.

The trade-off between signal fidelity, response time, and battery life is a long-standing challenge for wearable devices [91, 92]. In many healthcare applications, the wearable usually takes the role of data collecting device due to their limited energy and computing resources [93]. The collected data are transmitted to nearby mobile devices through wireless communications (e.g., Bluetooth, WiFi) to predict emergency events or upload to users' healthcare providers for further diagnosis. Though maintaining the collected signal fidelity is crucial [94], continuous wireless communication has a high cost on the battery life [95]. E.g., Bluetooth could consume up to several mWs [96], while WiFi could go as high as 10s of mW [97], depending on the data rate. As a result, many healthcare wearables have to reduce signal quality (i.e., by lowering data rate) and increase response latency (i.e., by increasing communication intervals) to improve battery lifetime [95].

2.2.2 Sparsity in biosignal monitoring

As biosignal events are often intermittent, continuously monitoring these events results in wasted energy, computing power, and memory. In this project, we consider events that are associated with EEG, EOG, and EMG, but the proposed solution would generally be applicable to other biosignals in multiple application domains.

Event Sparsity. We observe that the events of interest (e.g., seizures, microsleeps, etc.) are important but rarely happen. Several studies have reported that these events only occur less than 5% of the signal duration [98]. Thus, detecting these events on the device could help to cut a significant amount of energy needed to stream the signals out. However, detecting these events requires multiple signal modalities (i.e., EEG, EOG, EMG, etc.) and a complex algorithm, making it challenging to implement on resource-constrained devices. Our intuition is that we could decompose these complex events into smaller and generic patterns of interest (PoIs). For example, an epileptic seizure waveform could consist of EEG spike/polyspike and slow-wave (focal/generalised non-specific seizures), 3-Hz spike-and-wave discharges (absence seizures), and stiffing and convulsion patterns (tonic-clonic seizures). Similarly, we can decompose a microsleep event into alpha, theta wave, slow eye movements, and muscle contraction patterns on the EEG, EOG, and EMG signals. Thus, it is feasible to detect these patterns directly on the device with an efficient pattern recognition technique.

Signal Sparsity We also observe that the sparsity property also presents at the signal level. While biosignals are known to be non-sparse in time or frequency domains, they could have sparse representations in other domains (e.g., wavelets). Thus, we do not need all the collected samples to reconstruct the signal. The compressive sensing (CS) theory has been developed to exploit signal sparsity. It states that the number of signal measurements depends on inherent information contained in the signal and is much lower than the Nyquist rate [95]. The effectiveness of CS relies directly on finding a reliable domain with high sparsity. However, this is still an open challenge for non-stationary biosignals [99].

From these observations, we hypothesise that **by exploiting both event and signal sparsity, the amount of data reduction could be significant**, leading to a highly energy-efficient system. However, we must take great care in designing such a system. With the constrained computing resources of wearable devices, any additional energy spent on complicated algorithms could easily outweigh any benefits from the reduced wireless transmission.

The remaining questions are *(1) How can we develop the pattern detection models so that they can be both accurate and efficient? (2) How can we devise a compressive sensing method that could achieve both low sampling rate while maintaining high signal fidelity?* and *(3) How can we optimise the system to ensure the efficiency of additional computation?* We will discuss these questions in Chapter 4 of this thesis.

2.2.3 Low-power Microcontrollers

With the emergence of small-scale wearable devices, low-power microcontrollers (MCUs) play a critical role in providing computational power with low power consumption. Compared with other computing hardware, including mobile devices, wearable devices have more constraints towards MCU, including low battery capacity and limited processing resources. Unfortunately, health monitoring applications often require wearable devices to provide high-fidelity signals and a long operating time, which contradicts the abovementioned constraints. As wearable devices are almost unusable while charging, the low-power operation is an essential feature of these devices. Additionally, improvements in MCU resources do not follow Moore's law as stated in [100]. It points out that processing speed and RAM capacity has only increased from 8MHz to 48MHz and 10kB to 64kB, respectively, within almost a decade. Thus, computational efficiency is another vital factor for algorithms designed for wearables.

2.2.4 Pattern Recognition with Neural Networks

Pattern recognition helps identify similarities and regularities within the sensory data to output perceivable results. Conventionally, feature engineering and classification were commonly used to apply in both frequency and time domains and search for groups with the same characteristics. Then, conventional learning techniques such as Support Vector Machine (SVM), K-nearest neighbours, Random Forest, etc., will classify these features into appropriate groups. Recently, deep neural networks have achieved significant performance in many fields, leading to an increased focus on applying these methods in a sensory data stream. Prior works such as SqueezeNets [101], ShuffleNets [102], and MobileNets [103, 104] designed for resource-constrained devices showed the state-of-the-art performance in many applications with low resource overheads. However, those architectures are designed for "mobile"-level networks consisting of at least a few million parameters (a few MB), exceeding the extremely limited memory and storage budgets (tens of KB) of MCUs [105]. Therefore, in our work, we search for the lightweight MobileNet architecture that fits into the limited memory and storage requirements of our target MCUs while also showing good performance and efficiency.

2.2.5 Tiny Machine Learning

Popular ML frameworks such as PyTorch [106] and TensorFlow [107] are impossible to operate on tiny MCUs due to their high memory requirements. Hence, several ML frameworks designed for MCUs are developed to fill this void [108]. First, there exists a code generation approach that generates codes optimized for target hardware by taking model definition as input. For example, uTensor [109], TinyEngine [110], Embedded Learning Library (ELL) [111], STM32Cube.AI [112] are introduced. Although this approach shows promising results, it suffers from non-portability between different hardware platforms. Second, there is an interpreter-based approach. For example, TensorFlow Lite for Microcontrollers (TFLM) [113] uses an interpreter for executing models on MCUs. TFLM is highly portable compared to code generation approaches, and thus we adopt to use of TFLM in our work.

2.2.6 Compressive Sensing

The Nyquist–Shannon sampling theorem assisted signal processing in many applications; however, with the digital revolution, the signals based on this theory are bulky with too many samples, leading to significant burdens on storing, processing, and

communicating the captured data [114, 115]. A compression method called Compressive Sensing (CS) was developed and applied to signals and images. The data are reshaped and saved in a much smaller size. When needed, compressed data was decompressed to the original size. CS samples the signal by a much smaller number of samples than the Nyquist–Shannon theorem, which requires at least twice the highest frequency content of the signal [114, 115], by measuring minimum samples containing full information of the signals. CS method can be used in a wide range of applications, such as signal processing, biosignal processing, medical image processing [116], data mining [117], telecommunications [117], etc. [118]. However, CS contains two main challenges in wearable devices: (1) the signal variations among channels, subjects, and trials, and (2) finding an optimal sparse domain is non-trivial due to the variations among various biosignal patterns [118].

2.2.7 Related work

Machine learning on microcontrollers. Tiny Machine Learning [119], TinyML, provides the solution to bring powerful deep learning models into extremely resource-constrained devices such as microcontrollers. Various approaches [120, 121, 122] have been proposed to optimise and achieve a compact design that satisfies the extreme hardware constraints. MCUNet [120] mitigates the memory bottleneck of the CNN architecture with the patch-based inference approach, which only gradually loads small parts of the network into memory. FANN-MCU [121] provides toolkits for building energy-efficient networks on MCUs by employing multi-layer perceptrons and code generation. YONO [122] enables the operation of a multi-task system on an MCU by compressing multiple heterogeneous models and improving efficiency through in-memory operations. Unlike the prior works, our work further considers biosignals' sparsity and locality properties to facilitate system efficiency.

Compressive sensing in healthcare. Compressive sensing (CS) has been applied in processing biological data in mobile healthcare and telemonitoring to provide a faster, more accurate, and more energy-efficient system [123, 124, 125, 126]. In [124], the Discrete Cosine Transform domain is employed to compress the EEG signals and develop a classifier to detect seizure events. Similarly, the authors in [125] and [126] perform compressive sensing on electrocardiogram (ECG) signals by assuming there is a specific sparse domain such as Daubechies 2 or principal component analysis (PCA). However, it has been pointed out in [127, 128] that the use of a fixed sparse domain for biosignals reduces the compression effectiveness and creates large variations in reconstruction quality. We leveraged the advantage of CS and the new ability

to detect patterns directly on our hardware design to propose an efficient biosignal monitoring system for low-power wearable devices.

Low-power wearable platforms. Power consumption is a major concern in any wearable or Internet-of-Things devices. Thus, it has attracted much attention in recent platforms and OSes. Amulet [129], Mindo [130], Convergence [131], RIOT-OS [132], TinyOS [133], FreeRTOS [134] focused on leveraging the event-driven scheduling and low power modes to reduce the energy consumed by an MCU and others high-power components. While existing systems provide hardware or OS optimisations, they have not considered the sparsity of captured signals, which is our main contribution.

Commercialised biosensing platforms. There are several commercialized biosensing platforms on the market, such as Emotiv Epoch [23], Muse [26], and Neurosky MindWave [24]. They are equipped with 250-600 mAh batteries and last from 5 to 9 hours of continuous streaming. Since none of them provides compression or recognition ability, they rely on conventional methods such as downsampling or proprietary wireless protocols to prolong battery lifetime. In our PROS framework, we take advantage of on-chip intelligence and compressive sensing to boost energy efficiency further and enable real-time responses on the devices.

2.3 In-ear blood vessels activity and pressure monitoring.

2.3.1 Frequent blood pressure measurement

Blood pressure (BP) is one of the foremost vital signs measured when patients first arrive in the hospital, as BP can provide doctors with insight to initiate their diagnosis. For example, chronic kidney disease, sleep apnea, and adrenal and thyroid disorders can all cause high BP, while low BP indicates the possibility of heart or endocrine problems, dehydration, severe infection, or even blood loss. Additionally, uncontrolled elevated BP is a major symptom of many life-threatening diseases, such as hypertension, heart failure or stroke [135]. Until recently, the reliable way to measure BP was done by a health care practitioner (HCP) such as a physician, doctor, or nurse. The clinician wraps an arm cuff around the patient's upper arm and rapidly inflates the cuff with air. Once the cuff has reached maximum inflation, the systolic and diastolic pressures are determined by slowly releasing the air from the cuff and observing the pulse sound with a stethoscope over the brachial artery below the cuff.

Since the invention of digital BP devices, non-medical trained users can self-measure their BP at home, as an acoustic sensor can replace the stethoscope, and a pressure sensor with a DC pump can substitute the pressure gauge and hand pump.

However, these devices often cause discomfort and inconvenience for those who need **frequent BP monitoring**, such as hemodialysis (kidney failure) patients [136], individuals with undiagnosed white coat hypertension or undiagnosed masked hypertension [137], which have a prevalence of 15-30% [138] and 16.8% [139] in the US, respectively. There is also increased use of frequent BP monitoring for post-operative organ transplant recipients [140], with more than 30,000 solid organ transplants occurring every year [141]. In such cases, BP is measured every 30 minutes for 24 hours [142], while each hemodialysis session takes around four hours. Therefore, there is a significant need for an unobtrusive and comfortable BP monitoring approach.

2.3.2 Fundamental of BP measurement

Starting with a brief overview of existing blood pressure (BP) monitoring widely used today, we will point out current limitations, providing context for our novel approach. Measuring blood flow pressure can be done with both invasive and non-invasive methods. Although invasive approaches deliver highly accurate results, it is costly and only available in hospitals. Non-invasive techniques are far more useable, as their process is quick, low cost, and relatively simple. However, these non-invasive techniques are set to cycle or must be done manually, unlike continuous measurements provided by invasive arterial methods.

Artery deformation under the effects of cuff pressure grants the key signatures to estimate BP, as shown in Fig. 2.3. When the cuff pressure is equal to the systolic pressure (SBP), blood flow continues through the occluded artery, but only the highest arterial pressure can be detected. On the other hand, if the cuff pressure is lower than the diastolic pressure (DBP), the detected pulse is very weak. Using auscultatory methods, medical practitioners listen to pulse sound propagation through a stethoscope to determine BP. Oscillatory, on the other hand, was developed for the digital device by estimating BP from the change of pulse amplitude. It detects the Maximum Pulse Amplitude (MPA) A_M first and applies predefined fractions of the peak amplitude ratio A_M/A_S and A_M/A_D to detect where the systolic and diastolic pressure occurs and uses these values to infer the pressure. A_S and A_D are the amplitude of systole and diastole, respectively. Unlike auscultatory methods, oscillatory methods do not need to completely occlude the blood vessel in order to detect the

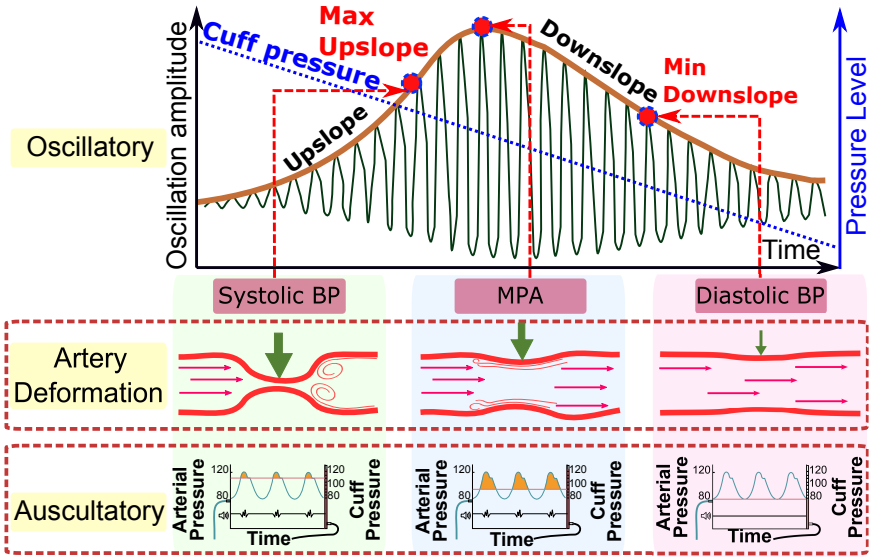


Figure 2.3: The response of blood artery to outer pressure that causes the measurement of BP.

systolic BP [143], which is well-suited for our balloon model. However, current oscillation ratios only apply to the arm or wrist BP measurement model. Therefore, they are not eligible for our in-ear case. Generating a new in-ear ratio requires a large-scale data set, including an invasive method to measure BP from inside the ear, which is infeasible. Instead, we propose a technique to measure BP without applying the characteristic ratios. To achieve this goal, we thoroughly examine the change of amplitude with respect to the change of cuff pressure. Then, we extract the key properties and formulate them into mathematical equations for processing. According to [144], during the deflation:

- Pulse amplitude increases when the cuff pressure is close to the systolic level. The increment increases more quickly when the pressure reaches and passes through the systolic point.
- At the systolic and diastolic cycle cross-section, the amplitude obtains its highest value (the MPA).
- Amplitude rapidly decreases once the pressure passes the MPA and moderately decreases once it reaches the diastole point. In other words, the DBP position occurs at the highest decreasing amplitude.

These observations provide key insights for composing the solutions to detect MPA, SBP, and DBP. In particular, the diastolic position is the minimum of the downslope

amplitude, and MPA is the peak of the amplitude, as shown in Fig. 2.3. We can derive the systolic location as the maximum upslope amplitude. However, sometimes our in-ear balloon pressure might not reach the systolic phase due to comfort requirements. Therefore, we have to rely on the relational equation between MPA, SBP, and DBP [145]:

$$P_M = \beta P_S + (1 - \beta) P_D, \tag{2.1}$$

where β is the systole ratio of the cardiac cycle and P_M, P_S and P_D are the MPA, SBP, and DBP respectively. Most literature reports β as a fixed value [146, 147] and is widely accepted, but each person can have a slightly different ratio dependent on age, gender, and health condition. Moreover, incorrect estimation of β increases the estimation error, as noticed in [148]. In our eBP system, we propose an adaptive estimation for β based on the pulse-wave form.

Realising the importance of frequent BP monitoring and the discomfort of current devices, we design eBP as ear-worn equipment to (1) capture the pulse signal inside the ear from a balloon-attached pulse sensor and (2) use this information to estimate the BP.

2.3.3 Photoplethysmography (PPG)

Photoplethysmography (PPG) is an optical method to measure blood volume changes in the microvascular bed of tissue [149]. The technique is non-invasive, inexpensive, and widely used in many clinical applications such as vascular diagnostics, pulse oximetry, or blood pressure measurement. A typical PPG system consists of a light source and a photodetector. The light source, e.g., LEDs, emits the light to the tissue bed. The transmitted or reflected light, which reflects blood volume variations, is captured by the photodetector. The current output of the photodetector is converted to digital values by a transimpedance amplifying circuit.

Green and infrared light are typically employed for PPG. Green light is the most popular light source due to its reliability and the ability to capture oxygen absorption in oxyhemoglobin and deoxyhemoglobin. On the other hand, infrared light can penetrate more deeply into the tissue. Thus, it is often used to measure deeply concentrated blood flow in muscles [150]. The digitalised PPG signal has a pulsatile AC component representing the heart rate. It is superimposed by a large DC component that varies slowly with respiration and vasomotor activity. In our blood pressure measurement application, we only consider the AC component of PPG signals.

2.3.4 Related work

Invasive methods. Measuring blood flow pressures can be separated into two categories: invasive and non-invasive. Although invasive approaches are only used in clinical settings when continuous BP monitoring is medically necessary [151], as it is expensive and inconvenient. Furthermore, there is potential for wave reflection at the catheter tip [152] when inserting the catheter into the artery [153]. Thus, expertise is required for invasive methods and therefore used as the ground truth for non-invasive validation [154].

Non-invasive methods. There are several non-invasive methods: auscultation (sphygmomanometry) [155], ultrasound [156], oscillometric [157], photoplethysmographic (PPG) [158], and palpatory [159]. Originally, the mercury sphygmomanometer was the gold standard for official BP measurement. However, the mercury ban has led to an increase in non-mercury devices [160]. Other existing techniques require the blood flow to be blocked completely. Therefore, it causes discomfort when measuring BP frequently. In the case of prolonged dialysis, patients hardly rest because the BP cuff constantly squeezes their arm. The most current BP monitoring devices consist of an arm cuff and a monitor that harnesses around the body, yet this often hinders the wearer's mobility [161]. Therefore, by moving the location of measuring BP to inside the ear, our device has a minimal impact on the users' mobility and comfort.

Cuff-less solutions. There have been prior attempts to build cuff-less, continuous BP monitoring devices [162, 163]. Continuous systems rely on the Pulse Transit Time [164] technique, which infers the BP from the time it takes the pulse to propagate from one point to another. It usually measures the time interval between the peaks of the pulse signal and uses reference peaks from an electrocardiogram (ECG). However, these systems are not accurate due to the low specificity ECG on BP. The recent failure of Quanttus is an example [165]. Another cuff-less approach tried capturing BP by pressing on the phone screen [166]. Though this approach is moderately accurate, it requires a user to maintain constant finger pressure to obtain a good measurement, which is poorly suited for frequent BP monitoring.

Finger-based BP measurements. To measure BP with PPG sensors, the Penaz finger-cuff attaches a PPG sensor under the cuff to detect the arterial pulsation and drive a servo-loop to maintain the equality of external pressure and internal artery pressure [160]. The SBP and DBP are obtained from external oscillations. Yet, this method yields inaccurate BP measurements when there is movement noise [167]. Recently, smartphone-based BP monitoring has increased as most smartphones have a PPG sensor [168]. While convenient, it does require continuous finger pressure. On

the other hand, Wang et al. [169] introduce the combination of a smartphone with built-in cameras and accelerometers to measure BP.

Ear-based BP measurements. Similar to eBP, others studied the ear as a place to measure blood flow pressure, measuring on the earlobe [170] or tragus [171]. Existing studies have used a microphone to detect blood flow [172], a time delay between two ears [173], a sealed air chamber between the end of the earplug and the eardrum [174], and a light-emitting element for irradiating a living body part [175]. Even though these works explain the device design or the system structure, they do not prove the device's performance by comparing it to gold-standard data.

2.4 Summary

This chapter discusses the background and related methods of head-based biosignal sensing techniques. We also highlight the current gaps in state-of-the-art literature, which we will contribute to with this thesis. In the following chapters, we will discuss the proposed wearable systems to enable the ability to reliably and efficiently sense head-based biosignals from the brain, eyes, facial muscles, and blood vessels. Finally, we will present the applications of our proposed systems in microsleep detection, long-term seizure monitoring, and frequent blood pressure measurement.

Chapter 3

WAKE: Sensing head-based biopotentials and microsleep detection with a behind-the-ear wearable

3.1 Introduction

This chapter presents WAKE, a novel behind-the-ear wearable device for microsleep detection. By monitoring biosignals from the brain, eye movements, facial muscle contractions, and sweat gland activities from behind the user's ears, WAKE can detect microsleep with a high temporal resolution. We introduce a Three-fold Cascaded Amplifying (3CA) technique to tame the motion artefacts and environmental noise for capturing high-fidelity signals. Through our prototyping, we show that WAKE can suppress motion and environmental noise in real-time by 9.74-19.47 dB while walking, driving, or staying in different environments, ensuring that the biosignals are captured reliably. We evaluated WAKE using gold-standard devices on 19 sleep-deprived and narcoleptic subjects. The Leave-One-Subject-Out Cross-Validation results show the feasibility of WAKE in microsleep detection on an unseen subject with average precision and recall of 76% and 85%, respectively.

3.1.1 Goals and Challenges

In this research, we explore the challenges of building a novel wearable physiological sensing device, called WAKE, for microsleep detection situated only behind the ears (BTE). WAKE captures the core biomarkers that are directly related to microsleep from the human head, namely brain waves (EEG), eye movements (EOG), facial muscle contraction (EMG), and skin conductivity (EDA), while being light-weighted and socially-acceptable. While WAKE is currently a standalone device, it could be integrated with earphones and headsets, which are already worn daily for listening to music or communication.

Challenges. To realise WAKE, we face the following key challenges: (1) heavy noise created by motion and coupled from the environment in daily use is the **long-standing challenge** limiting the practical uses of wearable biosignal sensing systems, as it is difficult to ensure high fidelity signals; (2) making a wearable, and socially-acceptable device that can capture microsleep is non-trivial because multiple sensors are usually needed to capture its core biomarkers; (3) microsleep detection from behind the ears is an unexplored topic where existing techniques cannot be applied directly; and (4) the BTE biosignals are weak and overlap with each other in the three-orders magnitude range.

Contributions: The key findings and contributions of this work are as follows:

1. We devise a Three-fold Cascaded Amplifying (3CA) hardware technique to make it more practical by ensuring high-fidelity signals while mitigating motion and environmental noise.

2. We identify and localize the minimum number of areas behind human ears where biomarkers from the brain, the eyes, facial muscles, and sweat glands can be captured reliably for microsleep detection.

3. We design and prototype a wearable, compact, and socially acceptable device that can capture multiple head-based physiological signals.

4. Using a wide range of microsleep biomarkers as features, we developed a hybrid model of a hierarchical classification model and EMG-event-based heuristic rule to detect users' microsleep.

5. We evaluate the proposed system using our custom-built prototype on 19 subjects. In Leave-One-Subject-Out Cross Validation (LOSOCV), the system obtains 76% precision and 85% recall, showing the feasibility of microsleep detection of WAKE on an unseen subject.

3.1.2 Impact of environmental noise and motion artefacts

Various noise and artefacts affect a wearable biosignal sensing system. Motion artefacts and electromagnetic interference from the environment are two major roadblocks to the practicality of the system. Several approaches have been proposed to address the issues of artefacts and noise, such as blind source separation with independent component analysis (ICA) or incorporating additional sensors such as inertial measurement units. These approaches, however, depend on a large number of electrodes

to provide spatial information, require significant computation and are challenging to implement in a real-time system [176]. Throughout our in-lab experiments using a PSG device, we found that environmental noise significantly impacts the original signal, while human motion artefacts completely distort the whole signal, making it not even usable. *This requires a novel solution to remove noise from the signals captured from wearable devices.* (Section. 3.4, 3.5)

3.2 Exploring Microsleep biomarkers from behind the ears

As mentioned previously, the ear is the intersection of multiple microsleep biomarker sources (e.g., the brain, the eyes, facial muscles, and sweat glands) and is also a natural harbour point where a wearable device could be worn. While recent work on ear-based biosensing has shown the feasibility of capturing individual biosignals (e.g., EEG [82], EOG [177], EEG/EMG [85], and EDA [86]) from the area around and behind the ears, monitoring microsleep-related biosignals with a wearable form-factor has not been explored before. Thus, it is unclear about (1) *where are the best places for EEG, EOG, EMG and EDA sensors to achieve both wearability and sensing sensitivity*, (2) *what is the minimum number of required electrodes*, and (3) *what are the unique characteristics of BTE signals*?

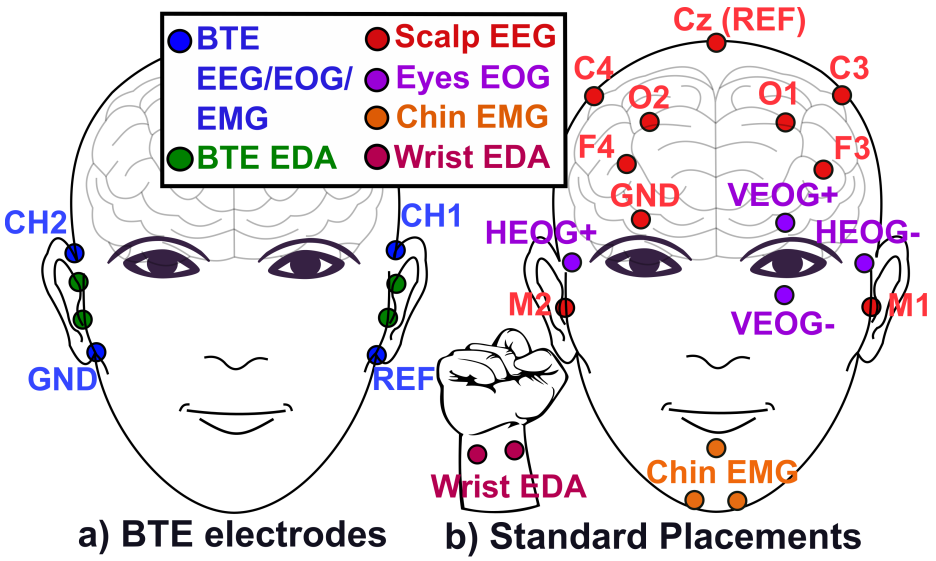


Figure 3.1: BTE vs. standard locations.

The BTE electrodes placements. From our study on the ear anatomy, we derive the best sensor placement locations for microsleep detection, as shown in Fig. 3.1a

and Fig. 3.2. At these locations, we can capture signals coming from the mid-brain area (EEG), eye movements (EOG), facial muscle contractions (EMG), and sweat gland activities (EDA). These sensor locations allow us to design a socially-acceptable wearable device that is well hidden behind the user's ears, just like commercial-off-the-shelf (COTS) earphones.

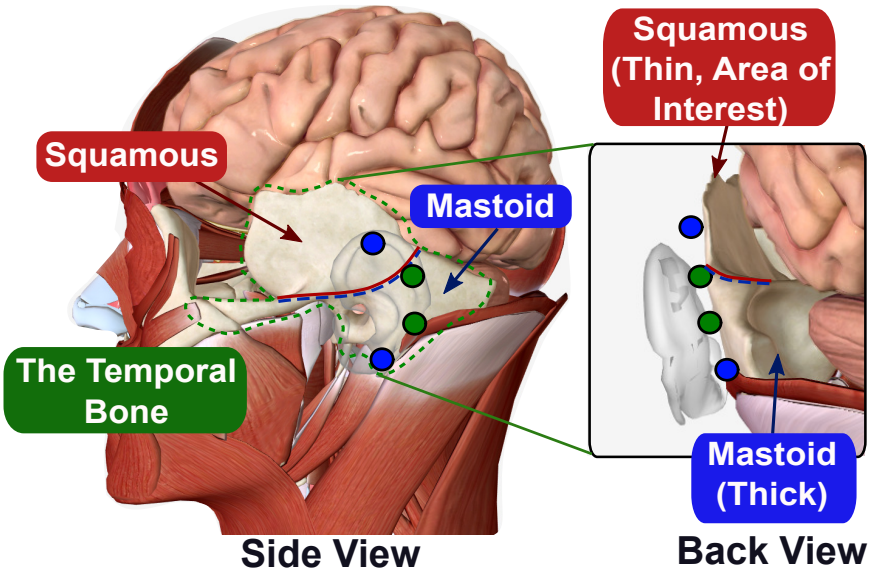


Figure 3.2: BTE anatomy.

Fig. 3.2 illustrates the anatomy of the temporal bone covering the whole BTE area. It consists of two major parts, i.e., the Squamous and Mastoid processes. To capture EEG generated by the mid-brain area, we would want to place the electrodes on the Squamous process, which is the thin upper part of the temporal bone. This makes electrodes as close to the brain as possible. Two electrodes, i.e., channel 1 on the left ear and channel 2 on the right ear, are used to capture EEG on both sides of the brain. To capture EOG, i.e., vertical EOG (vEOG) and horizontal EOG (hEOG), we need to maximize the vertical and horizontal distance between each pair of electrodes, respectively. Thus, we place the reference electrode on the Mastoid process, which is the thick lower part. With this setup, channel 1 can pick up eye blinks and up/down movements, while channel 2 can capture the eyes' left and right movements. Additionally, both channels can capture most of the facial muscle activities that link to the muscle group beneath the area behind the ears. Since EEG, EOG, and EMG are biopotential signals, we can use the same electrodes. Thus, we only need four electrodes, including two signal electrodes, a reference, and a common ground, to

capture them. Capturing EDA behind the ear is promising because it has high sweat gland density [56]. As sweat gland activities are not symmetric between two halves of the body, placing two electrodes on each ear is necessary to capture EDA reliably.

Examining BTE signals. Signals captured from BTE electrodes resemble the most important biomarkers of microsleep that we would expect from standard electrode placements (i.e., EEG, EOG, EMG, and EDA), as shown in Fig. 3.3. In particular, Fig. 3.3a presents the α rhythms seen on both BTE leads when the eyes are closed. Similarly, the same features, such as eye blinks, left gaze, right gaze, teeth grinding, and emotional arousal, can be captured with BTE electrodes, as shown in Fig. 3.3b, c, and d, respectively. Thus, the results confirm that we could capture the aforementioned microsleep features from only behind the ears.

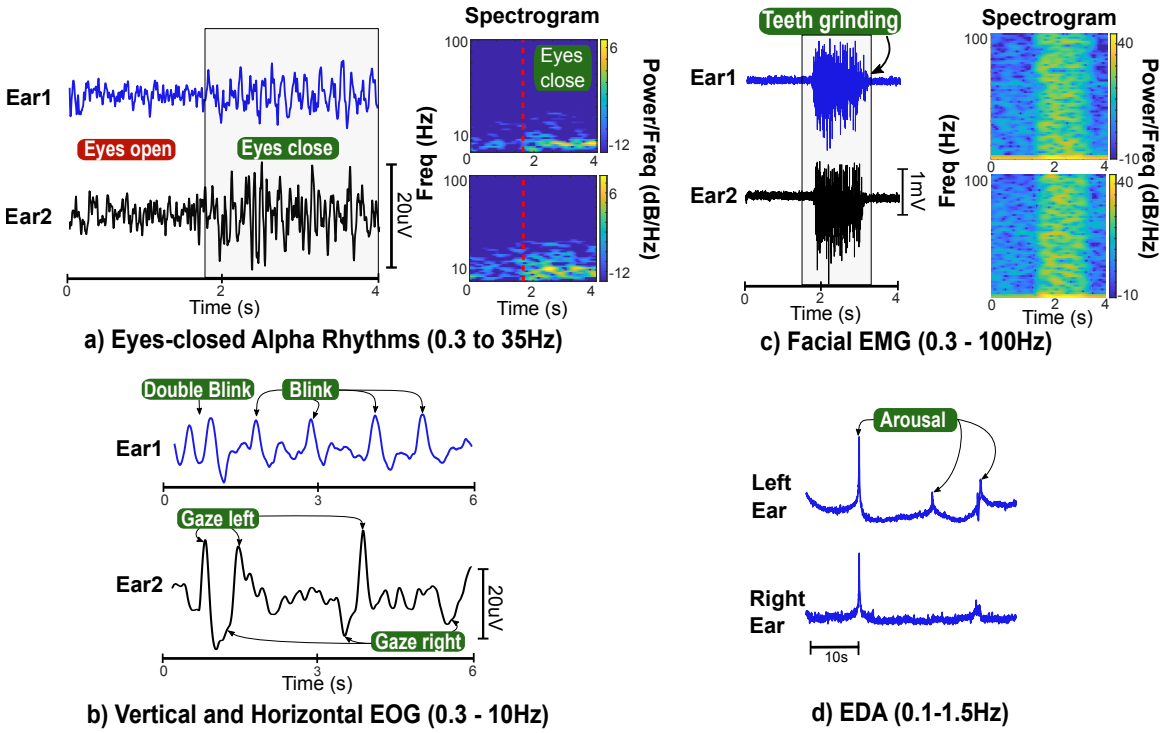


Figure 3.3: EEG, EOG, EMG and EDA signals captured from behind the ears.

There are unique challenges to BTE signals. First, BTE signals are much smaller than the ones expected with standard placements. Particularly, the amplitude of EEG and EOG captured from BTE are all less than 50uV, which is much smaller than standard placements (100-500uV)[178]. This is probably because BTE electrodes are far from the signal sources of EEG and EOG. Secondly, we notice a significant amplitude difference (i.e., three orders of magnitude) between BTE EEG/EOG and EMG signals, as BTE EMG events could be as strong as a few millivolts. Moreover,

the spectrogram in Fig. 3.3 shows that BTE EMG events have very strong power in all frequency bands from 0.3 to 100Hz. As we use the same BTE electrode to capture EEG, EOG, and EMG, addressing the overlap of the three signals is not trivial. Low amplitude BTE EEG/EOG signal overlapped with EMG making it challenging to ensure high fidelity microsleep features while being robust against environmental noise and motion artefacts.

3.3 System Overview

We design WAKE to include four main components (Fig. 3.4): (1) a motion mitigation sensing hardware using the 3-folds Cascaded Amplifying (3CA) technique, (2) a firmware adaptively amplifies the signals, (3) software algorithms running on a host device to process data from BTE sensors and detect user's microsleep, and (4) an ear-worn device designed for long-time usage.

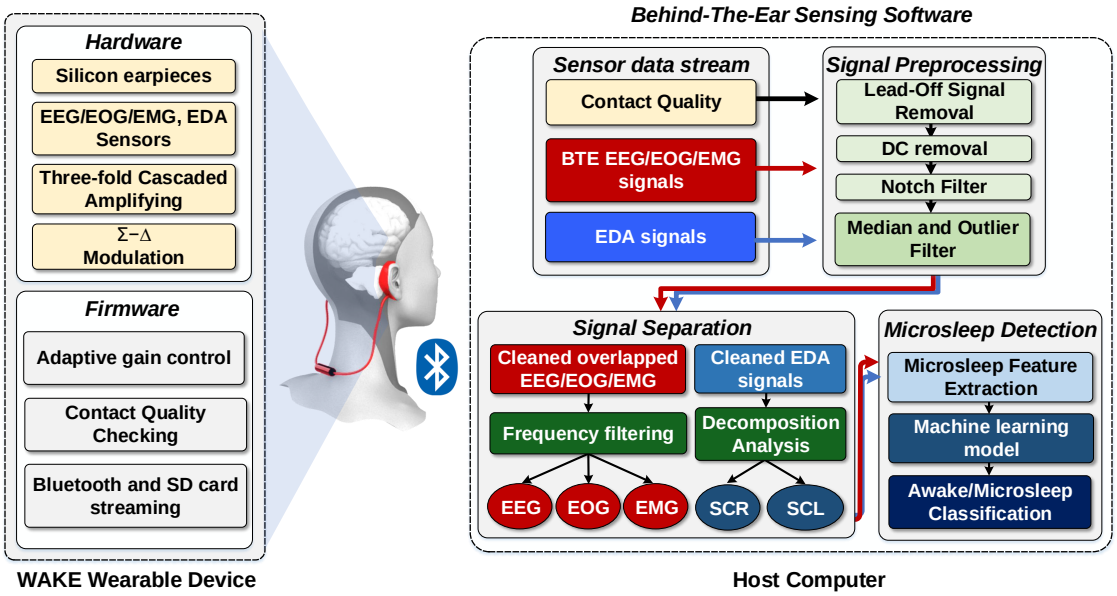


Figure 3.4: WAKE system overview.

WAKE hardware. We design a highly sensitive sensing circuit (Fig. 4.6) to capture the brain waves (EEG), polarisation signal created by eyeball activities (EOG), facial muscle contractions (EMG), and electrodermal activities (EDA). In WAKE, we derive an approach called 3CA, allowing the system to minimise the impact of motion artefacts and environmental noise in real time at hardware and firmware levels. The key idea is to utilise multiple buffering and amplifying stages with precision buffers

and instrumentation amplifiers to address the effects of electrode fluctuation, cable shaking, and environmental interference (Sec. 3.4).

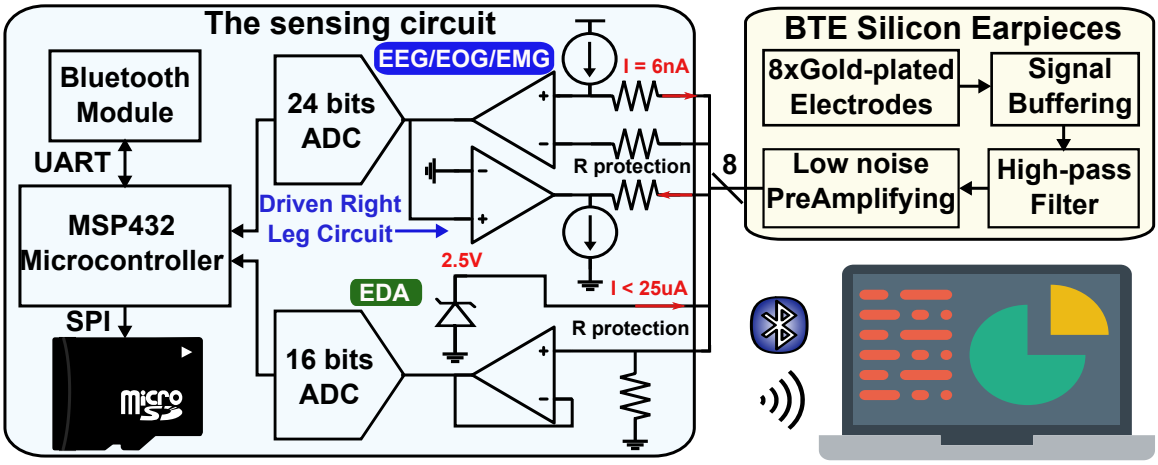


Figure 3.5: WAKE hardware module.

WAKE firmware. WAKE firmware is designed to control our sensing hardware so that data from four main sensors: EEG, EOG, EMG, and EDA, can be captured reliably (Sec. 3.5). The key challenges are that the signals are often weak and overlap each other. Thus, we design the firmware with three main components (1) adaptive gain control (AGC), (2) contact quality checking, and (3) Bluetooth and SD card streaming. AGC addresses the overlapping issue by dynamically changing the amplifier gain based on different signal types. Electrode contact quality is constantly monitored to detect and remove noisy signals created by loose electrodes. Finally, the collected data is streamed over Bluetooth and to an SD card for later analysis.

WAKE algorithms. WAKE algorithms are implemented on a host device (i.e., mobile phones, laptops, etc.). Upon receiving the signals from the WAKE ear-worn device, the data are separated into different streams and ready for further processing. Three main data streams were collected: BTE EEG/EOG/EMG signals, EDA signals, and contact impedance signals. During signal pre-processing, the DC component and electrical noise are removed by DC removal, notch, median, and outlier filters. The clean EEG, EOG, EMG, and EDA signals obtained from pre-processing are then used for microsleep classification. The features extracted from these signals are later used together with machine-learning algorithms to detect microsleep.

WAKE earpieces. WAKE system is designed for comfortable, reliable, cost-effective, and continuous collecting of behind-the-ear signals. To realise that goal,

we design the earpieces by carefully sketching the device architecture and then implementing them using off-the-shelf components. The earpiece materials were also carefully selected, ensuring good contact between the electrodes and the human skin as well as allowing it to be comfortably worn by users. We also validated and identified the most proper electrode materials that provide the highest sensitivity (Sec. 3.7).

In the next section, we will discuss our proposed solution to address one of the most important challenges of designing a reliable wearable device: "*how to cancel the noise created by human motion artefacts and coupled from the environment?*"

3.4 Mitigating Impact of Motions & Noise

noise created by motion and from the environment are important challenges that we must overcome to ensure WAKE's reliability and practicality. These noise span across all frequencies of interest and are highly unpredictable, making their removal non-trivial from the signal by software methods such as filtering or ICA. In literature, Active Electrodes (AE) [179] have been proposed to mitigate motion artefacts and environmental interference. However, conventional AE does not consider the unique characteristic of BTE signals, which are (1) weak EEG and EOG signal amplitudes, (2) signals overlapping with three orders of magnitudes difference, and (3) limited spaces for BTE electrodes. We propose a technique called Three-fold Cascaded Amplifying (3CA) on the electrical pathway of WAKE. Fig. 3.6 presented the model for the 3CA technique with three stages: (1) Stage 1 – Unity Gain Amplifying, (2) Stage 2 – Feed Forward Differential PreAmplifying (F2DP), and (3) Stage 3 – Adaptive Amplifying. The 1st and 2nd stages are implemented on our BTE earpieces, while the 3rd stage is implemented on the sensing circuit and its firmware.

Stage 1 – Unity-gain amplifying. The root cause of motion artefacts lies in the fluctuations of the wires and micro-movements of electrodes [180]. These fluctuations create changes in the electrical pathway resulting in measurement noise. We address the motion artefacts by introducing the first stage: unity-gain amplifying (a.k.a buffering). Considering the reference circuit model, as in Fig. 3.6, V_s is the signal source from the ears, Cw is inherent capacitance on signal cables, and Z_c is the skin-electrode contact impedance. V_o, A, Z_i, R_i, C_i, Z_o, C_p are output voltage; ideal voltage gain; input impedance, resistance, and capacitance; output impedance; and parasitic capacitance of each amplifier.

