



Ge, Yao (2024) *AI enabled RF sensing of Diversified Human-Centric Monitoring*. PhD thesis.

<https://theses.gla.ac.uk/84202/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>
research-enlighten@glasgow.ac.uk

AI enabled RF sensing of Diversified Human-Centric Monitoring

Yao Ge

SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF
DOCTOR OF PHILOSOPHY

JAMES WATT SCHOOL OF ENGINEER
COLLEGE OF SCIENCE AND ENGINEER



University
of Glasgow

1 2024

Abstract

This thesis delves into the application of various RF signaling techniques in Human-Centric Monitoring (HCM), specifically focusing on WiFi, LoRa, Ultra-wideband (UWB) radars, and Frequency Modulated Continuous Wave (FMCW) radars. Each of these technologies has unique properties suitable for different aspects of HCM. For instance, 77GHz FMCW radar signals demonstrate high sensitivity in detecting subtle human movements, such as heartbeat, contrasting with the capabilities of $2.4\text{GHz}/5\text{GHz}$ WiFi signals.

The research extends to both large-scale and small-scale Human Activity Recognition (HAR), examining how ubiquitous communication signals like WiFi and LoRa can be utilized for large-scale HAR, while radar signals with higher central frequencies are more effective for small-scale motions, including heartbeat and mouth movements.

The thesis also identifies several unresolved challenges in the field. These include the underutilization of spatial spectral information in existing WiFi sensing technologies, the untapped potential of LoRa technology in identity recognition, the sensitivity of millimeter-wave radar in detecting breathing and heartbeat against minor movements, and the lack of comprehensive datasets for mouth motion detection in silent speech recognition. Addressing these challenges, the paper proposes several innovative solutions:

- A comprehensive analysis of methodologies for RF-based HCM applications, discussing challenges and proposing potential solutions for broader healthcare applications using wireless sensing.
- Exploration of communication signals in HCM systems, especially focusing on WiFi and LoRa sensing. It introduces the continuous AoA-ToF maps method to enhance HCM system performance and the LoGait system, which uses LoRa signals for human gait identification, extending the sensing range to 20 meters.
- Development of a FMCW radar-based structure for respiration detection, incorporating an ellipse normalization method to adjust distorted IQ signals, reducing the root mean square error by 30% compared to baseline methods.
- Collection and analysis of a large-scale multimodal dataset for silent speech recognition and speech enhancement, including designing experiments to validate the dataset's utility in a multimodal-based speech recognition system.

University of Glasgow
College of Science & Engineering
Statement of Originality

Name: Yao Ge

Registration Number:

I certify that the thesis presented here for examination for a PhD degree of the University of Glasgow is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it) and that the thesis has not been edited by a third party beyond what is permitted by the University's PGR Code of Practice.

The copyright of this thesis rests with the author.

No quotation from it is permitted without full acknowledgement.

I declare that the thesis does not include work forming part of a thesis presented successfully for another degree.

I declare that this thesis has been produced in accordance with the University of Glasgow's Code of Good Practice in Research.

I acknowledge that if any issues are raised regarding good research practice based on review of the thesis, the examination may be postponed pending the outcome of any investigation of the issues.

Signature: Yao Ge

Date: 6th January 2024

List of Publications

Journal Paper

1. **Ge, Yao**, et al. "*Contactless WiFi Sensing and Monitoring for Future Healthcare- Emerging Trends, Challenges, and Opportunities.*" IEEE Reviews in Biomedical Engineering 16 (2022): 171-191.
2. **Ge, Yao**, et al. "*LoGait: LoRa Sensing System of Human Gait Recognition using Dynamic Time Wrapping.*" IEEE Sensors Journal (2023).
3. **Ge, Yao**, et al. "*A Comprehensive Multimodal Dataset for Contactless Lip Reading and Acoustic Analysis.*" Scientific Data (2023).
4. **Ge, Yao**, et al. "*Respiration Monitoring System with Ellipse Normalized methods using 77GHz FMCW radar.*" IEEE Sensors Journal (to be submitted)).
5. **Ge, Yao**, et al. "*A Clinical Trail towards Recent Remote Sensing for Healthcare.*", IEEE Journal of Biomedical and Health Informatics (to be submitted).
6. Li, S., Zhu, S., **Ge, Y.**, et al. "*Depth-guided Deep Video Inpainting*" IEEE Transaction on Multimedia (2023).
7. Song, Y., Taylor, W., **Ge, Y.**, Usman, M., Imran, M. A., & Abbasi, Q. H. "*Evaluation of deep learning models in contactless human motion detection system for next generation healthcare.*" Scientific Reports 12.1 (2022): 21592.

Conference Paper

1. **Ge, Yao**, et al. "*WiFi sensing of Human Activity Recognition using Continuous AoA-ToF Maps.*" 2023 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2023.
2. **Ge, Yao**, et al. "*Respiration detection of sedentary person using ubiquitous WiFi signals.*" 2022 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting (AP-S/URSI). IEEE, 2022.
3. **Ge, Yao**, et al. "*A doppler-based human activity recognition system using WiFi signals.*" 2021 IEEE Sensors. IEEE, 2021.
4. **Ge, Yao**, et al. "*Real-time human activity recognition system exploiting ubiquitous Wi-Fi signals.*" 2021 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting (APS/URSI). IEEE, 2021.
5. **Ge, Yao**, et al. "*Intelligent instruction-based IoT framework for smart home applications using speech recognition.*" 2020 IEEE International Conference on Smart Internet of Things (SmartIoT). IEEE, 2020.
6. Farooq, M., **Ge, Yao**, et al. "*Privacy-Preserving Speaker Recognition Using Radars for Context Estimation In Future Multi-Modal Hearing Assistive Technologies.*", IEEE International Radar Conference 2023, Sydney, Australia, 6-10 Nov 2023.

7. Li, Shibo, **Ge, Yao**, et al. "*Human activity recognition based on collaboration of vision and wifi signals.*" In 2021 International Conference on UK-China Emerging Technologies (UCET), pp. 204-208. IEEE, 2021.
8. Taha, Ahmad, **Ge, Yao**, et al. "*Indoor activity position and direction detection using software defined radios.*" In EAI International Conference on Body Area Networks, pp. 15-27. Cham: Springer International Publishing, 2021.
9. Farooq, M., Qayyum, A., **Ge, Yao**, et al. (2023) "*LoRa-based Privacy-Aware and Contactless Surveillance in Next-generation Smart Homes.*", IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting, Portland, Oregon, USA, 23–28 July 2023, pp. 1751-1752.
10. Chen, Z., Tang, C., **Ge, Yao**, et al. (2023) "*Integrating RF-Visual Technologies for Improved Speech Recognition in Hearing Aids.*", IEEE MTT-S International Microwave Biomedical Conference (IMBioC 2023), Leuven, Belgium, 11-13 September 2023.
11. Song, Y., Taylor, W., **Ge, Yao**, et al. (2021, August). "*Design and implementation of a contactless ai-enabled human motion detection system for next-generation healthcare.*" In 2021 IEEE International Conference on Smart Internet of Things (SmartIoT) (pp. 112-119). IEEE,2021.

List of Tables

2.1	General WiFi sensing applications	17
2.2	Signal processing techniques applied in literature for wireless sensing	20
2.3	Feature extraction and classification techniques applied in the literature for wireless sensing	23
3.1	Confusion matrix for all five classes with Transformer based activity classification	46
3.2	Recognition accuracy vs Activity type	53
3.3	Recognition accuracy vs User identity	54
4.1	Review of RF sensing works with comparison of our proposed work	59
5.1	Parameter set of FMCW radar	92
5.2	Pearson correlation coefficient (PCC) vs User identity	93
5.3	Mean absolute error (MAE) vs User identity	93
5.4	Root mean square error (RMSE) vs User identity	93
5.5	RMSE vs User identity without Gaussian estimation	94
5.6	RMSE vs radar placement direction	94
6.1	Dataset review of multimodal based speech recognition works	99
6.2	Radar parameter setup of UWB X4M03 and TI AWR2243	102
6.3	Parameter setup of Laser speckle detection system	102
6.4	Volunteer information sheet	104
6.5	Corpus list for single subject experiment. The index of each participant is identical to the user label.	105
6.6	Corpus list for supplementary experiments of changing the distance.	105
6.7	Corpus list for supplementary experiments of two-person scenario. The reading materials are referred from corpus publication [207].	106

List of Figures

1.1	Overview applications of RF sensing	1
2.1	Recent WiFi research used in healthcare	12
2.2	CSI matrices of MIMO-OFDM channels	14
2.3	Experiment setup in LoRa sensing	18
2.4	WiFi Sensing System Architecture	19
2.5	Example of PCA & LPF process for CSI amplitude signals	23
2.6	Geometry of the Fresnel zone	24
2.7	DFS representing human activity of leaning forward and back	25
2.8	BVP methodology proposed in [91]	25
2.9	IR-UWB impulse signals format	29
2.10	The plot of a single mmWave FMCW radar chirp in time and frequency domain.	31
2.11	Phasor diagram of theoretical plot of respiration and heartbeat reflected by 5.8 GHz WiFi signal	32
2.12	Typical vital detection system structure [137]	34
3.1	Experiment hardware connection of system	40
3.2	Filtered respiration signal.	40
3.3	CDFs of real-time frequency estimation	41
3.4	Frequency analysis of the two-person scenario	42
3.5	Layout of experiment setup	44
3.6	Doppler spectrum of indoor activity profile	44
3.7	CATM based HAR system of WiFi sensing	46
3.8	Raw CSI data within volunteer keep breathing evenly for 15 seconds of different subcarriers	47
3.9	The PDP filtered result of CSI.	48
3.10	ATM result of CSI while human jogging (a,b,c) and walking (e,f,g).	51
3.11	Res-LSTM Temporal neural network structure	52
3.12	Experimental setup of data collection, the signal of "Rx1", "Rx2", "Rx3" col- lected we labelled it as "D1", "D2", "D3"	53
3.13	Accuracy comparison of different WiFi HAR systems in 3 directions	55
3.14	Accuracy comparison of CATM and NT-CATM in 3 directions	56

4.1	Overview of proposed LoGait system that consists of three components: (1) LoRa Data Collection; (2) Preprocessing; and (3) Classification using DTW and kNN.	61
4.2	Amplitude and phase plots of raw LoRa signals.	64
4.3	Plot of CM results under static environment, with the red dashed line labelled for noise components and green one for LoRa chirp signals in (a) and (b). . . .	64
4.4	Comparison plots with/without Hampel filter and phase unwrapping, with circled outliers in red boxes. The first graph shows original gait signals and the second shows Hampel filtered and phase-unwrapped gait signals.	66
4.5	Picture of humanoid robot motion.	67
4.6	Gait profiles of humanoid robot motion (first graph) and human volunteer (second graph).	67
4.7	Comparison of LoRa signals under different scenarios of no-person, a person standing still, a person stepping shown in the first figure; and gait profiles of three users in the second figure.	68
4.8	Comparison of DTW distance matrix of (a) gait profiles from the first user and second user; (b) gait profiles from the first user.	70
4.9	Experimental setup of 3 scenarios.	71
4.10	Confusion matrix of gait recognition in 3 scenarios	72
4.11	Devices of LoRa signal modulation & demodulation sources	72
4.12	Experimental setup in Long distance scenario	73
4.13	Recognition accuracy via sample rate in 3 scenarios	74
4.14	Recognition accuracy of different distance algorithms for classification in 3 scenarios: LOS, NLOS, LD	75
4.15	DTW distance between two profiles from single subject via Transmitter Gain .	77
4.16	Recognition accuracy of LD scenario with different amount of walking subjects	78
5.1	Overview of preprocess methods of vital detection.	81
5.2	Comparison of RAM and accumulation MTI map. Red boxes indicate the static object, yellow boxes indicate the human location.	83
5.3	Normalization steps of phase signal in slices	84
5.4	Demodulated phase signal in time dimension	84
5.5	Demodulated phase signal in time dimension	84
5.6	Comparison of original rate and corrected rate from spectrum	87
5.7	Experiment setup in lab room	89
5.8	Experiment setup in lab: (a) and (b) are from front body, (c) is from body side. Red line represents the radar estimation result and blue link represents the ground truth. Gray part covers the range of ± 5 bpm.	90
5.9	Performance of PCC, MAE, RMSE in lab scenario with 10 subjects.	91
6.1	Schematic diagram of the Multimodal signals	98
6.2	Data collection setup with device label in the real scenario	100

6.3	Detailed setup schematic diagram for single person scenario from (a) Top view and (b) Front view (Laser’s location is not fixed due to the camera based signals process only require the laser directly point to skin of subjects. The UWB radar facing to subject directly was called ‘xe2’ in dataset folder, another is called ‘xe1’).	103
6.4	Detailed setup schematic diagram from top view of (a) second scenario which considers the distance of 120cm and 220cm between sensor and speaker, and (b) third scenarios which considers speech data collected under multiple speakers. Camera was set to collect left speaker and laser to right speaker.	103
6.5	The structure of the multimodal speech dataset	109
6.6	Multimodal data illustration including UWB and mmWave radar signal, laser, audio, image, and mouth skeleton points. From left to right columns, the first represents the volunteer is speaking of vowel ‘[ɔ:]’, second is speaking of word ‘bleed’, and the third is speaking of sentence ‘There’s been an accident’. The last row illustrates the camera vision of the volunteer’s mouth with the processed skeleton.	110
6.7	Speech separation scheme of mixed audio signals and laser-speckle signals using a NMF based method	113
6.8	Classification performance of human speech across 15 Words with confusion matrix of (a) mmWave radar (b) UWB radar, (c) video stream, (d) audio signal, (e) laser speckle signals.	114
6.9	Multi-modal sensor fusion scheme for word speech classification.	114
6.10	Classification performance of human speech across 15 Words with confusion matrix of (a) the fusion of video and UWB, (b) the fusion of audio and UWB, (c) the fusion of video, audio and UWB.	115
6.11	Classification performance of human speech across 10 Sentences with confusion matrix of (a) UWB signals, (b) the fusion of video and UWB, (c) the fusion of audio and UWB.	116
6.12	Comparative performance with or without pretrained model enhancement in speech recognition	116

Acknowledgements

My PhD journey, spanning three and a half years, was enriched by the invaluable support and guidance of numerous individuals, for which I am deeply grateful. My sincere thanks go to my academic colleagues, whose contributions were vital to the success of my project.