Since the biosignals are extremely weak (i.e., μV level), instrumentation amplifiers are usually used to amplify the signals, making them available for further processing.

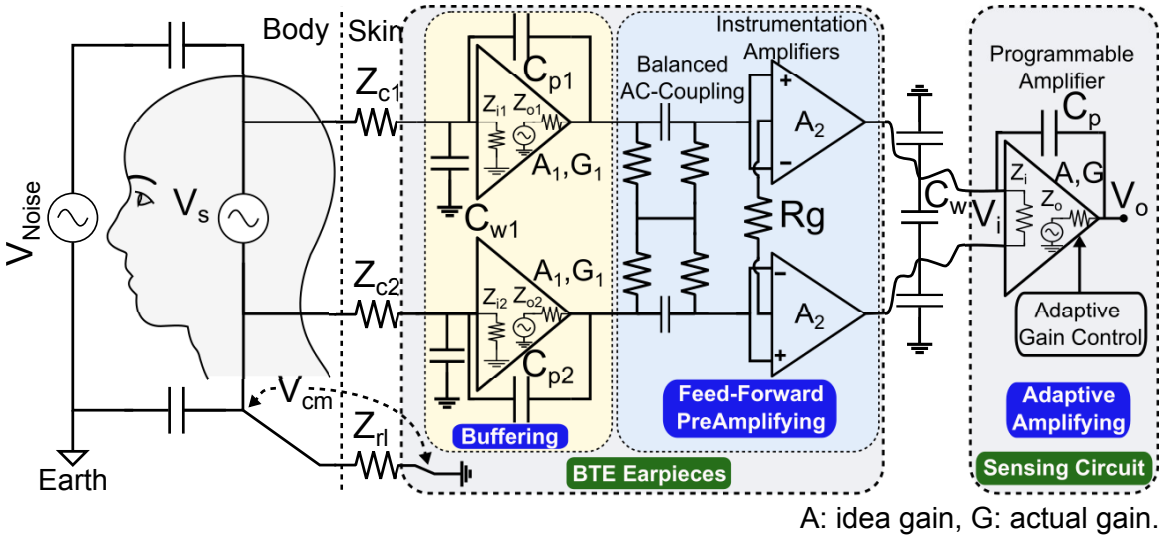


Figure 3.6: 3CA model.

When an instrumentation amplifier is used, we can model the effect of motion artefacts by using the voltage gain rule ($V_o = A * V_i$) and Kirchhoff's current and voltage laws (3.1) at the input of the amplifier:

$$(V_s - V_i)/(Z_{c1} + Z_{c2}) - (V_i)/(Z_i) + j\omega C_p(V_o - V_i) = 0. \tag{3.1}$$

By eliminating V_i from Eq. 3.1, we have a relationship among the actual gain ($G = V_o/V_s$) of the circuit, skin-electrode contact impedance (Z_{c1}, Z_{c2}), and the inherent capacitance on signal wires (C_w):

$$G = \frac{A}{1 + (Z_{c1} + Z_{c2})(\frac{1}{R_i} + j\omega(C_w + C_i - (A-1)C_p))}. \tag{3.2}$$

As motions happen, cable sway and electrode movements create fluctuations on C_w and $Z_{c1} + Z_{c2}$, respectively. This results in the instability of the actual gain (G). To minimise the fluctuation effect of C_w (generated by triboelectric processes and change of parasitic capacitance in the measurement network [181]), we can use an op-amp buffer for each electrode to convert the high impedance lines (Z_{c1}, Z_{c2}) to approximately zero ($Z_{o1}, Z_{o2} \approx 0$). After this process, we can mitigate the signal fluctuations in the transmission network. To analyse signal instability created by electrode movements, we devise a similar gain equation for the op-amp buffer in the first stage, i.e.,

$$G_1 = \frac{A_1}{1 + Z_{c1}(\frac{1}{R_{i1}} + j\omega(C_{w1} + C_{i1} - (A_1 - 1)C_{p1}))} = \frac{A_1}{1 + Z_{c1}\gamma}. \tag{3.3}$$

Ideally, the effect of Z_{c1} fluctuation can be removed if we can satisfy the following equation: $\gamma = 0$. While it is very challenging to achieve in practice, we still can make γ as close to 0 as possible. This could be done by using an ultra-high input impedance buffer, where $A_1 = 1, R_{i1} \to \infty$, and $C_{i1} \to 0$ in our first stage. Putting a buffer circuit directly on the electrodes is the best way to minimise C_{w1}. However, this is not desirable, as we have limited space for our BTE electrodes. We notice that as long as we keep C_{w1} small and stable, putting the circuit directly on the electrode is possible. This is done by fixing the connection between each electrode and its buffer in a stable structure to avoid triboelectric noise and shielding it by using a micro-coax shielded cable. By driving the shield with the same voltage as the signal from the output of the amplifier, we effectively minimise C_{w1}. Up to this point, the unity-gain amplifying stage can remove the impact of human motion artefacts.

The use of ultra-high input impedance buffers at this stage also brings another benefit, i.e., reducing environmental noise coupled into the measurement because of the imbalance among signal lines. When a common-mode voltage (V_{cm}) is introduced to the subject body by an environmental noise source (V_{noise}) (Fig. 3.6), a portion of V_{cm} leaks into our differential measurement and becomes noise (V_n) because of impedance mismatch, i.e., $V_n = V_{cm}(\overline{Z_c}/\overline{Z_i})((\Delta Z_c/\overline{Z_c}) + (\Delta Z_i/\overline{Z_i}))$ [182]. During a long-term measurement in practice, contact impedance can easily deteriorate (e.g., from <10kΩ to several hundreds of kΩ) because of drying gel or unstable contacts [183]. Junction-gate field-effect transistor (JFET) input amplifier is a good solution to address this issue thanks to its extremely high input impedance ($10^{12} - 10^{15}\Omega$). Thus, we can significantly reduce V_n, making our measurement more robust against impedance mismatch. The signals, however, need to go through another stage to remove all the environmental noise, as described in the following discussion.

Stage 2 – Feed Forward Differential PreAmplifying (F2DP). To ensure robustness against environmental interference, intuitively, we would want to preamplify our weak and overlapped BTE signals before driving the cables to our sensing circuit. Conventionally, if an amplifier with positive gain (> 1) is used, the equation $\gamma = 0$ cannot be satisfied, making the system prone to motion artefacts. Furthermore, electrode contact impedance mismatch, which is often seen in practice, leads to gain mismatches among electrodes, as shown in Eq. 3.3. Gain mismatches between two electrodes will allow more common-mode noise to be coupled into the system. By dividing into unity-gain and F2DP stages, we overcome this challenge since the input impedance of F2DP is effectively close to 0. Thus, the effect of contact impedance will not affect the gain in the next stages.

Inspired by the robustness against noise of balanced audio systems where preamplified differential signals are generated before transiting over wires, differential signalling is employed in our design. We apply the Feed-Forward (FF) differential amplifying technique, which has been shown by simulation in [184] to significantly increase the Common-Mode Rejection Ratio (CMRR), i.e., the ability to reject noise coupled from the environment. It could boost the CMRR by 49dB over the conventional driven right-leg circuit. However, the FF topology used in [184] is not practical because its stability suffers from gains mismatch when two different gain resistors are used in the proposed topology. Mismatch in these resistors causes the output common-mode level to move with the output signal, resulting in distortion [185]. Thus, we employ the cross-connection technique where only one gain resistor (i.e., R_g in Fig. 3.6) is needed to set the gain for two FF instrumentation amplifiers in our F2DP. After F2DP, fully differential and preamplified signals are produced, making them robust against environmental interference while driving the cables to the sensing circuit.

F2DP only works when the DC component is removed completely. We found that the traditional high pass RC filter approach is not efficient in removing the DC component (100x larger than signals of interest) because it introduces an additional ground path and component mismatch, which reduces the efficiency of rejecting the environmental noise of F2DP. Balanced AC-coupling topology [186] is a best-fit solution to overcome these challenges because it does not introduce any additional ground path and minimises the component mismatch since the pole and zero of the filter cancel themselves out. In particular, considering 'Balanced AC-coupling' components in Fig. 3.6, this topology does not include the ground path, thereby eliminating its side effects. Moreover, RC components are never precise in practice (approximately from 1% to 20% error for a capacitor), and their mismatch problem is difficult to solve. The chosen balanced AC-coupling topology dampens these mismatches by cancelling the redundant poles and zeros created by component mismatch [186].

Stage 3 – Adaptive Amplifying. After the previous stage, the system can reliably collect BTE signals; we are now solving the problem of our BTE signals themselves. The unique challenge that we need to address with our BTE signals is the significant amplitude range differences between EEG/EOG and EMG signals (i.e., from uV level to mV level for EEG/EOG and EMG signals, respectively). This challenge has not been considered in the traditional EEG system, as EEG electrodes are placed far away from EMG sources. The difference leads to signal saturation at the ADC on the sensing circuit when EMG signals are amplified with the same gain as EEG/EOG signals. The CMRR of the amplifier is presented by the following

equation: $CMRR = 10 * log \frac{A_d^2}{A_{cm}^2}$, where A_d and A_{cm} are the differential and common-mode gain, respectively. In an instrumentation amplifier, A_{cm} is a constant depending on the internal resistors. Thus, CMRR only depends on A_d. Since the difference between EMG and EEG/EOG could be as large as three orders of magnitude, setting the gain too low to avoid EMG saturation will also significantly lower CMRR (up to 60 dB), increasing the noise floor to a level where EEG/EOG signal cannot be captured. We found that the gain needs to be dynamically adjusted in real time so that both small EEG/EOG and large EMG signals are captured with high resolution.

3.5 Signal Processing

3.5.1 WAKE On-board processing.

Adaptive Gain Control. As aforementioned in Sec. 3.4, one important and unique challenge in ensuring high-fidelity BTE signals is the large difference in the amplitude range (which could be up to three orders of magnitude) between EEG/EOG and EMG signals. Thus, the analogue gain of our sensing circuit needs to adapt dynamically to the changes in signal amplitude.

Fortunately, we observe that (1) EMG events do not happen frequently, (2) EMG events can happen quickly with big amplitude changes, and (3) signal amplitude during an EMG event is stochastic and can vary significantly. By understanding these characteristics, we could design our AGC to (1) keep the gain at maximum for EEG/EOG signals while there are no significant EMG events, (2) react quickly to the abrupt increase of amplitude to detect EMG events but (3) react slowly to the decrease of amplitude while an EMG event is still happening to avoid gain oscillation. Peak Envelope Detector (PED) and Square Law Detector (SLD) are two popular AGC techniques [187] that fit our needs. We use PED because of its low computational complexity. If there is no EMG event, we use a small window size so that PED can react quickly, while a larger window size is used during an EMG event to avoid gain oscillation.

We can choose the PED window size based on the property of EMG signals, sampling rate, desired response time, and the device's computational resource. According to [188, 189], EMG signal has a frequency range of 1-500Hz and is most dominant between 50-150Hz. Thus, with the sampling rate of 1000Hz, we can cover the EMG signal range by using the small and large window sizes of 2 and 1000 samples, respectively. With the large window size of 1000 samples, our AGC can guarantee that the gain will be adjusted within one second after the EMG event has ended. This

response time is acceptable in our application as a microsleep can last at least a few seconds. The small window size of two samples can make our AGC very sensitive even with the fastest EMG signal, but it also increases our processor load. Since the dominant EMG signal power is in the range of 50-150Hz, we can increase the small window size to reduce the load. During a gain transition, the amplifier needs to be stabilized before new data can be obtained. We can interpolate missing samples with a lightweight linear interpolation. Fig. 3.7 shows an EMG event is captured without saturation by using AGC with the small and large window sizes of 10 and 1000 samples, respectively.

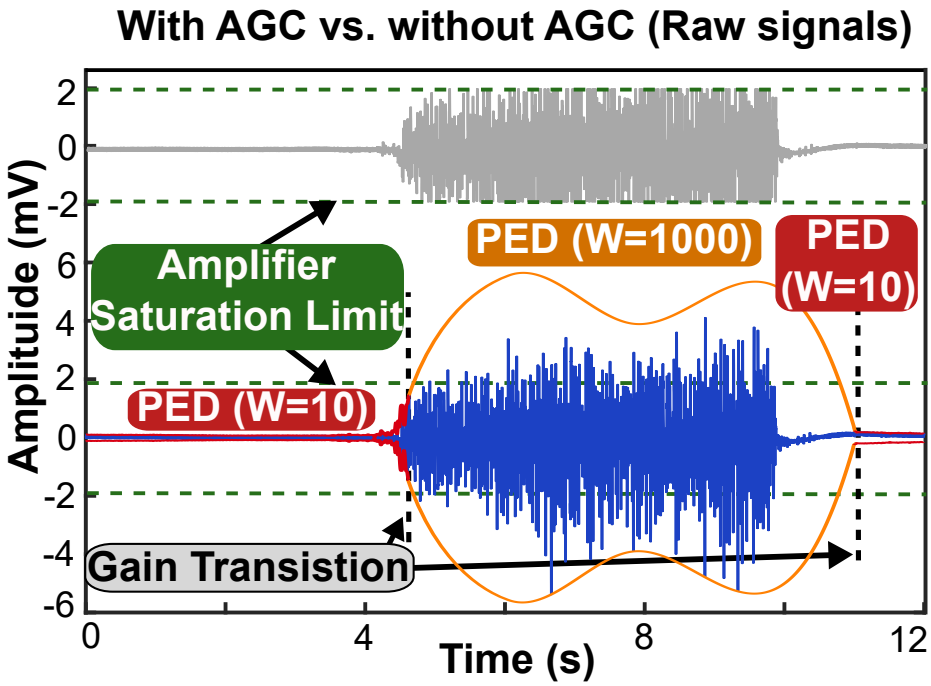


Figure 3.7: EMG saturation w/o AGC.

$\Sigma - \Delta$ **modulation.** To ensure high signal quality during the quantization process, we employ the $\Sigma - \Delta$ modulation, which could be found in high precision Analog-to-Digital Converters (ADCs). Low quantization noise is achieved by utilizing oversampling, noise shaping, digital filtering, and decimation, as illustrated in Fig. 3.8. With a given ADC having the resolution of b bits, the full signal scale FS, and the quantization error (ε) is assumed to be a stationary, random process; the quantization noise is a constant ($\sigma_\varepsilon{}^2 = FS^2/(3 * 2^{2b})$) [190]. Without a noise shaping function, the quantization noise spreads out uniformly over the Nyquist spectrum. Thus, the power spectral density is also constant, i.e., $S_e(f) = \sigma_e{}^2/f_s$ with f_s as

the sampling frequency. Oversampling (by a factor of K) widens the Nyquist spectrum, thereby reducing the quantization noise energy in the spectrum of interest, i.e., $S_e(f) = {\sigma_e}^2/(Kf_s)$. To further reduce the noise inside the spectrum of interest, we pass the signal through a noise-shaping function called the $\Sigma - \Delta$ modulator. Fig. 3.8 presents a first-order $\Sigma - \Delta$ modulator constructed as a negative feedback loop. By using Z-transform, we have the noise transfer function (NTF) of the loop to be $NTF(z) = 1 - z^{-1}$ [190]. Converting the NTF to the frequency domain by using trigonometric identities, we have $NTF^2(f) = 4sin^2(\pi f/fs)$. Thus, the new noise power spectral density is $S_e(f) = {\sigma_e}^2/(Kf_s) * |NTF(f)|^2$. Similarly, we can generalize the NTF equation for an Nth-order $\Sigma - \Delta$ modulator as $NTF^2(f) = [2sin(\pi f/fs)]^{2N}$. As illustrated in Fig. 3.8, a $\Sigma - \Delta$ modulator can shift the noise energy to the high frequency of the spectrum. A digital filter removes the noise power outside our spectrum of interest. The signal is then decimated to the required sampling rate before outputting the results. Thus, we can achieve a low noise floor for our measurement.

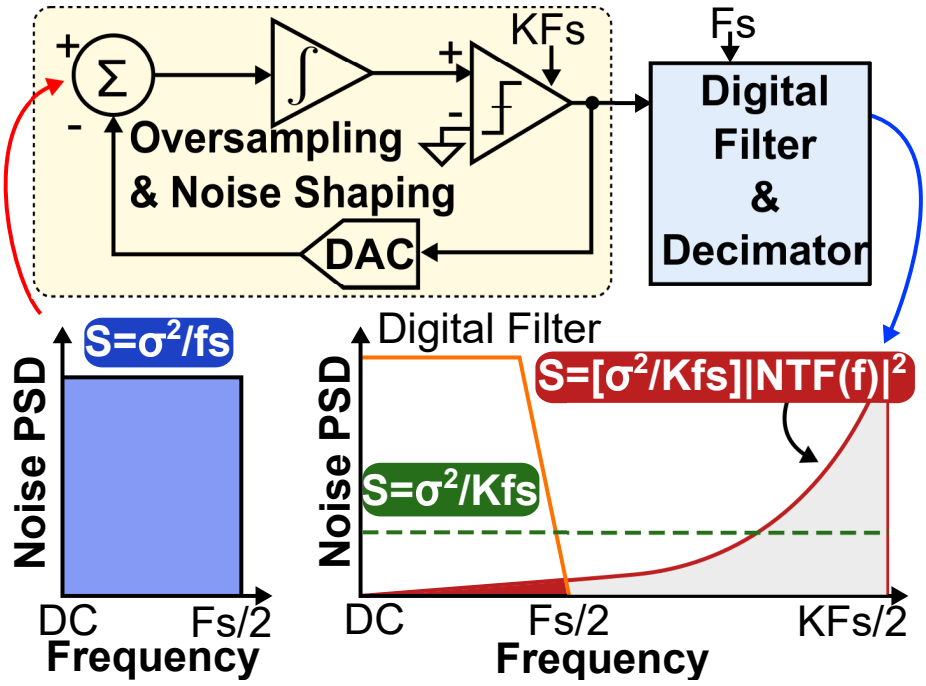


Figure 3.8: $\Sigma - \Delta$ modulation.

Contact quality checking. To monitor contact quality, we inject a small sinusoidal AC current (i.e., $I = 6nA$, 30Hz) through the skin-electrode contact. By measuring the response voltage, we can calculate the contact impedance by this equation: $Z_{30Hz} = V_{RMS,30Hz}/I - R_{protection}$. We follow the clinical standards as in [47], stating the acceptable upper limit of electrode impedance is $10k\Omega$ to achieve optimal

biosignals recording. This can be achieved with standard procedures of skin preparation and conductive gel application. As the electrode impedance can vary over time because the electrode contact loosens or the gel dries out, we can notify the user to adjust the contact or reapply the gel if the impedance is higher than our defined threshold.

3.5.2 WAKE Physiological Signals Extraction

In WAKE, each sensor data (EEG/EOG/EMG and EDA) is pre-processed at the host device corresponding to their own characteristics before putting it into the signal analyzing procedure. We show examples of changes in those signals between microsleep and awake states in Fig. 3.9. We apply to all sensor data the notch filter to remove 50/60Hz power line interference, linear trend removal to avoid DC drift, and outlier filters to remove spikes and ripples.

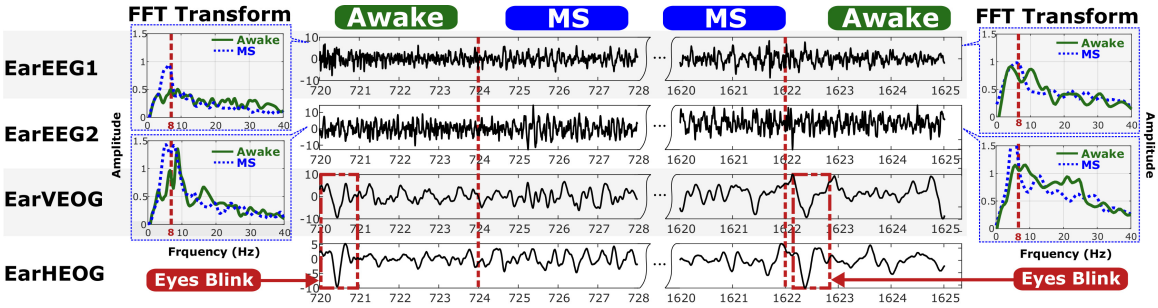


Figure 3.9: Signals captured by WAKE during the transition between awake and microsleep.

Collecting EEG/EOG/EMG signals. WAKE's mixed-biosignals include EEG, EOG, and EMG, which are in the frequency range of 4-35 Hz, 0.1-10 Hz, and 10-100 Hz, respectively. We apply different bandpass filters to split the mixed BTE biosignals into the signals at the frequency range of interest. In particular, we extract wakefulness-related EEG bands (i.e., θ, α, and β waves) using 4-8 Hz, 8-12 Hz, and 12-35 Hz bandpass filters, respectively. We extract horizontal EOG (hEOG) for eye movement and vertical EOG (vEOG) for eye blink using 0.3-10 Hz filters. A 10-100Hz bandpass filter and a median filter are then applied to the mixed signals to extract the EMG band and eliminate spikes and other excessive components.

Collecting EDA signal. EDA signal is the superposition of two different components, skin conductance response (SCR) and skin conductance level (SCL), at the frequency range of 0.05-1.5 Hz and 0-0.05 Hz, respectively. Even though EDA signals

have fast responses, they are very slow to decline to baseline. Thus, if another response happens right after the first response, the signal level will increase even more. Thus, frequency filtering is not effective in separating EDA signals. To address this, we employ a non-negative deconvolution technique proposed in [191] to decompose EDA into SCR and SCL components.

3.6 Algorithms

We present two classification methods: (1) feature engineering-based classification and (2) deep learning on raw data. Feature-based classification is built on well-studied microsleep features with off-the-shell machine-learning models. While this approach may help the learning procedure become more stable and interpretable with fewer data, processing features is labour-intensive. The second approach, i.e., deep learning, tackles this issue and has been shown to achieve state-of-the-art performance in microsleep detection [192, 193]. However, previous work only considers standard EEG and EOG placements. Thus, the use of deep learning on BTE signals remains unanswered.

3.6.1 Classification based on Feature Engineering

EMG active events detection: Microsleep appears when the body is relaxed. A strong EMG signal can have significant power across all frequency bands of interest (discussed in Sec. 3.2). It will contaminate our BTE EEG and EOG signals rendering them unusable. We detect the active event based on the sum of all frequency bands in the spectrogram. For each data signal, we use the first 10 seconds as the ground-based noise. Any data whose total spectrum energy is 10% larger than that of ground-based is an active event.

Feature Extraction: We divide the collected time series data from each source into fixed-size epochs. The selected features are extracted from each epoch for classification.

Temporal features: This category includes typical features used in the literature for time series data analysis in the temporal domain, namely mean, variance, min, max, Hjorth, skewness, and kurtosis. In microsleep detection, EOG, EMG, and EDA signals are often analyzed in the time domain due to their considerable variation in amplitude and lack of distinctive frequency patterns [194]. Those six temporal features are extracted from each of hEOG, vEOG, EMG, and EDA signals for a total of 24 temporal features. We use wavelet decomposition for the hEOG signal to extract

saccade features, namely mean/max velocity, mean/max acceleration, and amplitude range. Eyeblink features, namely mean/max amplitude, peak closing velocity, peak opening velocity, and closing time, are extracted from the vEOG signal.

Spectral features: The spectral features are extracted to analyze the characteristics of the EEG signal because brainwaves are generally available in discrete frequency ranges at different stages. Those features include the ratio of powers, absolute powers, θ/β, α/β, θ/α, and $\theta/(\beta+\alpha)$. Accordingly, 14 features are extracted from each channel of EEG providing 28 spectral features in total.

Nonlinear features: Bioelectrical signals show various complex behaviours with nonlinear properties. In particular, the chaotic parameters of EEG can be used for microsleep detection. The discriminant ability of nonlinear analyses of EEG dynamics is demonstrated through the measures of complexity such as correlation dimension, Lyapunov exponent, entropy, fractal dimension, etc. [195], with the last two features proven to be most informative. In this study, we extract these two nonlinear features for each of the two EEG channels (a total of four features).

Feature Selection: When all features are used altogether, irrelevant correlated features or feature redundancy can degrade the performance. Therefore, we adopt three feature selection methods, including Recursive Feature Elimination (RFE), L1-based, and tree-based feature selection, to select the set of most relevant features. RFE is a greedy optimization algorithm that removes the features whose deletion will have the least effect on training error. L1-based feature selection is used for linear models, including Logistic Regression and SVM. In our linear models, we use the L1 norm to remove features with zero coefficients. Finally, the feature importance ranking generated by the tree-based model is used to eliminate irrelevant descriptors.

Microsleep Classification: Various classification methods, from Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Logistic Regression (LR), and Decision Tree (DT) to ensemble methods like RandomForest or AdaBoost, have been proposed in the literature for awake and microsleep classification, each shown to be effective in specific settings [196]. To cope with the high complexity of our collected signals, we developed a hierarchical stack of three base classifiers. Our hierarchical model consists of a Random Forest classifier (with 50 estimators) in the first layer, an Adaboost classifier (with 50 tree estimators) in the second layer, and SVM (with RBF kernel) in the last layer. Specifically, for the first two layers, we only keep the predictions with high probabilities (> 0.7) and transfer the rest of the samples to the next layer. In the last layer, SVM classifies all of the remaining samples. We also apply a heuristic rule to the final prediction based on the knowledge that an EMG

event is likely to lead to an 'awake' event. The results of our empirical analysis are presented in Sec. 4.7, which highlights the overall accuracy of the performance and proves the efficiency of the proposed classification model.

3.6.2 Deep Learning on Raw Data

Deep neural networks (DNNs) is a branch of Artificial Neural Network which can model highly complex nonlinear functions. An advantage of DNNs compared to other machine learning algorithms is their ability to automatically learn features from raw inputs. Also, DNNs usually require fewer manual adjustments even though they need more data and computations, thus being easier to deploy and maintain.

Modified Sorsnet: while the deep learning community has extensively studied perceptual data (e.g., image, audio, text), brain signals are still under-explored. Recent works from sleep staging and microsleep detection have demonstrated the effectiveness of DNNs, especially Convolutional Neural Networks (CNNs), in learning meaningful patterns of EEG signals. We make use of a relevant CNN architecture proposed by [197] as it achieves state-of-the-art performance for sleep classification using a single EEG channel. This model contains 12 1-D convolutional blocks followed by two fully connected layers. Each block consists of a 1-D convolution layer, a BatchNorm layer and a ReLU activation function. As our data contain 24 1-D signals, we modify this architecture by stacking these signals into 24-channel signals and updating the input channel of the first convolutional layer. For training, we use cross-entropy loss [198], Adam optimiser [199] with $\beta_1 = 0, \beta_1 = 0.99$, and Lambda scheduler for the learning rate decay. We train the SorNet model with 200 epochs. To derive the best hyper-parameter, we apply the grid search for the learning rate in $[1e^{-4}, 0.1]$ and find out that the learning rate 0.002 works well for us in most of the cases. We use an additional validation set on the precision score to overcome the over-fitting and pick the best model.

3.7 Implementation

Earpieces' Material. We design the BTE biosensing earpieces by attaching electrical conductive material on top of a silicone base as illustrated in Fig. 3.10. The silicone material (Dragon Skin 10) is chosen so that the earpiece can fit with the curve created by the mastoid process while being comfortable and alterable with different user's ears (Fig. 3.10, 3.11). Furthermore, the chosen silicone material is skin safe

and does not create irritation to the user's ears. The silicone base is moulded based on the average size of the human ear [200]. To maintain good contact between the electrodes and skin, we put a memory wire inside the silicone base. The memory wire creates a grip on the wearer's ears, pushing the silicone against the skin. It helps the earpieces be usable for different human ear sizes and shapes.

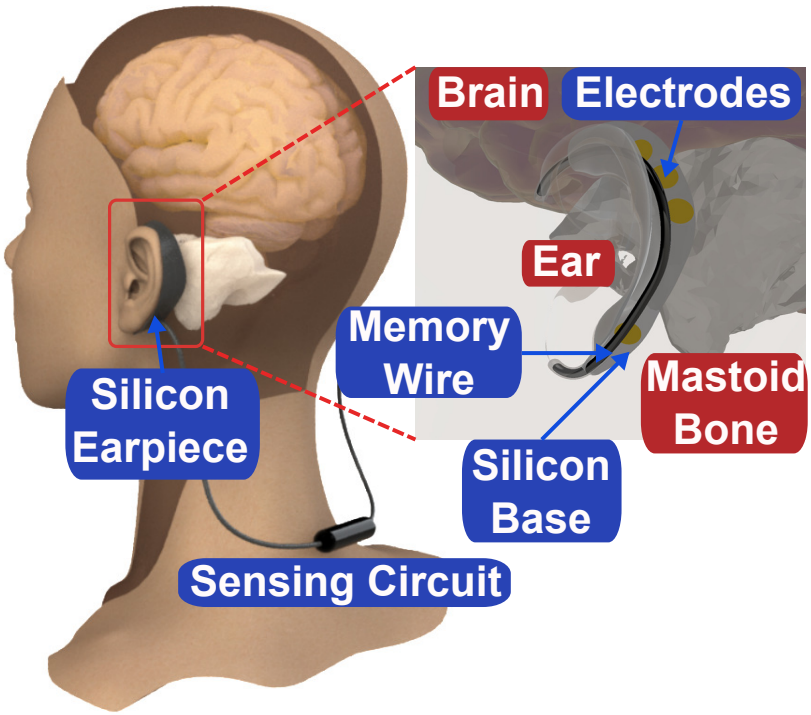


Figure 3.10: 3D model.

Electrodes' Material. We evaluated three different materials for the electrodes attached to the earpieces, including (1) silver fabric, (2) copper pad and (3) gold-plated copper pad. The silver fabric electrodes are highly conductive and can make good contact with the skin thanks to the flexibility of the fabric, but the silver gets tarnished quickly because of the skin oil and sweat. Thus, the contact quality degrades after several uses as the resistance increases dramatically from less than 1Ω to several hundred kΩ. Similarly, copper-based electrodes also degrade quality after several uses. We address that issue by plating gold liquid over the copper electrodes because the gold-plated electrodes are more resistant to skin oil and sweat. In addition, gold is well-known to be chemically inert. Thus, skin allergy to gold is extremely rare. The resistance of the gold-plated electrodes is always less than 1Ω. To enhance contact conductivity and adhesion, we apply Weaver's Ten20 conductive paste on the electrodes before wearing the earpieces. The amount of conductive paste should be

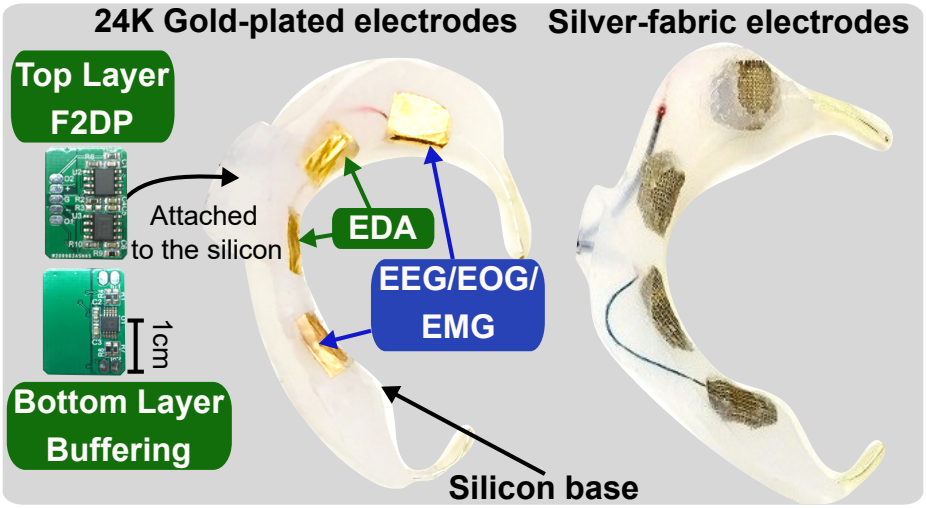


Figure 3.11: BTE silicon earpieces.

enough to cover each electrode's surface. The more we spread the paste, the more signals our electrodes will see. As the paste is water-soluble, we could easily clean it with water after each session. The contact impedance between the electrodes and the skin is also measured to be in the range from 5 to 10 kΩ at 30Hz with a proper skin preparation. This impedance value satisfies the clinical standards [47], which state the acceptable upper limit is 10kΩ, to achieve optimal recordings.

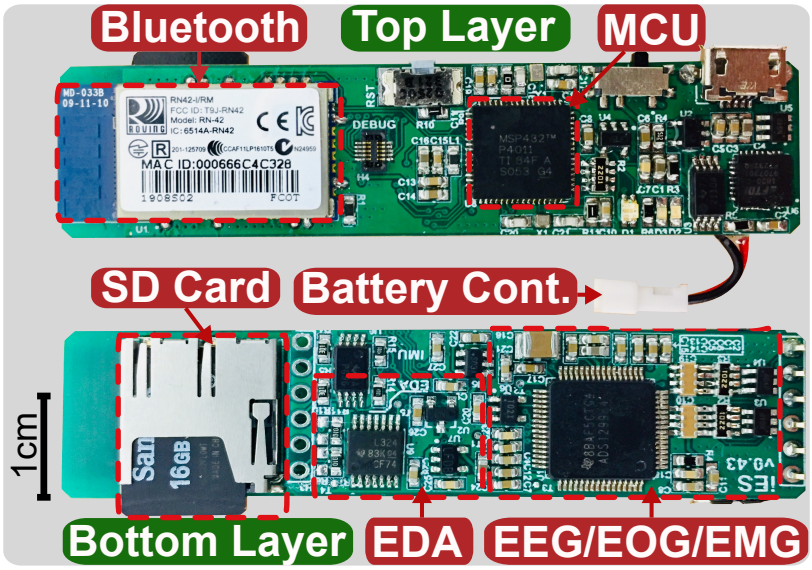


Figure 3.12: WAKE circuit.

Putting Things Together. We use the low-power, precision AD8244 JFET-input buffer to implement our Stage 1 of 3CA. The AD8244 device has unity gain,

very high input resistance (i.e., 20TΩ), and very low input capacitance (i.e., 12pF) so that the effect of motion and impedance mismatch can be minimised, as pointed out in Sec. 3.4. The precision instrumentation amplifier AD8222 is used to implement our Stage 2 (F2DP). The preamplifying gain is chosen at 100, so the full range of the ADC (i.e., -2.5V to 2.5V) is utilised. We use an ultra-low noise amplifier and 24-bit ADC chip ADS1299 to digitise the signals. The ADS1299 provide an integrated second-order $\Sigma - \Delta$ modulator. It samples the input signal at 1.024 MHz and shapes the noise across the Nyquist bandwidth (i.e., 0-512 kHz). A third-order digital low-pass Sinc filter is used to remove most of the noise at high frequency. The decimator downsamples the filtered signal to 1000 Hz and 250 Hz to be stored in an SD card and transmitted over Bluetooth, respectively. The main processing unit (MSP432) is used to (1) drive the analogue front end on the sensing circuit, (2) adjust the amplifier gain dynamically, and (3) stream data to a host device through Bluetooth.

3.8 Performance Evaluation

3.8.1 BTE Signals Sensitivity Validation

In this section, we compare the ability to capture EEG, EOG, EMG, and EDA with WAKE from BTE against the ground-truth devices from standard placements on the scalp, the eyes, the chin, and the wrist (Fig. 3.1). Ground-truth EEG, EOG, and EMG are measured by using an FDA-approved Lifeline Trackit Mark III device with electrodes placed at C3, C4, O1, O2, Cz, M1, M2, upper and lower parts of the left eye (VEOG), two sides of the left and right eyes (HEOG), and the chin (chin EMG), according to the International 10-20 system. Ground-truth EDA is measured by BioPac's BioNomadix Wireless EDA Amplifier system with electrodes placed on the left wrist. The data was collected for one hour while the subject sat on a couch. We calculate Normalized Cross-Correlation (NCC) between our BTE signals with the ground-truth ones to measure the similarity between them.

The measured signals are shown in Fig. 3.13. NCCs of EEG, EOG, EMG, and EDA are as follows: Ear1-C3: 0.35, Ear1-O1: 0.28, Ear2-C4: 0.44, Ear2-O2: 0.52, Ear1-VEOG: 0.47, Ear2-HEOG: 0.59, Ear1-chinEMG: 0.62, Ear2-chinEMG: 0.76, and EarEDA-WristEDA: 0.37. The results show that Ear2, i.e., the channel crossing right and left ears, has strong correlations with scalp EEG and horizontal EOG. Ear1, i.e., the channel placed on the left ear, has a moderate correlation with scalp EEG and a strong correlation with vertical EOG. Both Ear1 and Ear2 channels have strong

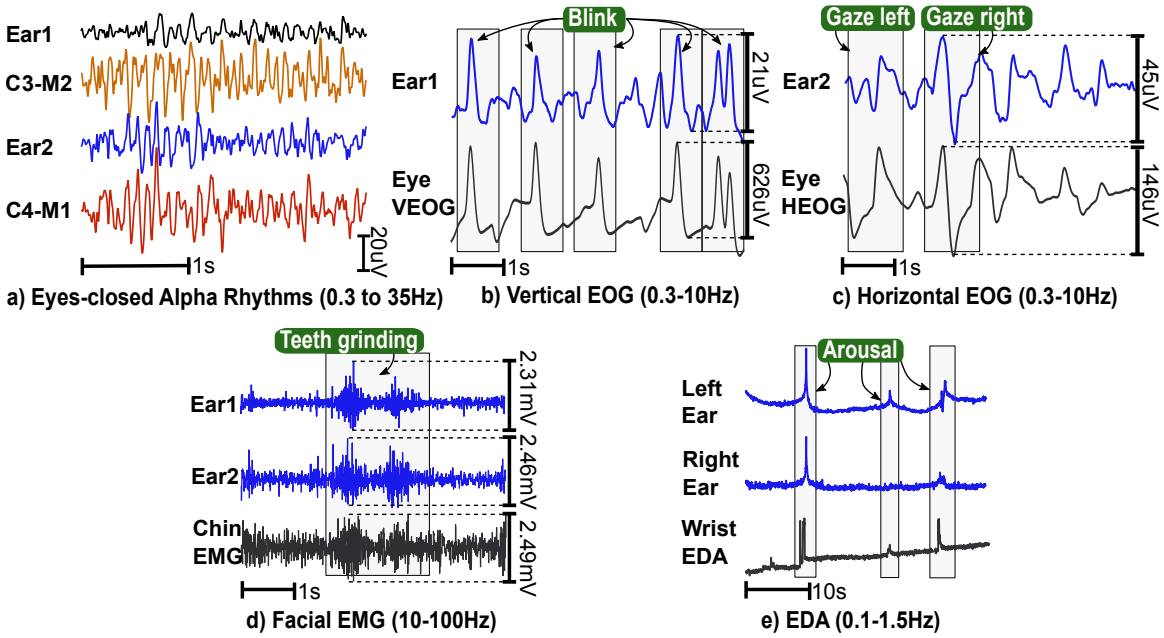


Figure 3.13: BTE EEG, EOG, EMG and EDA signals vs. ground-truth.

correlations with chin EMG. EDA on the left ear shows a moderate correlation with the signal from the wrist.

3.8.2 Noise Suppression Performance

Motion artefacts Mitigation. We evaluated the 3CA technique in two scenarios: (1) walking and (2) sitting in a car. Each evaluation is done in one hour. Evaluation (1) consists of ten minutes of standing stationary and 50 minutes of walking in a hallway. Evaluation (2) also consists of ten minutes of sitting in a car while it is parked in a parking lot and 50 minutes of driving on an urban road (40 mph). Evaluation (1) presents artefacts created by human motion, while evaluation (2) presents artefacts introduced from the environment while driving. Two pairs of electrodes are put as close as possible to the same ear of a subject so that the same signals can be obtained.

Without 3CA, the BTE EOG signals (i.e., eye blinks) are completely distorted by significant motion artefacts. The noise power introduced by motion is shown in Fig. 3.14. During standing and parking scenarios, BTE signals with and without 3CA have the same power. However, during walking and driving, 3CA reduces the noise power by 19.47 dB and 11.87 dB. Thus, the eye blink signals are captured reliably (Fig. 3.15, 3.16).

Environmental Noise Reduction. We evaluated the ability to minimise environmental noise in three different practical environments: (1) in an office, (2) at home,

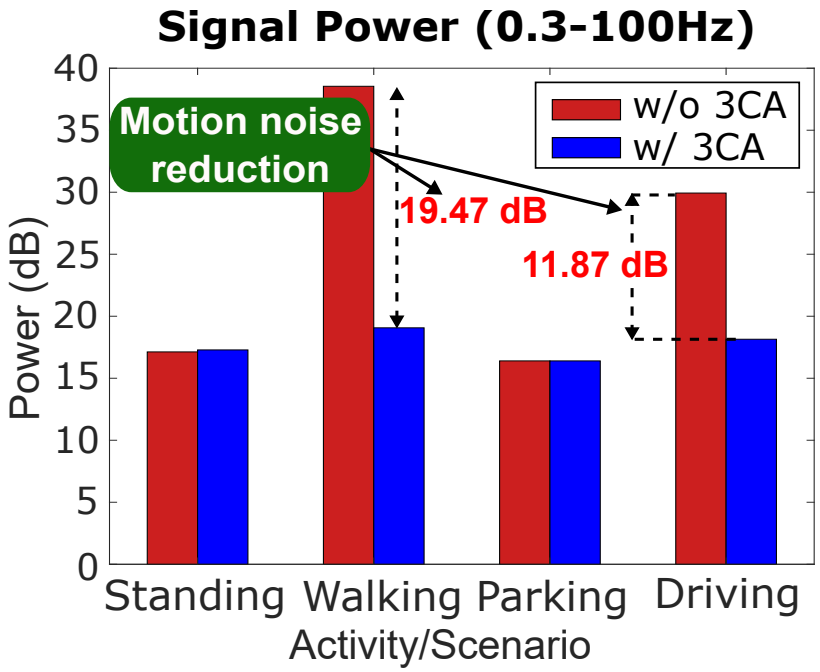


Figure 3.14: Motion noise power reduction.

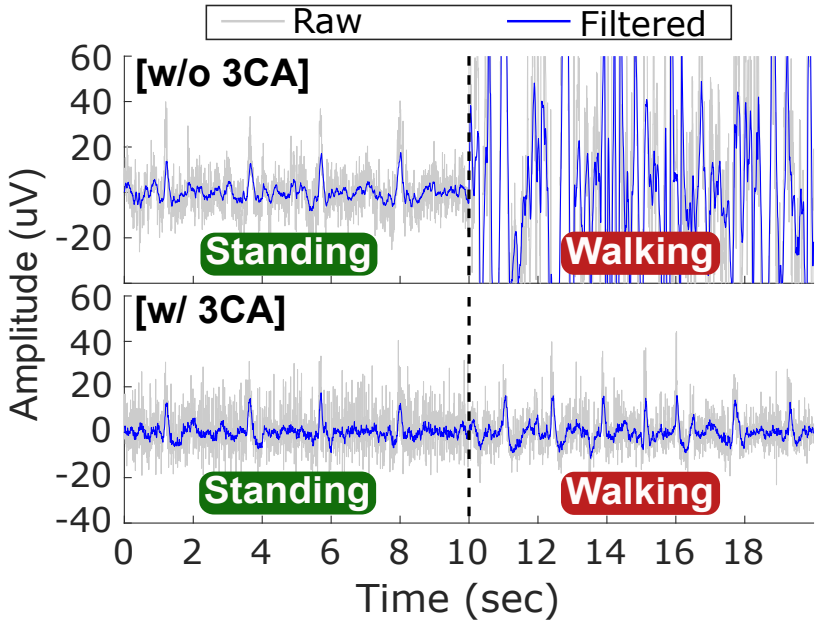


Figure 3.15: Walking motion noise suppression.

and (3) inside a car. The results of the noise spectrum are shown in Fig. 3.17, 3.18, and 3.19, where 3CA can reduce the noise power by 9.74 to 16.1 dB. We also found that the 60Hz noise and its harmonics coupled with the electrical power line are the main sources of noise while the subject is stationary. During motion, motion artefacts are the most significant noise source at frequency ranges 0.3-100Hz (Fig. 3.19).

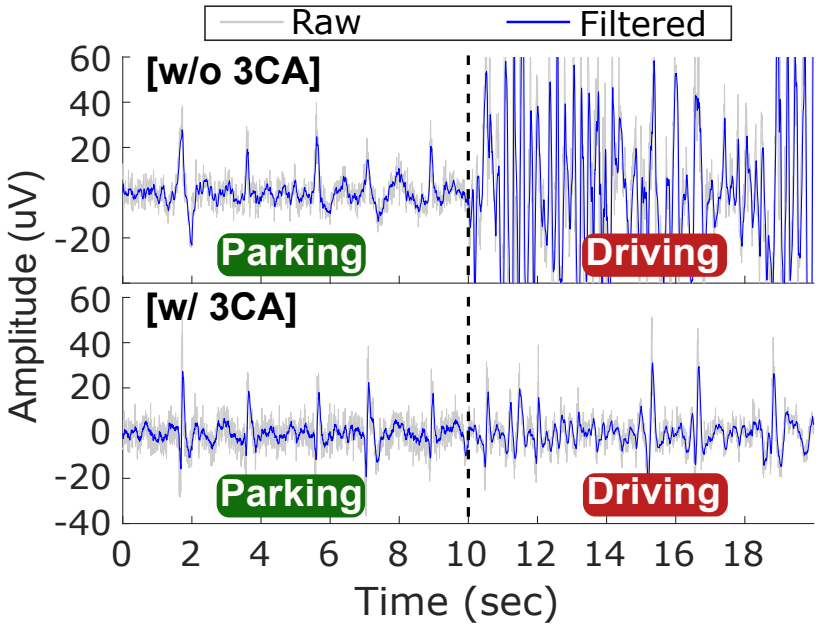


Figure 3.16: Driving motion noise suppression.

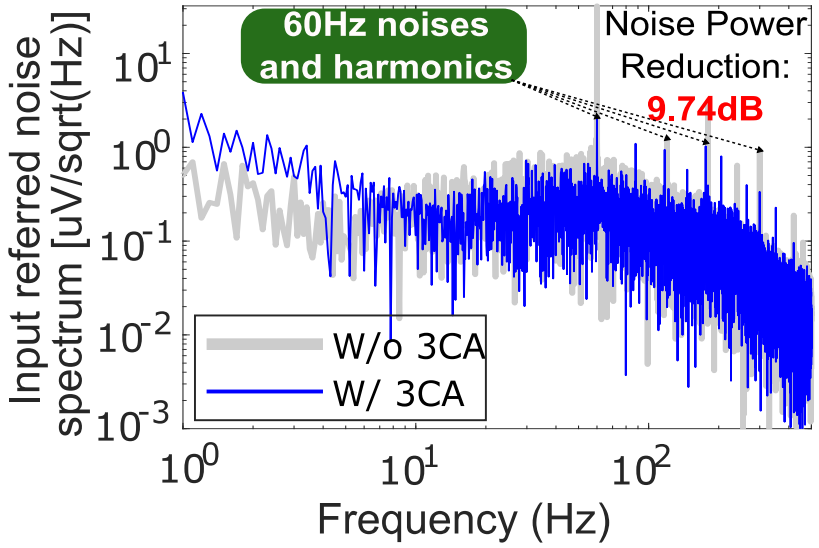


Figure 3.17: 3CA noise reduction (In an Office).

3.8.3 Microsleep Detection Performance

We evaluated WAKE's ability to detect microsleep by conducting the Maintenance of Wakefulness Test (MWT), which is the existing gold standard for quantifying microsleep [201]. We conducted three sets of experiments on the data of 19 subjects. In the first experiment, we perform the Leave-One-Subject-Out Cross-Validation (LOSOCV), i.e., we train each classifier on the set of 18 subjects and evaluate the unseen subject. The second and third sets deal with each individual subject, in which

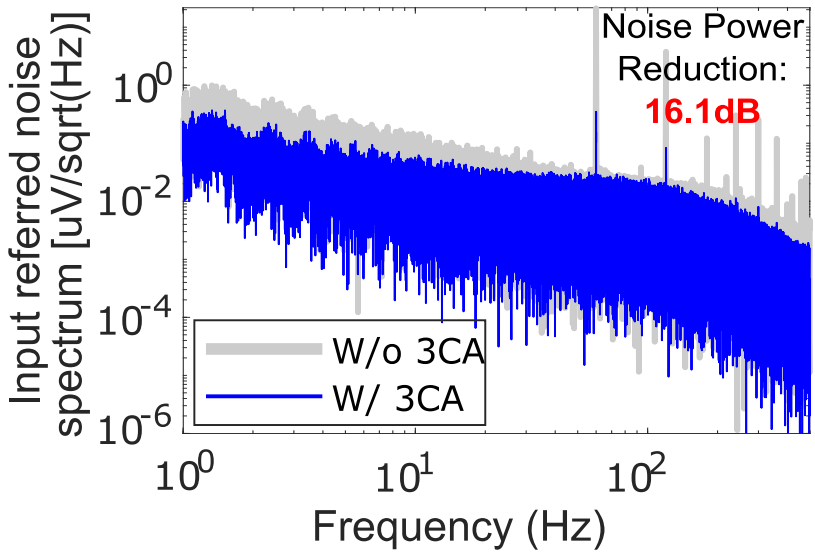


Figure 3.18: 3CA noise reduction (At Home).

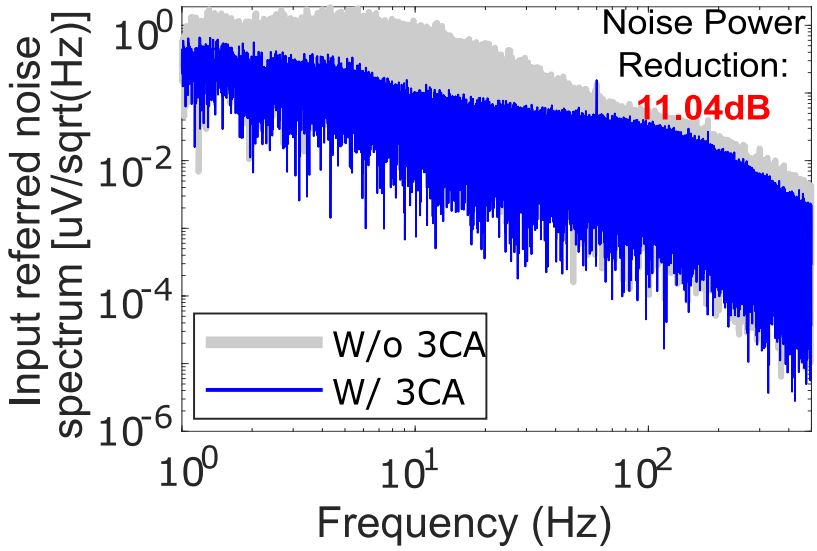


Figure 3.19: 3CA noise reduction (Inside a Car).

we provided a test-set and 10-folds cross-validation. Also, we conduct these sets of experiments on four different epoch sizes: 3s, 5s, 7s, and 9s.

Experimental Protocol. WAKE protocol has been thoroughly designed and approved by the Institutional Review Board. 19 sleep-deprived and narcoleptic subjects on the campus were recruited for the study. Participants' demographics are shown in Tab. 3.1. The Sleepiness Level of each subject was recorded by using the Epworth Sleepiness Scale (ESS). The ESS score is interpreted as <10, healthy level; 10-15, Excessive Daytime Sleepiness (EDS); and 16-24, Severe Excessive Daytime Sleepiness

Table 3.1: Demographic data of participants

Age	18 - 44 years old
Sleepiness Level	Healthy: 9, EDS: 8, SEDS: 1, Narcolepsy: 1
Gender Ratio	Male: 12, Female: 7

(SEDS). The subjects were advised to sleep for less than five hours (only applied to subjects at the healthy level) on the night before the study and also not to consume caffeine or alcohol products before the study so that their microsleep could be faithfully captured. During each MWT session, the subject was asked to try to stay awake in a sleepiness-inducing environment. We use an FDA-Approved Video-EEG system (Lifelines Trackit Mark III) to conduct PSG as the 'ground truth'.

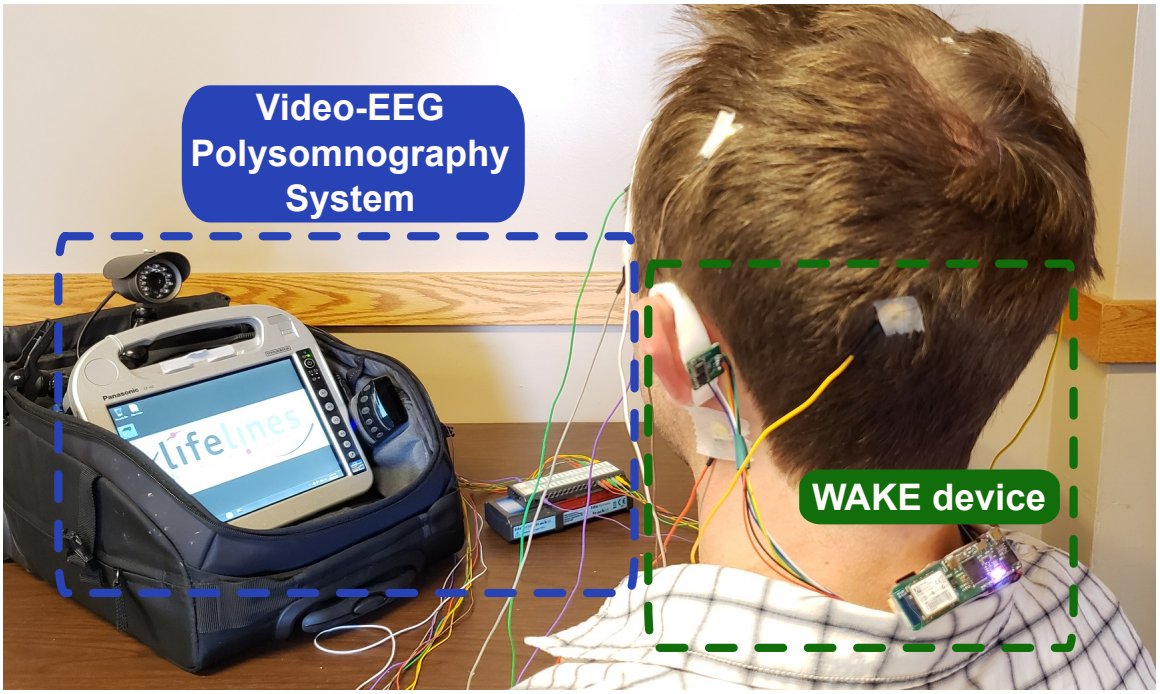


Figure 3.20: Experiment setup.

The MWT Protocol. We conducted two sessions of MWT for each subject with a maximum of 40 minutes each. The subject was asked to sit comfortably on a couch. The WAKE device and the 'ground-truth' PSG system were installed on them, as shown in Fig. 3.20. We minimised all the external factors that could affect the subject's drowsiness by blocking all the light and sound coming from outside of the experiment room. The room was dark, and its temperature was set at the subject's comfort level. The subject was asked to relax but try to keep themselves awake for as

long as possible, so they would not fall asleep voluntarily. The MWT started when the light in the experiment room was turned off. We woke the subject up after they fell asleep. The collected PSG data was sent out for scoring by two certified sleep experts. To handle the variation of the manual process, one expert scored while the other expert verified independently, and the differences were resolved by discussion. Awake and microsleep episodes were marked down by following the guideline of AASM for Sleep Study [202].

Classification Evaluation Metrics. We cast the problem of microsleep detection as a binary classification problem: positive class for microsleep epoch and negative class otherwise. Here, we briefly describe four indices of the confusion matrix: true positive (TP) is the number of actual positive epochs which are correctly classified; true negative (TN) is the number of actual negative epochs which are correctly classified; false positive (FP) is the number of actual negative epochs which are incorrectly classified as positive; false negative (FN) is the number of actual positive epochs which are incorrectly classified as negative. Given these notions, we can now define precision, sensitivity, and specificity scores as follow: $Precision = \frac{TP}{TP+FP}; Sensitivity = \frac{TP}{TP+FN}; Specificity = \frac{TN}{TN+FP}$. Due to the nature of our detection problem (the number of microsleep epochs is much less than that of awake ones), precision, sensitivity, and specificity are preferred over the accuracy index.

Data summary. Our dataset contains 19 subjects. We then segment and label each epoch based on the epoch size. For instance, with 5s epochs (80% overlap), our data consists of 35,558 and 8,845 samples for awake and microsleep states, respectively. The ratio of negative:positive is approximately 4:1, as an essence of rare microsleep events. This imbalance problem is known to severely affect the performance of popular classification algorithms. Thus, we downsampled the awake set to the same amount of microsleep data in each experiment and put this imbalance ratio (number of negative epochs/number of positive epochs) in the weighted cost during training. For example, in the first iteration of the LOSOCV experiment, we left subject 18 out for testing, and we pooled samples of all training subjects, which consisted of 32,778 negative samples and 8,572 positive samples. We downsampled the negative samples to 8,572 instances (same as the positive one) and used the weighted cost of 32,778:8,572 for training. For feature-based classification, we perform experiments on four epoch sizes: 3s, 5s, 7s, and 9s. With the recommended epoch size of 5s, we illustrate the promising results of deep neural networks on learning microsleep from the collected signals.

Set 1: Leave-one-subject-out cross validation: we alternatively trained our classifiers on the data pool of 18 subjects and evaluated the trained model on the remaining subject. The final scores are the average over these 19 iterations. Table 3.2 and Table 3.5 (row 1) present our results on this setting for feature-based learning and deep learning, respectively. The hierarchical classifier achieved the best performance among examined classifiers, obtaining approximately 0.76 in precision and just over 0.8 in specificity for all epoch sizes. The model has a slight variation in sensitivity scores in which the recommended size of 5 seconds results in the highest sensitivity of 0.85. This result is expected as large value shifts are known to happen across different subjects. Nevertheless, this result shows the feasibility of WAKE for microsleep detection on unseen subjects.

Table 3.2: Evaluation scores with Leave-one-subject-out cross validation setting using feature-based classification, over 4 epoch sizes (3s, 5s, 7s, 9s).

Epoch	**Precision**	**Sensitivity**	**Specificity**
3s	0.76	0.65	0.80
5s	0.76	0.85	0.81
7s	0.76	0.61	0.81
9s	0.75	0.58	0.81

Set 2: Test-set on each subject (Test-set): we applied stratified split onto data of each subject, dividing them into two parts with the ratio 75% (training): 25% (testing) with respect to the percentage of positive and negative samples. We then trained our classifiers on the training data and evaluated the performance on the test set. Table 3.3 and Table 3.5 (row 2) present our results on this setting for feature-based learning and deep learning, respectively. Among simple classifiers, RandomForest models with 20 estimators constantly achieved the best scores for each epoch size. Compared to Setting 1 (LOSOCV), this setting results in better-evaluated scores (above 0.8 for precision and nearly 0.9 for sensitivity), which can be explained by the high similarity between data within a certain subject.

Set 3: 10-fold cross validation: we conducted cross-validation for each subject's data and averaged the scores for the final results. Specifically, for cross-validation on a particular subject, we left 1/10 of the data for evaluation and trained on the remaining data. This procedure was performed ten times before we got the average scores as the representative. Table 3.4 and Table 3.5 (row 3) show our results on this setting for feature-based learning and deep learning, respectively. Similar to

Table 3.3: Evaluation scores with Test-Set validation setting using feature-based classification, over 4 epoch sizes (3s, 5s, 7s, 9s).

Epoch	**Precision**	**Sensitivity**	**Specificity**
3s	0.83	0.9	0.81
5s	0.87	0.9	0.87
7s	0.88	0.93	0.86
9s	0.89	0.95	0.88

Table 3.4: Evaluation scores with 10-fold cross validation setting using feature-based classification, over 4 epoch sizes (3s, 5s, 7s, 9s).

Epoch	**Precision**	**Sensitivity**	**Specificity**
3s	0.84	0.89	0.83
5s	0.88	0.89	0.96
7s	0.88	0.93	0.87
9s	0.9	0.94	0.90

the Test-set setting, our Random Forest classifiers were able to achieve high scores on precision, sensitivity, and specificity over all of the subjects.

With the current experiment setup, the results show that our proposed system could achieve the false alarm rate, i.e., $1-Specificity$, as low as 4% while maintaining the sensitivity of 89% (10-fold cross validation). This false alarm rate is acceptable in a natural setting such as driving [203]. As the negative-to-positive ratio in practice could be much higher than the current value of our experiment (i.e., 4:1), it could lead to an increased false alarm rate. We will tackle this issue in the next chapter by decomposing complex biosignals into patterns of interest and recognising them directly on wearable devices. Adaptive energy detectors are also used to filter out obvious background signals to reduce the false alarm rate further.

Table 3.5: Evaluation scores using Deep neural network on raw data with epoch size = 5s. We evaluated on 3 test settings.

Setting	**Precision**	**Sensitivity**	**Specificity**
LOSOCV	0.56	0.45	0.65
Test-Set	0.86	0.85	0.86
10-Folds CV	0.88	0.89	0.88

Compared to feature-based classification, deep learning models show promising and comparable performance, especially on test-set and k-fold classification. Fig. 3.21 presents the learning curves of training SorNet on the test-set setting. After 200

epochs, the training loss and precision curves reach a point of stability. The training loss, training precision, and validation precision converge to 0.015, 0.998, and 0.897, respectively. Our learning curves show that the network is sufficiently expressive and capable of learning the target classification function over the data. Poor performance of our deep networks on the leave-one-subject-out setting is expected due to the highly complex pattern of brain signals and the challenging cross-subject phenomenon. Regarding the size of an epoch, 5 seconds constantly leads to reliable performance: high precision and high sensitivity with respect to the average scores, though increasing epoch size may improve performance in some test settings.

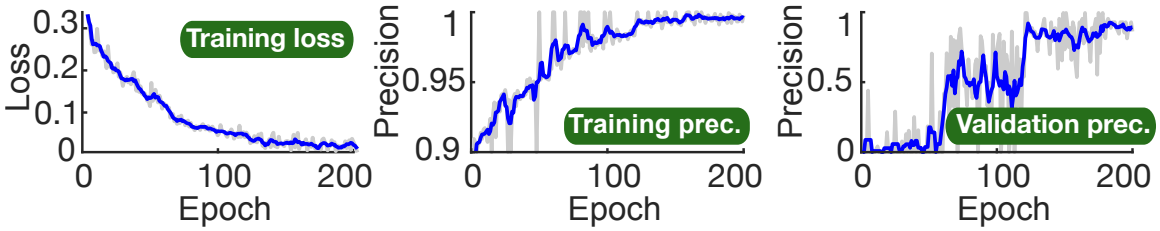


Figure 3.21: Learning curves of training SorNet on the test-set setting.

Performance on Mobile Devices: To evaluate the performance of our developed Machine Learning algorithm on a mobile device, we deploy both the features-based and deep neural network (DNN) models on a Samsung Galaxy S10 (Android 11, 1.95-2.73GHz Octa-core CPU, 128GB Flash, 8GB RAM). For each model, we run the classification 100 times and measure the maximum memory usage and inference latency. Table 3.6 presents our experimental results with the maximum memory usage and latency are 118MB and 72ms, respectively. The latency of the feature-based model is a bit higher than the DNN model since we could not find equivalent native Android libraries. Thus, it is deployed inside a virtual Linux environment on Android. Nevertheless, we can see from the results that the computational overhead of both feature-based and deep learning models is not significant, and the developed models can be readily deployed on mobile platforms.

Table 3.6: Classification performance on a Galaxy S10.

	Memory Usage	**Latency (avg./std.)**
Feature-based	74MB	72.4/19.5 ms
DNN	118MB	5.74/1.93 ms

3.8.4 Usability Analysis

WAKE Prototype Power Usage. We measured the power consumption of the WAKE prototype by using a Monsoon Power Monitor device with a sampling rate of 5 kHz. Each measurement was done in 180 seconds to get a stable result. It results in 900k data points. At 25^0C and 3.7V nominal battery voltage, the average power usage of WAKE is as follows: (1) Active state (real-time biosignals streaming with Bluetooth) consumes $241.5mW$, and (2) Idle state (no streaming with only MCU is kept running in idle mode while other components are turned off) consumes $51.60mW$. With a $600mAh$ Li-Po battery, the WAKE prototype can operate for 9.2 hours in Active and stay Idle for 43.1 hours. Further component-level measurements of usage power during Active were done by turning off each component one by one and repeating the measurements. Fig. 3.22 presents a full active power usage breakdown of the WAKE device. The sensing components (amplifiers and external ADCs) and Bluetooth communication module consume most of the system power with an average of $93.5mW$ and $85.2mW$, respectively. The storage module (uSD card) will increase an additional $90.2mW$ if it is turned on. The processing unit only consumes $62.8mW$. These numbers show the capability of our WAKE prototype to monitor the user's microsleep for a long duration. They could be further lowered by optimizing Bluetooth transmission parameters and taking advantage of the deep power-saving modes of the MCU.

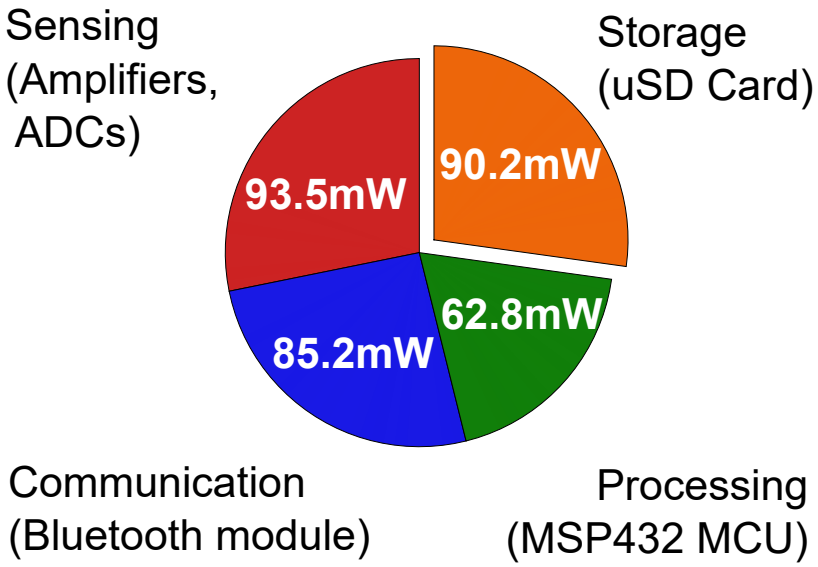


Figure 3.22: Active power usage.

WAKE Prototype Thermal Profiling. We conducted thermal measurements

for the processing unit of our WAKE prototype for 14 hours continuously. The measurement was designed to emulate the scenario where continuous microsleep monitoring is needed during normal working hours. It was divided into three states: (1) idle (the device waits for a Bluetooth connection), (2) streaming (the device streams the measured biosignals to both its onboard uSD Card and a Bluetooth-connected mobile device), (3) standby (the device stops data streaming but its Bluetooth connection is still available for future commands). The idle, streaming and standby states lasted for 1, 12, and 1 hour, respectively. Thermal data was measured by the internal temperature sensor of the processing unit and reported every 5s. Fig. 3.23 presents our measurement results. On average, the temperature of idle and standby states are 31.65 and 35.75 degrees Celsius, respectively. During streaming, the temperature increases to an average of 37.38 degrees, and the peak is 38.9 degrees. According to the standard of the American Society for Testing and Materials, 43 degrees Celsius is the threshold for prolonged use (i.e., $> 8h$) on human skin without creating any injury [204]. The temperature of our WAKE prototype is always below this threshold.

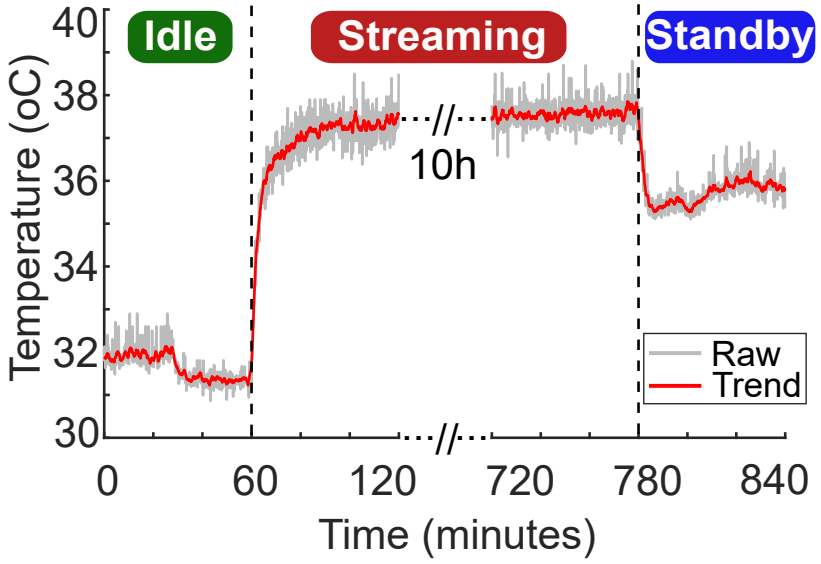


Figure 3.23: CPU core thermal profile.

User Study. We conducted a survey to evaluate WAKE's usability. We distributed our survey to the 19 subjects in our MWT study and 17 other people on the campus after they had used the WAKE device for at least two hours. In total, 36 people answered our survey. Fig. 5.21a presents the questions that we used to ask our participants' opinions on their experience with the WAKE device. The results show that over 85% of people felt comfortable with our WAKE device and were willing to

wear it during daily mental fatigue tasks, such as during driving, night-time working, etc. 91.6% of people agree that the WAKE device is more comfortable than the 'ground-truth' device used in PSG. 62.5% of people found it easy to use the WAKE device, while 16.7% had some difficulties with skin preparation and putting on the conductive paste.

We noticed that people with eyeglasses are most likely to be affected by wearing WAKE, as both devices need to be rested on users' ears. Thus, to evaluate the compatibility of WAKE and eyeglasses during daily activities, an additional user study on a population of eight people was conducted. In this study, we asked the users to wear both WAKE and eyeglasses during their daily working time for 3-4 hours. They were asked to wear WAKE before wearing eyeglasses so that the eyeglasses' temple tips could sit on top of the WAKE's silicone earpieces. The survey questions and results are presented in Fig. 5.21b. All users reported that they did not feel disturbed during their normal activity and could easily wear WAKE and eyeglasses. 75% of the users agreed that it was comfortable to wear both devices for long hours thanks to the softness of the silicon. Only two users had slight discomfort because of additional weight and the gripping force of WAKE earpieces.

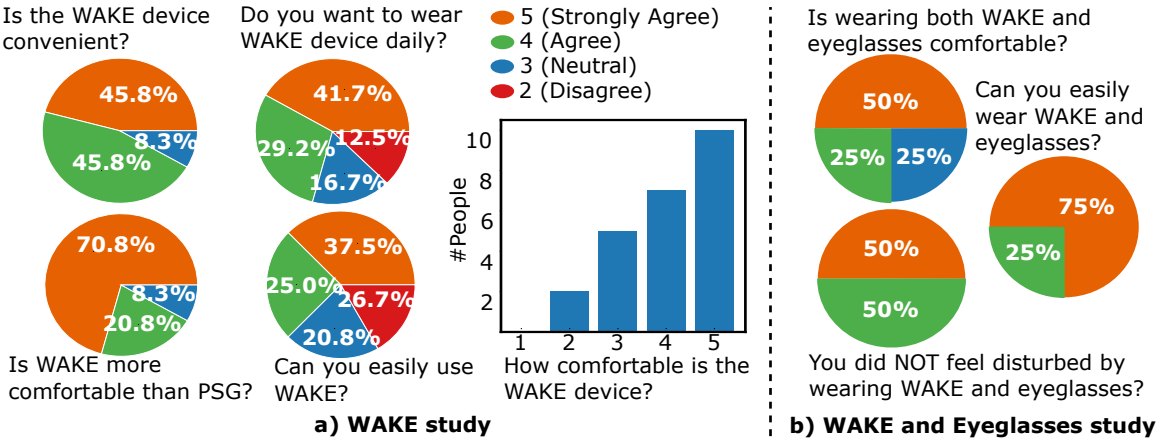


Figure 3.24: User study.

3.9 Summary

In this chapter, we presented WAKE, a novel compact, lightweight, and socially acceptable wearable device to detect microsleeps from behind the ears. We proposed the Three-fold Cascaded Amplifying technique to remove the impact of motion artefacts and environmental noise. We evaluated the motion, environmental noise suppression,

and microsleep detection performance on 19 subjects. WAKE can reduce noise power by 9.74 - 19.47 dB in different practical scenarios such as walking and driving. We develop a classification model based on the core biomarkers of microsleep captured by WAKE. WAKE achieves 76% precision and 85% recall in detecting microsleep in LOSOCV.

Chapter 4

PROS: efficient Pattern-dRiven cOmpressive Sensing for Low-Power Biosignal-based Wearables

4.1 Introduction

In this chapter, we explore the challenges of building a new event-driven compressive sensing framework, called PROS, that could enable highly energy-efficient wearables for biopotential-based applications. We develop PROS based on the sparsity nature of biosignals and events. Specifically, PROS consists of tiny pattern recognition primitives and a pattern-driven compressive sensing algorithm that work together to significantly reduce transmission rate while maintaining high fidelity signal. PROS also enables the ability to react to critical events immediately on the device.

4.1.1 Goals and Challenges

While we currently focus on EEG, EOG, and EMG biosignals and a head-worn form factor in this study, PROS is also applicable for a variety of healthcare wearable devices such as smartwatches, earphones, smart clothes, etc., where achieving continuous, high-fidelity biosignal streaming, low-latency responses, and long battery life is critical to their applications.

Challenges: To realise PROS, we face the following challenges: (1) biosignal events (e.g., seizures, microsleep, pain, etc.) require multimodal sensing channels and a complex algorithm (e.g., machine learning) to detect, which is not feasible on low computing resource wearable; (2) we lack a reliable domain with high sparsity to compress biosignals on the device effectively; (3) low power wearable devices have extremely constrained computing resource, i.e., an MHz microcontroller (MCU) and KBs of system memory, making it challenging to deploy advanced computations without consuming significant energy.

Contributions: To overcome the aforementioned challenges, we make the following contributions:

1. We identify the pattern primitives of biosignals such as EEG, EOG, and EMG and develop tiny recognition models (TinyPR) for continuous on-chip detection and low-latency responses.

2. We devise a pattern-driven compressive sensing (PDCS) technique to efficiently compress the captured signal pattern with appropriate wavelet domains, boosting the compression factor and recovered signal quality.

3. We design a hardware platform and employ optimisation techniques in both hardware and OS levels to support advanced signal processing and neural network operations of PROS.

4. The prototype of PROS is evaluated on two open datasets of 120 subjects. In a practical use case such as epileptic seizure detection, PROS can reduce the data rate by 24X, boost the power efficiency by more than 1200%, and enable real-time responses within 10s of milliseconds while maintaining high fidelity signals.

4.2 PROS System Overview

We design PROS with three objectives, (1) detect signal patterns of interest (PoIs) directly on-chip to eliminate most of the irrelevant signal, (2) compress the detected PoI by using the recognition information to reduce wireless transmission rate further, and (3) optimise hardware and OS operations to enhance system's efficiency. To achieve these goals, we develop three main components for PROS as illustrated in Fig. 4.1, (a) a firmware framework that detects and compress the PoIs by using our developed tiny embedded machine learning models and a pattern-driven compressive sensing algorithm, (b) a low-power hardware platform that accelerates advanced signal processing and embedded machine learning operations, and (c) a mobile app to recover the compressed PoIs for further processing.

Efficient Features Extraction Pipeline. We design a feature extraction pipeline based on the characteristics of biosignals to compute Mel-frequency cepstral coefficients (MFCCs) from the signals. As the computing resource is highly constrained, we tune the MFCC processing so that the output contains enough vital information

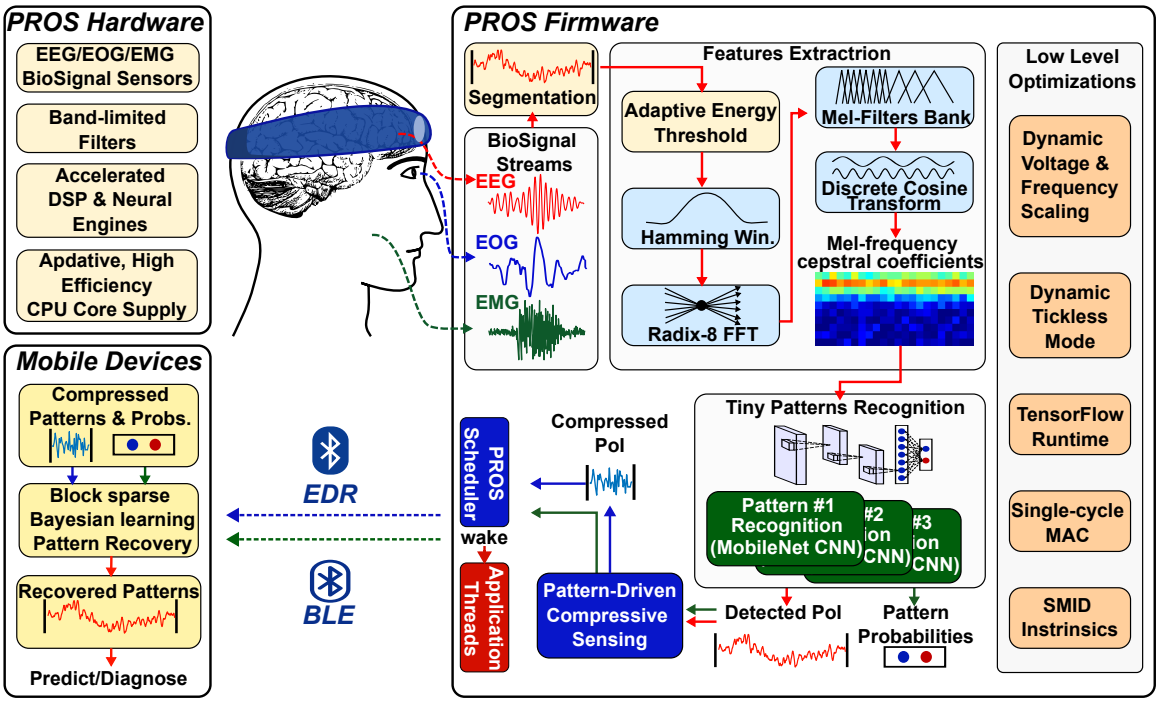


Figure 4.1: PROS system overview.

about each pattern while being small and lightweight. Additionally, we employ accelerated signal processing methods available on the hardware to boost the processing speed.

Tiny Pattern Recognition Primitives. We develop tiny pattern recognition (TinyPR) primitives to effectively detect patterns of interest (PoIs) from the captured biosignal streams. Each primitive only detects one PoI to ensure its efficiency and flexibility in different applications. Each primitive only needs a few KBs of memory and milliseconds of CPU time to operate. We use TensorFlow Lite Microcontroller (TFLM) runtime and vectorized neural operations to push the inference performance.

Pattern-Driven Compressive Sensing. To further reduce the amount of communication, we devise a novel compressive sensing technique to exploit the sparsity property of each PoI. We adaptively apply the optimal compression ratio and wavelet domain transformation based on the pattern recognition information. We use a random binary compression technique to compress the signal with minimal system overhead on the wearable device. To reconstruct the signal, we employ a state-of-the-art Block Sparse Bayesian Learning recovery algorithm combined with pattern recognition information to reduce the required compressed samples by taking advantage of the sparsity of biosignal. As a result, we could achieve a high compression factor and reconstruct the signal with high fidelity on the mobile application.

Hardware and OS Optimisations. To further enhance the processing efficiency, we implement hardware and OS optimisations such as (i) dynamic voltage and frequency scaling, (ii) dynamic tickless mode, and (iii) adaptive energy thresholding. We also develop a scheduler that provides configurations and wakes the application's threads when subscribed PoIs are detected.

PROS Hardware and Mobile Application. We design a low-power hardware platform from the ground up to support PROS. To enable advanced optimisations on the firmware, we equip it with an energy-efficient signal processing and neural network processor and an adaptive, high-efficiency CPU core supply. We also developed a lightweight signal reconstruction algorithm on mobile devices to reconstruct the compressed PoI with high fidelity. The reconstructed PoI could be used for further processing or diagnosis.

4.3 Tiny pattern recognition primitives

This section presents our end-to-end pipeline, called TinyPR, for recognising biosignal patterns. The key contributions of TinyPR are (1) identifying the generic biosignal pattern primitives that are feasible to be efficiently recognised on the low-power hardware and (2) providing a design strategy that can be both accurate and lightweight for those pattern primitives. The developed pattern recognition models can be served as building blocks for biopotential-based applications requiring on-chip pattern recognition. We first highlight key challenges and insights into the design of our framework.

4.3.1 Key challenges and designs

As per our system requirements, the target recognition model should be highly expressive to detect the biosignal patterns but also resource-efficient for the MCUs' deployment. This expressive-efficiency trade-off poses a critical challenge for our system design.

Detecting biosignal patterns has remained challenging, despite some positive outcomes in preliminary works [205, 206]. Biosignals are highly irregular and heterogeneous [207] due to the complexity and intrinsic properties of biosystems, causing the difficulty in understanding and detecting the interest patterns [205, 208]. For instance, recent works [206, 208] find that most existing approaches are ineffective for learning patterns for clinical analysis and event detection. Besides, the scarcity of interest patterns [209, 210] in biosignals makes the learning even harder: the training data is heavily imbalanced. The resource restriction of MCUs adds another challenge

to our design. With limited computing resources in terms of memory, operations, and computation capacity, MCUs require the inference system to have low memory footprints (e.g., a few KBs) and low inference latency.

Existing methods of biosignal learning are mainly based on either the deep learning approach or feature-based machine learning approach [206, 211]. While achieving high recognition performance and being easier to implement on hardware, deep learning models are usually too large for MCUs. On the other hand, simple feature-based learning models are more resource-efficient but not sufficiently and robustly effective at detecting complex patterns [206]. In this work, we propose the combination of the feature-based approach with deep learning: utilising an informative feature extractor to reduce the burden of learning domain knowledge features. Moreover, we can significantly reduce models' sizes by leveraging quantisation techniques without degrading the recognition performance [212].

Subject variation is also a challenge for biopotential-based systems. While biopotential signals vary among people, our intuition is that they contain similar patterns due to the typical structure of the human body. For example, eye blink signals usually have two opposite consecutive peaks corresponding to the closing and opening phases of the eyelids; alpha brainwaves typically have cyclical or rhythmic changes with a frequency from 8 to 12Hz when the brain neurons become synchronised in a relaxed state. Therefore, our intuition is that if we train the TinyPR models to target common and generic signal patterns of interest, these models could generalise well to unseen subjects. We present the detection performance evaluations of our developed TinyPR models in detail in Sec. 4.7.

4.3.2 Pattern Recognition as the Rare Event Detection Problem

Most target patterns rarely occur in biosignals. For instance, seizure events usually account for less than 5% in EEG recording data [209]. This results in the highly skewed distribution of training data. Standard methods for event detection and feature selection may not work well with the imbalanced data [213] because they tend to learn features only from the major classes (background signals) and may easily misclassify the minor classes (targeted patterns).

Therefore, we cast our pattern identification problem as the rare event detection problem [214]. Solving this problem requires adopting either supervised or unsupervised techniques for rare-event detection [213]. The latter requires large models with an enormous amount of unlabeled data, which is not feasible for deploying MCUs.

Hence, we focus on the supervision approach to design a more lightweight classification model. In particular, considering the target patterns as positive and the rest patterns as negative, the problem becomes a binary classification task. We note that data distribution is highly skewed as positive data is much smaller than negative data. To deal with this issue, we apply the SMOTE [215] method to upsample the positive patterns. The next section will present the design of our feature extractor and binary classification model.