I extend my profound gratitude to my PhD supervisor, Prof. Qammer H. Abbasi, for his relentless efforts and invaluable guidance in both my academic work and personal life. His mentorship has been a cornerstone of my journey, providing not only expert academic direction but also life lessons that have shaped my personal and professional growth.

I am equally thankful to my supervisory team, Prof. Muhammad Ali Imran and Prof. Jonathan Cooper, for their invaluable advice and the opportunity to be a research assistant.

Special appreciation is due to Prof. Shuyuan Zhu, who was instrumental during the writing of my first conference paper, and to Prof. Wenjin Wang, who generously offered me the chance to work in his lab, a highlight of my research experience.

I am indebted to my colleagues at the Communication, Sensing and Imaging Lab - Mr. Shibo Li, Mr. Jingyan Wang, Mr. Muhammad Farooq, Mrs. Hira Hameed, Dr. Chong Tang, Mr. Zikang Chen, Dr. Haobo Li, Dr. William Taylor and others - for their insightful discussions, advice, and participation in my experiments. Additional thanks go to Mr. Yingen Zhu and Mr. Dongmin Huang for their assistance and suggestions in Dr. Wenjin's lab.

I express my gratitude to the University of Glasgow Mobility Scholarship for funding my exchange program at the Southern University of Science and Technology, a pivotal experience in my academic journey.

Lastly, but certainly not least, my heartfelt thanks to my parents for their unwavering financial support and encouragement throughout my PhD. Their belief in my endeavors has been a constant source of motivation.

Declaration

I, Yao Ge, declare that, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Yao Ge

Abbreviations

- HCM - Human Centric Monitoring
- RF - Radio Frequency
- Wi-Fi - Wireless Fidelity
- HAR - Human Activity Recognition
- LOS - Line Of Sight
- NLOS - Non Line Of Sight
- IoT - Internet of Things
- LoRa - Long-Range
- FMCW - Frequency Modulated Continuous Wave
- JCAS - Joint Communication And Sensing
- UWB - Ultra-Wide Bandwidth
- RFID - Radio Frequency Identification
- CSI - Channel State Information
- RSSI - Received Signal Strength Indicator
- MIMO - Multiple Input Multiple Output
- OFDM - Orthogonal Frequency Division Multiplexing
- NIC - network interface card
- CFR - channel frequency response
- SDR - software-defined radio
- USRP - universal software radio peripheral
- WARP - wireless open-access research platform
- FFT - Fast Fourier transform
- CIR - channel impulse response
- STFT - Short-time Fourier transform
- DWT - Discrete wavelet transform
- PCA - Principal Component Analysis
- ICA - Independent Component Analysis
- AoA - Angle of Arrival
- PDP - Power Delay Profile
- IFFT - Inverse Fast Fourier Transform
- DFS - Doppler Frequency Spectrum
- BVP - Body Velocity Profile
- DTW - Dynamic Time Wrapping

- MCFS - Multi-Cluster Feature Selection
- KNN - K-Nearest Neighbors
- SVM - Support Vector Machines
- EMD - Earth Mover's Distance
- DNN - Deep Neural Network
- CNN - Convolutional Neural Network
- LSTM - Long Short-Term Memory
- ABLSTM - Attention-based Bi-directional Long Short-Term Memory
- VGG - Visual Geometry Group
- FCN - Fully Convolutional Network
- ns - nanosecond
- ps - picosecond
- ETSI - European Telecommunications Standards Institute)
- FCC - Federal Communications Commission
- ToF - Time of Flight
- CW - Continuous Wave
- LFMCW - Linear Frequency Modulated Continuous Wave
- SFCW - Stepped Frequency Continuous Wave
- SCG - Seismocardiography
- ECG - Electrocardiogram
- CATM - Continuous AoA-ToF Maps
- PD - peak detection
- ACF - Auto-Correlation Function
- CDF - Cumulative Distribution Function
- SFO - Sampling frequency offset
- PDD - Packet Detection Delay
- MUSIC - Multiple Signal Classification
- CM - Conjugate Multiplication
- DTW - Dynamic Time Warping
- ISS - Inter-Symbol Spacing
- LD - long distance
- HR - Heart Rate
- RR - Respiratory Rate
- RAMs - Range-Angle Maps
- CFAR - Constant False Alarm Rate
- ROI - Range of Interest
- DBSCAN - Density-Based Spatial Clustering of Applications with Noise
- MTI - Moving Target Indicator
- VMD - Variational Mode Decomposition
- IMF - Intrinsic Mode Functions
- PCC - Pearson Correlation Coefficient

- MAE - Mean Absolute Error
- RMSE - Root Mean Square Error
- SNR - Signal to Noise Ratio
- ASR - Automatic Speech Recognition
- SSR - Silent Speech Recognition
- EEG - Electroencephalogram
- NTP - Network Time Protocol
- NMF - Negative Matrix Factorization
- SDR - Signal-to-Distortion Ratio
- SIR - Signal-to-Interference Ratio

Contents

Abstract	ii
List of Publications	iv
List of Tables	vi
List of Figures	ix
Acknowledgements	x
Declaration	xi
Abbreviations	xii
1 Introduction	1
1.1 Background of Human-Centric Monitoring	2
1.2 Motivation and Objectives	3
1.2.1 Comparison of RF sensing and other approaches	3
1.2.2 Diverse Applications of RF sensing	5
1.2.3 Objectives	6
1.3 Main Contributions	7
1.4 Organization of thesis	8
2 Literature review of RF based HCM System	9
2.1 Related Work of Joint Sensing and Communications	10
2.2 WiFi sensing	11
2.2.1 Technical background of WiFi sensing	12
2.2.2 The Evolution of WiFi Sensing for healthcare	14
2.3 Review of LoRa sensing technologies	16
2.4 Review of signal processing techniques	19
2.4.1 Signal Processing Techniques	19
2.4.2 WiFi Sensing Algorithms	23
2.5 Review of Radar sensing applications in vital detection	27
2.5.1 Introduction of Radar based sensing techniques	27
2.5.2 Applications with radar technology	31

2.5.3	Vital signs detection modelling with Radar	32
2.6	Summary	35
3	WiFi Sensing of Human Activity Recognition	37
3.1	Preliminaries	38
3.2	Respiration detection with sanitized CSI phase	39
3.2.1	Experimental Setup	39
3.2.2	Respiration estimation	40
3.2.3	Evaluation	41
3.3	Methodology and System Design of Doppler WiFi HAR	42
3.3.1	System Setup	43
3.3.2	Processing of WiFi CSI data	43
3.3.3	Evaluation of Doppler based recognition system	45
3.4	Continuous AoA-ToF maps (CATM) Methodology	46
3.4.1	Calibration technique	46
3.4.2	AoA-ToF Maps Construction	48
3.4.3	Temporal classification neural network structure	51
3.5	Evaluation of CATM system	52
3.5.1	Experiment Setup	52
3.5.2	Overall evaluation and Discussion	53
3.5.3	Section Conclusion	55
3.6	Summary	56
4	LoRa based Remote Sensing of Human Gait Identification	58
4.1	Introduction	58
4.1.1	Contributions	60
4.2	Methodology	60
4.2.1	Preliminaries	60
4.2.2	Preprocessing methods of LoGait system	63
4.2.3	Gait analysis	65
4.2.4	DTW based Gait Recognition	68
4.3	Evaluation	70
4.3.1	Experiment Setup	70
4.3.2	Overall performance and Discussion	74
4.4	Summary	77
5	Radar based Respiration detection with Ellipse Normalization	79
5.1	Introduction	79
5.2	Methodology	80
5.2.1	Preprocessing methods	80
5.2.2	Estimation of respiration rate	87
5.3	Evaluation	89