4.3.3 Informative Feature Extraction

Powerful prior knowledge via informative feature extraction can significantly reduce the complexity of recognition models. Mel Frequency Cepstral Coefficients (MFCC), together with Wavelets transform, are the two most common approaches used for extracting biosignal features [216]. Since the computing resource and energy on low-power microcontrollers (MCU) are highly constrained, we only pick the features that are informative while being resource-efficient. MFCC features fit well with these criteria as multiple previous works [216, 217] have proved that MFCC features are reliable in detecting biosignal (EEG/EOG/EMG) events. Furthermore, there are available components in the optimised firmware library, such as ARM-CMSIS, for efficient implementation. Efficient performance is critical for low-power MCUs since heavy processing can easily outweigh any benefits of data reduction.

As MFCC is initially used for audio signals, we configure its components to extract useful features from biosignal data. We note that most of the information in biosignals (EEG, EOG, EMG) locate at the low-frequency bands ($<$ 300Hz) [218, 219]. We, therefore, use only ten bands among 39 features of MFCC to extract essential features, further helping reduce the input size of the recognition model. We use Hamming window to slice the signals into slicing frames. Note that sudden chop-off at the frame's edge can lead to a noisy signal because of the sudden amplitude drop. Hence, we gradually drop amplitude near the edge of frames. We apply the Discrete Cosine Transform to extract features in the frequency domain and triangular Mel-scale filter banks to transform the signal to the Mel-scale power spectrum. Given these features, we can now build an efficient classifier.

Though it is possible to extract meaningful features with autoencoders automatically, it is not efficient on low-power microcontrollers. It has been pointed out in [220] that directly extracting features would be much more energy and computationally efficient by taking advantage of the accelerated library of the targeted hardware. Thus,

we design our TinyPR models around optimised signal processing and neural operations provided by TinyML frameworks such as TensorFlow Lite Microcontroller [220] and CMSIS-NN [221].

It is also important to note that while MFCC could extract temporal and spectral features well, these features might not be sufficient for all applications. Thus, we envision that PROS serves as an open framework where multiple processing pipelines and pattern recognition models could be developed for various applications.

4.3.4 Efficient Design for Recognition Model

We build a deep classification model on top of the extracted MFCC features to complete the recognition framework. The resource constraints pose two questions for our design: how to design the best-fit model given particular conditions on memory and power and how to efficiently run the model on MCUs. We wish to achieve these objectives without degrading the recognition performance.

Efficient Architecture. Recent works of TinyML [119], or machine learning for edge devices, provide potential solutions to our problem. TinyML aims to shrink sizeable deep learning models (millions to billions of parameters) into tiny models of a few KBs, mainly by changing the network topology to remove the redundant parameters [103, 104], reducing the input size, or loading only parts of the network to the memory to address the memory bottleneck [120, 222]. However, existing models are not directly applicable to our PROS system because the shrunk models' sizes are still relatively larger than our desiderata, and the designs are primarily specific for image signals instead of biosignals. Therefore, we derive a simple yet powerful architecture for our system based on the recent advances of TinyML [119].

The critical component of our architecture is the block of depthwise convolution (DW-Conv) and pointwise convolution (PW-Conv) [103], which has been proven helpful in multiple resource-aware models, such as MobileNets [103, 104] and MicroNets [108]. DW-Conv is a type of spatial convolution that applies independently on each channel of inputs. PW-Conv uses a 1×1 kernel to iterate every point, further linearly combining DW-Conv outputs. Compared to the standard convolution, DW-Conv and PW-Conv require much smaller numbers of parameters, thus being more computationally effective [223]. Also, these operations are supported by the micro deep learning framework TFLMicro [113].

Our architecture consists of a convolutional layer as the input layer, followed by a sequence of DW-PW-Conv blocks, a Dropout layer, and a linear layer. Each DW-PW-Conv block is a stack of a DW-Conv and a PW-Conv with batch normalization and

a Relu activation. Under different systems, we control models' sizes by varying the number of DW-PW-Conv blocks and channels' sizes to fit the MCUs' requirements. In particular, we apply the search approach in MobileNetV2 [104] to search for the architecture's configurations achieving the best trade-off of efficiency and recognition accuracy. For the deployment on MCUs, we use the TensorFlow Lite Microcontroller framework [224] to compress the model into the numeric domain, reducing the memory footprint and speeding up the computation.

Post-training Dynamic Range Quantisation. To further reduce the model's size for inference, we apply the dynamic range quantisation technique [225]. While the floating-point format is used for parameters of most deep learning models to achieve better precision during training, it may be costly to store floating-point numbers, especially on low-memory edge devices. Quantisation techniques help solve this issue by converting trained parameters into another number representation. For instance, converting the commonly used float32 format to the int8 format helps save 24 bits. Weights are converted back into float32 format during the inference for better classification performance. We find that the recognition performances in metric-wise are nearly identical to the original ones after this transformation.

Memory complexity. Our final models have only a few KBs in size, highly optimised compared to MobileNetV1 and MobileNetV2 (16.9 MBs and 4 MBs, respectively). We attribute this tremendous compression mainly to the use of an informative MFCC feature set: with a small size (10 features (Section 4.3.3)) and with a low dimension (22 dimensions). The input size of classification models reduces from $224 \times 224 \times 3$ (for images in MobileNets) down to 10×22, leading to small numbers of convolution channels and layers required to learn the feature representation. As a result, our smallest models have only nearly 3.5K parameters in total.

Inference with Confidence. Together with producing accurate predictions, an essential requirement for recognition models in practice is to provide confidence in the prediction. Inspired by the clinical procedure in diagnostics, we impose the confidence level on the pattern recognition result. Together with each classification's output (binary value), our model produces a confidence score representing the certainty of the prediction. This score is generated by thresholding the soft-max scores of the binary classes. The application can choose the threshold to make a trade-off between sensitivity or specificity depending on its requirements.

At this stage, we could eliminate most of the irrelevant signals. However, as we still need to transmit the captured PoI signals, we need to compress the data to reduce the transmission rate further.

4.4 Pattern-Driven Compressive Sensing

This section discusses the challenges and our proposed Pattern-driven Compressing Sensing (PDCS) technique to reduce the amount of wireless communication in our system. While downsampling is a popular technique to reduce the data rate, it has been shown that it could significantly degrade the quality of biosignal analysis and induce higher noise and aliasing [226]. In this study, we employ the compressive sensing (CS) theory as it could avoid signal degradation while requiring minimal system processing and memory overheads, both of which are critical for low-power biopotential-based wearables [227]. It bases on a fundamental assumption that biosignals have sparse representations in a transformed domain such as frequency or time-frequency (e.g., wavelets) [228, 118]. Thus, sampling the signal based on the fastest frequency component based on Nyquist–Shannon theory is redundant [229].

The key contribution of PDCS is the ability to incorporate pattern recognition information to build an efficient data-driven compressive sensing method. Conventionally, compressive sensing techniques are deployed on low-power devices due to the simplicity of the compression. The performance, however, depends heavily on the choice of the compression ratio and sparse domain basis. Since pattern recognition information was unavailable in previous works [230, 95] due to energy and computational resource constraints, the compression ratio and sparse domain basis are often chosen and tuned offline based on pre-collected data and apply to the whole signal during runtime. It leads to significant variations and inconsistent performance with non-stationary biosignals such as EEG [99]. By enabling energy-efficient on-chip pattern recognition, we can recognise and apply different compression ratios and sparse domain basis for each signal pattern in real time.

It is also important to note that the merit of PDCS is complementary to TinyPRs. For example, assuming TinyPRs could reduce the transmission rate by M times by eliminating the irrelevant signal and PDCS compress the detected signal by N times on average, we will have the total compression ratio of $M \times N$. Furthermore, the theoretical computational (and energy) cost of PDCS is much lower than running a TinyPR model on the wearable device, i.e., only one matrix-vector multiplication versus a convolutional neural network inference, making the return on investment of PDCS significant.

4.4.1 PDCS framework design

We design our PDCS framework as illustrated in Fig. 4.2. PDCS is a digital CS design where we perform compression after digitalisation. This design has the advantage that we could use precision, high-rate ADC (e.g., $\Sigma - \Delta$ modulated ADCs [231]) to avoid high-frequency noise and aliasing. PDCS has four important steps as follows.

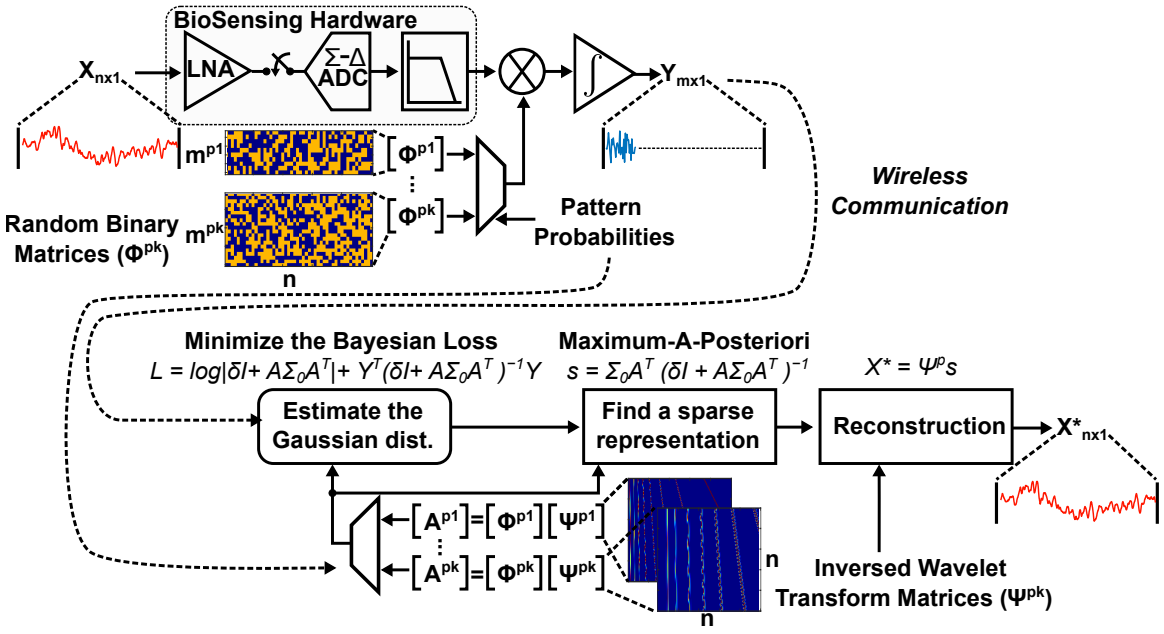


Figure 4.2: Pattern-driven Compressing Sensing framework.

First, we identify the domain and the transformation basis $\Psi_{n,n}$ where the input signal $X_{n,1}$ has a sparse representation $s_{n,1}$, i.e., $X = \Psi s$. As sampling with CS is based on the inherent information contained in the signal rather than its frequency bandwidth, the higher sparsity of the representation s, the less information is presented in the signal. Hence, a lower number of measurements is needed. The sparsity and density are defined as the percentage of zero and non-zeroes values in s, respectively. $Sparsity = 1 - Density$.

Second, we choose an i.i.d random distribution to construct our measurement matrices Φ. We construct multiple Φs for various compression factors (CFs). To ensure the compressed signal can be successfully recoverable, the coherence (μ) between Φ and Ψ, i.e., $\mu = \sqrt{n} * \max_{1 \leq i,j \leq n} |\langle \Phi_i, \Psi_j \rangle|$, is employed [118]. Lower μ ($1 < \mu < \sqrt{n}$), means more efficient compression. Random measurement matrices such as Gaussian, Bernoulli, Binary, etc. have low μ with any basis [232, 233]. Thus, they are employed as universal encoders.

Third, on the wearable device, we compress the captured biosignals ($X_{m,1}$) based on pattern recognition information from TinyPR primitives (Sec. 4.3) and desired CFs ($CF = n/m$) for each pattern (p), i.e., $Y_{m,1} = \Phi^p_{m,n} X_{n,1}$, with m and n, are the sizes of the compressed (Y) and the original signal (X), respectively. m should be much smaller than n for the compression to be effective. We transmit Y together with its recognition information to help with the recovery.

Fourth, at the receiver side (e.g., tablets, phones), we find a sparse representation $s_{n,1}$ by minimising a Bayesian loss function. Using the received pattern recognition information, we dynamically apply different basis functions ($\Psi^p s$) to the Block Sparse Bayesian Learning (BSBL) algorithm to get the optimal results. The original signal is recovered by $\hat{X} = \Psi^p s$.

We tune CF based on the acceptance loss of the recovered signal. The configurations are evaluated on sample datasets to ensure satisfying accuracy. We measure the loss of the CS method by using the Structural SIMilarity index (SSIM) [234]. We employ SSIM in this study since it has better performance on structured signals [234]. The higher SSIM is better. SSIM = 1 means perfect recovery.

4.4.2 Sparsity variations among patterns

Finding the optimal domain where biosignals have sparse representations is the most crucial task and the most non-trivial one. Previous works on compressive sensing with biosignals show the feasibility of biosignals such as EEG, EOG, and EMG to have sparse representations in time-frequency domains such as Gabor, Spline, and Wavelets domains [235, 236, 237, 99]. However, as they do not take into account individual signal pattern structure, many studies have reported large variations in the reconstruction accuracy among different channels and trials [99, 238].

Fig. 4.3 confirms the significant sparsity variations among different biosignal patterns in the same Daubechies 2 (*db*2) wavelet domain. Six biosignal patterns are extracted from an open biosignal dataset [239]. They include (1) eye blink (EOG), (2) spike-and-wave (EEG), (3) absence seizure (EEG), (4) chewing (EMG), (5) tonic-clonic seizure (EEG), and (6) muscle contraction (EMG). We apply the same discrete wavelet decomposition with seven levels. By keeping the recovery similarity index, i.e., SSIM, to be at least 0.9 between the original and recovered, we can find the minimum number of wavelet coefficients that are needed to reconstruct the original signal.

The *db*2 mother wavelet has a high structural correlation with eye blinks patterns. Thus, fewer wavelet coefficients are needed to reconstruct the original signal with

only 4% density. On the other hand, the db2 wavelet works poorly with chewing, tonic-clonic seizure, and muscle contraction patterns. Their densities are 15, 22, and 36%, respectively. Up to 9X can be observed in the density difference among these biosignals; hence, finding the optimal wavelet domain is a significant challenge that we need to address.

4.4.3 Optimal wavelet domains search

In this study, we assume that a universal wavelet domain for all the biosignal or even each signal group, such as EEG, EOG, or EMG, might not exist. However, there exists **an optimal wavelet domain for individual biosignal pattern**. Thus, by knowing the pattern of the interested signal, we could choose the appropriate sparse wavelet domain for each pattern to get the best compression factor. This is not possible in conventional CS systems [99, 240, 236, 230] where we lack the pattern recognition ability from biosignal streams. Hence, we have to trade off between signal fidelity (i.e., by using the smallest CF) or compression factor (i.e., by accepting the loss with low sparsity patterns). Sec. 4.3 discusses how we overcome this challenge by capturing pattern information directly on low-power hardware. The next step is to find the optimal wavelet domain for each biosignal pattern of interest.

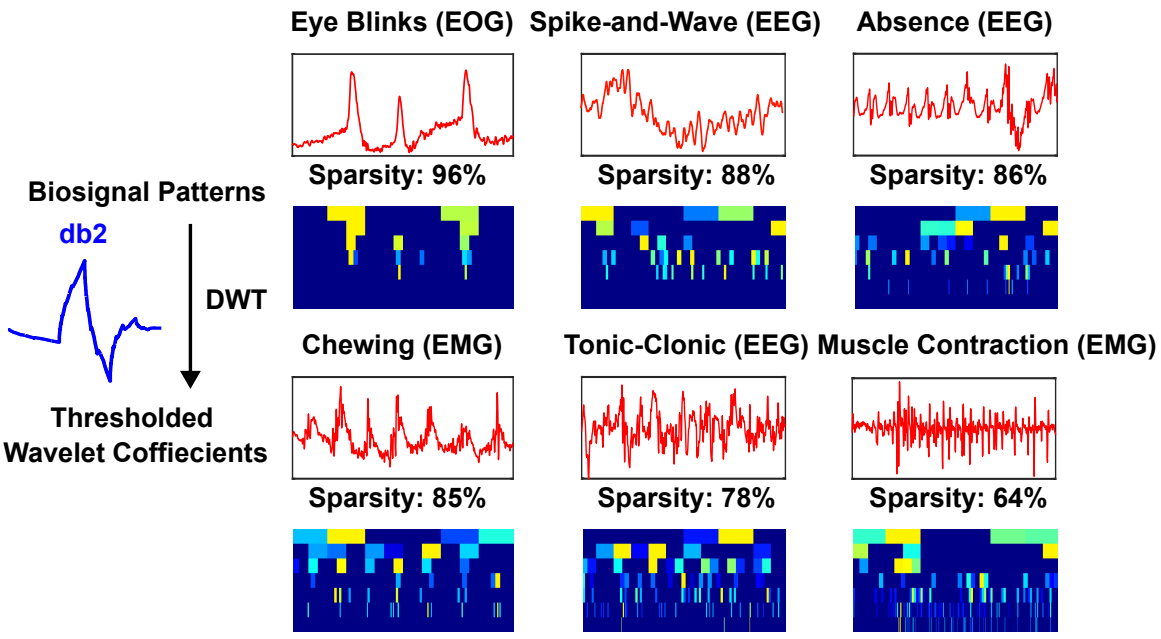


Figure 4.3: Sparsity variations among patterns in a wavelet (db2) domain.

There are several quantitative metrics in literature to choose the optimal wavelet domain, such as maximum cross correlation [241], mean squared error [242], con-

tinuous wavelet coefficients [243], minimum description length [244], etc., that are used for biosignals such as EEG, EOG, EMG, or ECG. They are based on the intuition that the optimal wavelet domain will have the highest similarity between its transformation basis and the input signal [245]. They, however, could not tell us the sparsity of a signal pattern, making it difficult to estimate the compression factors. Furthermore, some studies also point out that similarity-based methods might not always result in optimal wavelet domains [245]. To alleviate this issue, we propose another selection metric called Maximum Sparsity Index (MSI). We define MSI as the maximum percentage of discrete wavelet coefficients that are not significant to reconstruct the signal.

Algorithm 1: Optimal wavelet domains search

```
input : ssim_thr /*Minimum desired recovery quality*/
        wavelets_list /*Wavelet domains search space*/
output: best_W /*a wavelet domain with the highest MSI*/
best_W ← None;
best_MSI ← 0;
for W in wavelets_list do
    MSIs ← None
    for X in signal_list do
        coeffs ← DWT(X, W)
        for thr in thresholds_range do
            s ← thresholding(coeffs, thr)
            X̂ ← IDWT(s)
            if SSIM(X, X̂) ≥ ssim_thr then
                MSI ← zeros(s)/len(s)
                break
        MSIs.append(MSI)
    if best_MSI < avg(MSIs) then
        best_MSI ← avg(MSIs)
        best_W ← W
return best_W;
```

Listing 1 presents our search algorithm. Since there could be an infinite number of wavelet domains [246], we only pick out 70 mother wavelet functions in six families such as Daubechies (db1-15), Coiflet (coif1-5), Fejér-Korovkin (fk4-fk22), Symlet (sym2-15), Biorthogonal Spline (bior1.1-6.8), Reverse B-Spline (rbior1.1-6.8), that are commonly used for biosignals [247, 248, 249]. For each mother wavelet function, we apply Discrete Wavelet Transform (DWT) to the input signal (X) with five

decomposition levels to get its wavelet coefficients ($coeffs$). The number of decomposition levels is chosen to extract all the frequency information inside the input biosignals [250]. As $coeffs$ is near sparse (i.e., the coefficients that are significantly larger than zero are sparse), we iteratively apply different thresholds to get a sparse representation (s).

We then quantify the quality of the reconstructed signal ($\hat{X}$) from s by applying the Inverse Discrete Wavelet Transform (IDWT) and calculating the SSIM index. Only the ones with $SIMM \geq ssim\_thr$ are kept. The $ssim\_thr$ we used for optimal wavelet domain search is 0.9. From our preliminary evaluations, this is sufficient for the signal to maintain its quality similar to the original (as discussed in Sec. 4.7). Note that this threshold is adjustable depending on the application's requirements. The sparest wavelet representation is the one that has the largest threshold. We calculate MSI by finding the ratio of non-zeroes components in s. The optimal wavelet domain is the one that has the smallest average MSI for all the input signals of the same pattern group. Finally, we repeat the same process to find optimal wavelet domains for all the patterns.

It is important to note that we only use DWT and IDWT to quantify patterns' sparsity, not running them on either the wearable or mobile device. After knowing the optimal domains, we can construct different Φ and Ψ matrices for individual patterns and store them on wearable and mobile devices. However, the conventional compressive sensing theory would require the compressed sample size to be around four-time the density of a sparse representation [233], making it very challenging to work on near-sparse biosignals. E.g., a muscle contraction pattern (Fig. 4.3) with 36% density will not work as it requires the compressed signal to have 1.44X more samples than the original signal.

4.4.4 Recovery with Pattern Information and Block Sparse Bayesian Learning

We devise an efficient reconstruction algorithm based on received pattern recognition information and the Block Sparse Bayesian Learning (BSBL) technique [251] as illustrated in Fig. 4.2. BSBL technique help to address the issue of a high compressed sampling rate by taking into account the temporal sparsity and correlation among signal blocks.

To apply the BSBL technique, we consider a window of signal (s) of size N as a series of blocks of size d, i.e.,

$$s = [\ \underbrace{s_1\ \ldots\ s_d}_{s^T[1],\ \text{1st block}},\ \underbrace{s_{d+1}\ \ldots\ s_{2d}}_{s^T[2]},\ \ldots\ , \underbrace{s_{N-d+1},\ \ldots\ , s_N}_{s^T[N/d]}\ ]^T \tag{4.1}$$

A signal with few blocks that are non-zeroes is called a block-sparse signal. This study assumes that the biosignal patterns are block-sparse in their respective optimal wavelet domains.

Each block (s_i) in the signal is modelled as a combination of two multivariant Gaussian distributions, i.e., the noiseless signal $p(s_i; \gamma_i, B_i) \sim N(0, \gamma_i B_i)$, and the noise vector $p(n; \delta) \sim N(0, \delta I)$. γ_i and B_i are the block sparsity control parameter and mutual correlation matrix of the i-th block, respectively. δ is a positive scalar representing the noise and I is the identity matrix. We estimate the parameters γ_i, B_i, and δ, by applying Type-II-maximum likelihood procedure to minimise the following cost function [252, 251],

$$L = log|\delta I + \Phi\Psi\Sigma_0(\Phi\Psi)^T| + Y^T(\delta I + \Phi\Psi\Sigma_0(\Phi\Psi)^T)^{-1}Y \tag{4.2}$$

where $\Sigma_0 = diag\{\gamma_1 B_1, ..., \gamma_{N/d} B_{N/d}\}$. In contrast to the conventional BSBL technique, we dynamically apply different wavelet basis (Ψ) to the Bayesian learning process based on received pattern information and its optimal wavelet domain. After the learning has converged, we find s by using Maximum-A-Posteriori estimation, i.e., $s = \Sigma_0(\Phi\Psi)^T(\delta I + \Phi\Psi\Sigma_0(\Phi\Psi)^T)^{-1}$. The signal is reconstructed as $\hat{X} = \Psi s$.

Till this point, we could significantly reduce the wireless transmission rate. However, we might reach the stage where communication is no longer the bottleneck. Thus, we will need to look elsewhere to further increase energy efficiency.

4.5 Hardware and OS Optimisations

As PROS performs neural network inferences continuously in the background, processing efficiency is critical. We will discuss in this section the hardware and OS optimisation techniques that we have adopted from the state-of-the-art to push the processing efficiency of PROS further.

Dynamic Voltage & Frequency Scaling. DVFS technique improves energy efficiency by reducing the operating frequency and voltage of the CPU core based on the workload's demand [253, 254]. We could formulate the energy consumption of a CPU core as, $E_{cpu} = (CV^2 f + VI_{static})T_{run} + VI_{static}T_{sleep}$. where C, V, f, I_{static}, T_{run}, and T_{sleep} are total gate capacitance, operating voltage, switching frequency, static leakage current, running and sleep time, respectively. As switching frequency

is directly related to the operating voltage, i.e., $f \propto (V - V_{threshold})^{1.3}$ [254], we can significantly reduce the power consumption by lowering f, which also lowers V. DVFS, however, has a point of diminishing return [255]. When we decrease f, the time required for completing a task (T_{run}) increases, leading to increased static energy consumption due to I_{static}. We confirm this phenomenon on an ARM MCU. As we can observe from Fig. 4.4, the power efficiency of the CPU core increases up to 30%, i.e., from 7.1 to 9.1 DMIPS/mW (Dhrystone Million Instructions per Second per milliwatt) when we reduce f_{max} from 120 to 26 MHz and V from 1.14 to 1.05V. However, this scaling is not linear as the static power becomes dominant at lower fs, i.e., the efficiency drop to 0.5 DMIPS/mW with $f = 100kHz$.

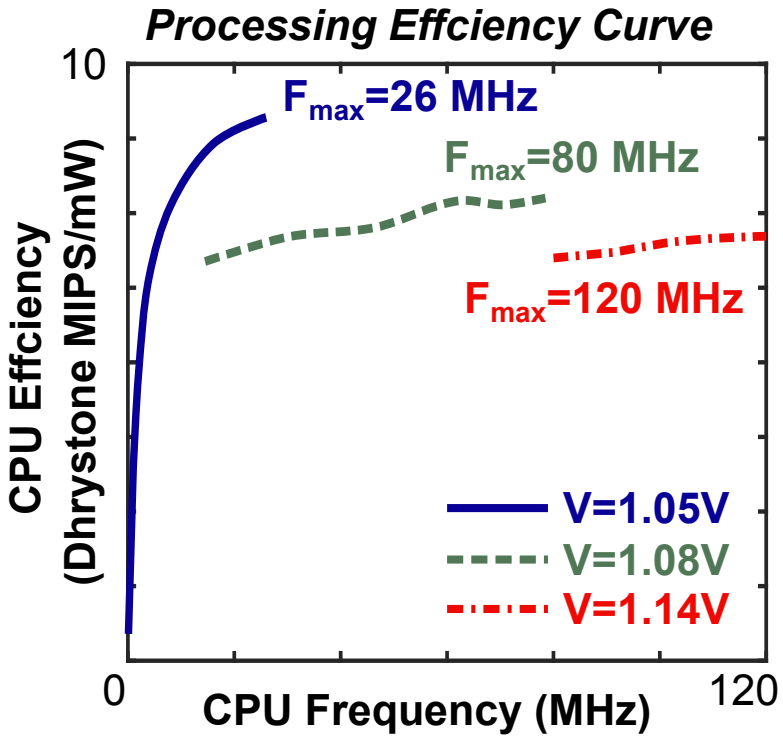


Figure 4.4: Efficiency Curve.

To address this issue, we develop a lightweight DVFS algorithm inside the PROS kernel. It is based on the principle that the CPU should run at the maximum frequency of the lowest possible voltage while still meeting the deadline (T_d). The deadline can either be the time when a signal window is returned by the DMA (Direct Memory Access) peripheral or the desired value set by the developer. It follows three steps as follows. First, we set f and V at the maximum values (f_{max}, V_{max}) and run all the background processing (e.g., tinyPR primitives, preprocessing, com-

pressive sensing, etc.) required by the application to measure the CPU time (t_0). Second, we estimate the lowest possible CPU frequency that still meets the deadline, i.e., $f_{min} = \lceil T/t_0 * f_{max} \rceil$. From f_{min}, we can find the lowest possible voltage range (V_{min}) that could support f_{min}. Finally, we set the CPU frequency to the maximum f, supported by V_{min}. This is the optimal frequency for our workload. Depending on the application's workload dynamic, we can run DVFS once at the system startup or run it every scheduling cycle.

Dynamic Tickless Mode. Many OSes such as Linux or FreeRTOS [134] use a global hardware timer generating periodic ticks (e.g., 100 or 1000 ticks per second). This is a nice and simple timebase for OS tasks such as scheduling or synchronizations [256]. However, it negatively impacts low power performance as the CPU is constantly wakened up from its sleep mode every the timer interrupts fires. This leads to significant energy loss due to constantly waking up. Fig. 4.5 illustrates the energy consumed by switching back and forth between wake and sleep mode every 1 ms will outweigh any energy saved by putting the CPU to sleep.

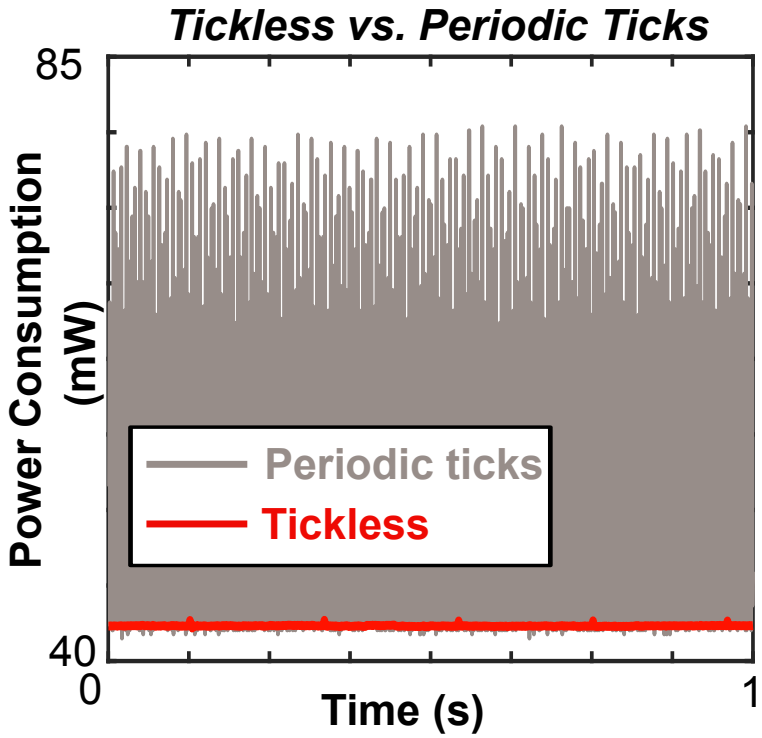


Figure 4.5: Tickless Sleep.

To address this issue, we employ the dynamic tickless mode (dyntick) [256] for PROS. Dyntick eliminates the periodic timer interrupts when the system is idle. The

CPU is put into sleep mode until the next task is ready to run or an interrupt is fired. Since the kernel still needs to wake up when its tasks are ready, we implement a low-power timebase (e.g., the real-time clock peripheral on ARM Cortex-M MCUs) that can still run while the CPU is in sleep mode. We set the alarm on this low-power timebase to wake up the CPU when its tasks are ready. We also use it to track how much time the CPU has slept to adjust the kernel timebase. This significantly reduces the energy wasted due to constantly waking up while maintaining the OS kernel's proper operations.

Adaptive Energy Threshold. Our tinyPR primitives (Sec. 4.3) are powerful tools to recognise PoIs. However, they might be too expensive to run on obvious background signals. Thus, we apply a lightweight adaptive energy threshold method, which is quite effective in eliminating non-stationary background noise in speech recognition systems [257, 258]. After a signal segment of size n has been captured, we calculate its energy by $E = \frac{1}{n}\sum_{i=1}^{n}|x(i)^2|$ and compare with a threshold value. The signal segment with a lower energy level is eliminated. We adaptively update the threshold value (λ) based on m previous measurements of background and PoI signals by $\lambda = \alpha\sum_{j=1}^{m}\frac{1}{j}E_{bg} + \beta\sum_{j=1}^{m}\frac{1}{j}E_{PoI}$ [257]. As the definition of background signals varies from one application to another, we will need to adjust α and β accordingly.

PROS Abstractions. To provide a friendly interface for application developers, we wrap up all underlying processing procedures with the PROS scheduler. The scheduler provides the application with the interfaces to (1) set up and configure the TinyPR primitives needed by the application and (2) wake up the application threads for real-time responses when a subscribed PoI is detected. It also handles background operations such as running TinyPR primitives and the PDCS algorithm.

We wrap the tinyPR models, pre-processing pipeline, and PDCS algorithm as C++ classes and implement the PROS scheduler as a FreeRTOS task. At the initial state, the developer can declare the TinyPR models, confidence threshold, compression ratio, and their mapping to application tasks. During runtime, if the output probability of the positive class is over the defined threshold, the scheduler will notify the subscribed tasks for execution. Direct task-to-task notification of FreeRTOS is employed to ensure efficiency. The notified task could request access to the signal data buffers, but it will need to make a copy before they are overwritten.

4.6 Implementation

PROS Firmware Framework. We implement PROS based on the FreeRTOS real-time kernel, which provides the base OS functionalities: preemptive task scheduling, dynamic memory management, and synchronisations. We implement additional optimisation modules: DVFS, and dynamic tickless sleep mode, then integrate them into the FreeRTOS kernel. We train our TinyPR primitives on an Nvidia RTX 3090 GPU and use the TensorFlow Lite Microcontroller to perform inferences on PROS hardware. The neural network operations, MFCC calculation, and adaptive energy detector are accelerated by SIMD (Single Instruction Multiple Data) and single-cycle MAC (Multiplication-and-Accumulation) instructions. We use pre-generated binary matrices stored in MCU's FLASH to perform the PDCS algorithm. We also implement the optimal wavelet search algorithm in MATLAB.

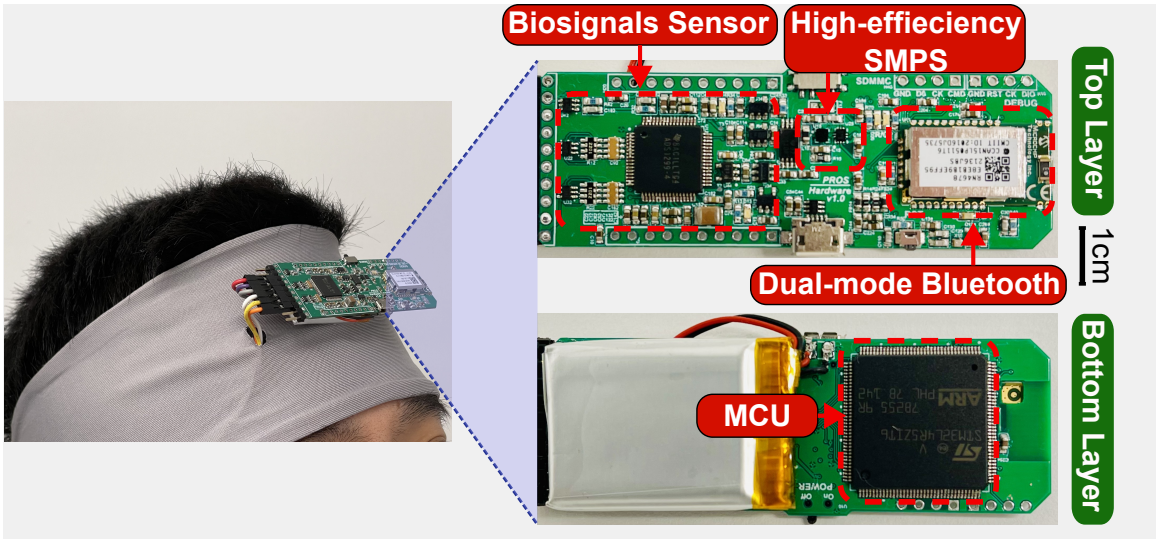


Figure 4.6: PROS hardware platform.

PROS hardware and mobile apps. We build a hardware prototype (Fig. 4.6) to support all the operations of PROS. Specifically, it contains an ARM Cortex-M4F MCU (STM32L4R5, 2MB FLASH, 640KB RAM) with four efficiency modes, accelerated DSP, and neural engines. To support DVFS, we bypass the internal regulator with an adaptive Switched Mode Power Supply (SMPS), providing the CPU voltage ranging from 1.00 to 1.35V. We use a dual-mode Bluetooth module (RN4678) to provide wireless communication with Bluetooth EDR (Enhanced Data Rate) and BLE 5 protocols. It supports throughput up to 48 and 256 kbps with BLE and EDR, respectively. We put together our hardware in a headband form factor.

Finally, we deploy the signal reconstruction algorithm to two mobile platforms, i.e., Galaxy S20 and Surface Go 2.

4.7 Evaluations

4.7.1 Datasets Preparation.

In this study, we use two open biosignal datasets, namely, TUSZ (Epileptic Seizures EEG events) and TUAG (EOG and EMG events), to develop and evaluate our TinyPR primitives and PDCS algorithms. The collection protocol was approved by Temple University Hospital IRB [259]. From the TUSZ dataset, we pick out a subset of 60 subjects with three important epileptic seizure patterns, i.e., (1) spike-and-sharp-wave (SPSW) patterns, (2) 3-Hertz spike-and-wave discharges (ABSZ), and (3) muscle stiffing and convulsions patterns (TCSZ), which represent focal/generalised non-specific, absence, and tonic-clonic seizures. Similarly, we pick up 60 subjects together with three EOG and EMG patterns, i.e., (1) eyes movements (EYEM), (2) chewing (CHEW), and (3) muscle contractions (MUSC), from the TUAG dataset. The chosen patient data were collected in various clinical settings such as the epilepsy monitoring unit, intensive care unit, emergency rooms, and routine EEG sessions. They contain both inpatients and outpatients with ages from five to 83 years old. As the sampling rate of the datasets varies (from 250 to 1024 Hz), we uniformly resample all the data to 500 Hz, which is the optimal data rate for EEG analysis [226].

We set aside ten subjects from each dataset for the unknown-subjects evaluations. The remaining are split into ten folds for training (80%), validating (10%), and testing (10%). We then segment the raw signal into non-overlapping windows. We use the duration of three seconds to cover sufficient pattern information. For the TUSZ dataset, it results in 2,099,479 data points (background: 2059148, SPSW: 34128, TCSZ: 3651, ABSZ: 2552) for the 10-fold cross-validation and 138,340 data points (background: 136292, SPSW: 1752, TCSZ: 220, ABSZ: 76) for the unknown-subjects evaluation. With the TUAG dataset, we have 495,338 data points (background: 440,479, EYEM: 15488, CHEW: 2993, MUSC: 36378) for the 10-folds cross-validation and 11,993 data points (background: 9467, EYEM: 99, CHEW: 104, MUSC: 2323) for the unknown-subjects evaluation.

Per each window, we apply the MFCC approach with our parameters' configuration (see Sec. 4.3.3) to extract the feature set of 10 MFCC features. The window is positive if the target pattern appears and negative otherwise. With each fold,

we apply an oversampling technique, SMOTE [215], to enrich the amount of positive data for the training set. Upsampling is not applied to validation, test, and unknown-subjects sets to keep the original distribution of samples.

4.7.2 TinyPR Primitives.

Classification metrics. As we cast our TinyPR primitives as binary classification models (Sec. 4.3). We use four indices of the confusion matrix: true positive (TP) is the number of actual positive segments which are correctly classified; true negative (TN) is the number of the actual negative segments that are correctly classified; false positive (FP) is the number of actual negative segments that are incorrectly classified as positive; false negative (FN) is the number of actual positive segments which are incorrectly classified as negative. With these notions, we define the sensitivity, specificity, and G-Mean scores as follows, $Sens = \frac{TP}{TP+FN}; Spec = \frac{TN}{TN+FP}; G-Mean = \sqrt{Sens * Spec}$. We plot the Receiver Operating Characteristic (ROC) curve to quantify the trade-off between sensitivity and specificity. We also report Area Under the Curve (AUC) as an additional performance metric.

10-fold cross-validation. We use standard Adam optimiser [199] with learning rate of $1e-2$ and $(\beta_1, \beta_2) = (0.5, 0.999)$ for optimisation. The classification loss is cross-entropy. We train each model with batch size 32 for 200 epochs. We evaluate the G-Mean score of models on the validation set to select the best model.

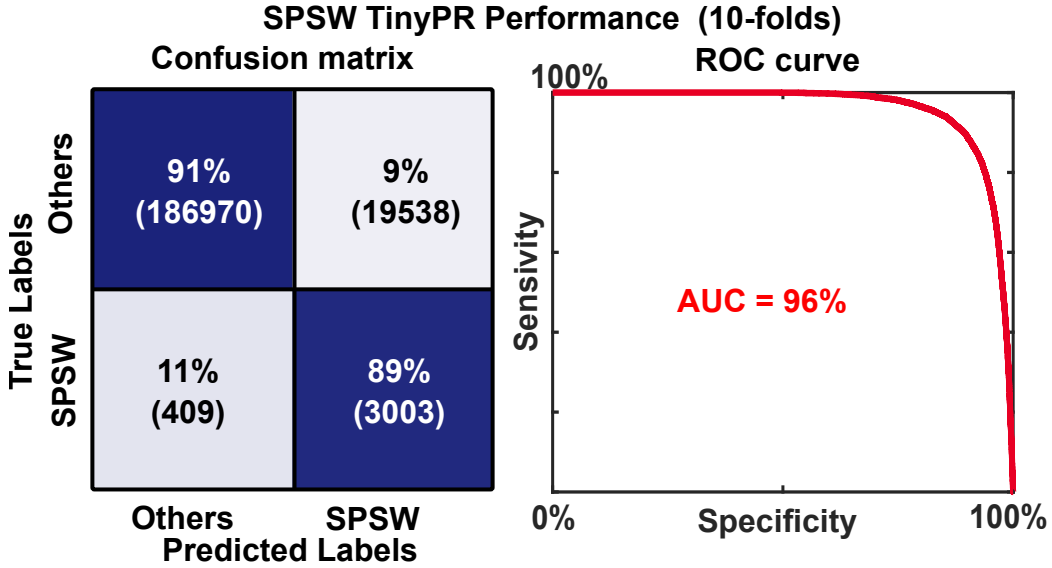


Figure 4.7: SPSW TinyPRs (tiny configuration) performance in 10-fold cross-validations.

We present the results on 10-fold cross-validations in Fig. 4.8. All the results are from the tiny configuration. Fig. 4.7 shows an example of a normalized confusion matrix and ROC of the SPSW TinyPR primitive. From the confusion matrix, we could see that the TinyPR model can effectively eliminate 91% (specificity) of the irrelevant negative windows while being able to capture 89% (sensitivity) of the positive ones. The G-Mean and AUC scores are 90% a 96%, respectively. Fig. 4.8 summarizes all the results. Among these, the ABSZ TinyPR has the highest scores, with 98% and 99% for G-Mean and AUC, respectively. The EYEM model has the lowest scores (78% G-Mean, 87% AUC) due to low its low signal amplitude and large variations between vertical and horizontal movements.

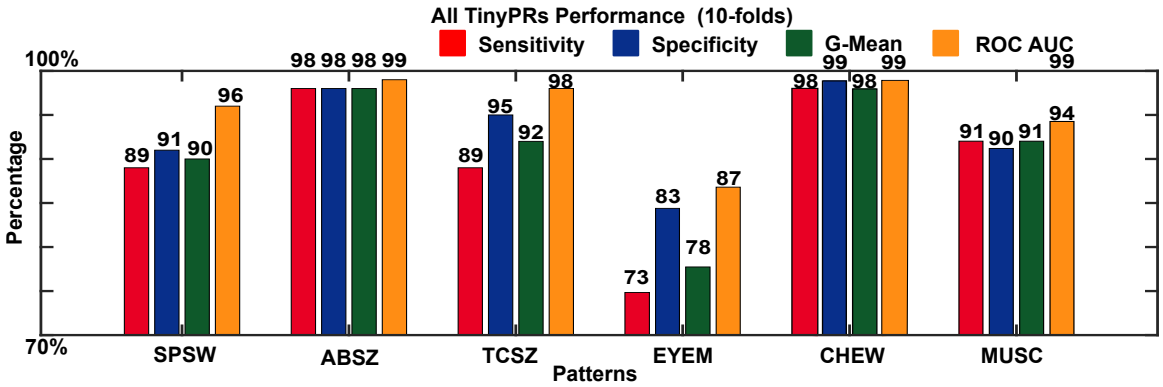


Figure 4.8: TinyPRs (tiny configuration) performance in 10-fold cross-validations.

Unknown subjects evaluation. To further evaluate the ability to work on unknown subjects, we use the best TinyPR primitives, chosen by their G-Mean scores, from our 10-fold validations to test on the unknown-subjects set. Fig. 4.9 and 4.10 present our results. Fig. 4.9 shows that our SPSW TinyPR primitive could achieve 90%, 83%, 86%, and 94% of specificity, sensitivity, G-Mean, and AUC, respectively. Similarly, Fig. 4.10 presents the results of all TinyPR primitives. The EYEM model has the lowest G-Mean and AUC at 75% and 85%, while all other models can achieve more than 85%. These results show the feasibility of our developed TinyPR primitives to work even on people that the models have not encountered.

Model and features sizes. To quantify the effects of model sizes, we evaluate different configurations, such as tiny, small, and medium, on the same SPSW pattern. We see that the training time will converse quicker with larger model sizes, i.e., < 50 epochs with the medium configuration versus > 100 epochs with the tiny configuration. However, the results do not significantly improve, i.e., 1-2% variation, in the G-Mean score. Similar results are also observed with larger feature sizes, i.e.,

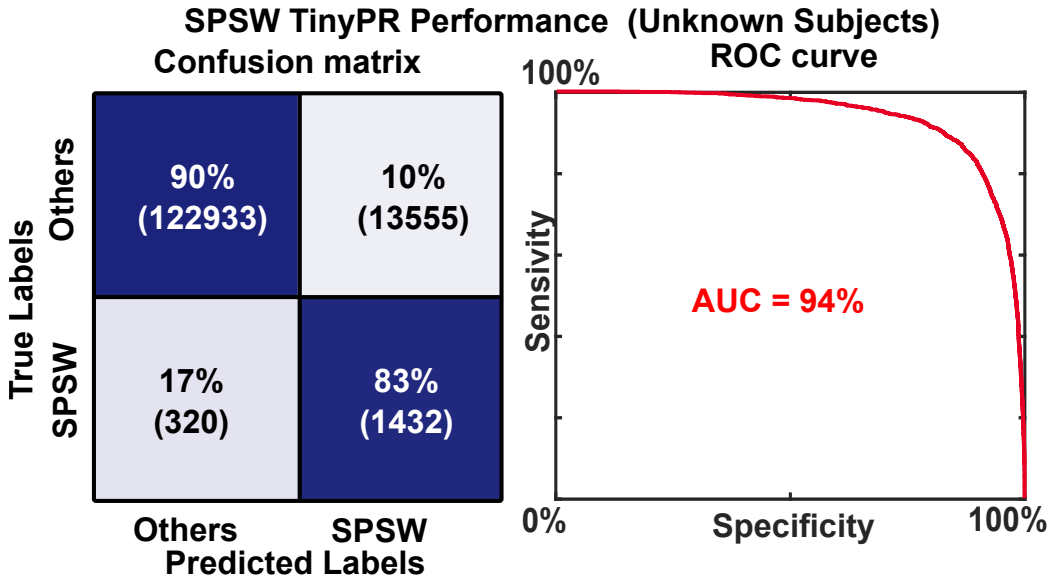


Figure 4.9: SPSW TinyPRs (tiny configuration) performance in unknown subjects evaluations.

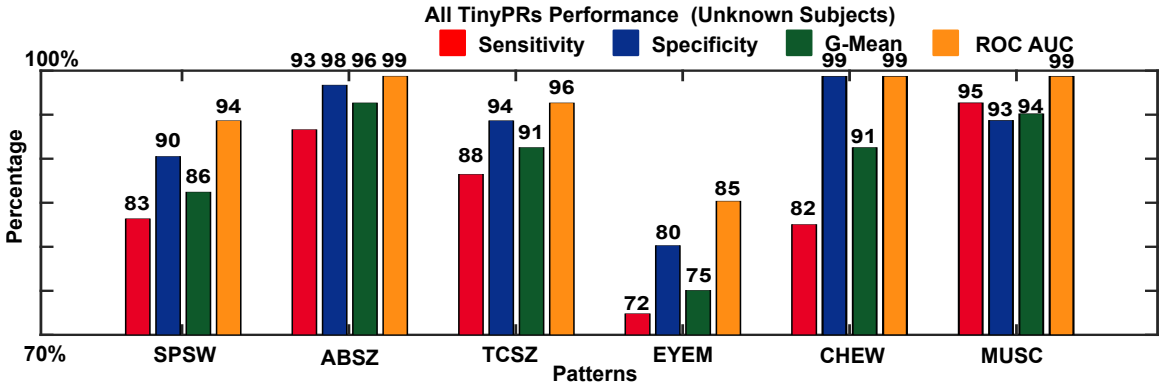


Figure 4.10: TinyPRs (tiny configuration) performance in unknown subjects evaluations.

20 vs 10 MFCC features. Since we only need to train the model once, we can afford a longer training time to achieve smaller model and feature sizes. Smaller sizes will significantly reduce the latency and memory footprint during runtime. After quantization, we observe a reduction of 63% with our tiny configuration, i.e., from 11KB (PyTorch) to 5KB (TFLM). The loss after quantization is minimal, with <1% of the G-Mean score.

4.7.3 Pattern-driven Compressive Sensing.

Optimal wavelet domains search. We conduct the optimal wavelet domains search (Sec. 4.4.3) for all the patterns on the training dataset. Fig. 4.11 presents the

results for the ABSZ pattern. The results confirm our intuition that the choice of wavelet domain is significantly important. For the ABSZ pattern, the global maxima and minima of average MSI are 91.2% and 61.9% with *bior*4.4 and *rboi*3.1 wavelet domains, respectively. This means there are more than 4X differences between the density of the two domains, making CS unusable with the latter. Interestingly, we observe that several local maxima in each wavelet family have similar results to the global maxima. E.g., the *db*3 domain has an average MSI of 90%, which is only 1% lower. This shows overlapping among wavelet domains, which one might exploit to further improve recovery latency by using simpler wavelet domains. Table 4.1 summarizes the optimal wavelet domains for all patterns. SPSW has the highest MSI (92.6%) with *bior*6.8 domain, while TCSZ has the lowest (71.4%) with *sym*14 due to high frequency and stochastic muscle components.

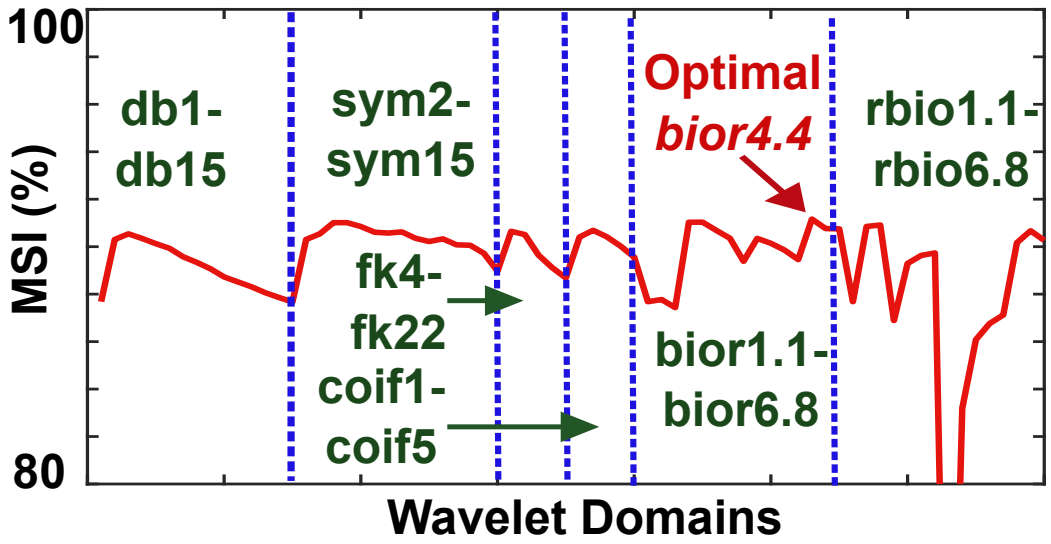


Figure 4.11: ABSZ optimal domain search.

To measure the computational cost, we perform the search on a Linux workstation (8-cores Core-i7 3.6GHz, 128GB RAM, MATLAB R2022a). MATLAB Parallel toolbox is used to speed up the process. With eight parallel workers, the search consumes 6981MB of memory and could take from 38 (CHEW) to 247 (MUSC) hours to finish. As we only need to run the search once for each pattern, it will not affect the real-time performance during the deployment.

Compression factors tuning. After knowing the optimal wavelet domains, we conduct evaluations to quantify the recovery signal quality with different compression factors (CFs). We run the evaluations on the whole training dataset and note down the CFs and the average SSIMs values in Table 4.1. The CFs and SSIMs reflect the

Table 4.1: Recovery quality with different CFs.

Pattern	Wavelet Domain	MSI (%)	SSIM with different CFs (w=3s)				
			1.5X	2X	3X	4X	5X
SPSW	bior6.8	92.6	0.99	0.98	0.96	0.94	0.89
ABSZ	bior4.4	91.2	0.99	0.97	0.94	0.89	0.82
TCSZ	sym14	71.4	0.84	0.81	0.57	0.39	0.31
EYEM	sym5	89.6	0.98	0.97	0.91	0.84	0.80
CHEW	bior4.4	84.0	0.93	0.93	0.84	0.78	0.71
MUSC	sym5	79.7	0.92	0.88	0.70	0.60	0.50

MSI value of each pattern. For the SPSW pattern (MSI = 92.6%), we can achieve the CF of more than 5X without having the average SSIM drop below 0.85. For patterns with low MSI, such as TCSZ (71.4%), we could only achieve the CF of 1.5-2X without deteriorating the recovered signal.

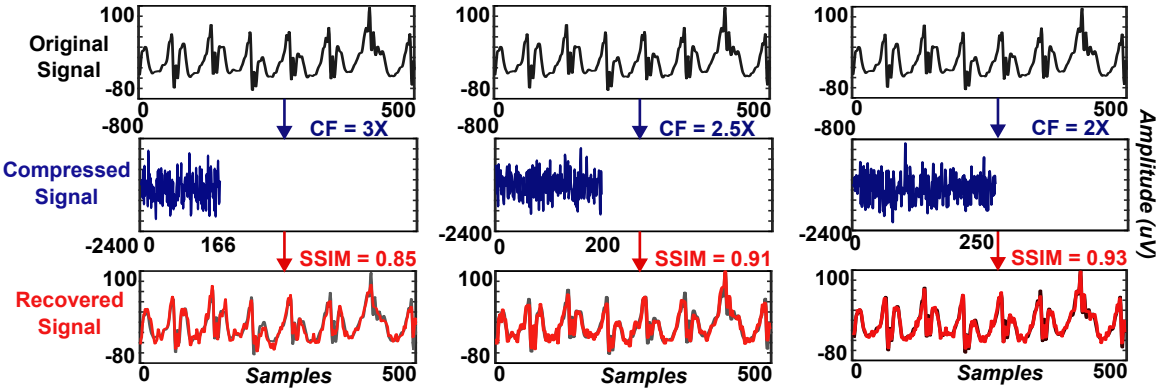


Figure 4.12: ABSZ recovery quality in *bior*4.4 domain.

We visualize the recovered ABSZ signal quality with different SSIMs in Fig. 4.12. We could see that when SSIM is over 0.85, the recovered signal looks very similar to the original one. When SSIM is over 0.93, we cannot visually spot the differences without extensively zooming in. This fits with the literature that the recovery starts to be indistinguishable by human eyes when $SSIM \geq 0.92$ [260].

Comparison with previous works. In previous works on compressive sensing (CS) such as [251, 261, 262, 263], static and pre-defined CFs and sparse recovery domains (e.g., Discrete Cosine Transform) are used due to the lack a pattern recognition capability on low-power hardware. In [263], three state-of-the-art CS algorithms, namely, DCT-based BSBL-BO [251], DCT-based l_1 [262], and Block-CoSaMP [261] are compared on the EEGLab dataset [264] (32 channels, 80 3-s EEG windows). With a CF of 2X, only DCT-based BSBL-BO achieves a satisfying SSIM of 0.85,

while DCT-based l_1 and Block-CoSaMP could only reach SSIMs of 0.45 and 0.48, respectively. In contrast with previous works, PROS enables the ability to recognise signal patterns of interest directly on the sensing hardware. This ability helps the proposed PDCS algorithm to apply optimal CF and sparse recovery domains for individual patterns, leading to more efficient compression.

4.7.4 Hardware runtime performance.

We deploy the developed TinyPR primitives and PDCS algorithm to our PROS hardware (ARM-Cortex M4F MCU, 2MB FLASH, 640KB RAM, GCC -Ofast) to measure their memory footprints, processing latency, and energy consumption. We also deploy the PDCS recovery algorithm on two mobile devices, i.e., (1) Galaxy S20 (Octa-core 2.2-2.7GHz Cortex-A55) and (2) Surface Go 2 (Dual-core 1.7GHz Intel Pentium) to measure the processing performance. We use the Otti Arc profiler to measure energy consumption with the sampling rate at 4000Hz. Fig. 4.13 presents our experiment setup.

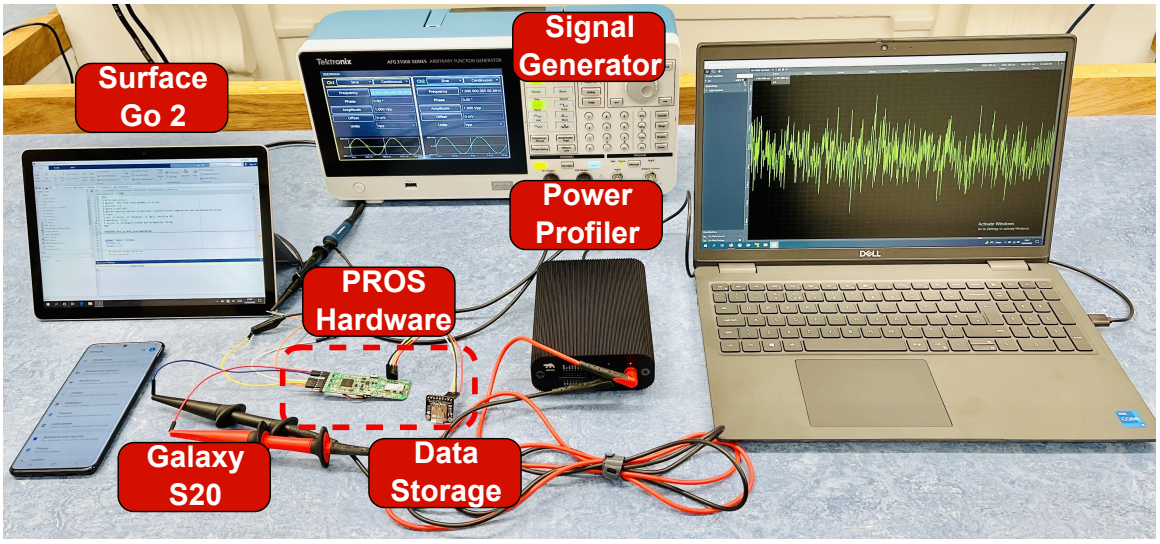


Figure 4.13: Runtime experiment setup.

Memory footprints. With the tiny configuration, each TinyPR primitive only consumes 5KB of FLASH (0.2% of system FLASH) and 30KB of RAM (4.5% of system RAM). The consumed RAM could be dynamically reused if we do not need to run multiple TinyPR primitives concurrently. Our PDCS consumes from 4-30KB of FLASH (0.2-1.4% of system FLASH) to store random binary CS matrices.

Processing latency. On our PROS hardware, the latency of MFCC calculation, one TinyPR inference, and compressive sensing could be as low as 6, 26, and 1ms

when we clock the MCU at 120MHz, respectively. When we clock the MCU at 24MHz, the results are 29, 126, and 4ms. On mobile devices, we run the recovery algorithm on 2000 different signal samples. The average latency is 50 and 94ms on the Galaxy S20 and Surface Go 2. The results show that PROS can respond to critical events within milliseconds on the hardware. The response time on the phone only depends on wireless communication latency, as the additional processing of our recovery algorithm is minimal.

In a real-time setting, biopotential signal windows are continuously buffered and need to be processed every few seconds, e.g., three seconds windows in our evaluation. Our PROS prototype could process the signals in real time since the whole processing latency on both the wearable and mobile devices could be as low as 88ms. This show the feasibility of PROS in real-time applications.

Energy consumption. We measure the energy consumption of each operation on PROS hardware with and without DVFS. When DVFS is not used, the MFCC calculation, one TinyPR inference, and compressive sensing consume 0.6, 2.4, and 0.1mJ. When DVFS is used, the results are 0.4, 2, and 0.07mJ. Thus, we could see that our DVFS and SMPS could increase the energy efficiency by 30-50%.

To analyse the system's total energy consumption, let's assume that for each 3s biosignal window, we only need to detect the SPSW pattern, our tinyPR model can eliminate 90% (P) of the irrelevant signals, and the PDCS algorithm has the compression factor (CF) of 4X. Without PROS, the total energy (E_t) to transmit a biosignal window through Bluetooth is measured at 64mJ (56mJ for Bluetooth transmission and 8mJ for data loading between the MCU and Bluetooth module.). With PROS, the total energy to process a biosignal window on average can be calculated as $E_{pros} = P*E_{tinyPR}+(100\%-P)*(E_{tinyPR}+E_{PDCS}+E_t/CF) = 4.6mJ$ (14X reduction from the baseline, i.e., E_t). We can also see that the proportion between transmission and processing energy is shifted from 88%/12% (E_t) to 30%/70% (E_{pros}). Thus, the processing energy now becomes the dominant factor. When we enable DVFS, we can further optimise the processing efficiency and the total energy to process a biosignal window is reduced to $E_{pros+dvfs} = 3.9mJ$ (16X reduction). The actual performance depends on the workload setting and applications' configurations on the number of needed patterns and required sensitivity/specificity. In Sec. 4.7.5, we will present a more practical workload for the seizure detection use case.

Comparison with open-source platforms. We conduct the processing latency (L) and energy consumption (E) measurements of our proposed TinyPR models and PDCS algorithm on an Arduino Nano BLE Sense and a Raspberry Pi Zero v1.3

(Fig. 4.14). Table 4.2 summarises the results. Since Raspberry Pi is not designed for low-power applications, it has the largest overhead. The performance of PROS on both the Arduino and Raspberry could be further improved by optimising the processing software.

Table 4.2: Open-source platforms evaluations.

Platform	CPU/RAM	TinyPR (L/E)	PDCS (L/E)
Arduino Nano	64MHz/256KB	50ms/4.2mJ	12ms/1.0mJ
RaspPi Zero	1GHz/512MB	92ms/108mJ	4ms/3.6mJ
PROS	120MHz/256KB	26ms/2.4mJ	1ms/0.1mJ

We also measure the power consumption of a commercialised biosensing platform, i.e., OpenBCI. The average consumption while streaming is 146 mW. Thus, with 250, 320, and 500mAh LiPo batteries, it could last for 6.3, 8.1, and 12.7 hours, respectively. With PROS hardware and the workload discussed in Sec. 4.7.5, we could increase the battery life to 84, 107, and 168 hours.

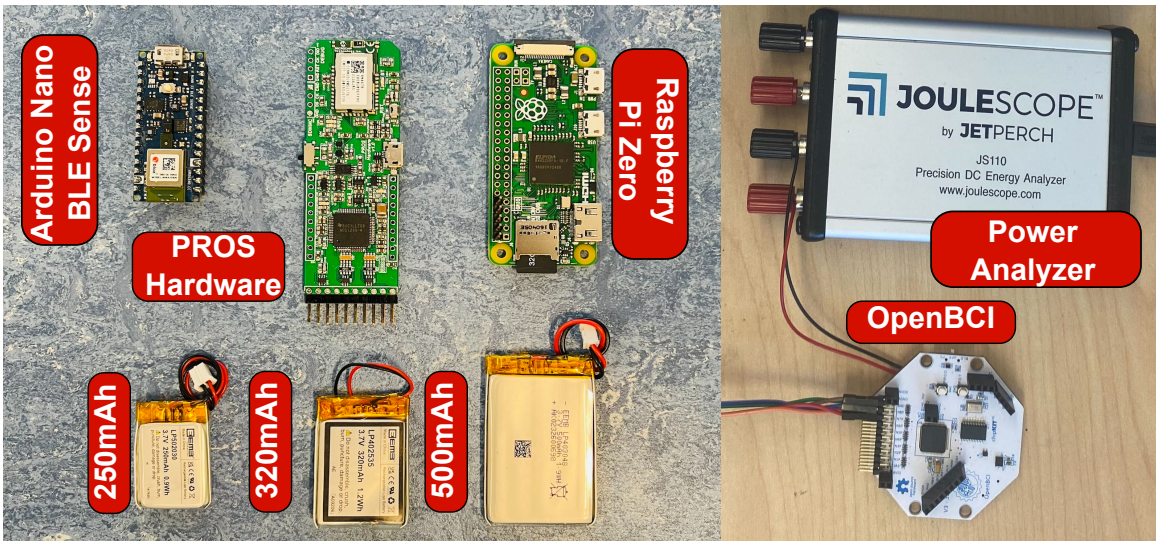


Figure 4.14: Open-source hardware and biosensing platforms.

4.7.5 Epileptic seizures detection use case.

We conducted our experiment to quantify the significance of PROS in detecting epileptic seizures. We choose the seizure detection use case because of its high-fidelity data requirement and challenging local processing on the device.

Previous studies [265, 266] have pointed out that detecting non-motor seizures is non-trivial and requires medical experts to analyse and diagnose the captured signals. Thus, maintaining high-fidelity signals is essential. Furthermore, local seizure detection on wearable devices is challenging when considering the constrained computing resource of low-power microcontrollers. In literature, neural networks such as VGG [267] or ResNet [268] are feasible for detecting seizures with good accuracy. However, such networks are too large to be run on an MCU with limited memory (¡1 MB of SRAM) [269]. Even if we can extensively prune the network to run on MCUs, the accuracy will degrade significantly, leading to unusable results) [269]. Thus, sending signals to a nearby offload device is still necessary for further analysis, diagnosis or classification.

In this case study, we focus on focusing on three crucial seizure types, i.e., tonic-clonic, absence, and focal/generalised non-specific seizures. They require three seizure-related patterns, i.e., TCSZ, ABSZ, and SPSW, respectively. To ensure practicality, we use the unknown-subjects dataset (10 subjects, Sec. 4.7.1). This results in 208,246,500 samples, i.e., 29 hours of data. The CFs for TCSZ, ABSZ, and SPSW are set at 2, 4, and 5X, respectively. Since tonic-clonic seizures have the highest risk of fatality [270], we put TCSZ at the highest priority, followed by ABSZ and SPSW.

The results show that PROS can reduce the number of transmission data by 24X (8,807,370 vs 208,246,500 samples). Our TinyPR primitives can pick up more than 85% seizure signals and eliminate 86% non-seizure ones. The recovered signal is high-fidelity with the average SSIMs for TCSZ, ABSZ, and SPSW: 0.93, 0.92, and 0.93, respectively. By transmitting all the signals to a mobile device with Bluetooth EDR, the wearable device consumes 15.7kJ. With PROS, the device only consumes 1.71kJ, giving a boost of 818%. Interestingly, since PROS significantly reduces the transmission data, we could use a lower-rate protocol such as BLE, which is impossible with the original amount of data throughput. By using BLE and PROS, the device only consumes 1.15kJ in total, boosting the energy efficiency up to 1265%. Thus, with a 500mAh Li-Po battery (Fig. 4.6), the device could last for a whole week while continuously monitoring seizure events. Finally, PROS could respond to deadly tonic-clonic seizures directly on the device within as low as 32ms. This is especially important when the mobile device might not be available, e.g., during charging or out of communication range. These results show the feasibility that PROS could significantly improve users' experience and even reduce fatalities.

4.8 Summary

In this study, we propose PROS, an efficient pattern-driven compressive sensing framework for low-power biosensing wearables, by exploiting the sparsity of biosignals. In a practical use case such as epileptic seizure detection, PROS significantly boosts the energy efficiency and enables real-time response to critical events while maintaining high fidelity signal.

Chapter 5

eBP: Sensing in-ear blood vessels to monitor blood pressure with a wearable computer.

5.1 Introduction

This chapter presents a device called eBP to measure BP from inside the user's ear, aiming to minimise the measurement's impact on users' normal activities while maximising its comfort level. eBP has three key components: (1) a light-based pulse sensor attached to an inflatable pipe that goes inside the ear, (2) a digital air pump with a fine controller, and (3) a BP estimation algorithm. In contrast to existing devices, eBP introduces a novel technique that eliminates the need to block the blood flow inside the ear, which alleviates the user's discomfort.

We prototyped eBP custom hardware and software and evaluated the system through a comparative study on 35 subjects. The study shows that eBP obtains the average error of 1.8 mmHg and -3.1 mmHg and a standard deviation error of 7.2 mmHg and 7.9 mmHg for systolic (high-pressure value) and diastolic (low-pressure value), respectively. These errors are around the acceptable margins regulated by the FDA's AAMI protocol, which allows average errors of up to 5 mmHg and a standard deviation of up to 8 mmHg.

5.1.1 Goals and Challenges

In this work, we propose a novel wearable system that captures BP inside the ear called eBP. eBP aims to address the discomfort and inconvenience issues with its discreet design, quiet components, and convenient location.

Challenges: realising eBP has the following challenges:

1. In-ear BP monitoring is an unexplored topic in which many of the existing techniques cannot be applied. Even the feasibility of the technique has yet to be confirmed.

2. The mechanism enabling the use of an inflatable balloon to measure BP from inside the ear is non-trivial. When the balloon inflates, the sensor should attach firmly to the ear canal and not slide out. In addition, applying insufficient pressure will result in an inaccurate BP measurement, while applying too much pressure may cause discomfort or hurt the ear canal.
3. The in-ear pulse signals are weak and buried under noise. In addition, the motion artefacts are difficult to remove and can impact BP measurement accuracy.
4. BP measurements are sensitive to the contact quality (i.e., pressure) between sensor and in-ear skin; yet maintaining consistent contact pressure is difficult.

Contributions: In this work, we make the following contributions. First, we propose a novel concept of in-ear frequent BP monitoring and show that it is not only feasible but also comfortable. Second, we propose a blocking-free optical-oscillometric approach to allow the in-ear sensor to measure important parameters in BP measurements (i.e., systolic amplitude and diastolic amplitude). Third, we prototype a device with a custom-built circuit and hardware/software components for in-ear BP measurements. In particular, we customise an off-the-shelf catheter to safely insert it into the ear canal with a light pulse sensor attached. In addition, we build a portable in-ear wearable device to control the light pulse sensor and the catheter to capture the pulse signal (i.e., BP) accurately and reliably. Fourth, we devise an algorithm to process and qualify the highly noisy pulse signals captured from inside the ear to ensure high-quality BP measurements. Lastly, we conducted a study with 35 users and verified the performance of the proposed system using a FDA-approved BP measurement device (KonQuest KBP-2704A).

5.1.2 Applications and Impacts

While eBP is currently a standalone device, with the continued trend of incorporating biometric monitoring into devices [271, 272, 273, 274, 275] that are worn on a daily basis, there would be minimal behavioural changes required on the part of the wearer to benefit from eBP. As ear-worn devices are becoming increasingly popular [276, 277, 278], eBP could potentially be integrated into a headphone or hearing aid, both of which are ubiquitous as the World Health Organization reports that approximately 466 million people worldwide suffer from disabling hearing loss [279] and more than 365 million headphones were sold in 2017 in the US alone [280]. In addition, our proposed BP calculation algorithm can be applied to make existing cuff devices more comfortable. In the case of hardware design, the use of a medical balloon to deliver a

sensor into the ear can widely benefit other applications. For example, it can improve the contact points and the conductivity of electrodes for the in-ear sensing area.

5.2 System Overview

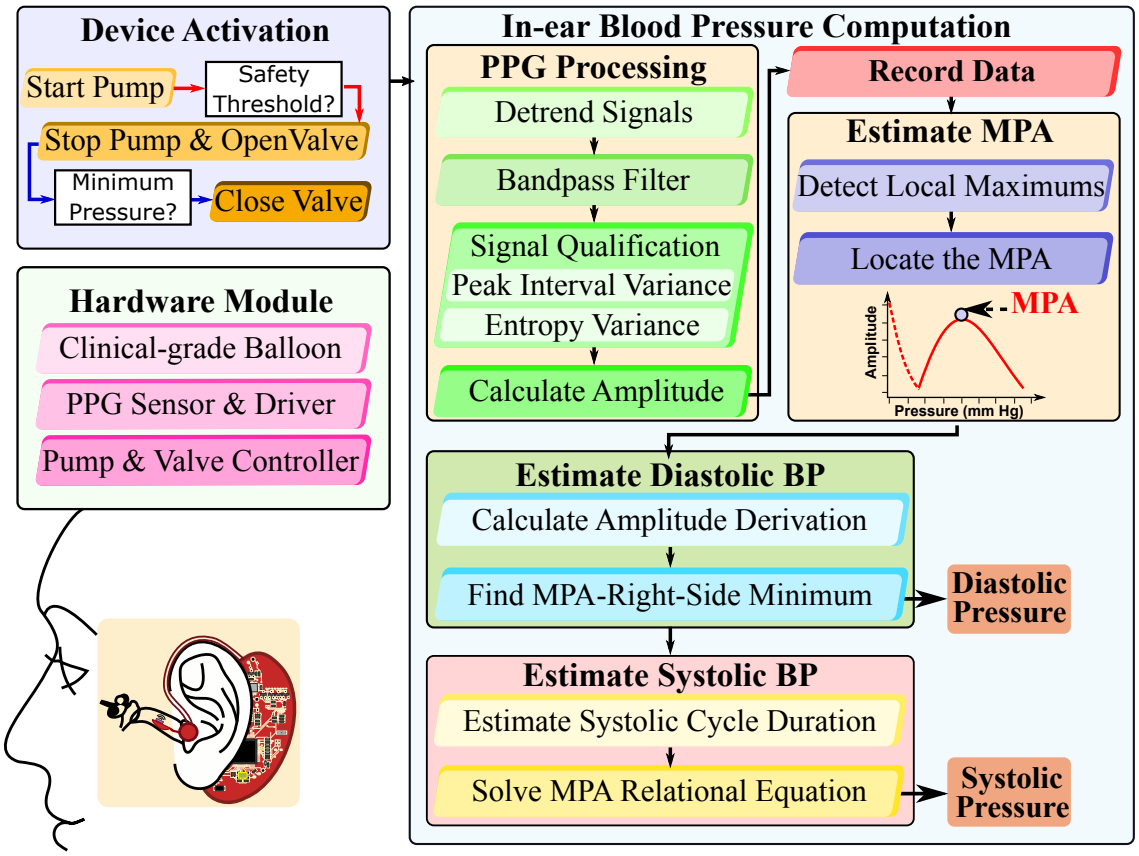


Figure 5.1: eBP system.

Learning from what was previously mentioned, we designed our system as shown in Fig. 5.1 to address the challenges. In this section, we explain the goal of our design in detail, the system's components, as well as the core algorithms. First of all, there are three essential goals that our design needs to satisfy:

1. allows frequent monitoring with minimal impact on the user's mobility,
2. be comfortable, unobtrusive and easy to use, and
3. obtains a high-accuracy BP prediction.

5.2.1 Motivation and challenges for in-ear BP monitoring:

Commercialised devices are mostly cumbersome and require a piped cuff to be attached to the user's limb, which limits the user's movements. Mounting the sensors

onto glasses [281, 163] can significantly improve flexibility, but it may be aesthetically unappealing and highly obtrusive [282]. Therefore, we explore the option of measuring BP from inside the ear for portability, aesthetic appearance, comfort, and social acceptance. realising these objectives, many challenges must be overcome.

First, the systolic and diastolic ratio was invented for the arm cuff BP and cannot be adapted directly into our design. Consequently, we propose a new approach to identify the SBP and DBP without using this characteristic fixed ratio.

Second, the traditional approach uses a cuff to compress the artery on the arm or wrist for pulse measurements. However, this is challenging to repeat inside the human ear, as the in-ear balloon can inflate in any direction, which affects the pulse measurement. Fortunately, by analysing the human ear anatomy, we found that the two sides of the ear canal are structured differently. In particular, one side (S1) of the canal includes an artery covered by muscles where the pulse signal could be captured, while the other side (S2) is right next to a bone. When the in-ear balloon is inflated, the balloon expands and obtains contact with the artery at S1 and does not expand at S2. This mechanism creates a similar environment to the traditional approach used on the arm/wrist; therefore, we can derive an algorithm for BP measurement using the amplitude and pressure relationship from pulse measurement theory.

Third, as the balloon inflates and shrinks across multiple measurements, the solution of adhering the pulse sensor to the balloon surface can wear out and be damaged, raising the concern that the sensor can fall off inside the ear. Therefore, we develop a meticulous procedure to attach the sensor by firmly wiring and securing it on the balloon surface.

Fourth, the superficial artery is relatively small compared to the brachial artery. The sensor should face the artery in order to obtain a clear pulse signal. However, placing the pulse sensor in the ear is not an easy-handled task, as visibility is restricted. To solve this issue, we develop a pulse-signal qualification model accumulated with our design to precisely justify when the pulse signal is detected.

Last but not least, hypersensitive skin in ear-canal spurs safety and patient-experience concerns. One-third of the outer ear canal's skin is from 1 to 1.5mm thick, while the inner two-thirds is only 0.1mm in thickness, as shown in Fig. 5.2. The sensing area mandates our efforts to customise the design by carefully excluding electronic components that have toxic materials or sharp edges.

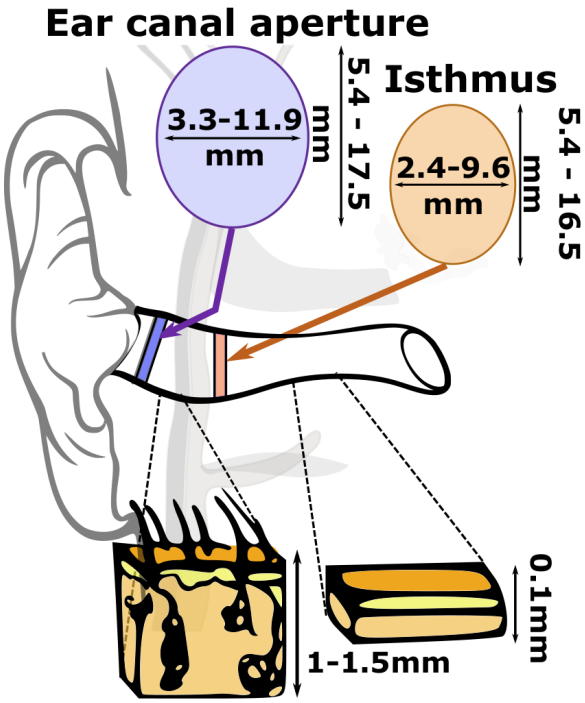


Figure 5.2: Ear canal anatomy.

5.2.2 Proposed solutions:

Non-ratio approach for the calculation of systolic and diastolic BP: Unlike the oscillometric method, we do not apply the fixed-ratio BP because there is no valid ratio for inside the ear. For safety purposes, our pressure may not cover the SBP range. Therefore, we aim to estimate the pressure in diastole first according to its minimal downslope amplitude. Then, we substitute the DBP into the MPA-relational equation to estimate the SBP. Moreover, we propose a personalised approach to estimate the systolic fraction β instead of using the common fixed ratio. A thorough explanation of the technique is presented in Section 5.4.1.

In-ear pulse sensor with the flexible circuit: eBP uses the light-based sensing technique, named Photoplethysmography (PPG) [283], to capture the superficial pulse (BP value). The optical sensor is small and sustainable enough to be attached to the balloon. However, state-of-the-art BP sensing technology is often designed on a printed hard circuit board. When the sensor is placed on the balloon, its surface might create sharp contact, which may hurt the user's ear. We overcome this problem by designing a flexible BP sensing circuit (Section 5.3.2). This flexible circuit adapts to the balloon's deformation, making the device comfortable to use for a long period of time.

High-quality elastic balloon: The balloon, which serves our specific purpose, needs to satisfy the following criteria: bio-compatible, safe, highly elastic, strong and resilient to cleaning. To satisfy these conditions, we customise an off-the-shelf medical balloon often used for bladder catheterisation (Section 5.3.2).

In-ear PPG signal qualification: The in-ear PPG power is weaker than that of the finger, wrist, or arm. Therefore, after basic preprocessing, we employ two techniques for signal qualification, including the **Peak Interval Variation** and **Entropy Variance** (Section 5.5) to eliminate bad data chunks and identify the correct position inside the ear to place the sensor. For conventional signal filtering, we process every 50 milliseconds with DC removal and a bandpass filter. This procedure helps eliminate noise and other unwanted band signals to disclose only the pulse waveform. With data that qualifies for this criteria, we calculate their amplitude using our modified peak-to-peak technique.

In-ear PPG signal processing:

1. *Modified peak-to-peak amplitude calculation*: Current peak-to-peak calculation is inconsistent for real-time processing due to the random order of peaks and bottoms. We propose a solution by adding a verification module to ensure order consistency.
2. *Drift removal for Mean Arterial Pressure detection:* During the first few seconds of the balloon deflating, a large drift away from the calibrated pressure causes the false detection of maximum amplitude. We have developed a solution to detect the mean arterial pressure based on its local maxima property regardless of the appearance of the drift. (Section 5.5).

Ear-worn air pump and draining components: Air pump and draining components are designed to inflate and shrink the balloon with a predefined configuration. We target the miniaturized components to develop the air pump. The controller will process the signal and detect whether the pressure is sufficient. Then, following the information, it will decide if more air should be pumped in or if the valve should be opened to reduce pressure. The final product will be worn outside the ear. (Section 5.3.1).

5.3 System Hardware Design

5.3.1 Design considerations

The in-ear BP measuring module consists of three main components: (1) a PPG sensor attached to an in-ear balloon, (2) an air pump and a valve controller to control the

balloon having good contact with the in-ear skin, and (3) a central sensing algorithm to compute the PPG value by controlling both the above mentioned two components for robust BP monitoring using an in-ear sensor. The detailed architecture of the module is shown in Fig. 5.3. Let's discuss the requirements of each component in detail.

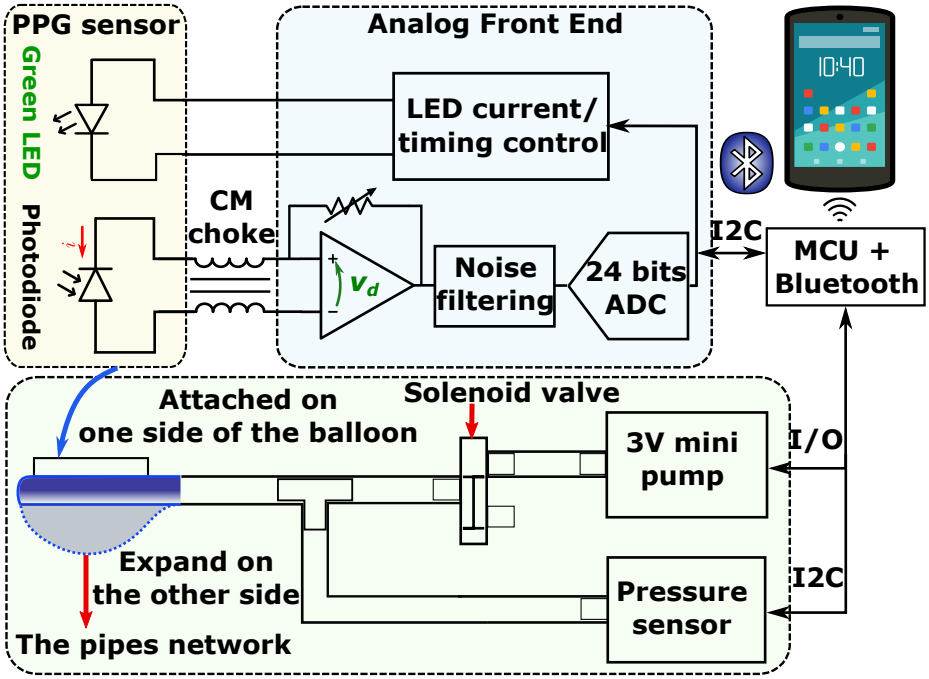


Figure 5.3: In-ear BP module design.