5.3.1	Data collection	89
5.3.2	Overall Performance	92
5.4	Summary	95
6	Data collection of multimodal RF sensing of human speech recognition	96
6.1	Background & Summary	96
6.1.1	Literature survey of radar-enabled speech recognition	98
6.2	Methods	99
6.2.1	Data acquisition scheme	100
6.2.2	Speech voice	101
6.2.3	Mouth skeletal points	101
6.2.4	mmWave FMCW radar	101
6.2.5	Laser-speckle system	101
6.2.6	Participant	104
6.2.7	Setup of data collection	104
6.3	Data Record	107
6.3.1	Data storage structure	107
6.3.2	Raw Data Statement	107
6.3.3	Processed Data Statement	108
6.4	Technical validation	108
6.4.1	Signals analysis	108
6.4.2	Speech separation task assisted by laser signal	112
6.4.3	Multimodal speech recognition	113
6.5	Summary	117
7	Conclusions and Future Expectations	118
7.1	Conclusions	118
7.2	Future Research	121

Chapter 1

Introduction

The concept of Human-Centric Monitoring (HCM) represents a framework within our research that emphasizes the development and deployment of monitoring technologies designed to prioritize human needs, behaviors, and well-being. This approach underscores the importance of not only advancing technological innovation but also ensuring that such advancements are made with a deep sense of empathy, respect for individual privacy. Which means the monitoring method should provide intelligent healthcare monitoring and supports for human users in a noncontact way and also keep the users' privacy without picturing them in vision. RF sensing is a potential technique in this field, with multiple known applications, as Fig. 1.1 illustrates.

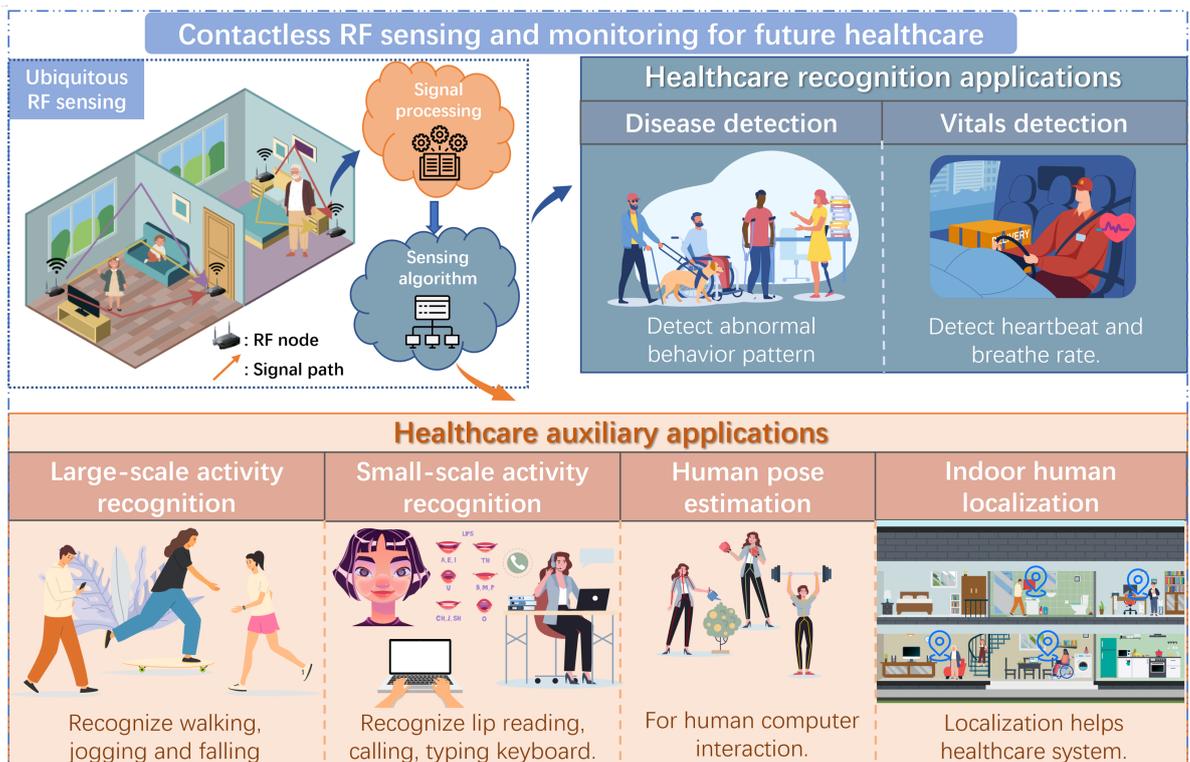


Figure 1.1: Overview applications of RF sensing

The key challenges we address in HCM is exploring the scheme and sensing performance of different RF modulated signal. Our work explores the future applications of HCM in RF based remote sensing, including multiple technical trends such as WiFi sensing, LoRa sensing, and Radar sensing. We are actively engaged in researching and developing innovative solutions that could redefine how monitoring technologies interact with human environments, aiming to make these interactions more intuitive, helpful and, above all, centered around the well-being and preferences of individuals.

1.1 Background of Human-Centric Monitoring

The increasing integration of technology and intelligent systems into daily life, particularly for monitoring and support of healthcare, has become a significant trend. Existing healthcare monitoring instruments predominantly rely on contact based or camera based methods. However, concerns over privacy invasion have heightened consumer interest in noncontact methods for monitoring daily activities. The COVID-19 pandemic has further underscored the necessity for such technologies. During the pandemic, contactless health monitoring systems have proven more effective in minimizing virus transmission via contact points, while also ensuring the safety of healthcare personnel. Wireless Health Monitoring Systems, for instance, can detect normal and abnormal breathing patterns by analyzing lung movements and heartbeats, particularly in potential COVID-19 patients. These systems also identify abnormal activities, such as falls or difficulty walking, which are critical for the urgent care of the elderly.

HCM systems for human healthcare have become increasingly popular, driven in part through our knowledge economy as well as the significant improvements in our longevity and living standards. In healthcare applications, such systems can provide individuals with the capability of long-term detection of daily activities and variations in vital signs, all in the privacy of our homes. With simple, long-term, and continuous health monitoring in the daily home environment, it is possible to record the signs of illness and physiological deterioration that cannot be detected during a short formal clinical consultation. Such monitoring systems can also be combined with deep learning and can be used to monitor behavior, including emotional states and mental well-being. Such information can be integrated into smart homes to support our daily lives.