The PPG sensor. The PPG sensor consists of a LED (transmitter) shining onto an artery and a photodiode (receiver) capturing the reflected light. Since the human ear canal can be as small as 2.4mm in diameter, both the LED and photodiode need to be miniaturised to fit the ear canal. Second, the artery inside the ear canal is hidden deep below the skin. Thus, the LED in our PPG sensor needs to be bright enough so that the light can deeply penetrate the skin deeper and reach the artery. In addition, the photodiode also needs to be highly sensitive to pick up the small reflected light. Fortunately, the PPG measurement inside the ear does not suffer from the effect of ambient light. In particular, we found that the noise term coming from ambient light can be minimised.

The pump/valve controlling module. To precisely control the air balloon, the pump/valve controlling module needs to ensure the consistency of the contact force for maintaining high-quality PPG measurements. The module includes a digital pressure sensor for measuring the contact force, a mini air pump, a solenoid valve for filling

and venting the air and a network of soft silicon pipes connecting them, as illustrated in Fig. 5.3. When the measurement starts, the mini pump will be activated to fill the balloon with air, and while the valve is closed, the air is kept inside the network. When the PPG sensor has reached the ear canal wall, the pump is stopped, and the balloon is kept in an inflated state for PPG measurement. Pressure values inside the channel are sampled and streamed to our MCU. After the measurement is completed, the valve will be opened to release the contained air and let the balloon return to its normal state.

The in-ear balloon. PPG sensing requires tight contact points between the sensors and the skin for accurate measurement. However, keeping the sensor in contact with the skin at all times might generate discomfort for the user for long-term use. To overcome this challenge, the sensor is designed to be flexible and operates as an in-ear balloon to only contact the human skin tightly when conducting PPG measurements. Specifically, we mount the PPG sensor on top of a small balloon that can be pumped up or vented out via a controller module. Furthermore, the circuit for the PPG sensor needs to deform as the balloon inflates flexibly. It must be engineered on the side of the balloon where it can remain stable as the balloon inflates and deflates.

Since the balloon needs to be inserted into the ear canal, it needs to be carefully designed and must satisfy the following requirements: (1) *bio-compatibility*, it must not contain any substances that could be harmful to the human body in its material; (2) *safety and robustness*, as the balloon needs to sustain a certain amount of air pressure for the measurement and also accommodate various sizes of human ear canals, (3) *user's comfort*, the human ear canal is thin and sensitive; thus, the balloon needs to be soft to make the user feel comfortable while providing enough pressure for our measurement. In addition, as the balloon inflates and shrinks across multiple BP measurements, it needs to have high durability elasticity. Additionally, the inflation and shrinking rate of the balloon need to be consistent to maintain reliable measurements. Last but not least, it needs to be strong and resilient to scrubbing when cleaning so that the sensor can be reused multiple times.

5.3.2 Implementation

Proprietary design of the balloon attaching PPG sensor. We designed our in-ear sensor by integrating a PPG sensor with the balloon of a Foley catheter made by POIESIS MEDICAL [284], as illustrated in Fig. 5.4. The Foley catheter is created from 100% medical silicon so it can be safely and comfortably inserted inside the body [284]. We found that the SFH7050 PPG sensor from OSRAM [285] is the

best fit for the small size of the ear canal. It has a size of 4.7 mm x 2.5 mm x 0.9 mm and performs highly accurate measurements due to its special design for the crosstalk blocking technique [285]. The PPG sensor includes a super-bright green LED transmitter, having a wavelength of 525 nm, which is a commonly used wavelength in PPG measurement. The receiver is a sensitive photodiode that captures the reflected lights from the LED transmitter and infers PPG values. Section 5.4 describes the algorithm to calculate PPG in detail. The sensor is driven by a specialized analogue front-end IC (AFE4404) [286] from TI. This AFE provides an accurate built-in LED driver and a timing controller, an ultra-low noise trans-impedance amplifier with a wide-range programmable gain, and a precision 24-bit ADC to control the SFH7050 sensors. To ensure high fidelity signals, a common-mode (CM) choke coil is used as an analogue low-pass filter before the input of the AFE. It helps to suppress CM noise induced by electromagnetic interference over the wires between the sensor and the AFE. The digital data from the AFE is then streamed to our MCU (MSP430F5529) [287] through I2C communication with a sampling rate of 320 Hz.

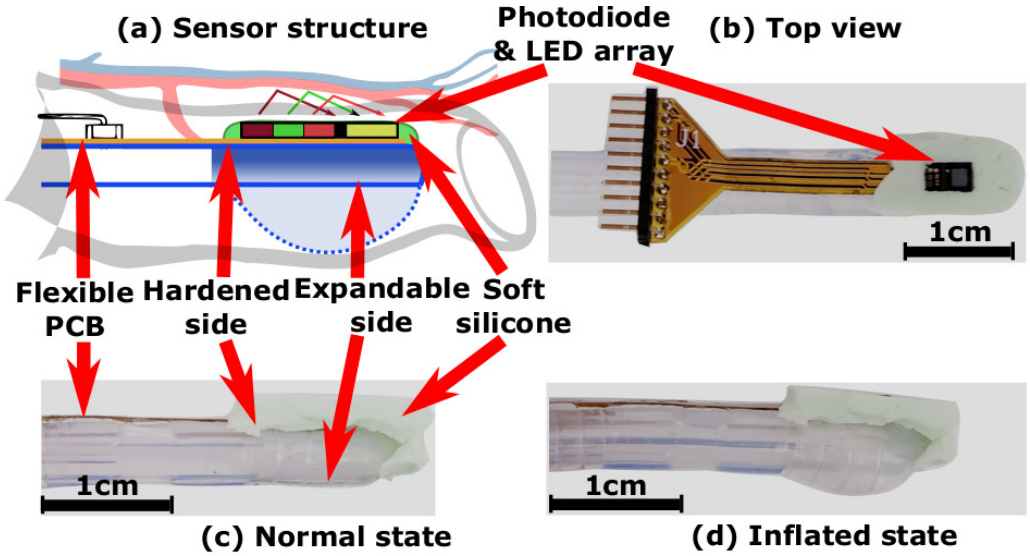


Figure 5.4: In-ear PPG sensor and balloon design.

Our in-ear balloon is hardened, so it will only expand on one side. The other side must be kept stable for the mounted LEDs and photodiode. The PPG sensor is soldered on a thin layer (0.1 mm) of flexible PCB. The PCB is then integrated on top of the balloon catheter by using a thin layer of liquid silicone gel. After curing for one hour at 80^{o}C, the bonding between the PPG sensor and catheter surface becomes hardened and stays robust. Furthermore, to make the sensing unit more comfortable inside the ear, we coated Smooth-On Ecoflex 00-30 soft silicone [288] around the edge

of the sensor, covering all sharp corners. The surface of the sensor was kept flat by using a glass slide, which is removed once the Ecoflex is cured. Thus, the flat surface of the sensor offers a better sensing ability.

The prices of a catheter balloon and the SFH7050 PPG sensor are $16.99 and $2.82, respectively. In total, manufacturing a disposable in-ear balloon sensor costs approximately $19.81, which is cheaper than a disposable BP handcuff [289, 290]. The cost can be reduced further through large-scale production.

In-ear balloon pressure monitoring. In literature, it has been shown by experiments that the relationship among pressure inside the balloon, its volume, and diameter are nonlinear [291]. Especially when the diameter is in the range from 7 mm to 9 mm inside the ear canal, the pressure has an initial peak called the equilibrium point accompanied by a slow balloon expansion as the constituent polymer makeup of the balloon is altered. After the balloon has reached its equilibrium point, the pressure inside the balloon will keep stable or reduce, even if its volume increases. Fig. 5.5 shows the relationship between the pressure and diameter of the silicone-based balloon used in our study, which matches the results in existing studies [292, 291]. As a result, we cannot rely on pressure values to know whether the balloon has reached the wall of the ear canal or not. Instead, the quality of PPG signals is observed, and the pump will be stopped when we observe clear PPG signals.

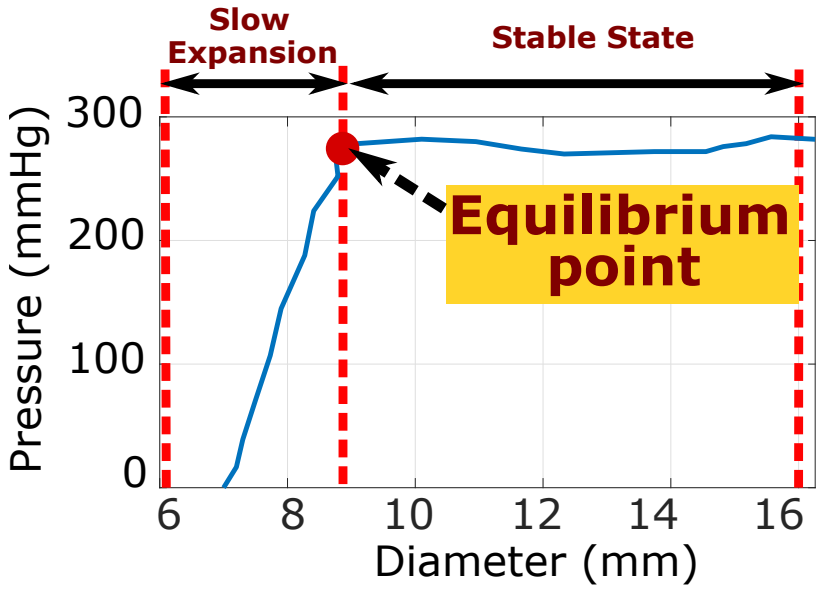


Figure 5.5: The pressure vs. diameter curve of the balloon.

In addition, an over-threshold protection mechanism is implemented to stop pumping air when the pressure inside the balloon is over the threshold. Since the size of

each person's ear canal is different, with its diameters in the range from 2.4 to 17.5 mm, we want to continuously and slowly inject the air until one side of the balloon touches the skin of the user's ear canal and partially blocks the artery. However, the balloon also has limitations regarding how much air it can hold. Thus, we do not want to inject too much air into it, making it permanently deformed or causing it to burst. From the balloon datasheet [284] and an experimental burst test in [291], the failure pressure of the silicone-based balloon is between 15 and 20 psi. Thus, the pressure inside the balloon is continuously monitored by the MSP430 MCU, and the pump will be stopped if the pressure reaches more than ten psi, as a rule of thumb. This addresses the challenge of different ear canal sizes while maintaining the safety of our system.

Concerning safety, we use the Ecoflex soft silicone design across our evaluation, including a few trials at the beginning. This is because the diameter of the balloon, when expanding inside the ear canal, does not exceed the break-out point of the silicone gel. When the diameter of the inflated balloon is larger than 30 mm, we start seeing some cracks on the silicon at the connection between the Ecoflex and the balloon. Since the ear canal diameter is only from 7mm to 9mm, which is much smaller than the cracking point, the balloon is safe to be inflated inside the ear.

Central processing controller. The central controller, as shown in Fig. 5.6 is responsible for (1) communicating to mobile devices through Bluetooth to receive commands and report sensing data, (2) driving the analogue front-end IC to collect the PPG measurement to the sensor, and (3) controlling the pump/valves to control the balloon pressure for accurate PPG measurement. Overall, Fig. 5.7 presents the eBP prototype depicting the integration of the in-ear pulse sensor with the main module.

Power consumption. All components in our designed module are chosen to operate with low power consumption and also have small sizes. These are very important in a "wearable" scenario where the user still has mobility when they wear the device, whereas it is not possible with a conventional cuff-based BP measuring device. Additionally, less power consumption, in turn, will reduce the capacity of the Li-Po battery, which is the main limiting factor to minimise the weight of our designed module. During the measurement, the MCU, AFE, pressure sensor, and Bluetooth module consume a maximum of only 4.6 mA, 325 uA, 1.7 mA, and 30 mA, respectively. The LED transmitter, valve, and mini pump draw ten mA, 110 mA, and 150 mA. Thus, our module consumes around 303 mA while the BP measurement is running. On the other hand, only 4.95 mA is drawn when our measurement is

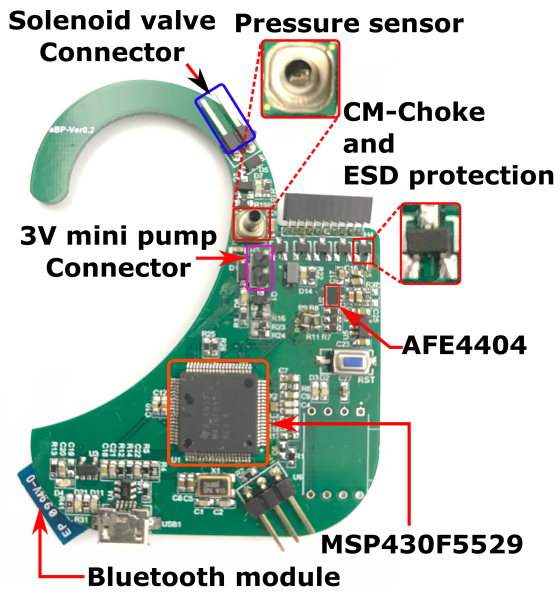


Figure 5.6: eBP hardware.

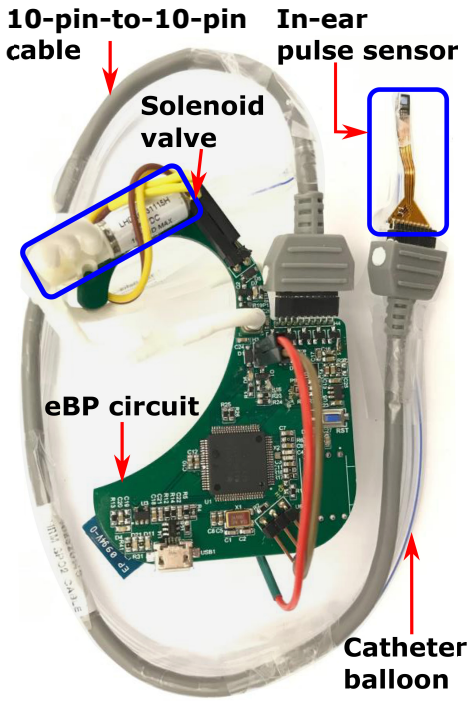


Figure 5.7: eBP prototype.

not running. Thus, a 400 mA Li-Po battery is used, providing up to 1.3 hours of continuous measurement, equivalent to roughly 80 measurements. However, running the device all the time is not practical or necessary. Instead, the users usually only need to measure their BP a few times per day. If the system is not running any measurement, it can last for more than two days (53 hours) in an idle state.

Integration with COTS earphones. While our current implementation of eBP is still a proof-of-concept prototype, we envision that eBP could be further miniaturised and integrated into COTS earphones. With this form factor, eBP could frequently measure blood pressure while the users listen to music without obstructing their daily activity. It is also socially acceptable as all the measuring components are well hidden inside and behind the user's ear.

5.4 In-ear Blood Pressure Estimation Algorithms

This section presents the algorithm to **measure the systolic BP with a partially blocked artery, which does not depend on the fixed-BP ratio**. In conventional oscillometry, after detecting the MPA as the highest pulse amplitude (A_M) (Fig. 5.8 (b)), the peak amplitude fractions are applied to estimate the location of systolic amplitude (A_S) and diastolic amplitude (A_D). From [293], the ratio of A_S and A_M is 0.61, whereas the ratio between the A_D and A_M is 0.74. However, this ratio is not

consistent, as [294, 295] reports them to be 0.55 and 0.85, respectively. In addition, these ratios are for the arm cuff BP monitoring devices. Moreover, deriving an in-ear oscillation ratio requires a huge data set and an invasive method for measuring BP from inside the ear as a reference. Our method is independent of the ratio and does not completely occlude the artery. In fact, the pressure only needs to be slightly higher than MPA. eBP determines the MPA and DBP first, from direct measurements and then infers SBP indirectly.

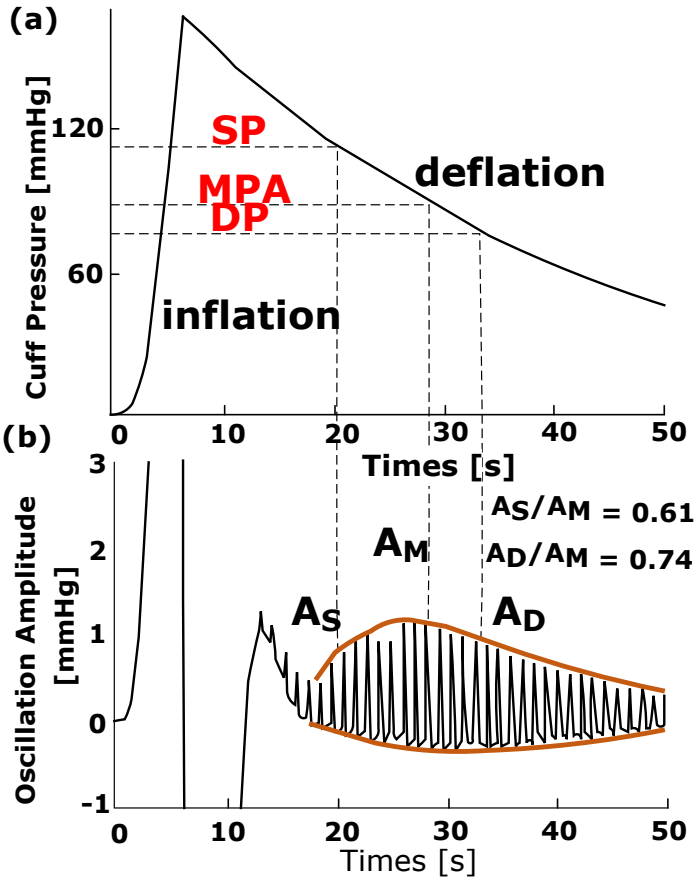


Figure 5.8: Amplitude vs. Pressure.

5.4.1 Systolic BP Measurement

To estimate the SBP (P_S) given the pressure of MPA (P_M) and diastole (P_D), we apply Eq. 2.1, where β is the systole ratio of the cardiac cycle. Most literature reports β as a fixed value [146, 147], and this is widely accepted, but it is developed for the brachial artery, not for the artery inside the ear. In our eBP system, we propose an adaptive estimation for β. To achieve this goal, we first explain the derivation of the Eq. 2.1. By considering one PPG cycle, we can formulate MPA as follows:

$P_M = \sum_{i=1}^{n} P(i)/n$ in discrete form or $P_M = \frac{1}{\tau}\int_0^{\tau} P(t)dt$ in continuous form. By assuming systole belongs to the interval (0, $\tau\beta$) and diastole is from ($\tau\beta$, τ), P_M is the total pressure average of systolic and diastolic pressure.

$$P_M = \frac{1}{\tau}\int_0^{\tau\beta} P(t)dt + \frac{1}{\tau}\int_{\tau\beta}^{\tau} P(t)dt \tag{5.1}$$

Then, we multiply the first term and second term by β and $1-\beta$, respectively.

$$P_M = \beta\left[\frac{1}{\tau\beta}\int_0^{\tau\beta} P(t)dt\right] + (1-\beta)\left[\frac{1}{\tau(1-\beta)}\int_{\tau\beta}^{\tau} P(t)dt\right] \tag{5.2}$$

$\frac{1}{\tau\beta}\int_0^{\tau\beta} P(t)dt$ is the average of SBP and $\frac{1}{\tau(1-\beta)}\int_{\tau\beta}^{\tau} P(t)dt$ is the average of DBP. Eq. 2.1 is equivalent to Eq. 5.2 by substituting $P_S = \frac{1}{\tau\beta}\int_0^{\tau\beta} P(t)dt$ and $P_D = \frac{1}{\tau(1-\beta)}\int_{\tau\beta}^{\tau} P(t)dt$. In one cycle, we detect the peak and two bottom points and then subtract their position in sequence as shown in Fig. 5.9. Δt_D, Δt_S and Δt_C are the duration of diastolic, systolic and the whole cycle, respectively. The systolic fraction is $\beta = \Delta t_S/\Delta t_C$. Given a frame of n cycles, we can compute β by averaging all β_i in the frame. Since our system runs in real-time, we only collect the first ten cleanest frames to estimate the systolic fraction. However, we first need to determine MPA and DBP prior to estimating the SBP. Their estimation procedures are described in the following sections.

5.4.2 Mean Arterial Pressure Detection

Mean arterial pressure represents the pulse pressure or the highest PPG amplitude (i.e., the MPA point). The precise location depends on the quality of the amplitude. In this section, we reveal the issue of using peak-to-peak to calculate the amplitude of real-time data and propose our improvement. In addition, sometimes, a sharp drift occurs when switching from inflation to deflation. The drift's amplitudes are higher than that of the real MPA, leading to a false detection problem. We also present our solution to address the issue of drift in this section.

To compute the peak-to-peak amplitude of a window sample containing multiple pulses, conventional wisdom calculates the amplitude every cycle and averages them. Fig. 5.10 (a) demonstrates when the first peak appears before the first bottom, then, the amplitude is calculated by subtracting the peak from the right-side bottom. However, when the bottom occurs first, the left-side bottom is subtracted, as shown in Fig. 5.10 (b). Since the appearance of the peak or bottom is variable, the amplitude obtained from each frame is inconsistent. To obtain a unique calculation, we omit the

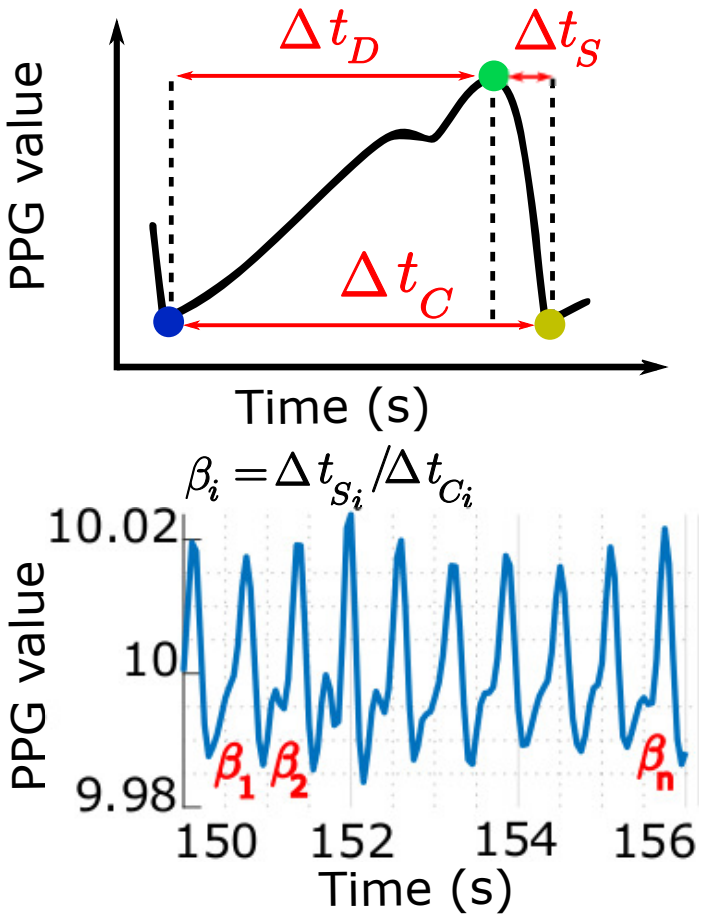


Figure 5.9: Systolic fraction β detection.

first bottom so that Fig. 5.10 (b) can convert into Fig. 5.10 (a). Let $P = \{p_i, i = \overline{1,n}\}$ and $B = \{b_i, i = \overline{1,m}\}$ represent the position sets of peaks and bottoms of the PPG signal X, respectively. The amplitude of each cycle denoted as amp_i, can be derived as follows:

$$amp_i = \begin{cases} X(p_i) - X(b_i), & p_i < b_i \\ X(p_i) - X(b_{i+1}), & otherwise \end{cases} \tag{5.3}$$

Fig. 5.11 (a) demonstrates the PPG signal variation with respect to reducing pressure. The bottom panel displays a PPG signal sample from 150^{th} to 170^{th} seconds. Fig. 5.11 (b) shows the corresponding amplitude using the peak-to-peak method.

Drift Removal: To detect the correct MPA point instead of ones belonging to the drift, we impose additional criteria leveraging the local maxima property. Specifically, MPA is not only the maximum amplitude point, but it also indicates the pulse amplitude transient state of increasing to decreasing as shown in Fig. 5.8 [293] and 5.11. In contrast, points within the drift are not local maximums. Therefore, we employ the following steps to precisely detect the MPA.

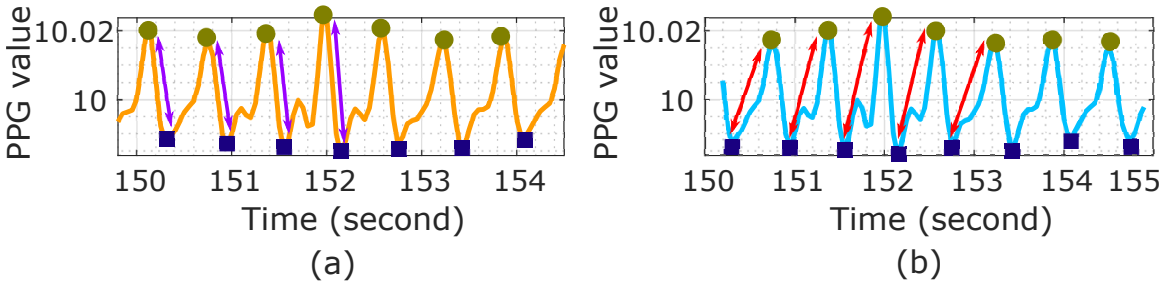


Figure 5.10: Illustration of the inconsistency of conventional peak-to-peak computation.

1. Detect pulse amplitude's local maximums. This step confirms the removal of points belonging to the drift.
2. The highest value of local maximums corresponds to the MPA location.

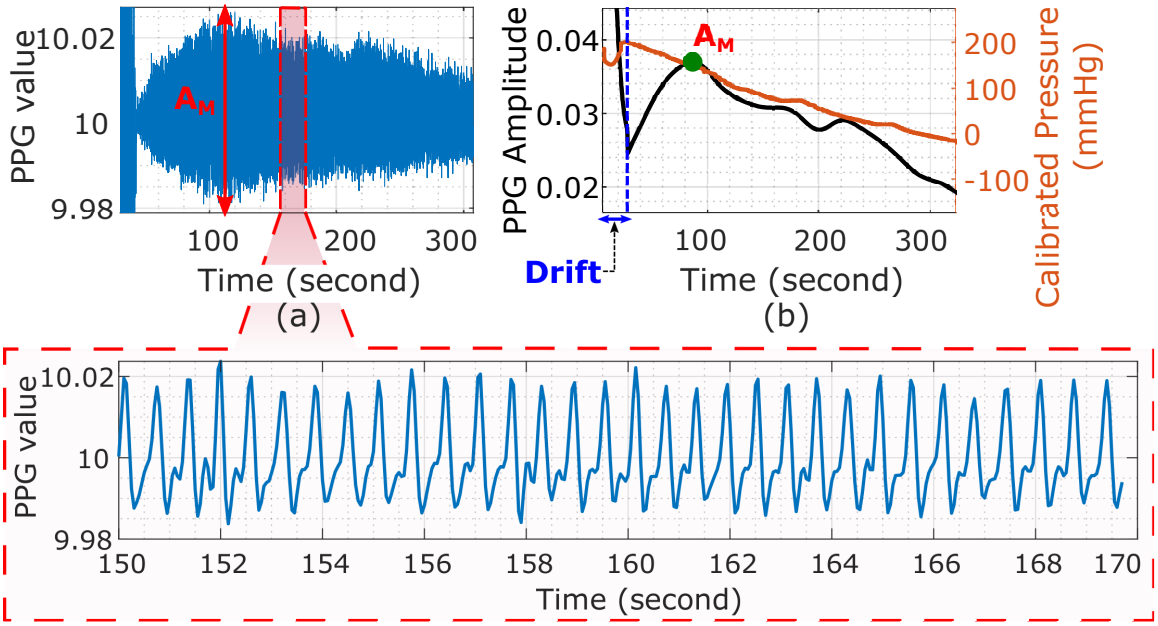


Figure 5.11: In-ear PPG signal (a) with corresponding amplitude and pressure (b).

5.4.3 Diastolic BP Measurement

Since the oscillation ratios cannot be applied to eBP, we propose a non-ratio method to estimate the DBP. When the pressure passes the MPA point, it will reach the point of diastole that yields the following unique signature: according to [144], amplitude rapidly decreases once the pressure passes the MPA and moderately decreases once it reaches the diastole point. In other words, the DBP position occurs at the highest decreasing amplitude. We can formulate this as the minimum of the first derivative

amplitude. Fig. 5.12 illustrates this idea. The dashed orange line represents the PPG amplitude, the solid purple is the first derivative, the dotted blue depicts the calibrated pressure, and the grey one is the PPG signal. In this example, the drift does not occur. Thus the MPA is the local maxima of the amplitude, which corresponds to the 0 points of the first-order derivative. After the MPA, a rapid decrease is observed until the 43^{rd} second, which corresponds to the minimal first-order derivative and indicates the location of DBP.

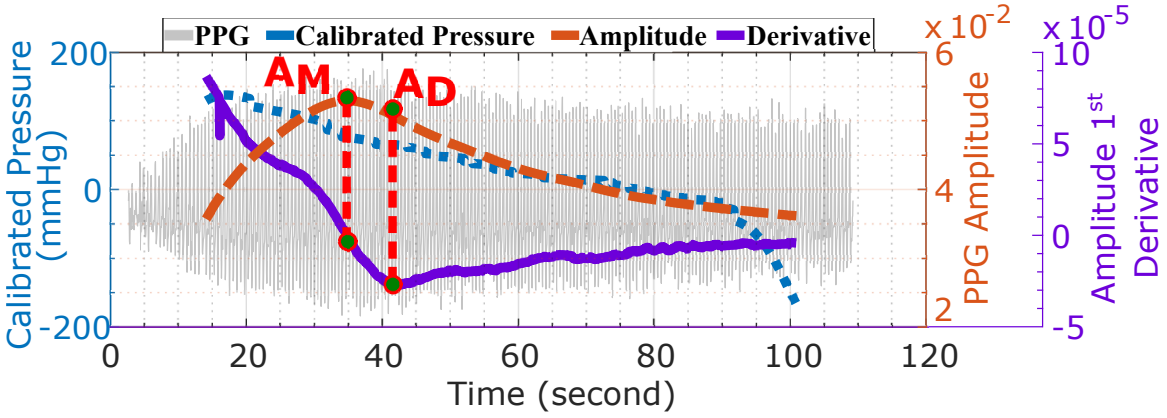


Figure 5.12: First derivative of PPG amplitude discloses diastolic BP.

5.5 PPG Qualification and Signal Processing

Due to the sensitivity of the pulse signal inside the ear, we propose a set of criteria to qualify the correct pulse shape. This criterion serves to remove non-pulse data chunks and detect the correct location to measure BP. In this section, we also present key signal-processing techniques. We process the recording signal in real-time using a sliding window of approximately 4 seconds with 90% overlap. Before extracting the BP measurement, we pre-process each window as follows:

Signal Qualification. Our signal qualification aims to omit non-pulsatile data and movement noise from the user and the in-ear sensor probe:

Peak Interval Variability (PIV): The ideal PPG signal has a fixed peak-to-peak interval, however, as the PPG signal is quasi-periodic, there is interval variability. The PIV is calculated as the standard deviation of all peak intervals in the current window. In normal conditions, the PIV falls under a certain threshold. However, the sensor is prone to movement noise, which disrupts the PPG trend and the corresponding peak interval. This leads to a drastic change in the peak-to-peak intervals with increasing

PIV, which informs us that the current processing window is contaminated by noise. In our application, a PIV threshold of 1.1 is sufficient to detect the distortion.

Algorithm 2: PPG signal Entropy Variance

input : x /* Current window signal with n samples */
B /* List of bottom indices */
output: vS /* Variance Entropy of current window */

```
ls ← []; /* List entropy */
for each sample b_i in B do
    u ← getSubSignal(x,b_i, b_{i+1}); ls[i] ← calcEntropy(u);
vS ← calcVariance(ls);
return vS;
```

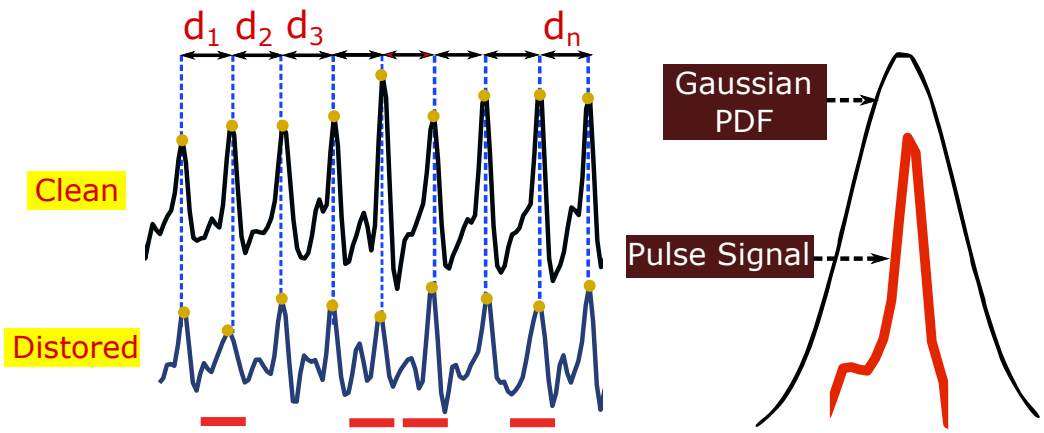


Figure 5.13: Peak Interval Variance versus Entropy variance.

Entropy variance: The PIV qualifies the pulse signal based on quasi-periodic properties. This is sufficient to calculate the heart rate but inadequate to compute the PPG amplitude. Consider the two data chunks in Fig. 5.13; both the clean and distorted signals have similar PIV scores when their peaks almost overlap each other. However, the distorted signal peak height, or amplitude, is inconsistent. The red underline marks non-pulse-shaped signals. Therefore, a rule is needed that restricts the signal based on the pulse shape for our motion elimination model. Among different techniques, entropy can classify pulse waveform [296, 297, 298] by quantifying how much the signal probability density function differs from a uniform distribution. The entropy S of a signal x is defined as $S(x) = -\sum_{i=1}^{n} x[i]^2 log_e(x[i]^2)$ that correlates to the change in signal shape. A chunk of PPG data is clean whenever all pulses have similar entropy. In other words, the entropy variance should be smaller than a threshold of 230, which, according to our experiment, is the minimum value to have a

clean and stable PPG signal. Algorithm 2 summarizes the estimation of a PPG signal entropy window. Given a window signal with n samples and a bottom indices list, we isolate each pulse by collecting data between b_i and b_{i+1}. The entropy is calculated for all extracted pulses and stored in a list ls. Finally, we obtain the entropy variance by computing the variance of ls.

We propose two metrics, including Peak Interval Variability and Entropy variance, to assess the signal quality. The PPG signal is considered clean data if it satisfies both criteria and will be processed to estimate BP.

Bandpass Filter. We apply the finite impulse response (FIR) bandpass filter in the 0.42-3.33 Hz range to obtain the PPG amplitude. The PPG peak is essential to predict the amplitude and BP, which occurs between 0.42Hz and 3.33 Hz corresponding to the human heart rate range of 25 to 200 beat-per-minutes [299].

Gaussian Filter. High-frequency noise present in PPG amplitude can corrupt the true location of MPA and diastole, thereby increasing the error rate. A comparison between Gaussian and Median filtering [300] is conducted to identify the optimal amplitude smoothing technique. To quantify the granularity level η, we use the standard deviation of differences $\eta = \sigma(\frac{dx}{dt})$ with x as the interested signal. Fig. 5.14 (a) shows that Gaussian filtering is more fine-grained than Median filtering. In particular, the granularity level η of Gaussian filtered signals is one order of magnitude higher than the Median kernel, thereby confirming our choice. We compute the Gaussian kernel h via the following equation: $h[i] = e^{-(i-w_h/2)^2/2\sigma_h^2}, i = \overline{1, w_h}$, where w_h is the filter size and $\sigma_h = w_h/5$.

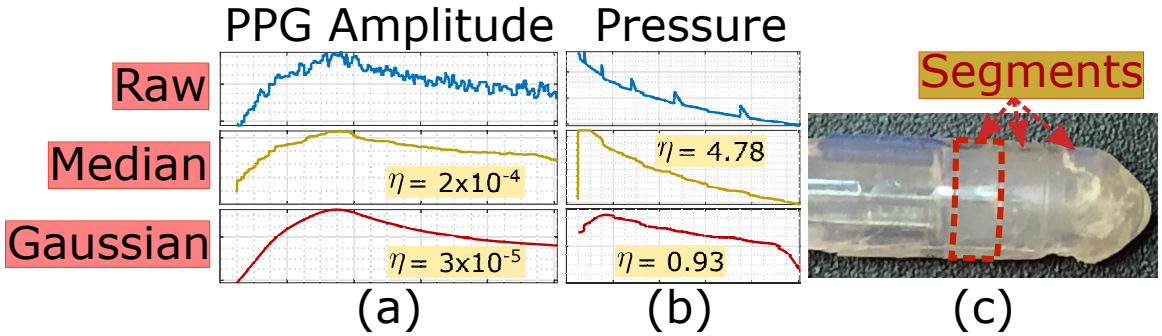


Figure 5.14: Results of median and Gaussian smoothing filter on amplitude (a) and pressure trend (b), and the balloon structure (c).

Pressure Spikes Correction: During deflation, pressure changes should yield a linear trend. However, the balloon's structure is composed of multiple segments (Fig. 5.14 (c)), which induces a faster contraction when crossing the border between

segments. This event causes spikes in the pressure deflation raw signal, as shown in Fig. 5.14 (b). Unlike the amplitude, there is minimal difference between the two filtering approaches granularity score (Fig. 5.14 (b)). Therefore, we use the Median filter to address those spikes, as small drifts are still observed at either end following Gaussian filtering.

PPG Peak Detection. Each heartbeat interval creates one local peak, which can be differentiated using a peak detection algorithm within each predefined non-overlapping subwindow. A signal sample s_i is considered a peak if it is the maximum of that window.

Up until this point, we have described the complete system design. In the following section, we present our system performance through an intensive evaluation.

5.6 Evaluation

In this section, we present the set of experiments conducted to evaluate the overall performance of eBP and demonstrate the feasibility of using our BP device frequently in daily life. We first present the key results of performing BP measurements using the eBP system. Then, we evaluate different factors that can affect eBP's performance. Finally, we analyze the users' experience survey when using eBP.

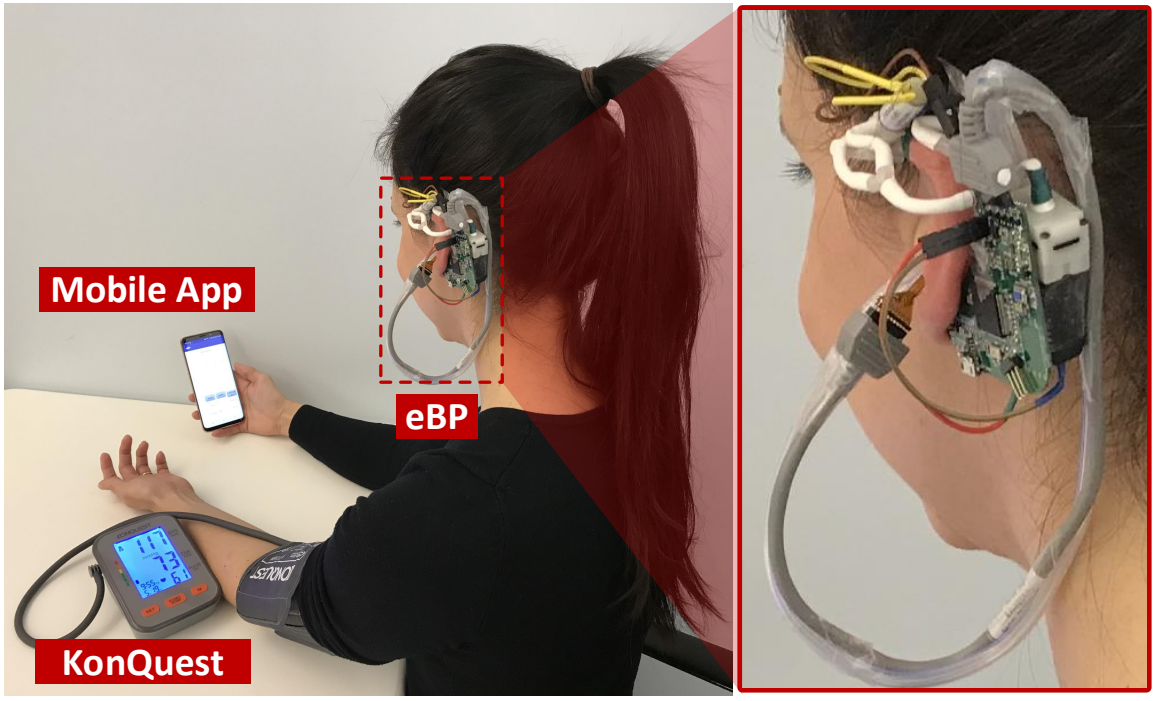


Figure 5.15: Experiment setup to compare eBP with the KonQuest device.

5.6.1 Experimental Methodology

We obtained the IRB approval to conduct experiments for the evaluation of eBP. The participant demographics are shown in Table 1. We tested eBP alongside an FDA-approved, gold standard, arm-cuff BP measurement device (KonQuest KBP-2704A [301]) (Fig. 5.15). For assessment, we use the metric that is widely accepted by other BP studies, which consists of bias or mean error μ, a precision or standard deviation (SD) σ error and a Pearson correlation coefficient ρ.

This experiment is tested on 35 participants of both genders and various ages. eBP participants place the in-ear balloon inside their ears. Next, the cuff of the KonQuest device is wrapped around the upper arm of each participant. We simultaneously measure the BP of each participant from our Android app running on the Samsung Galaxy S9 and the gold standard BP device. This process is repeated twice and takes about twenty minutes. We sterilise our device with an alcohol wipe, between each experiment, by softly cleaning the balloon tip and sensor. During the experiment, the participant has to sit still to comply with the standard BP measurement protocol and ensure correct readings are obtained from the KonQuest device. In addition, the balloon needs to be mounted in the right position so that it will not fall out. It turns out that only the P2 location in Fig. 5.20 (a) can hold the sensor properly because the tragus helps to keep the sensor tight.