These systems are heralded for their ability to facilitate long-term, continuous health monitoring in the home environment. They are capable of detecting daily activities and variations in vital signs, capturing subtle signs of illness or physiological changes that might be missed during short clinical visits. Additionally, these systems' capability to

integrate with deep learning technologies allows for the monitoring of behavioral patterns, including emotional states and mental well-being, which can be further incorporated into smart home systems to enhance daily living. The thesis particularly focuses on RF sensing, comparing its advantages to other prevalent monitoring methods.

1. **RF based Sensors:** Radio Frequency (RF) sensors work on the principle of detecting changes in RF waves caused by human presence and movement. By analyzing disruptions in RF signals caused by human movement, it can track activities and even detect vital signs without physical contact with the person being monitored. They offer the benefit of being contactless and can be less intrusive than wearable sensors.
2. **Wearable Sensors:** These sensors are attached to the body and can provide continuous monitoring of physiological parameters like heart rate, body temperature, and oxygen levels. Although highly accurate, the requirement to wear them constantly can sometimes be intrusive or uncomfortable for long-term monitoring.
3. **Camera based Imaging:** This method uses cameras to visually monitor patients. It can be used for movement analysis, fall detection, and even assessing changes in facial expressions or skin color. However, camera based systems often face privacy concerns and can be expensive and complex to install, especially for comprehensive coverage.
4. **Acoustic based Solutions:** These involve using sound sensors or microphones to monitor health can detect sounds related to breathing, coughing, or even changes in voice that might indicate health issues. Like camera based systems, acoustic solutions can also raise privacy concerns and may be limited in their ability to provide detailed physiological data.

Each of these methods offers unique benefits and challenges, with considerations regarding privacy, accuracy, cost, and ease of integration into daily life. The detailed comparison is discussed in the following sections.

1.2 Motivation and Objectives

1.2.1 Comparison of RF sensing and other approaches

Broadly, current sensing and monitoring systems can be divided into those using contact based sensors, including wearable [1] and contactless systems [2], [3]. Besides wearable devices, the contactless monitoring approaches can be divided into visual based sensing RF signals based sensing. RF signals at frequencies between 30 kHz & 300 GHz, comprise electromagnetic waves called radio waves (as are widely used in radar systems, including household and commercial behaviour recognition [4]).

Wearable devices offer a diverse array of functionalities, including but not limited to monitoring heartbeat and respiration rates, oxygen saturation levels, electromyographic signals, among many others [5], [6]. However, the effectiveness of wearable technology in capturing health information critically depends on the user's compliance with wearing the device consistently. Neglect in this regard can lead to significant drawbacks, especially in critical applications such as fall detection. Moreover, the issue of reusability poses a challenge, especially considering the risk of transmitting contact-borne viruses like COVID-19 if the devices are not properly sanitized. Consequently, technology adoption faces hurdles particularly among vulnerable populations, such as the elderly or individuals with disabilities. Moreover, the alternatives like Wi-Fi based sensing systems, which leverage Network Interface Cards (NICs), present more economical solutions in certain contexts.

In contactless sensing methods, camera based sensing applications have proven their precision [7]. However, several disadvantages make it difficult, in some scenarios, to rely on such systems, including:

- Complexity of the system and high cost due to computational requirements for multiple cameras to cover areas of activity.
- Privacy concerns due to the capturing and storage of images, which unauthorized users can access in a low-security system.

More specifically, there are four main advantages that RF sensing possesses of which are concluded:

- **Noninvasive and Unobtrusive Monitoring:** Unlike wearable or camera based systems, RF based sensing does not require direct contact or line-of-sight (LOS) with the subject. This feature enables continuous, unobtrusive monitoring, particularly beneficial in healthcare, elderly care, and security applications. Meanwhile, compared with wearable sensing technologies, ambient RF sensing has the advantage of reducing the risk of contact transmission infections. Because it is capable of the contactless measurement of vital signatures and macro-health indicators in nonline-of-sight (NLOS) environments.
- **Through-Wall and Occlusion Capabilities:** RF signals can penetrate walls and non-metallic objects, allowing for monitoring in occluded or visually inaccessible areas. This capability is crucial in scenarios like emergency response, where immediate information about people in obscured locations can be life-saving.
- **Operational in Various Environmental Conditions:** RF based systems are less affected by environmental factors such as lighting conditions or weather, unlike camera based systems, which can struggle in poor light or obstructed views.
- **Preservation of Anonymity:** RF based sensing can track and analyze movements and vital signs without capturing facial or other identifiable features, inherently preserving more anonymity than visual monitoring systems.

1.2.2 Diverse Applications of RF sensing

RF based sensing has journeyed far beyond its initial role in noninvasive monitoring. Today, it stands at the forefront of technological innovation, particularly in the domain of joint communication and sensing. This evolution not only reflects the versatility of RF sensing but also underscores the need for continued research to harness its full potential.

The early allure of RF sensing was rooted in its ability to provide noninvasive, through-obstruction monitoring, and its adaptability to different environmental conditions. These features made it an ideal choice for applications ranging from healthcare monitoring to security surveillance. However, recent advancements have expanded its scope dramatically:

- In healthcare, RF sensing now plays a crucial role in continuous patient monitoring, offering a noncontact method to track vital signs, detect falls in elderly care, and even assist in early diagnosis of respiratory conditions. This progression from basic monitoring to complex health analytics exemplifies the technology's growing impact.
- In the realm of smart homes and the Internet of Things (IoT), RF sensing has evolved to become a central component for creating responsive environments. These systems can now discern human presence, activities, and even predict user needs, making everyday interactions with technology more intuitive and seamless.
- The rise of Joint Communication and Sensing (JCAS) brings a major leap in RF sensing research, which is its integration with communication technologies. This amalgamation, known as joint communication and sensing, is reshaping the landscape for both fields:
 1. Dual Functionality in One System: The integration of communication and sensing into a single RF system presents an elegant solution that is both cost-effective and efficient. This convergence means fewer hardware components are required, reducing the overall complexity and cost of deployment.
 2. Application in Autonomous Navigation: In autonomous vehicles and drones, the combination of communication and sensing is pivotal. It enables these systems to navigate complex environments while maintaining constant communication with control systems and other vehicles. This integration is crucial for ensuring the safety and reliability of autonomous systems, particularly in densely populated urban areas.
 3. Urban and Infrastructure Applications: The concept extends to urban planning and infrastructure maintenance. For instance, integrated RF systems can provide comprehensive traffic monitoring, which is essential for managing congestion and planning city infrastructure. Additionally, these systems can play a critical role in structural health monitoring, offering real-time data on the condition of bridges, buildings, and other critical structures, which is vital for ensuring public safety.

1.2.3 Objectives

The overall objective of this thesis is to explore multiple RF sensing techniques scheme for the applications their modulation matched, for future HCM system. More specifically, detailed objectives are listed in below.