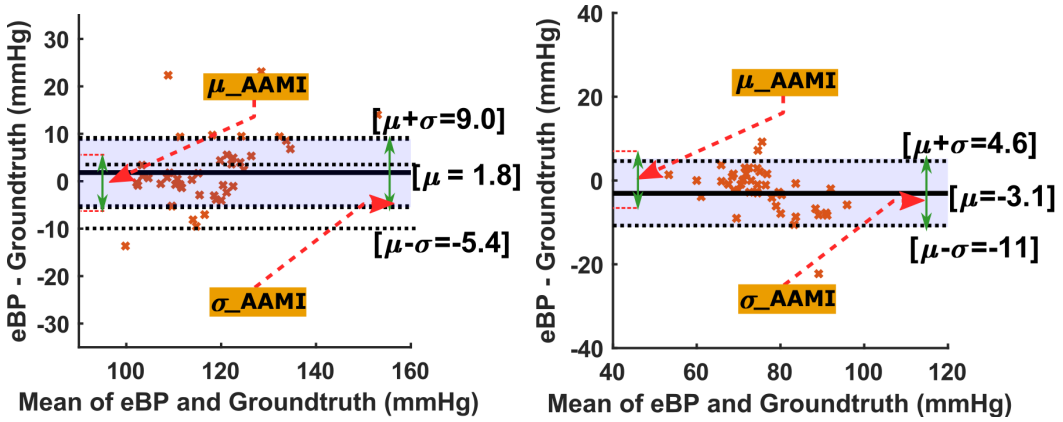
Table 5.1: Demographic description of participants

Demographic data of study population	
Age (years)	18 - 35 years old
Blood Pressure	Systolic: 93-146, Diastolic: 53-113
Gender Ratio	Male: 24, Female: 11

5.6.2 System Performance

In this section, we evaluate eBP performance and showcase the comparative results between eBP and the KonQuest KBP-2704A.

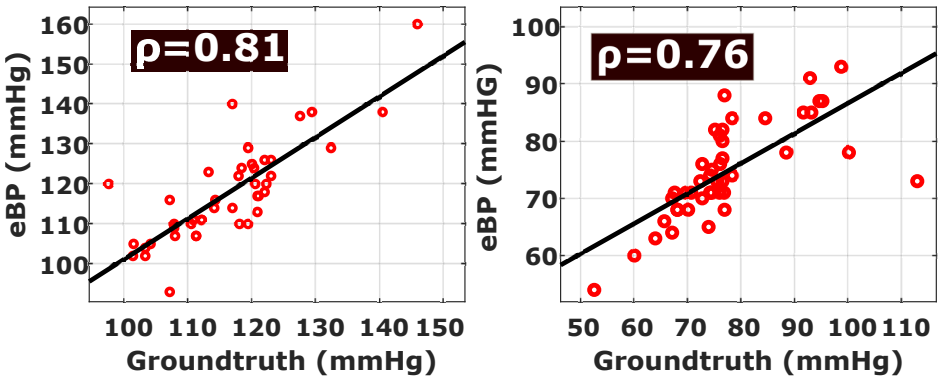
Fig. 5.16 shows the Bland-Altman diagram that describes the average error between eBP and the ground truth. Consequently, the mean and SD error of SBP and DBP are 1.8 mmHg and 3.1 mmHg, which is within the Association for the Advancement of Medical Instrumentation's (AAMI) requirement ($\mu_{AAMI} < 5mmHg$) [302]. In addition, our SD error for SBP and DBP also satisfies the criteria where



(a) SBP (b) DBP

Figure 5.16: Bland-Altman plot comparing eBPś measurements and groundtruth .

$\sigma_{AAMI} < 8mmHg$ [302]. On the other hand, Fig. 5.17 displays the Pearson correlation coefficients of SBP and DBP measurements by eBP and the KonQuest device. We select five participants' data for calibration using a polynomial regression model. There were some error cases where the measurement was performed without taking sufficient stability. For example, the balloon fell out of the ear because of sweat, movement, or the ear canal was too narrow. When the balloon falls out, there is no valid pulsatile waveform detected. As a result, the system cannot predict the BP, thus, providing no data for the evaluation. The correlation shown of 0.81/1.0 for the SBP and 0.76/1.0 for the DBP represents that our system's prediction is highly correlated to that of the FDA-approved device.



(a) SBP (b) DBP

Figure 5.17: Pearson correlation coefficients of eBPś estimation and groundtruth.

5.6.3 Power consumption

We measure the power consumption of both eBP hardware module and eBP app (installed on a Samsung Galaxy S9) in two scenarios: (1) during BP measurement and (2) without BP measurement. eBP hardware module power consumption is measured using Monsoon Power Monitor. eBP app power consumption is measured using the AccuBattery application. Note that the power measurement of eBP hardware is done in 1 minute, while it takes 9 minutes to obtain a reliable measurement from AccuBattery app. eBP hardware consumes 1279.28mW and 31.34mW during BP measurement and without BP measurement, respectively. eBP app consumes 1406 mW and 1119 mW during eBP measurement and without BP measurement, respectively. In summary, eBP hardware consumes 1247.94 mW (1279.28 mW - 31.34 mW) to operate the pump, the valve, the LED, and the microcontroller. eBP app consumes 287 mW (1406 mW - 1119 mW) for BP calculation.

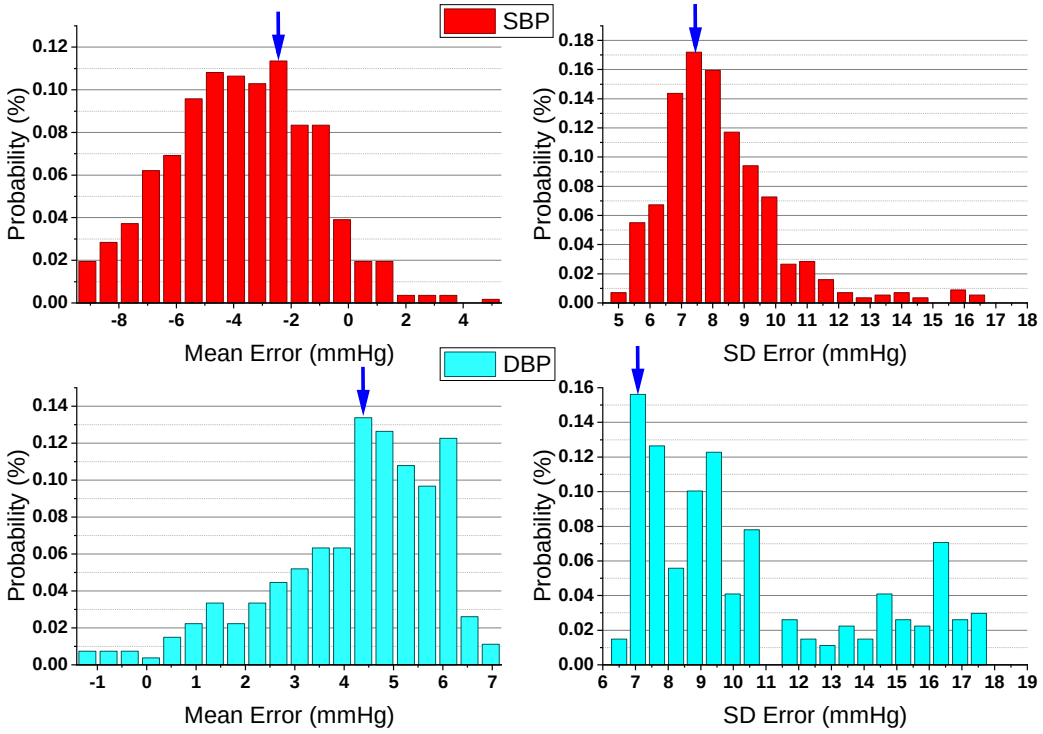


Figure 5.18: Mean and SD error in cross-validation.

5.6.4 Prediction Stability

We conducted experiments to verify the robustness of our calibration procedure based on polynomial fitting. We replicated the process by taking 250 randomly picked times

from the learning set. Finally, we explore the frequencies of mean and SD error as shown in Fig. 5.18. Overall, the highest frequencies of both SBP and DBP mean error falls between 4-5 mmHg which satisfies AAMI standards. Similarly, the highest frequency of SD errors is less than 8mmHg, which also qualifies the AAMI protocol. In addition, 9 out of 35 candidates proceeded with ten data collection times to calculate the intraclass correlation coefficient (ICC). Fig. 5.19 shows the ICC result of each candidate. The average ICC of SBP and DBP are 0.8 and 0.76, respectively.

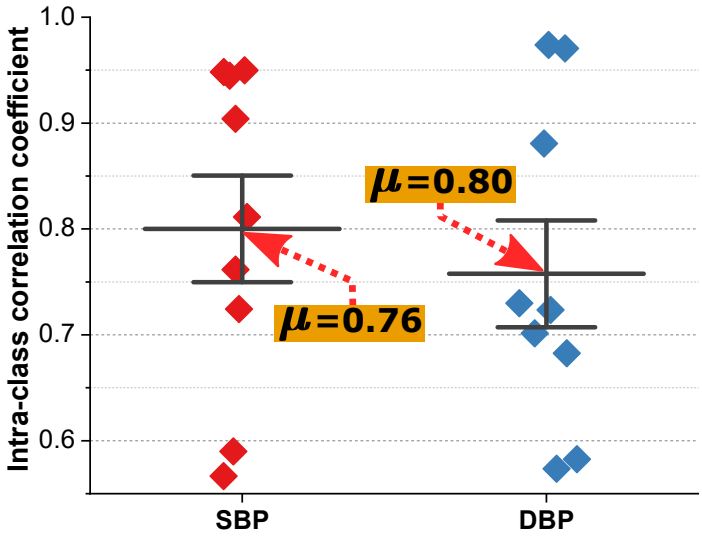


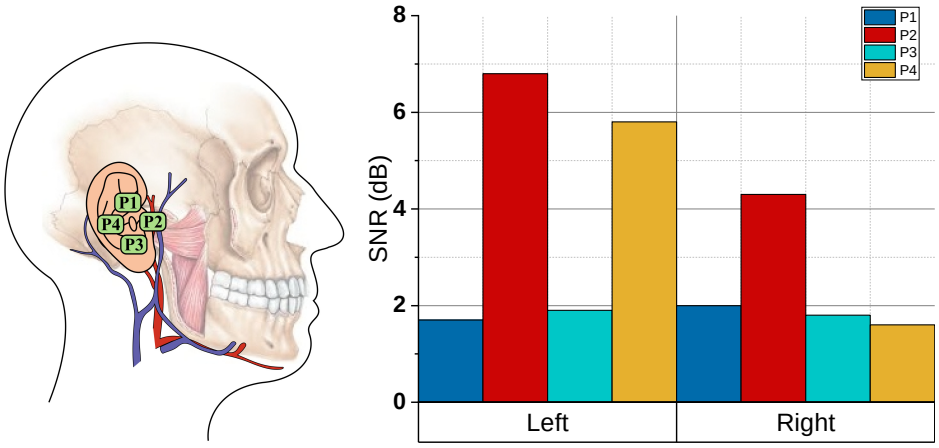
Figure 5.19: Intraclass correlation coefficient of eBP and ground truth.

5.6.5 Optimal sensor location

In this section, we evaluate PPG signal quality when being captured from inside the ear. Fig. 5.20 (a) illustrates the four most common positions our sensors are located in. We collect 10 seconds of data from each participant separately from the main test for this evaluation. Fig. 5.20 (b) shows that the signal power is highest at position P2 on the left side.

5.6.6 User Experience Survey

We asked participants (students and school staff) who participated in the study to complete a questionnaire on their experience with the eBP once the experiment was complete. The survey included questions on the users' perspective of eBP's convenience, portability, ease of use, and a comparison of eBP to the digital device. The survey uses a rating scale from 1 to 5, corresponding with "Very Unsatisfied" to "Very Satisfied".



(a) Setup (b) SNR

Figure 5.20: The SNR of PPG signals at different sensor locations.

After we explained how to use eBP to record a BP measurement, users were then asked to place the device inside their ears. As shown in Fig. 5.21, more than 60% of the users reported that they were satisfied or highly satisfied that they could easily use eBP by themselves. As this experiment was the first time all participants had their BP measured from inside their ears, some users were initially confused about the operating procedure eBP. We believe that after repetitive use, people would be more comfortable using eBP by themselves. Additionally, 65% of the users found eBP's portability to be satisfactory or very satisfactory. The results also show that more than 80% of the users were satisfied or very satisfied with the convenience of eBP. When the participants compared the use of eBP alongside the commercial digital BP device, 75% of the users considered eBP easier to use. Overall, participants agreed with the novelty of measuring BP in the ear and were satisfied with eBP. However, there were 3 cases where users reported a little scratch due to the soft silicone falling off. Also, 7 out of 35 participants did not like the noise when the balloon partially slid out of the ear canal during an inflated state.

5.7 Summary

In this chapter, we presented eBP a new method to capture BP from inside the ear that measures the artery BP from the superficial artery near the ear canal. Existing techniques which measure BP on the arm or wrist cannot be applied to measure BP from inside the ear as they require fixed systolic and diastolic detection ratios. We developed our model to estimate the in-ear BP by observing the behaviour of

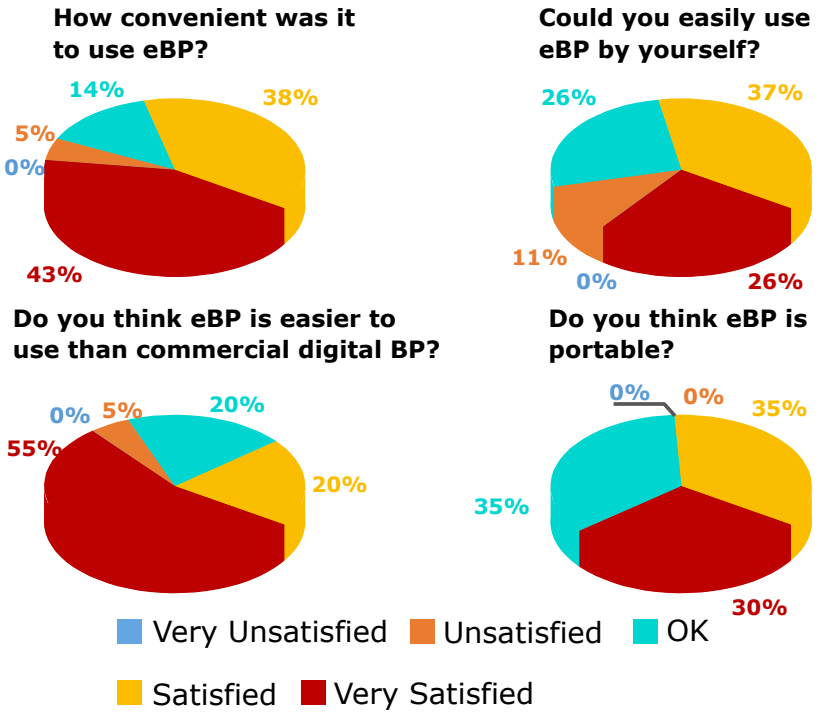


Figure 5.21: User experience questionnaire results.

pulse amplitude. Therefore, no constant parameters are required in our proposed model. In this paper, we also introduce a technique to customise an off-the-shelf catheter to become our in-ear pulse pressure sensor. We built custom hardware and software for eBP and evaluated the system through a comparative study on 35 subjects. The study shows that eBP obtains an average error of 1.8 mmHg and -3.1 mmHg and standard deviation error of 7.2 mmHg and 7.9 mmHg for systolic (high-pressure value) and diastolic (low-pressure value), respectively. These errors are within the acceptable margins regulated by the FDA's AAMI protocol, which allows an average BP difference of up to 5 mmHg and a standard deviation of up to 8 mmHg. These promising results not only show the feasibility of an in-ear blood monitoring concept but also open up the possibility of making current gold standard cuff-based BP measurement more comfortable.

Chapter 6

Conclusions

6.1 Concluding Remarks

In this thesis, we explore the challenges of enabling human physiological sensing by leveraging head-worn wearable computers to improve human healthcare and prevent life-threatening conditions that can be used in our daily life. In particular, we present our studies on techniques that utilise the unique advantages of wearable computers to facilitate new sensing capabilities to capture various biosignals from the brain, the eyes, facial muscles, sweat glands, and blood vessels. We also propose techniques to address motion artefacts and environmental noise in real time at the electrical level. Furthermore, we develop a framework to enable long-term and high-fidelity biosignal monitoring with efficient on-chip intelligence and a novel pattern-driven compressive sensing algorithm. Finally, we have demonstrated the usability of the proposed methods in three practical use cases such as microsleep detection, blood pressure measurement, and epileptic seizure monitoring.

6.2 Future Work

So far, we have only looked at a small set of possible applications that could be enabled by leveraging head-worn wearable computers. Furthermore, important topics such as security and privacy, user's long-term comfort, form factor, etc., have yet to be considered in the current state. While we have made efforts to make the proposed systems practical, the proposed designs and implementations have some limitations that are open for future investigation. In this section, we will discuss the potential research directions enabled by the proposed systems.

■ **Head-based biosignal sensing with a behind-the-ear wearable.** In this project, we have demonstrated the ability to capture multiple biosignals originating

from the brain, eyes, facial muscles, and sweat glands on the human head. In the future, we will aim to investigate the following issues.

In-the-wild evaluation. With the promising results from our in-lab evaluation, we aim for a larger-scale out-lab evaluation. One of the key challenges is the limitation of the existing ground truth for microsleep detection. Up to now, the gold standard to objectively assess microsleep is based on polysomnographic (PSG) data, which can only be conducted in a controlled environment. Ground truth based on pupil dilation or eye-tracking is a potential direction worth exploring.

Impact of sweat condition. While WAKE can address motion and environmental noise, there are several artefacts posing challenges to the real-world usability of a wearable system like WAKE. For instance, sweating and hydration can introduce noise into the measurement. Addressing these artefacts is the question that we will explore to enhance the practicality of WAKE.

Optimizing WAKE device. The current prototype is designed with off-the-shelf materials and components. Hence, it is challenging to ensure the manufacturing quality of our customised earpieces, electrodes, and sensing circuits.

Trade-off between classification performance and latency. Though highly complex prediction models may improve accuracy, they usually take a larger amount of computations because of the complex processes. We, thereby, make use of simpler off-the-shell classifiers, which are deployed and optimised commonly in latency-sensitive libraries. Another key part of our system is to use highly informative expert-based features which are simple and efficient to compute. These settings help us reduce the processing and inference time, shortening the latency toward warning the user.

Feasibility of deep learning on raw data. Deep neural networks (DNNs) are designed to extract features automatically from raw bio-signal data. This is a huge advantage compared to other classical machine learning algorithms, which require labour-intensive domain-crafted features. Shown in Table 3.5, DNNs achieve high detection accuracies for both test-set and 10-fold CV evaluations, which are as good as the performance of feature-based models (Tables 3.4, 3.3). However, we observe a non-trivial drop in performance in the LOSOCV setting. Notice that the LOSOCV is highly challenging because bio-signal data come from different subjects. We hypothesise that this drop is due to a large variance of inter-subject signals, outliers in physiological signals, and the lack of training data for DNNs. On the other hand, classical algorithms utilise embedded prior knowledge in hand-designed features to generalise better in the low-data regime. Nevertheless, our results illustrate the promise

of DNNs in detecting micro-sleep using our WAKE data and leave an open question on improving its generalisation over different subjects.

■ **Efficient pattern-driven compressive sensing framework.** We have presented a framework to design a highly efficient biosignal sensing wearable by exploiting the sparsity of the captured signals. We are currently looking into the following directions.

Improving TinyPR performance. The performance of TinyPR primitives directly links to the quality of the training data. Extensive and high-quality training data will result in more accurate pattern recognition models. We envision that the crowdsourcing effort of the community can address the challenge of high-quality training data. The more data we collect, the better TinyPR primitives can be built, and more healthcare wearable applications will be enabled.

Noise and artefacts. As motion, environmental, and muscle artefacts could contaminate the biosignal streams, pre-processing is needed to ensure signal integrity. We assume that motion and environmental noise could be mitigated on the sensing hardware by employing techniques such as active amplifying, sigma-delta modulation, and digital filters as proposed in previous works [179, 303]. For muscle activities, it is up to the application to decide whether they are signals of interest or unwanted artefacts. Thus, PROS provides a pattern primitive (MUSC) to detect muscle contractions. The application can further process the signal if needed.

Dataset's limitations. In this project, we chose the biosignal datasets that provide practical clinical settings, so we could gauge the effectiveness of PROS on various patients' conditions. The clinical settings, however, are relatively stable and will not reflect all the usage patterns in daily life, such as motion and environmental noise, wearing position, etc. We would love to investigate this further in our future work when real-life data becomes available.

Extending and sharing ability. We envision that PROS serves as a framework where the community could develop support for various biosensors after its release. The amount of potential biosensors and pattern primitives is significant. E.g., facial expressions (EMGs), emotions (electrodermal activity), coughing patterns (acoustic), and many more. Sharing the processing pipeline, such as TinyPR models or PDCS, among multiple applications is another exciting direction we are looking into to enhance the efficiency further.

Other considerations for daily usage. While the main focus of PROS is on battery lifetime, other factors could impact the user's experience. First, wearability is important since monitoring applications such as seizure detection rely on long-term

measurements to detect sudden attacks. Second, data privacy is another critical factor. Federated learning could tackle this issue by enabling multiple edge devices to collaborate and build a common model without exchanging local data. Finally, closed-loop control algorithms could be developed between mobile and wearable devices so that PROS could adapt to the changing conditions over time.

■ **Sensing in-ear blood vessels to monitor blood pressure with a wearable system.** This study proposes a wearable system that can sense blood vessel activities inside the ear and use the captured signals to enable comfortable and frequent blood pressure monitoring. There are several open issues that we would need to address in our future work.

Optimising the balloon design: Our current off-the-shelf medical balloon shape does not respond linearly to pressure changes, which causes several issues. One of these is the appearance of spikes, mentioned in Sec. 5.4, that we can remove with median filtering. Since the balloon's intended purpose is a urinary catheter, it has a high level of stiffness and thus demands strong pressure to break the equilibrium point. In addition, unlike the cuff, the elasticity of the balloon quickly recoils to its original state at the onset of deflation. Therefore, it inserts a large amount of pressure into the pressure sensor that sometimes overwhelms the device. As a result, improving the linearity of the balloon's material is our priority. The new material should satisfy the linearity property while being comfortable but still complying with the safety requirement as the current catheter balloon does.

Sensor placement: Changing the location of the PPG sensor inside the ear can affect the signal quality and system performance. Therefore, first of all, a better tragus-mounting mechanism is required to keep the sensor stable inside the ear. Then, SNR, PIV, and entropy variance are utilised to evaluate the signal quality of different sensor placement locations to obtain the cleanest signal.

High idle current draw: The idle current draw is 4.95 mA. 92% (4.6 mA/4.95 mA) of this value is consumed by the microcontroller, which is still running in its active mode. It is possible to further reduce the current draw, which can be as low as 2.1 uA, by using the standby mode of the microcontroller.

Leveraging on-chip processing data: Continuously streaming raw data over Bluetooth requires a significant amount of power (30 mA). This consumed power can be minimised by off-loading signal processing to the microcontroller and taking advantage of the dedicated DSP core. Therefore, on-chip processing can be utilised to lessen power consumption, which leads to smaller battery usage and hence reduce the size and weight of the device.

Miniaturizing the ear-worn circuit: The size of the current eBP prototype is not optimal. Therefore, design optimisation, including the use of smaller components and reducing power consumption, is required to miniaturise the eBP's main circuit.

Systole ratio estimation in atrial fibrillation (AF): Our system needs a correct estimation of the systolic ratio in the cardiac cycle to calculate BP. Therefore, when the pulse is irregular, as in an AF patient, it can affect the accuracy of the eBP measurement. Moreover, the system might not be able to return a BP result due to the failure to get a stable estimation of the systolic ratio. One possible solution is to reduce the window size and increase the number of selected cleanest frames, e.g. 100, so that it has more chances to capture clean data. In addition, a specific triggering mechanism needs to be designed to update the number of cleanest frames when an atrial fibrillation patient is detected. Further studies with clinical IRB are required to understand the performance of AF patients.

Large scale evaluation: Currently, eBP is evaluated on 35 participants from our institution. A more extensive study with a large number of participants and patients should be conducted to explore the practical effectiveness of our system.

Safety control and monitoring: So far, we have configured a safety threshold for the pressure and used a low-power DC pump to protect users. However, this solution only prevents the air from pumping into the balloon continuously. We also need a mechanism for supervising the eBP placement to ensure that users do not place it too deep inside their ears.

6.3 Long-term Vision

Tracing back to Mark Weiser's original vision of ubiquitous computing in 1991, the most profound technologies are the ones that disappear and weave into our daily life until they become indistinguishable from it [304]. I envision that in the near future, hundreds of wearable computers will "invisibly" attach and vanish into our human bodies. They could sense and modulate our biological processes and signals to help improve healthcare and well-being in a way we can never achieve with conventional medical devices. There is still a lot of work that we need to do, and this thesis is a first step towards that goal.

Bibliography

[1] Precedence Research. *Wearable Technology Market - Global Industry Analysis, Size, Share, Growth, Trends, Regional Outlook, and Forecast 2022-2030.* https://tinyurl.com/ja3kdaa4.

[2] Fortune Business Insights Pvt. Ltd. *Latest Research 2020: Wearable Medical Devices Market Witness Astonishing Growth at 24.7% CAGR to Reach USD 139,353.6 Million by 2026.* https://tinyurl.com/va7u3vap.

[3] Chris Falkous and Julianne Callaway. *Wearable Technology in Life Insurance.* https://tinyurl.com/3ypnb3de.

[4] Chris Falkous and Julianne Callaway. *Wearable Technology in Life Insurance.* https://tinyurl.com/3ypnb3de.

[5] Romit Roy Choudhury. Earable computing: A new area to think about. In *Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications*, pages 147–153, 2021.

[6] C-M Tsai, S-L Chou, Elliot N Gale, and Willard D McCall. Human masticatory muscle activity and jaw position under experimental stress. *Journal of oral rehabilitation*, 29(1):44–51, 2002.

[7] Ulf Lundberg, Roland Kadefors, Bo Melin, Gunnar Palmerud, Peter Hassmén, Margareta Engström, and Ingela Elfsberg Dohns. Psychophysiological stress and emg activity of the trapezius muscle. *International journal of behavioral medicine*, 1(4):354–370, 1994.

[8] K Kohyama, L Mioche, and P Bourdio3. Influence of age and dental status on chewing behaviour studied by emg recordings during consumption of various food samples. *Gerodontology*, 20(1):15–23, 2003.

[9] Laurence Mioche, Pierre Bourdiol, Jean-Francois Martin, and Yolande Noël. Variations in human masseter and temporalis muscle activity related to food texture during free and side-imposed mastication. *Archives of Oral Biology*, 44(12):1005–1012, 1999.

[10] Kaoru Kohyama, Laurence Mioche, and JEAN-FRANCOIS MARTIN. Chewing patterns of various texture foods studied by electromyography in young and elderly populations. *Journal of Texture Studies*, 33(4):269–283, 2002.

[11] Xiao-Wei Wang, Dan Nie, and Bao-Liang Lu. Emotional state classification from eeg data using machine learning approach. *Neurocomputing*, 129:94–106, 2014.

[12] Klaus-Robert Müller, Michael Tangermann, Guido Dornhege, Matthias Krauledat, Gabriel Curio, and Benjamin Blankertz. Machine learning for real-time single-trial eeg-analysis: from brain–computer interfacing to mental state monitoring. *Journal of neuroscience methods*, 167(1):82–90, 2008.

[13] J Sarnthein, A Morel, A Von Stein, and D Jeanmonod. Thalamic theta field potentials and eeg: high thalamocortical coherence in patients with neurogenic pain, epilepsy and movement disorders. *Thalamus & Related Systems*, 2(3):231–238, 2003.

[14] C De Deyne, Michel Struys, Johan Decruyenaere, J Creupelandt, E Hoste, and Francis Colardyn. Use of continuous bispectral eeg monitoring to assess depth of sedation in icu patients. *Intensive care medicine*, 24(12):1294–1298, 1998.

[15] Clete A Kushida, Michael R Littner, Timothy Morgenthaler, Cathy A Alessi, Dennis Bailey, Jack Coleman Jr, Leah Friedman, Max Hirshkowitz, Sheldon Kapen, Milton Kramer, et al. Practice parameters for the indications for polysomnography and related procedures: an update for 2005. *Sleep*, 28(4):499–523, 2005.

[16] Konrad E Bloch. Polysomnography: a systematic review. *Technology and health care*, 5(4):285–305, 1997.

[17] Michael Murias, James M Swanson, and Ramesh Srinivasan. Functional connectivity of frontal cortex in healthy and adhd children reflected in eeg coherence. *Cerebral Cortex*, 17(8):1788–1799, 2007.

[18] LM Jonkman, C Kemner, MN Verbaten, H Van Engeland, JL Kenemans, G Camfferman, JK Buitelaar, and HS Koelega. Perceptual and response interference in children with attention-deficit hyperactivity disorder, and the effects of methylphenidate. *Psychophysiology*, 36(4):419–429, 1999.

[19] R Bernier, G Dawson, S Webb, and M Murias. Eeg mu rhythm and imitation impairments in individuals with autism spectrum disorder. *Brain and cognition*, 64(3):228–237, 2007.

[20] B Rael Cahn and John Polich. Meditation states and traits: Eeg, erp, and neuroimaging studies. *Psychological bulletin*, 132(2):180, 2006.

[21] *actiCHamp*. https://goo.gl/uagGNX.

[22] *BIOPAC Electrode Cap*. https://goo.gl/GGBrJv.

[23] *Emotiv brainwear*. https://goo.gl/uagGNX.

[24] *NeuroSky MindWave*. https://goo.gl/cEf7fi.

[25] *BrainLink Pro*. https://goo.gl/hrxuAb.

[26] *Muse*. https://goo.gl/5zwtcJ.

[27] *Kokoon*. https://goo.gl/tr3QJV.

[28] *Kokoon*. https://goo.gl/ubW3T7V.

[29] *Neuroon Open*. https://goo.gl/5WyTuj.

[30] *Neuroon Open*. https://goo.gl/iY3Ber.

[31] Xiaoyang Dong, Wen Ye, Yunliang Tang, Jun Wang, Linyang Zhong, Jing Xiong, Haiping Liu, Guohui Lu, and Zhen Feng. Wakefulness-promoting effects of lateral hypothalamic area–deep brain stimulation in traumatic brain injury-induced comatose rats: Upregulation of α1-adrenoceptor subtypes and downregulation of gamma-aminobutyric acid β receptor expression via the orexins pathway. *World Neurosurgery*, 152:e321–e331, 2021.

[32] Nasser Zangiabadi, Lady Diana Ladino, Farzad Sina, Juan Pablo Orozco-Hernández, Alexandra Carter, and José Francisco Téllez-Zenteno. Deep brain stimulation and drug-resistant epilepsy: a review of the literature. *Frontiers in neurology*, page 601, 2019.

[33] Ayhan Yoruk, John D Bisognano, and John P Gassler. Baroreceptor stimulation for resistant hypertension. *American journal of hypertension*, 29(12):1319–1324, 2016.

[34] J.F. Pagel. *Excessive Daytime Sleepiness*, May 2009. https://www.aafp.org/afp/2009/0301/p391.html.

[35] Marco Hafner, Martin Stepanek, Jirka Taylor, et al. Why sleep matters—the economic costs of insufficient sleep: a cross-country comparative analysis. *Rand health quarterly*, 6(4), 2017.

[36] Katrin Uehli, Amar J Mehta, et al. Sleep problems and work injuries: a systematic review and meta-analysis. *Sleep medicine reviews*, 18(1):61–73, 2014.

[37] Drew Dawson, Amy C Reynolds, et al. Determining the likelihood that fatigue was present in a road accident: a theoretical review and suggested accident taxonomy. *Sleep medicine reviews*, 2018.

[38] American Academy of Sleep Medicine. *Economic burden of undiagnosed sleep apnea in U.S. is nearly $150B per year.* https://tinyurl.com/yxkfe5zh.

[39] Poul Jennum, Stine Knudsen, et al. The economic consequences of narcolepsy. *Journal of Clinical Sleep Medicine*, 5(03):240–245, 2009.

[40] American Addiction Centers. *Narcolepsy and How Substance Abuse Effects It*, 2018. https://tinyurl.com/yxh4hh7z.

[41] RAND Europe. *Lack of Sleep Costing UK Economy Up to £40 Billion a Year.* https://tinyurl.com/3x9bpfau.

[42] Kazumi Takahashi, J-S Lin, and Kazuya Sakai. Neuronal activity of orexin and non-orexin waking-active neurons during wake–sleep states in the mouse. *Neuroscience*, 153(3):860–870, 2008.

[43] Takeshi Sakurai. The neural circuit of orexin (hypocretin): maintaining sleep and wakefulness. *Nature Reviews Neuroscience*, 8(3):171, 2007.

[44] Malik TR Peiris et al. Frequent lapses of responsiveness during an extended visuomotor tracking task in non-sleep-deprived subjects. *Journal of sleep research*, 15(3):291–300, 2006.

[45] Amit Paul et al. Variability of driving performance during microsleeps. 2005.

[46] Laura Higgins and Bernie Fette. Drowsy driving. 2012.

[47] The PSG Reading Center. Procedure manual for polysomnography. https://tinyurl.com/yc9ptdjz.

[48] Nitish V Thakor. Biopotentials and electrophysiology measurement. *Measurement, Instrumentation, and Sensors Handbook*, pages 64–1, 2017.

[49] Chetan S Nayak and Arayamparambil C Anilkumar. Eeg normal waveforms. statpearls, 2020.

[50] Orval Hobart Mowrer, Theodore C Ruch, and Neal E Miller. The corneo-retinal potential difference as the basis of the galvanometric method of recording eye movements. *American Journal of Physiology-Legacy Content*, 114(2):423–428, 1935.

[51] Yan Li, Carmen CY Poon, and Yuan-Ting Zhang. Analog integrated circuits design for processing physiological signals. *IEEE Reviews in Biomedical Engineering*, 3:93–105, 2010.

[52] Ulrika Svensson. *Blink behaviour based drowsiness detection: method development and validation*. Statens väg-och transportforskningsinstitut, 2004.

[53] Alan Macy. The handbook of human physiological recording. *Referenced*, 7:2017, 2015.

[54] Shanette A Go, Krista Coleman-Wood, and Kenton R Kaufman. Frequency analysis of lower extremity electromyography signals for the quantitative diagnosis of dystonia. *Journal of electromyography and kinesiology*, 24(1):31–36, 2014.

[55] Raman K Malhotra and Alon Y Avidan. Sleep stages and scoring technique. *Atlas of sleep medicine*, pages 77–99, 2013.

[56] Caroline J Smith and George Havenith. Body mapping of sweating patterns in male athletes in mild exercise-induced hyperthermia. *European journal of applied physiology*, 111(7):1391–1404, 2011.

[57] Wolfram Boucsein. *Electrodermal activity*. Springer Science & Business Media, 2012.

[58] Yu Mike Chi, Tzyy-Ping Jung, and Gert Cauwenberghs. Dry-contact and noncontact biopotential electrodes: Methodological review. *IEEE reviews in biomedical engineering*, 3:106–119, 2010.

[59] Dongyeol Seok, Sanghyun Lee, Minjae Kim, Jaeouk Cho, and Chul Kim. Motion artifact removal techniques for wearable eeg and ppg sensor systems. *Frontiers in Electronics*, 2:685513, 2021.

[60] Nicola Soldati, Vince D Calhoun, Lorenzo Bruzzone, and Jorge Jovicich. Ica analysis of fmri with real-time constraints: an evaluation of fast detection performance as function of algorithms, parameters and a priori conditions. *Frontiers in human neuroscience*, 7:19, 2013.

[61] Ţarălungă Dragoș Daniel and Mihaela Neagu. Cancelling harmonic power line interference in biopotentials. *Compendium of New Techniques in Harmonic Analysis; IntechOpen: London, UK*, page 19, 2018.

[62] Michal Teplan et al. Fundamentals of eeg measurement. *Measurement science review*, 2(2):1–11, 2002.

[63] U.S. Department of Health and Human Services. *Types of Sleep Studies*. https://tinyurl.com/y232zaac.

[64] Zutao Zhang and Jiashu Zhang. A new real-time eye tracking based on nonlinear unscented kalman filter for monitoring driver fatigue. *Journal of Control Theory and Applications*, 8(2):181–188, 2010.

[65] Esra Vural, Mujdat Cetin, Aytul Ercil, et al. Drowsy driver detection through facial movement analysis. In *International Workshop on Human-Computer Interaction*, pages 6–18. Springer, 2007.

[66] Carskadon Roehrs et al. Daytime sleepiness and alertness. In *Principles and Practice of Sleep Medicine: Fifth Edition*. Elsevier Inc., 2010.

[67] Christian Guilleminault, Michel Billiard, Jacques Montplaisir, et al. Altered states of consciousness in disorders of daytime sleepiness. *Journal of the neurological sciences*, 26(3):377–393, 1975.

[68] Vivian Genaro Motti and Kelly Caine. Users' privacy concerns about wearables. In *International Conference on Financial Cryptography and Data Security*, pages 231–244. Springer, 2015.

[69] The Guardian. *Google Glass security failings may threaten owner's privacy.* https://tinyurl.com/y4xopov9.

[70] Investopedia. *How and Why Google Glass Failed.* https://tinyurl.com/y3bqxtlm.

[71] Martin Golz, Adolf Schenka, Florian Haselbeck, et al. Inter-individual variability of eeg features during microsleep events. *Current Directions in Biomedical Engineering*, 5(1):13–16, 2019.

[72] Abdul Baseer Buriro, Reza Shoorangiz, et al. Predicting microsleep states using eeg inter-channel relationships. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2018.

[73] Boon-Leng Lee et al. Standalone wearable driver drowsiness detection system in a smartwatch. *IEEE Sensors journal*, 16(13), 2016.

[74] Vigo. Vigo smart headset. https://tinyurl.com/y3jodh6u.

[75] BlinQ. Blinq system. https://tinyurl.com/yxlkcsr5.

[76] Arun Sahayadhas et al. Detecting driver drowsiness based on sensors: a review. *Sensors*, 12(12):16937–16953, 2012.

[77] Walter Karlen. Adaptive wake and sleep detection for wearable systems. Technical report, EPFL, 2009.

[78] Chin-Teng Lin, Che-Jui Chang, et al. A real-time wireless brain–computer interface system for drowsiness detection. *IEEE transactions on biomedical circuits and systems*, 4(4):214–222, 2010.

[79] Anh Nguyen, Raghda Alqurashi, Zohreh Raghebi, Farnoush Banaei-Kashani, Ann C Halbower, and Tam Vu. A lightweight and inexpensive in-ear sensing system for automatic whole-night sleep stage monitoring. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*, pages 230–244. ACM, 2016.

[80] Valentin Goverdovsky, Wilhelm von Rosenberg, Takashi Nakamura, David Looney, David J Sharp, Christos Papavassiliou, Mary J Morrell, and Danilo P Mandic. Hearables: Multimodal physiological in-ear sensing. *Scientific reports*, 7(1):6948, 2017.

[81] Takashi Nakamura, Yousef D Alqurashi, Mary J Morrell, and Danilo P Mandic. Automatic detection of drowsiness using in-ear eeg. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–6. IEEE, 2018.

[82] Martin G Bleichner and Stefan Debener. Concealed, unobtrusive ear-centered eeg acquisition: ceegrids for transparent eeg. *Frontiers in human neuroscience*, 11:163, 2017.

[83] Ying Gu, Evy Cleeren, Jonathan Dan, Kasper Claes, Wim Van Paesschen, Sabine Van Huffel, and Borbála Hunyadi. Comparison between scalp eeg and behind-the-ear eeg for development of a wearable seizure detection system for patients with focal epilepsy. *Sensors*, 18(1):29, 2018.

[84] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Nathaniel Watson. Contactless sleep apnea detection on smartphones. In *Proceedings of the 13th annual international conference on mobile systems, applications, and services*, pages 45–57. ACM, 2015.

[85] Phuc Nguyen, Nam Bui, et al. Tyth-typing on your teeth: Tongue-teeth localization for human-computer interface. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, pages 269–282. ACM, 2018.

[86] Wen Wen, Daisuke Tomoi, Hiroshi Yamakawa, Shunsuke Hamasaki, Kaoru Takakusaki, Qi An, Yusuke Tamura, Atsushi Yamashita, and Hajime Asama. Continuous estimation of stress using physiological signals during a car race. *Psychology*, 8(07):978, 2017.

[87] Unsoo Ha, Yongsu Lee, Hyunki Kim, Taehwan Roh, Joonsung Bae, Changhyeon Kim, and Hoi-Jun Yoo. A wearable EEG-HEG-HRV multimodal system with simultaneous monitoring of tES for mental health management. *IEEE transactions on biomedical circuits and systems*, 9(6):758–766, 2015.

[88] Abdelkareem Bedri, Richard Li, et al. Earbit: using wearable sensors to detect eating episodes in unconstrained environments. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, 1(3):37, 2017.

[89] Shengjie Bi, Tao Wang, et al. Auracle: Detecting eating episodes with an ear-mounted sensor. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):1–27, 2018.

[90] Zheng Lou, Lili Wang, Kai Jiang, Zhongming Wei, and Guozhen Shen. Reviews of wearable healthcare systems: Materials, devices and system integration. *Materials Science and Engineering: R: Reports*, 140:100523, 2020.

[91] Alexander J Casson. Wearable eeg and beyond. *Biomedical engineering letters*, 9(1):53–71, 2019.

[92] Aleksandr Ometov, Viktoriia Shubina, Lucie Klus, Justyna Skibińska, Salwa Saafi, Pavel Pascacio, Laura Flueratoru, Darwin Quezada Gaibor, Nadezhda Chukhno, Olga Chukhno, et al. A survey on wearable technology: History, state-of-the-art and current challenges. *Computer Networks*, 193:108074, 2021.

[93] Shyamal Patel, Hyung Park, Paolo Bonato, Leighton Chan, and Mary Rodgers. A review of wearable sensors and systems with application in rehabilitation. *Journal of neuroengineering and rehabilitation*, 9(1):1–17, 2012.

[94] Bin Hu, Hong Peng, Qinglin Zhao, Bo Hu, Dennis Majoe, Fang Zheng, and Philip Moore. Signal quality assessment model for wearable eeg sensor on prediction of mental stress. *IEEE transactions on nanobioscience*, 14(5):553–561, 2015.

[95] Dharmendra Gurve, Denis Delisle-Rodriguez, Teodiano Bastos-Filho, and Sridhar Krishnan. Trends in compressive sensing for eeg signal processing applications. *Sensors*, 20(13):3703, 2020.

[96] Bluetooth SIG Working Groups. *Bluetooth Core Specification 4.0*. https://tinyurl.com/2e25vsxu.

[97] Fernando Moreno-Cruz, Víctor Toral-López, Antonio Escobar-Molero, Víctor U Ruíz, Almudena Rivadeneyra, and Diego P Morales. trench: ultra-low power wireless communication protocol for iot and energy harvesting. *Sensors*, 20(21):6156, 2020.

[98] Shelagh JM Smith. Eeg in the diagnosis, classification, and management of patients with epilepsy. *Journal of Neurology, Neurosurgery & Psychiatry*, 76(suppl 2):ii2–ii7, 2005.

[99] Amir M Abdulghani, Alexander J Casson, and Esther Rodriguez-Villegas. Compressive sensing scalp eeg signals: implementations and practical performance. *Medical & biological engineering & computing*, 50(11):1137–1145, 2012.

[100] Amit Levy, Bradford Campbell, Branden Ghena, Daniel B Giffin, Pat Pannuto, Prabal Dutta, and Philip Levis. Multiprogramming a 64kb computer safely and efficiently. In *Proceedings of the 26th Symposium on Operating Systems Principles*, pages 234–251, 2017.

[101] Amir Gholami, Kiseok Kwon, Bichen Wu, Zizheng Tai, Xiangyu Yue, Peter Jin, Sicheng Zhao, and Kurt Keutzer. SqueezeNext: Hardware-Aware Neural Network Design. pages 1638–1647, 2018.

[102] Ningning Ma, Xiangyu Zhang, Hai-Tao Zheng, and Jian Sun. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design. pages 116–131, 2018.

[103] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[104] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018.

[105] Edgar Liberis, Łukasz Dudziak, and Nicholas D. Lane. nas: Constrained neural architecture search for microcontrollers. In *Proceedings of the 1st Workshop on Machine Learning and Systems*, EuroMLSys '21, page 70–79, New York, NY, USA, 2021. Association for Computing Machinery.

[106] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[107] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G.

Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. Tensorflow: A system for large-scale machine learning. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, pages 265–283, Savannah, GA, November 2016. USENIX Association.

[108] Colby Banbury, Chuteng Zhou, Igor Fedorov, Ramon Matas, Urmish Thakker, Dibakar Gope, Vijay Janapa Reddi, Matthew Mattina, and Paul Whatmough. Micronets: Neural network architectures for deploying tinyml applications on commodity microcontrollers. *Proceedings of Machine Learning and Systems*, 3, 2021.

[109] uTensor. *uTensor*, 2022. https://github.com/uTensor/uTensor.

[110] Ji Lin, Wei-Ming Chen, Yujun Lin, john cohn, Chuang Gan, and Song Han. Mcunet: Tiny deep learning on iot devices. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 11711–11722. Curran Associates, Inc., 2020.

[111] Microsoft. *Embedded Learning Library*, 2022. https://github.com/microsoft/ELL.

[112] STMicroelectronics. *STM32Cube.AI*, 2022. https://www.st.com/content/st$_c$om/en/ecosystems/artificial−intelligence−ecosystem − stm32.html.

[113] Robert David, Jared Duke, Advait Jain, Vijay Janapa Reddi, Nat Jeffries, Jian Li, Nick Kreeger, Ian Nappier, Meghna Natraj, Tiezhen Wang, Pete Warden, and Rocky Rhodes. Tensorflow lite micro: Embedded machine learning for tinyml systems. In A. Smola, A. Dimakis, and I. Stoica, editors, *Proceedings of Machine Learning and Systems*, volume 3, pages 800–811, 2021.

[114] H. NYQUIST. Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers*, 47:617–644, 1928.

[115] C E Shannon. Communities in the presence of noise, proceeding of the institute of radio engineers. 1949.

[116] Christian G. Graff and Emil Y. Sidky. Compressive sensing in medical imaging. *Applied Optics*, 54:C23, 3 2015.

[117] Richard G. Baraniuk, Emmanuel Candes, Michael Elad, and Yi Ma. Applications of sparse representation and compressive sensing. *Proceedings of the IEEE*, 98:906–909, 2010.

[118] Mahdi Khosravy, Nilanjan Dey, and Carlos A Duque. *Compressive sensing in healthcare*. Academic Press, 2020.

[119] Partha Pratim Ray. A review on tinyml: State-of-the-art and prospects. *Journal of King Saud University-Computer and Information Sciences*, 2021.

[120] Ji Lin, Wei-Ming Chen, Han Cai, Chuang Gan, and Song Han. Mcunetv2: Memory-efficient patch-based inference for tiny deep learning. *arXiv preprint arXiv:2110.15352*, 2021.

[121] Xiaying Wang, Michele Magno, Lukas Cavigelli, and Luca Benini. Fann-on-mcu: An open-source toolkit for energy-efficient neural network inference at the edge of the internet of things. *IEEE Internet of Things Journal*, 7(5):4403–4417, 2020.

[122] Young D. Kwon, Jagmohan Chauhan, and Cecilia Mascolo. Yono: Modeling multiple heterogeneous neural networks on microcontrollers. In *Proceedings of the 21th International Conference on Information Processing in Sensor Networks*, IPSN '22, 2022.

[123] Khalid Abualsaud, Massudi Mahmuddin, Ramy Hussein, and Amr Mohamed. Performance evaluation for compression-accuracy trade-off using compressive sensing for eeg-based epileptic seizure detection in wireless tele-monitoring. In *2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC)*, pages 231–236. IEEE, 2013.

[124] Khalid Abualsaud, Massudi Mahmuddin, Mohammad Saleh, and Amr Mohamed. Ensemble classifier for epileptic seizure detection for imperfect eeg data. *Scientific World Journal*, 2015, 2015.

[125] Luisa F. Polania, Rafael E. Carrillo, Manuel Blanco-Velasco, and Kenneth E. Barner. Compressed sensing based method for ecg compression. pages 761–764, 2011.

[126] A. Singh, L. N. Sharma, and S. Dandapat. Multi-channel ecg data compression using compressed sensing in eigenspace. *Computers in Biology and Medicine*, 73:24–37, 6 2016.

[127] Mir Mohsina and Angshul Majumdar. Gabor based analysis prior formulation for eeg signal reconstruction. *Biomedical Signal Processing and Control*, 8(6):951–955, 2013.

[128] Phuong Thi Dao, Anthony Griffin, and Xue Jun Li. Compressed sensing of eeg with gabor dictionary: Effect of time and frequency resolution. In *2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pages 3108–3111. IEEE, 2018.

[129] Josiah Hester, Travis Peters, Tianlong Yun, Ronald Peterson, Joseph Skinner, Bhargav Golla, Kevin Storer, Steven Hearndon, Kevin Freeman, Sarah Lord, et al. Amulet: An energy-efficient, multi-application wearable platform. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*, pages 216–229, 2016.

[130] Chin-Teng Lin, Chun-Hsiang Chuang, Chih-Sheng Huang, Shu-Fang Tsai, Shao-Wei Lu, Yen-Hsuan Chen, and Li-Wei Ko. Wireless and wearable eeg system for evaluating driver vigilance. *IEEE Transactions on biomedical circuits and systems*, 8(2):165–176, 2014.

[131] Elise Saoutieff, Tiziana Polichetti, Laurent Jouanet, Adrien Faucon, Audrey Vidal, Alexandre Pereira, Sébastien Boisseau, Thomas Ernst, Maria Lucia Miglietta, Brigida Alfano, et al. A wearable low-power sensing platform for environmental and health monitoring: The convergence project. *Sensors*, 21(5):1802, 2021.

[132] Emmanuel Baccelli, Cenk Gündoğan, Oliver Hahm, Peter Kietzmann, Martine S Lenders, Hauke Petersen, Kaspar Schleiser, Thomas C Schmidt, and Matthias Wählisch. Riot: An open source operating system for low-end embedded devices in the iot. *IEEE Internet of Things Journal*, 5(6):4428–4440, 2018.

[133] Philip Levis, Samuel Madden, Joseph Polastre, Robert Szewczyk, Kamin Whitehouse, Alec Woo, David Gay, Jason Hill, Matt Welsh, Eric Brewer, et al. Tinyos: An operating system for sensor networks. In *Ambient intelligence*, pages 115–148. Springer, 2005.

[134] Amazon. *FreeRTOS - Real-time operating system for microcontrollers.* https://www.freertos.org/index.html.

[135] Hbp and the cardiovascular system. https://tinyurl.com/ybv4vlst.

[136] Dana C. Miskulin and Daniel E. Weiner. Blood pressure management in hemodialysis patients: What we know and what questions remain. *Seminars in Dialysis*, 30(3):203–212, 2017.

[137] Eoin O'Brien. Ambulatory blood pressure monitoring in the management of hypertension. *Heart*, 89(5):571–576, 2003.

[138] S.S. Franklin, L. Thijs, T.W. Hansen, E. O'Brien, and J.A. Staessen. White-coat hypertension: New insights from recent studies. *Hypertension*, 62(6):982–987, December 2013.

[139] G. Ogedegbe, C. Agyemang, and J. E. Ravenell. Masked hypertension: evidence of the need to treat. *Current hypertension reports*, 12(5):349–355, October 2010.

[140] G. V. Ramesh Prasad. Ambulatory blood pressure monitoring in solid organ transplantation. *Clinical Transplantation*, 26:185–191, December 2012.

[141] Department of Health and Human Services. *Organ Donation and Transplantation*, 2018.

[142] P. E. Drawz, M. Abdalla, and M. Rahman. Blood pressure measurement: clinic, home, ambulatory, and beyond. *American journal of kidney diseases : the official journal of the National Kidney*, 60(3):449–462, April 2012.

[143] L. A. Geddes, M. Voelz, C. Combs, D. Reiner, and C. F. Babbs. Characterization of the oscillometric method for measuring indirect blood pressure. *Annals of Biomedical Engineering*, 10(6):271–280, Nov 1982.

[144] Guest commentary: How blood-pressure devices work. https://tinyurl.com/y5rmewuf.

[145] P. D. Baker, D. R. Westenskow, and K. Kück. Theoretical analysis of non-invasive oscillometric maximum amplitude algorithm for estimating mean blood pressure. *Medical and Biological Engineering and Computing*, 35(3):271–278, May 1997.

[146] M. Mafi, S. Rajan, M. Bolic, V. Z. Groza, and H. R. Dajani. Blood pressure estimation using maximum slope of oscillometric pulses. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3239–3242, Aug 2012.

[147] Soojeong Lee, Gwanggil Jeon, and Gangseong Lee. On using maximum a posteriori probability based on a bayesian model for oscillometric blood pressure estimation. *Sensors (Basel, Switzerland)*, 13:13609–23, 10 2013.

[148] Daniel Moran, Yoram Epstein, Gad Keren, Arie Laor, Jack Sherez, and Yair Shapiro. Calculation of mean arterial pressure during exercise as a function of heart rate. *Applied human science : journal of physiological anthropology*, 14:293–295, 12 1995.

[149] John Allen. Photoplethysmography and its application in clinical physiological measurement. *Physiological measurement*, 28(3):R1, 2007.

[150] Denisse Castaneda, Aibhlin Esparza, Mohammad Ghamari, Cinna Soltanpur, and Homer Nazeran. A review on wearable photoplethysmography sensors and their potential future applications in health care. *International journal of biosensors & bioelectronics*, 4(4):195, 2018.

[151] Langesaeter Eldrid, Leiv Arne Rosseland, and Audun Stubhaug. Continuous invasive blood pressure and cardiac output monitoring during cesarean delivery: A randomized, double-blind comparison of low-dose versus high-dose spinal anesthesia with intravenous phenylephrine or placebo infusion. *Anesthesiology: The Journal of the American Society of Anesthesiologists*, 109(5):856–863, 2008.

[152] H. Kanai, M. Iizuka, and K. Sakamoto. One of the problems in the measurement of blood pressure by catheter-insertion: Wave reflection at the tip of the catheter. *The Journal of pediatrics*, 8(5):483–496, 1970.

[153] Jerome Segal. Blood flow measurement catheter, March 1988. US Patent 4,733,669.

[154] Ivan G. Horvath, Adam Nemeth, Zsofia Lenkey, Nicola Alessandri, Fabrizio Tufano, Pal Kis, Balazs Gaszner, and Attila Cziraki. Invasive validation of a new oscillometric device (arteriograph) for measuring augmentation index, central blood pressure and aortic pulse wave velocity. *Journal of hypertension*, 28(10):2068–2075, 2010.

[155] Gareth Beevers, Gregory YH Lip, and Eoin O'Brien. Blood pressure measurement: Part i-sphygmomanometry: factors common to all techniques. *Bmj*, 322(7292):981–985, 2001.

[156] A. K. Reddy, G. E. Taffet, S. Madala, L. H. Michael, M. L. Entman, and C. J. Hartley. Noninvasive blood pressure measurement in mice using pulsed doppler ultrasound. *Ultrasound in medicine biology*, 29(3):379–385, 2003.

[157] S. Wassertheurer, J. Kropf, T. Weber, M. Van der Giet, J. Baulmann, M. Ammer, B. Hametner, CC. Mayer, B. Eber, and D. Magometschnigg. A new oscillometric method for pulse wave analysis: comparison with a common tonometric method. *Medical and Biological Engineering and Computing*, 24(8):498–504, 2010.

[158] G. Fortino and V. Giampà. Ppg-based methods for non invasive and continuous blood pressure measurement: an overview and development issues in body sensor networks. IEEE Medical Measurements and Applications Proceedings (MeMeA), pages 10–13, 2010.

[159] Dinesh Sahu and M Bhaskaran. Palpatory method of measuring diastolic blood pressure. *Journal of anaesthesiology, clinical pharmacology*, 26(4):528, 2010.

[160] G. Ogedegbe and T. Pickering. Principles and techniques of blood pressure measurement. *Cardiology clinics*, 28(4):571–586, 2010.

[161] Mion Junior D. Nobre, F. Ambulatory blood pressure monitoring: Five decades of more light and less shadows. *Arquivos brasileiros de cardiologia*, 106(6):528–537, June 2016.

[162] R. Wang, W. Jia, Z. Mao, R. J. Sclabassi, and M. Sun. Cuff-free blood pressure estimation using pulse transit time and heart rate. In *2014 12th International Conference on Signal Processing (ICSP)*, pages 115–118, Oct 2014.

[163] Microsoft files patent for eyeglasses that measure blood pressure. https://tinyurl.com/yazqxqjo.

[164] M. Lokharan, K. C. Lokesh Kumar, V. Harish Kumar, N. Kayalvizhi, and R. Aryalekshmi. Measurement of pulse transit time (ptt) using photoplethysmography. In James Goh, Chwee Teck Lim, and Hwa Liang Leo, editors, *The 16th International Conference on Biomedical Engineering*, pages 130–134. Springer Singapore, 2017.

[165] Health-tracking startup fails to deliver on its ambitions. https://tinyurl.com/hxtudnt.

[166] Anand Chandrasekhar, Chang-Sei Kim, Mohammed Naji, Keerthana Natarajan, Jin-Oh Hahn, and Ramakrishna Mukkamala. Smartphone-based blood pressure monitoring via the oscillometric finger-pressing method. *Science translational medicine*, 10(431):eaap8674, 2018.

[167] N. Kizilova. Review of emerging methods and techniques for arterial pressure and flow waves acquisition and analyses. *International Journal of Biosen Bioelectron*, 4(4):179–187, April 2018.

[168] Aman Gaurav, Maram Maheedhar, Vijay N Tiwari, and Rangavittal Narayanan. Cuff-less ppg based continuous blood pressure monitoring-a smartphone based approach. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, pages 607–610, 2016.

[169] Edward Jay Wang, Junyi Zhu, Mohit Jain, Tien-Jui Lee, Elliot Saba, Lama Nachman, and Shwetak N. Patel. Seismo: Blood pressure monitoring using built-in smartphone accelerometer and camera. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 425:1–425:9, 2018.

[170] Waldemar Greubel, Albrecht AC Von Muller, Hubertus Von Stein, and Rudolf Wieczorek. Method of continuous measurement of blood pressure in humans, August 1993. US Patent 5,237,997.

[171] Kimihisa Aihara, Shoichi Hayashida, Shinji Mino, Hiroshi Koizumi, Naoe Tatara, Taisuke Oguchi, and Junichi Shimada. Blood pressure meter, June 2011. US Patent 7,963,923.

[172] Albrik Levick Gharibian. Smart blood pressure measuring system (sbpms), August 2016. US Patent App. 14/757,077.

[173] James A Nolan and Trevor J Moody. Method and apparatus for continuous noninvasive monitoring of blood pressure parameters, December 1999. US Patent 6,004,274.

[174] Arie Ariav. Ear probe particularly for measuring various physiological conditions particularly blood pressure, temperature and/or respiration, September 2006. US Patent App. 11/373,280.

[175] Yuji Uenishi, Eiji Higurashi, Kazunori Naganuma, Shouichi Sudo, Junichi Shimada, Kimihisa Aihara, Hiroshi Koizumi, Naoe Tatara, Shoichi Hayashida, Shinji Mino, et al. Living body information detection apparatus and blood-pressure meter, November 2010. US Patent App. 12/847,985.

[176] Andrew D Nordin et al. Dual-electrode motion artifact cancellation for mobile electroencephalography. *Journal of neural engineering*, 2018.

[177] Antoine Favre-Félix et al. Absolute eye gaze estimation with biosensors in hearing aids. *Frontiers in neuroscience*, 13:1294, 2019.

[178] Bernard Grundlehner et al. Ambulatory eeg monitoring. 2019.

[179] Jiawei Xu, Srinjoy Mitra, Chris Van Hoof, et al. Active electrodes for wearable eeg acquisition: Review and electronics design methodology. *IEEE reviews in biomedical engineering*, 10:187–198, 2017.

[180] Evangelia-Regkina Symeonidou et al. Effects of cable sway, electrode surface area, and electrode mass on electroencephalography signal quality during motion. *Sensors*, 18(4):1073, 2018.

[181] AB Simakov et al. Motion artifact from electrodes and cables. 2010.

[182] AC Metting Van Rijn et al. High-quality recording of bioelectric events. *Medical and Biological Engineering and Computing*, 28(5), 1990.

[183] Emily S Kappenman and Steven J Luck. The effects of electrode impedance on data quality and statistical significance in erp recordings. *Psychophysiology*, 47(5):888–904, 2010.

[184] Esrafil Jedari, Rashid Rashidzadeh, Mitra Mirhassani, et al. Two-electrode ecg measurement circuit using a feed forward cmrr enhancement method. In *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*, pages 1–4. IEEE, 2017.

[185] R. Juszkiewicz. *Achieving a Fully Differential Output Using Single-Ended Instrumentation Amplifiers*. https://tinyurl.com/y4klzq6h.

[186] James C Huhta and John G Webster. 60-hz interference in electrocardiography. *IEEE Transactions on Biomedical Engineering*, 1973.

[187] Richard G Lyons. *Understanding digital signal processing, 3/E.* Pearson Education India, 2004.

[188] Carlo J De Luca. Surface electromyography: Detection and recording. *DelSys Incorporated*, 10(2):1–10, 2002.

[189] Muhammad Zahak Jamal. Signal acquisition using surface emg and circuit design considerations for robotic prosthesis. *Computational Intelligence in Electromyography Analysis-A Perspective on Current Applications and Future Challenges*, 18:427–448, 2012.

[190] Joshua D Reiss. Understanding sigma-delta modulation: the solved and unsolved issues. *Journal of the Audio Engineering Society*, 56(1/2):49–64, 2008.

[191] Mathias Benedek and Christian Kaernbach. Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology*, 47(4):647–658, 2010.

[192] Alexander Malafeev, Anneke Hertig-Godeschalk, David R Schreier, Jelena Skorucak, Johannes Mathis, and Peter Achermann. Automatic detection of microsleep episodes with deep learning. *Frontiers in neuroscience*, 15:564098, 2021.

[193] SKB Sangeetha, Sandeep Kumar Mathivanan, V Muthukumaran, N Pughazendi, Prabhu Jayagopal, Md Salah Uddin, et al. A deep learning approach to detect microsleep using various forms of eeg signal. *Mathematical Problems in Engineering*, 2023, 2023.

[194] S. Özšen et al. Examining the effect of time and frequency domain features of eeg, eog, and chin emg signals on sleep staging. In *2010 15th National Biomedical Engineering Meeting*, pages 1–4, April 2010.

[195] Lan lan Chen, Yu Zhao, Jian Zhang, and Jun zhong Zou. Automatic detection of alertness/drowsiness from physiological signals using wavelet-based nonlinear features and machine learning. *Expert Systems with Applications*, 42(21):7344 – 7355, 2015.

[196] B. Sen et al. A Comparative Study on Classification of Sleep Stage Based on EEG Signals Using Feature Selection and Classification Algorithms. *Journal of Medical Systems*, 38(3):1–21, 2014.

[197] Arnaud Sors, Stéphane Bonnet, et al. A convolutional neural network for sleep stage scoring from raw single-channel eeg. *Biomedical Signal Processing and Control*, 42:107–114, 2018.

[198] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[199] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[200] Patrick K Sullivan, Michael J Brucker, and Jagruti Patel. A morphometric study of the external ear: Age and sex related differences, 2010.

[201] Bernie Y Sunwoo, Nicholas Jackson, Greg Maislin, Indira Gurubhagavatula, Charles F George, and Allan I Pack. Reliability of a single objective measure in assessing sleepiness. *Sleep*, 35(1):149–158, 2012.

[202] Conrad Iber et al. *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*. American Academy of Sleep Medicine Westchester, IL, 2007.

[203] Louis Tijerina, Mark Gleckler, Duane Stoltzfus, S Johnston, MJ Goodman, and WW Wierwille. A preliminary assessment of algorithms for drowsy and inattentive driver detection on the road. Technical report, 1998.

[204] A. Han. Thermal management and safety regulation of smart watches. In *2016 15th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*. IEEE, 2016.

[205] Li Deng and Dong Yu. Deep learning: methods and applications. *Foundations and trends in signal processing*, 7(3–4):197–387, 2014.

[206] Nagarajan Ganapathy, Ramakrishnan Swaminathan, and Thomas M Deserno. Deep learning on 1-d biosignals: a taxonomy-based survey. *Yearbook of medical informatics*, 27(01):098–109, 2018.

[207] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1):18–31, 2011.

[208] Thomas M Deserno and Nikolaus Marx. Computational electrocardiography: revisiting holter ecg monitoring. *Methods of Information in Medicine*, 55(04):305–311, 2016.

[209] Nizar Islah, Jamie Koerner, Roman Genov, Taufik A Valiante, and Gerard O'Leary. Machine learning with imbalanced eeg datasets using outlier-based sampling. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 112–115. IEEE, 2020.

[210] Qi Yuan, Weidong Zhou, Liren Zhang, Fan Zhang, Fangzhou Xu, Yan Leng, Dongmei Wei, and Meina Chen. Epileptic seizure detection based on imbalanced classification and wavelet packet transform. *Seizure*, 50:99–108, 2017.

[211] David Belo, João Rodrigues, João R Vaz, Pedro Pezarat-Correia, and Hugo Gamboa. Biosignals learning and synthesis using deep neural networks. *Biomedical engineering online*, 16(1):1–17, 2017.

[212] Amir Gholami, Sehoon Kim, Zhen Dong, Zhewei Yao, Michael W Mahoney, and Kurt Keutzer. A survey of quantization methods for efficient neural network inference. *arXiv preprint arXiv:2103.13630*, 2021.

[213] A Lazarevic, Jaideep Srivastava, and Vipin Kumar. Data mining for analysis of rare events: A case study in security, financial and medical applications. In *Pacific-asia conference on knowledge discovery and data mining*, 2004.

[214] Stijn Luca, Peter Karsmakers, Kris Cuppens, Tom Croonenborghs, Anouk Van de Vel, Berten Ceulemans, Lieven Lagae, Sabine Van Huffel, and Bart Vanrumste. Detecting rare events using extreme value statistics applied to epileptic convulsions in children. *Artificial intelligence in medicine*, 60(2):89–96, 2014.

[215] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[216] A Moura, S Lopez, I Obeid, and J Picone. A comparison of feature extraction methods for eeg signals. In *2015 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pages 1–2. IEEE, 2015.

[217] G. N. Rajesh. Analysis of mfcc features for eeg signal classification. 2019.

[218] Radek Martinek, Martina Ladrova, Michaela Sidikova, Rene Jaros, Khosrow Behbehani, Radana Kahankova, and Aleksandra Kawala-Sterniuk. Advanced bioelectrical signal processing methods: Past, present, and future approach—part iii: Other biosignals. *Sensors*, 21(18):6064, 2021.

[219] Haryong Song, Yunjong Park, Hyungseup Kim, and Hyoungho Ko. Fully integrated biopotential acquisition analog front-end ic. *Sensors*, 15(10):25139–25156, 2015.

[220] Pete Warden and Daniel Situnayake. *Tinyml: Machine learning with tensorflow lite on arduino and ultra-low-power microcontrollers*. O'Reilly Media, 2019.

[221] Liangzhen Lai, Naveen Suda, and Vikas Chandra. Cmsis-nn: Efficient neural network kernels for arm cortex-m cpus. *arXiv preprint arXiv:1801.06601*, 2018.

[222] He Li, Kaoru Ota, and Mianxiong Dong. Learning iot in edge: Deep learning for the internet of things with edge computing. *IEEE network*, 32(1):96–101, 2018.

[223] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

[224] Marcia Sahaya Louis, Zahra Azad, Leila Delshadtehrani, Suyog Gupta, Pete Warden, Vijay Janapa Reddi, and Ajay Joshi. Towards deep learning using tensorflow lite on risc-v. In *Third Workshop on Computer Architecture Research with RISC-V (CARRV)*, volume 1, page 6, 2019.

[225] Yunhui Guo. A survey on methods and theories of quantized neural networks. *arXiv preprint arXiv:1808.04752*, 2018.

[226] Hongkui Jing and Morikuni Takigawa. Low sampling rate induces high correlation dimension on electroencephalograms from healthy subjects. *Psychiatry and clinical neurosciences*, 54(4):407–412, 2000.

[227] Thales Wulfert Cabral, Mahdi Khosravy, Felipe Meneguitti Dias, Henrique Luis Moreira Monteiro, Marcelo Antônio Alves Lima, Leandro Rodrigues Manso Silva, Rayen Naji, and Carlos Augusto Duque. Compressive sensing in medical signal processing and imaging systems. In *Sensors for health monitoring*, pages 69–92. Elsevier, 2019.

[228] Yaakov Tsaig and David L Donoho. Extensions of compressed sensing. *Signal processing*, 86(3):549–571, 2006.

[229] Robert J II Marks. *Introduction to Shannon sampling and interpolation theory*. Springer Science & Business Media, 2012.

[230] Daibashish Gangopadhyay, Emily G Allstot, Anna MR Dixon, Karthik Natarajan, Subhanshu Gupta, and David J Allstot. Compressed sensing analog front-end for bio-sensor applications. *IEEE Journal of Solid-State Circuits*, 49(2):426–438, 2014.

[231] Pervez M Aziz, Henrik V Sorensen, and J Vn der Spiegel. An overview of sigma-delta converters. *IEEE signal processing magazine*, 13(1):61–84, 1996.

[232] David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

[233] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *IEEE signal processing magazine*, 25(2):21–30, 2008.

[234] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[235] Shaou-Gang Miaou and Shu-Nien Chao. Wavelet-based lossy-to-lossless ecg compression in a unified vector quantization framework. *IEEE Transactions on Biomedical Engineering*, 52(3):539–543, 2005.

[236] Selin Aviyente. Compressed sensing framework for eeg compression. In *2007 IEEE/SP 14th workshop on statistical signal processing*, pages 181–184. IEEE, 2007.

[237] Fred Chen, Anantha P Chandrakasan, and Vladimir Stojanović. A signal-agnostic compressed sensing acquisition system for wireless and implantable

sensors. In *IEEE Custom Integrated Circuits Conference 2010*, pages 1–4. IEEE, 2010.

[238] Muhammad Ali Qureshi and Mohamed Deriche. A new wavelet based efficient image compression algorithm using compressive sensing. *Multimedia Tools and Applications*, 75(12):6737–6754, 2016.

[239] Neural Engineering Data Consortium. *Temple University EEG Dataset*. https://tinyurl.com/38vjv4u3.

[240] Monica Fira, V Maiorescu, and Liviu Goras. The analysis of the specific dictionaries for compressive sensing of eeg signals. In *Proceedings of the Ninth International Conference on Advances in Computer-Human Interactions, Venice, Italy*, pages 24–28, 2016.

[241] L Yang, MD Judd, and CJ Bennoch. Denoising uhf signal for pd detection in transformers based on wavelet technique. In *The 17th Annual Meeting of the IEEE Lasers and Electro-Optics Society, 2004. LEOS 2004.*, pages 166–169. IEEE, 2004.

[242] Angkoon Phinyomark, Chusak Limsakul, and Pornchai Phukpattaranont. Evaluation of mother wavelet based on robust emg feature extraction using wavelet packet transform. In *Proceedings of ANSCSE 13 13th International Annual Symposium on Computational Science and Engineering*, pages 333–339, 2009.

[243] J Rafiee, MA Rafiee, N Prause, and MP Schoen. Wavelet basis functions in biomedical signal processing. *Expert systems with Applications*, 38(5):6190–6201, 2011.

[244] MASK Khan, TS Radwan, and MA Rahman. Wavelet packet transform based protection of three-phase ipm motor. In *2006 IEEE International Symposium on Industrial Electronics*, volume 3, pages 2122–2127. IEEE, 2006.

[245] Wai Keng Ngui, M Salman Leong, Lim Meng Hee, and Ahmed M Abdelrhman. Wavelet analysis: mother wavelet selection methods. In *Applied mechanics and materials*, volume 393, pages 953–958. Trans Tech Publ, 2013.

[246] Marie Farge. Wavelet transforms and their applications to turbulence. *Annual review of fluid mechanics*, 24(1):395–458, 1992.

[247] Jingwei Too, AR Abdullah, Norhashimah Mohd Saad, N Mohd Ali, and H Musa. A detail study of wavelet families for emg pattern recognition. *International Journal of Electrical and Computer Engineering (IJECE)*, 8(6):4221–4229, 2018.

[248] M Sanjeeva Reddy, B Narasimha, E Suresh, and K Subba Rao. Analysis of eog signals using wavelet transform for detecting eye blinks. In *2010 International Conference on Wireless Communications & Signal Processing (WCSP)*, pages 1–4. IEEE, 2010.

[249] Feifei Qi, Wenlong Wang, Xiaofeng Xie, Zhenghui Gu, Zhu Liang Yu, Fei Wang, Yuanqing Li, and Wei Wu. Single-trial eeg classification via orthogonal wavelet decomposition-based feature extraction. *Frontiers in Neuroscience*, 15, 2021.

[250] Noor Kamal Al-Qazzaz, Sawal Hamid Bin Mohd Ali, Siti Anom Ahmad, Mohd Shabiul Islam, and Javier Escudero. Selection of mother wavelet functions for multi-channel eeg signal analysis during a working memory task. *Sensors*, 15(11):29015–29035, 2015.

[251] Zhilin Zhang and Bhaskar D Rao. Extension of sbl algorithms for the recovery of block sparse signals with intra-block correlation. *IEEE Transactions on Signal Processing*, 61(8):2009–2015, 2013.

[252] Michael E Tipping. Sparse bayesian learning and the relevance vector machine. *Journal of machine learning research*, 1(Jun):211–244, 2001.

[253] Marilyn Wolf. *The physics of computing*. Elsevier, 2016.

[254] Nam Sung Kim, Todd Austin, David Baauw, Trevor Mudge, Krisztián Flautner, Jie S Hu, Mary Jane Irwin, Mahmut Kandemir, and Vijaykrishnan Narayanan. Leakage current: Moore's law meets static power. *computer*, 36(12):68–75, 2003.

[255] Etienne Le Sueur and Gernot Heiser. Dynamic voltage and frequency scaling: The laws of diminishing returns. In *Proceedings of the 2010 international conference on Power aware computing and systems*, pages 1–8, 2010.

[256] Suresh Siddha, Venkatesh Pallipadi, and AVD Ven. Getting maximum mileage out of tickless. In *Proceedings of the Linux Symposium*, volume 2, pages 201–207. Citeseer, 2007.

[257] Qiyue Zou, Xiaoxin Zou, Ming Zhang, and Zhiping Lin. A robust speech detection algorithm in a microphone array teleconferencing system. In *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221)*, volume 5, pages 3025–3028. IEEE, 2001.

[258] S Gökhun Tanyer and Hamza Ozer. Voice activity detection in nonstationary noise. *IEEE Transactions on speech and audio processing*, 8(4):478–482, 2000.

[259] Iyad Obeid and Joseph Picone. The temple university hospital eeg data corpus. *Frontiers in neuroscience*, 10:196, 2016.

[260] Jeremy R Flynn, Steve Ward, Julian Abich, and David Poole. Image quality assessment using the ssim and the just noticeable difference paradigm. In *International Conference on Engineering Psychology and Cognitive Ergonomics*, pages 23–30. Springer, 2013.

[261] Richard G Baraniuk, Volkan Cevher, Marco F Duarte, and Chinmay Hegde. Model-based compressive sensing. *IEEE Transactions on information theory*, 56(4):1982–2001, 2010.

[262] Daibashish Gangopadhyay, Emily G Allstot, Anna MR Dixon, and David J Allstot. System considerations for the compressive sampling of eeg and ecog biosignals. In *2011 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pages 129–132. IEEE, 2011.

[263] Zhilin Zhang, Tzyy-Ping Jung, Scott Makeig, and Bhaskar D Rao. Compressed sensing of eeg for wireless telemonitoring with low energy consumption and inexpensive hardware. *IEEE Transactions on Biomedical Engineering*, 60(1):221–224, 2012.

[264] Arnaud Delorme and Scott Makeig. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21, 2004.

[265] Elisa Bruno, Pedro F Viana, Michael R Sperling, and Mark P Richardson. Seizure detection at home: Do devices on the market match the needs of people living with epilepsy and their caregivers? *Epilepsia*, 61:S11–S24, 2020.

[266] Steven C. Schachter. *Diagnosing Epilepsy.* https://tinyurl.com/yt6ncpse.

[267] Ali Emami, Naoto Kunii, Takeshi Matsuo, Takashi Shinozaki, Kensuke Kawai, and Hirokazu Takahashi. Seizure detection by convolutional neural network-based analysis of scalp electroencephalography plot images. *NeuroImage: Clinical*, 22:101684, 2019.

[268] Afshin Shoeibi, Marjane Khodatars, Navid Ghassemi, Mahboobeh Jafari, Parisa Moridian, Roohallah Alizadehsani, Maryam Panahiazar, Fahime Khozeimeh, Assef Zare, Hossein Hosseini-Nejad, et al. Epileptic seizures detection using deep learning techniques: a review. *International Journal of Environmental Research and Public Health*, 18(11):5780, 2021.

[269] Hongyu Miao and Felix Xiaozhu Lin. Enabling large neural networks on tiny microcontrollers with swapping. *arXiv preprint arXiv:2101.08744*, 2021.

[270] Michael R Sperling. Sudden unexplained death in epilepsy. *Epilepsy currents*, 1(1):21–23, 2001.

[271] Josiah Hester, Travis Peters, Tianlong Yun, Ronald Peterson, Joseph Skinner, Bhargav Golla, Kevin Storer, Steven Hearndon, Kevin Freeman, Sarah Lord, Ryan Halter, David Kotz, and Jacob Sorber. Amulet: An energy-efficient, multi-application wearable platform. In *Proceedings of the ACM Conference on Embedded Networked Sensor Systems*, pages 216–229, 2016.

[272] Anh Nguyen, Raghda Alqurashi, Zohreh Raghebi, Farnoush Banaei-kashani, Ann C. Halbower, and Tam Vu. A lightweight and inexpensive in-ear sensing system for automatic whole-night sleep stage monitoring. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*, pages 230–244. ACM, 2016.

[273] Fahim Kawsar, Chulhong Min, Akhil Mathur, Marc Van den Broeck, Utku Günay Acer, and Claudio Forlivesi. esense: Earable platform for human sensing. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, pages 541–541, 2018.

[274] N. Merrill, M. T. Curran, J. Yang, and J. Chuang. Classifying mental gestures with in-ear eeg. In *2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pages 130–135, June 2016.

[275] D. Looney, P. Kidmose, C. Park, M. Ungstrup, M. L. Rank, K. Rosenkranz, and D. P. Mandic. The in-the-ear recording concept: User-centered and wearable brain monitoring. *IEEE Pulse*, 3(6):32–42, 2012.

[276] The complete guide to hearable technology in 2019. https://tinyurl.com/y3hdvr8e.

[277] The future is ear: Why "hearables" are finally tech's next big thing. https://tinyurl.com/y9y6yeln.

[278] Vinaya Manchaiah, Brian Taylor, Ashley L. Dockens, Nicole R. Tran, Kayla Lane, Mariana Castle, and Vibhu Grover. Applications of direct-to-consumer hearing devices for adults with hearing loss: a review. *Clinical Interventions in Aging*, 12:859–871, 2017.

[279] World Health Organization. Deafness and hearing loss. https://www.who.int/en/news-room/fact-sheets/detail/deafness-and-hearing-loss, 2018.

[280] GfK. Global unit sales of headphones and headsets from 2013 to 2017 (in millions). www.statista.com/statistics/327000/worldwide-sales-headphones-headsets, 2019.

[281] Christian Holz and Edward Wang. Glabella: Continuously sensing blood pressure behavior using an unobtrusive wearable device. *Interactive, Mobile, Wearable and Ubiquitous Technologies*, page 23, September 2017.

[282] Why google glass failed and why apple watch could too. https://tinyurl.com/y6mpvaqp.

[283] Kirk Shelley and S Shelley. *Pulse Oximeter Waveform: Photoelectric Plethysmography*, pages 420–423. 01 2001.

[284] Poiesis medical duettetm dual-balloon 2-way urinary catheter. http://www.poiesismedical.com/products/duette/.

[285] Biomon sensor sfh7050, osram opto semiconductors. https://tinyurl.com/y7g6yrbn.

[286] Afe4404: Ultra-small integrated afe for wearable, optical heart-rate monitoring and biosensing. https://tinyurl.com/yy88eu5n.

[287] Msp430f5529: 25 mhz mcu with integrated usb phy, 128kb flash, 8kb ram, 12bit/14 channel adc, 32bit hw multiplier. https://tinyurl.com/qa5lyoe.

[288] Ecoflex 00-30, smooth-on. https://tinyurl.com/y77pfgnun.

[289] Medline disposable vinyl bp cuff with marquette connector, 2-tube, adult, 5/bx. https://tinyurl.com/y46u38z4.

[290] Disposable nibp cuff. https://tinyurl.com/y4gojld8.

[291] John M Mathis, John D Barr, Charles A Jungreis, and Joseph A Horton. Physical characteristics of balloon catheter systems used in temporary cerebral artery occlusion. *American journal of neuroradiology*, 15(10):1831–1836, 1994.

[292] Troy J Cross, Sophie Lalande, Robert E Hyatt, and Bruce D Johnson. Response characteristics of esophageal balloon catheters handmade using latex and nonlatex materials. *Physiological reports*, 3(6), 2015.

[293] Jiankun Liu, Jin-Oh Hahn, and Ramakrishna Mukkamala. Error mechanisms of the oscillometric fixed-ratio blood pressure measurement method. *Annals of biomedical engineering*, 41, 11 2012.

[294] Jiankun Liu, Hao-Min Cheng, Chen-Huan Chen, Shih-Hsien Sung, Mohsen Moslehpour, Jin-Oh Hahn, and Ramakrishna Mukkamala. Patient-specific oscillometric blood pressure measurement. *IEEE Transactions on Biomedical Engineering*, 63(6):1220–1228, 2016.

[295] G. Drzewiecki, R. Hood, and H. Apple. Theory of the oscillometric maximum and the systolic and diastolic detection ratios. *Annals of Biomedical Engineering*, 22(1):88–96, Jan 1994.

[296] R. R. Coifman and M. V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38(2):713–718, March 1992.

[297] S. Tong, Z. Li, Y. Zhu, and N. V. Thakor. Describing the nonstationarity level of neurological signals based on quantifications of time-frequency representation. *IEEE Transactions on Biomedical Engineering*, 54(10):1780–1785, Oct 2007.

[298] N. Selvaraj, Y. Mendelson, K. H. Shelley, D. G. Silverman, and K. H. Chon. Statistical approach for the detection of motion/noise artifacts in photoplethysmogram. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4972–4975, Aug 2011.

[299] Know your target heart rates for exercise, losing weight and health. https://tinyurl.com/yauzz8hm.

[300] Steven W. Smith. *The Scientist and Engineer's Guide to Digital Signal Processing*. 1997.

[301] Konquest kbp-2704a. https://tinyurl.com/y4njpa4s.

[302] Guocai Tao, Yan Chen, Changyun Wen, and Min Bi. Statistical analysis of blood pressure measurement errors by oscillometry during surgical operations. *Blood pressure monitoring*, 16(6), December 2011.

[303] Nhat Pham, Tuan Dinh, Zohreh Raghebi, Taeho Kim, Nam Bui, Phuc Nguyen, Hoang Truong, Farnoush Banaei-Kashani, Ann Halbower, Thang Dinh, et al. Wake: a behind-the-ear wearable system for microsleep detection. In *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*, pages 404–418, 2020.

[304] Mark Weiser. The computer for the 21 st century. *Scientific american*, 265(3):94–105, 1991.