1. Understanding RF sensing evolution by comprehensive review.
 - **Issue:** The rapid evolution of RF sensing from basic monitoring to complex integrated systems necessitates a comprehensive understanding of its adaptability and potential applications.
 - **Aim:** To analyse the progression of RF sensing technology, emphasizing its versatility and the critical need for continuous research and development to unlock further potential.
2. Implement multiple frequency band signals for HCM with different frequency band.
 - **Issue:** The increasing demand for accurate and noninvasive healthcare monitoring solutions calls for the exploration of advanced signal processing techniques.
 - **Aim:** To investigate the application of low-frequency and high-frequency signals, including WiFi and LoRa for communication, and UWB and FMCW for radar based sensing, in enhancing healthcare monitoring systems.
3. Improving HAR of WiFi sensing using spatial feature.
 - **Issue:** Existing human activity recognition systems often lack the granularity and accuracy needed for precise monitoring in indoor environments.
 - **Aim:** To develop and validate an advanced HAR system utilizing WiFi sensing, focusing on novel feature extraction methods and lightweight neural networks, to achieve fine-grained recognition of human activities.
4. Developing gait recognition with LoGait system
 - **Issue:** Traditional gait recognition technologies with JCAS face limitations in short range, particularly in challenging indoor and through-wall scenarios.
 - **Aim:** To pioneer a new approach in gait recognition using LoRa sensing technology, capable of accurately recognizing human gait over long distances and in various indoor conditions.
5. Improving Radar respiration detection by calibrating the tiny motion of human body.
 - **Issue:** Noninvasive and accurate monitoring of respiratory rates remains a challenge, especially in settings where direct contact or line-of-sight is obstructed.
 - **Aim:** To introduce a novel radar based respiration detection system employing ellipse normalization and Gaussian based rate estimation, enhancing detection reliability in nonline-of-sight and challenging indoor environments.
6. Creating RVTALL, a multimodal dataset for speech recognition

- **Issue:** The lack of comprehensive datasets integrating diverse data types hampers progress in speech recognition research, particularly in multimodal and lip-reading studies.
- **Aim:** To compile and validate the RVTALL dataset, encompassing RF, visual, audio, and laser data, to facilitate groundbreaking research in speech recognition, focusing on multimodality and lip-reading accuracy.

1.3 Main Contributions

The motivation for research in RF sensing is driven by its evolution from a straightforward monitoring tool to a complex system capable of integrated communication and sensing. This progression not only highlights the technology's adaptability and potential, but also underscores the importance of continued research and development in this field.

The thesis is structured to discuss the capability of multi-modalities signals in health-care applications, including communication based WiFi and LoRa signals, radar based UWB and FMCW (Frequency Modulated Continuous Wave) radar signals. It is worth to mention that all data collection with human subjects has been approved with the ethical approval to conduct these experiments was obtained by the University of Glasgow's Research Ethics Committee (approval no.: 300200232, 300190109).

- The third chapter focuses on the development and evaluation of HAR systems using WiFi sensing. It introduces novel feature extraction methods using Doppler features and Continuous Angle of Arrival-Time of Flight (AoA-ToF) maps, employing commercial WiFi devices for fine-grained human activity recognition. Key contributions include leveraging AoA-ToF maps in HAR, designing lightweight temporal neural networks for the system, and establishing a dataset for validation. The work also compares the developed system with existing methods, demonstrating its effectiveness in recognizing and classifying human activities in indoor environments.
- The fourth chapter's main contribution lies in introducing the LoGait system, a new approach for human gait recognition using LoRa sensing technology. This novel system is notable for its use of LoRa signals to recognize human gait over a significant distance of 20 meters, a first in this field. It involves a specialized preprocessing pipeline for processing LoRa signals to extract specific gait features. The system's effectiveness is validated through the collection and analysis of LoRa signals in various indoor scenarios, including challenging environments like through-wall and long-distance settings. This advancement in gait recognition technology showcases LoGait's potential in diverse applications.

- The fifth chapter focuses on the development of a radar based respiration detection system using ellipse normalization. It introduces innovative methods for processing FMCW radar signals to accurately detect and monitor respiratory rates. The system employs a set of signal processing techniques, including ellipse normalization and Gaussian based rate estimation scheme, to enhance the accuracy and reliability of respiratory detection, especially in challenging indoor environments. This research contributes significantly to the field of noncontact health monitoring, offering a new approach to remote respiration detection in different scenarios using radar technology.
- The sixth chapter primarily discusses the development of RVTALL, a multimodal dataset for speech recognition. This dataset integrates Radio Frequency, visual text, audio, laser, and lip landmark information, making it a comprehensive resource for speech recognition research. It includes data types such as 7.5GHz Channel Impulse Response from UWB radars, 77GHz FMCW data from mmWave radar, and visual and audio information, along with lip landmarks and laser data. The dataset, collected from 20 participants, contains about 400 minutes of annotated speech profiles, including vowels, words, and sentences. It is validated for use in studies focusing on lip reading and multimodal speech recognition, offering new perspectives in the field.

1.4 Organization of thesis

The paper is organized as follows: Chapter 2 introduces the literature review of RF based HCM system and reviews different techniques applied in HCM systems, including signal processing techniques and recognition algorithms. Chapter 3 presents the WiFi sensing system, including basic breathing detection and improved CATM scheme for human HAR. Chapter 4 presents the proposed LoGait system, adopting LoRa signals for human identification applications. Chapter 5 introduces the radar based respiration detection system and presents the improved calibration method with FMCW radar. And Chapter 6 presents the multimodal dataset and the related applications we proposed by the dataset, which may contribute to the research community in silent human speech recognition. While finally, Chapter 7 gives the conclusion of whole chapters and discusses the technical and ethical challenges based on the recent researches in the thesis. Meanwhile, the chapter provides future perspectives associated with RF sensing in HCM.

Chapter 2

Literature review of RF based HCM System

Following the introductory discussion about the increasing popularity of sensing and monitoring systems in healthcare, driven by advancements in our knowledge economy and improved living standards, the chapter further elaborates on the diverse landscape of these technologies. The chapter presents a comprehensive survey of the relevant literature in the field of RF sensing in JCAS and radar scenarios, and other types of monitoring systems, highlighting their respective advantages and limitations. Currently, RF sensing research in healthcare is being mainly developed for use in non-hospital environments, driven by two trends: vital sign detection and activity detection (see Fig. 2.1). Vital sign detection system aims to monitor the movement of the lungs and heart in humans using RF signals to recognize the respiration and heartbeat rate, in real-time. For activity detection task, alarms for critical events such as falling, and other specific actions that can cause severe and fatal consequences of human-beings has been studied in academia, and industry [8]. Moreover, due to the sensitivity of radio frequency signals to human activities, the system can also be used for security applications such as identity recognition [9].

In this chapter, we focus on a detailed review which explores the application of multiple types of RF sensing in such HCM applications, demonstrating its principle and relative advantages over other monitoring systems, including WiFi sensing, LoRa sensing, radar sensing and other sorts of contactless sensing methods.

2.1 Related Work of Joint Sensing and Communications

In recent years, joint communication and sensing (JCAS) using wireless signals have been widely studied for different future smart home systems and other sensing applications [10]–[12]. Sensing with such RF signals is not in itself a new concept in research. The phenomena depend upon the analysis using the radio signal transmission and reception parameters, using the same principles developed to detect the presence of objects in aircraft radar and sonar systems. In wireless transmission systems, the transmission signal’s attenuation is inevitable. due to path loss, shadowing, and multi-path fading [13]. On the other hand, these attenuations on the wireless channel can be used to map the physical environment where the RF signals are propagating, providing the theoretical underpinning principle for contactless sensing. Using a communication wireless signal is the most cost-efficient way to perform RF sensing since it is easily accessible in most indoor spaces.

JCAS is a concept that involves combining sensing and communication functions in a single device or system. This approach can lead to increased efficiency of communication and cost savings, as well as improved performance of sensing. This concept is becoming increasingly important as the IoT and other connected devices continue to grow in popularity. Recently, the main sensing techniques in this area are about WiFi, Radio Frequency Identification (RFID) and UWB.

The growing interest in JCAS based sensing systems is due to their practical deployments in indoor settings, as well as their ability to gain responses from monitored persons. Using ambient wireless signals like LoRa and WiFi rather than a camera provides three advantages. First, it preserves the users’ privacy—without requiring them to record videos of daily life activities. Secondly, it resolves the limitations of video based analysis by allowing sensing through-wall and dark spaces. Last but not the least, the utilisation of ambient signals like LoRa and WiFi reduces the cost. Although it requires computation resources to support the sensing algorithm like edge computing, the system’s transceiver units are provided by current communication facilities.

Typically, WiFi is one of the popular topics in this field due to the high cost-effectiveness compared to other sensors. In [14], [15], two WiFi based human activity recognition system are proposed. Both systems achieved over 95% accuracy in recognizing human activities. WiFi is also used in gait recognition [16], [17], the WiDIGR proposed in [16] achieved 78.28% accuracy in 6 subjects gait recognition. UWB based systems are also commonly used in different situation, including gesture recognition [18], gait recognition [19], [20] and human activity recognition [21]. In [22], RFID is also used in recognizing the fluid taking gesture. However, all mentioned systems mentioned above are limited with the

sensing distance. LoRa based systems have only a few works in sensing topic but has shown its availability in long-range detection [23], [24]. As Tab. 4.1 shows, the maximum sensing range of WiFi, UWB and RFID system is limited in 8m, 3m and 2.5m respectively. In comparison, LoRa system proposed in [23] has extended the sensing range to 35m for human tracking and 25m for respiration monitor. Another work presents the similar result but tested through-wall respiration detection [24]. In next section, we specifically introduce the architecture of different sensing technologies.

2.2 WiFi sensing

This section presents the existing research studies focused on the technical background of WiFi monitoring systems with channel state information (CSI) and received signal strength indicator (RSSI) and descriptions of different tracks of WiFi sensing technology. Meanwhile, human activity recognition based on other RF sensing technology is introduced.

Specification of WiFi sensing in healthcare At present, WiFi sensing research in healthcare is being mostly developed for use in non-hospital environments, driven by two trends: vital sign detection and activity detection (see Fig. 2.1). Vital sign detection system aims to monitor the movement of the lungs and heart in humans using WiFi signals to recognize the respiration and heartbeat rate, in real-time. For activity detection task, alarms for critical events such as falling, and other specific actions that can cause severe and fatal consequences to human-beings has been studied in academia, and industry [8].

Generally, such systems use WiFi devices alongside intelligent classification algorithms to monitor and predict human subjects' movements. In the same context, WiFi signals are also used to report, over the internet, the activity status and/or vitals of the monitored subjects to the medical specialist and families or carers. Beneficiaries of such valuable real-time data and information are the IoT systems [25]. For example, vital signs detection in a smart home can help IoT systems adjust the temperature, humidity, and other environmental factors automatically to improve the quality of the user's experience [26], [27]. At present, RF sensing has been applied in the home. For example, Linksys sells a WiFi router and provides a service called "Linksys Aware," which enables WiFi devices to perceive the signals' vibration around the house. Although there have been numerous research studies conducted in this field, it is difficult to replace wearable and visualized

healthcare applications due to their high reliability and efficiency. However, as academia and industry continue to optimize sensing technology, and as it becomes more reliable and accurate for the healthcare monitoring of human beings, we can expect to see changes from the current situation.

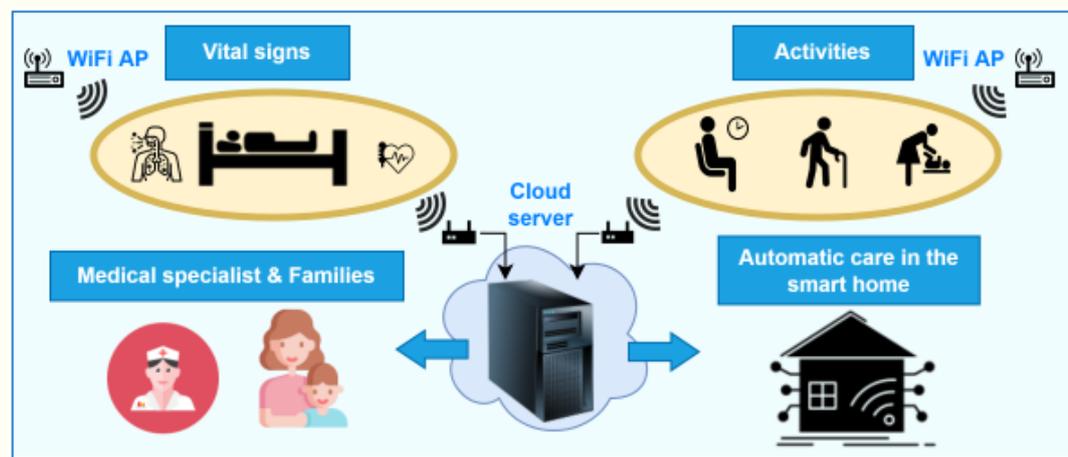


Figure 2.1: Recent WiFi research used in healthcare

2.2.1 Technical background of WiFi sensing

With the rapid advances in communication and network technology, it is possible to assume the broad deployment of WiFi devices across society. Multiple-input multiple-output (MIMO) systems using orthogonal frequency division multiplexing (OFDM) technology, which supports the IEEE 802.11n protocol, provide high throughput transmission mode to serve the high data rate requirements. In such a system, disturbance of physical objects is capable of bringing different extent variation of wireless information on different subcarriers, which provided conditions for the generalization of wireless sensing based on WiFi signals. This section, therefore, discusses some of the primary techniques used to perform WiFi sensing.

2.2.1.1 Received Signal Strength Indicator (RSSI)

The RSSI technique has been widely used for the localization of individuals. In MIMO systems, the RSSI is represented by the superposition of the strength of all the received signals. Most network devices can perform this task, including network interface cards (NICs), as they are easily accessible. An RSSI based detection system depends on the magnitude changes of RSSI levels caused by the activity. However, due to multi-path

fading and time dynamics, its performance under complex conditions is significantly impacted. Early WiFi sensing systems that have been used for commercial localization are primarily dependent on RSSI without fine-grained information. Hence, they cannot be used to recognize complex human behaviour [28].

2.2.1.2 Channel State Information

CSI is the channel property of the wireless communication link. It represents channel frequency response (CFR) for each subcarrier between transmitter and receiver, which describes the fading factor of the signal on every transmission path, *i.e.* the value of every element in channel gain matrix \mathbf{H} (sometimes called channel matrix or channel fading matrix). In WiFi systems, the CSI signals can be obtained from the physical layer on the commercial IEEE 802.11A/G/N wireless network card based on OFDM. For each subcarrier, the WiFi channel is modeled by $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$, Where \mathbf{y} stands for the received signal, \mathbf{x} is the transmitted signal, \mathbf{n} is the noise component. The receiver computes the CSI matrix with the pre-defined signal \mathbf{x} and the received signal \mathbf{y} . However, in reality, WiFi systems' estimation of CSI is affected by multipath fading. The CSI matrix of a given subcarrier with frequency f and time t can be represented as [29]:

$$\mathbf{H}(f, t) = e^{-j2\pi\Delta ft} (\mathbf{H}_s(f) + \sum_{i=1}^{N_d} a_i(f, t) e^{-j2\pi d_i(t)\lambda}) \quad (2.1)$$

Where $e^{-j2\pi\Delta ft}$ is the random phase shift due to the hardware/software error of the WiFi system; \mathbf{H}_s represents the CSI signals from all the static paths (including the signals in line of sight (LOS) areas and those reflected off the stationary objects). The rest of the expression is the summation of signals from all dynamic paths (including signals reflected from the dynamic objects). N_d is the index of the dynamic path, $a_i(f, t)$ represents the complex attenuation factor and the initial phase of the i^{th} path; $e^{-j2\pi d_i(t)\lambda}$ represents the phase change of i^{th} path; $d_i(t)$ and λ are the length of the i^{th} path and the wavelength of the WiFi signal, respectively. The CSI value can adapt the communication system to the current channel conditions and guarantee high reliability and high rate communication in multi-antenna systems. With MIMO and OFDM technologies, the size of the CSI matrix is constructed in 3 dimensions, with N transmitter antennas, M receiver antennas, and K subcarriers. The CSI packet is transmitted as $N \times M \times K$, with the packet index t (see Fig. 2.2). The propagation performance of wireless signals through both the direct path and the multiple reflection paths will show the physical space environment, including any object and the human body. Compared to RSSI values, the CSI offers a fine-grained representation of activity. Hence recent device-free WiFi sensing studies favour CSI, instead of RSSI [30].

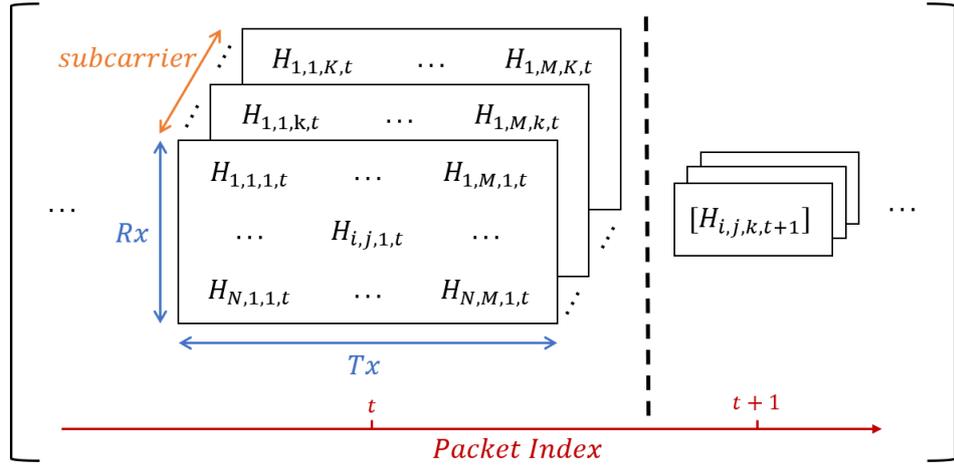


Figure 2.2: CSI matrices of MIMO-OFDM channels

2.2.2 The Evolution of WiFi Sensing for healthcare

2.2.2.1 Hardware platform development of WiFi sensing

For the past few years, research studies on CSI measurement from WiFi signals have been emerging for different sensing applications. In a WiFi system, CSI is essentially a data format used to represent the CFR sampling of the sub-carriers granularity in the system's frequency band, obtained from the physical layer of the commercial IEEE 802.11n wireless network card, based on OFDM technology. Based on the WiFi devices, the researchers first developed an open-source CSI tool driver using the Intel 5300 NICs [31]. This CSI tool enables 30 subcarriers in a 20 MHz channel bandwidth for CSI collection from commercial off-the-shelf (COTS) WiFi devices. This driver provides a quick and low-cost method to establish the WiFi sensing platform. In another study, the authors of [32], [33] have implemented their system based on the Qualcomm Atheros NICs offered by [34], which has 114 CSI subcarriers, hence a higher resolution compared to the Intel 5300 CSI tool. In [33], the results of the comparative study have shown that the higher the number of subcarriers, the higher the sensing accuracy. Other sensing devices include the Wi-ESP which has a reduced cost and is smaller in size compared to the previously mentioned COTS WiFi router [35]. Besides NICs, software-defined radio (SDR) platforms are commonly used to measure CSI, such as the universal software radio peripheral (USRP) and the wireless open-access research platform (WARP) [36]–[38].

2.2.2.2 WiFi Sensing Applications towards Healthcare

Based on the foundation of the open-source WiFi sensing driver's development demonstrated in Sec. 2.2.2.1, researchers have started to propose several methods and applications based on WiFi sensing. This section provides a general overview of WiFi sensing development trends in healthcare. Tab. 2.1 shows some popular applications of WiFi sensing in recent years. For the convenience of demonstration, different tasks are separated into two parts, human activity recognition, and vital signs monitoring. In this case, we define the classification and analysis of all active motion based on torso movement as human activity recognition. From another perspective, vital signs are necessary to maintain regular human activity and are therefore not directly controlled by consciousness and torso movement for the vast majority of time. So, we differentiate it from general human activity recognition.

For the human activity recognition applications in healthcare, we divide them into two types: healthcare auxiliary and healthcare recognition, based on the aspects of monitoring requirement of instant and long-term feedback. In case of healthcare auxiliary applications in an indoor environment, the literature covers daily activity recognition [29], [33], [39]–[44], and other specific activity recognition such as falling [45], smoking [46], sedentary behaviour [47], pose estimation [48]–[51], keystroke [52] and mouth motion [53]. As for the daily activity types, most papers consider: walking, running (or jogging), sitting, pushing and dragging, jumping, squatting, opening the door, and other actions that people always take in daily life. Through these instant activity monitoring methods, the alarm of dangerous accidents like falling can be transferred to the nearest community hospital and families to take an instant action to prevent delayed medical attention, especially for elderly people [54]. At the same time, these approaches are helpful for a disabled person to improve self-care capability through contactless interactive smart controlling methods of gestures recognition and pose estimation. For another range of the healthcare, recognition approaches, it mainly covers the detection of the diseases through long-term gait monitoring, for paraparesis detection [55] and Parkinson detection [56]. These works train the specific model to learn the gait difference of healthy people and patients for diseases recognition. Nevertheless, because datasets from disabled person are difficult to obtain, relatively few works have been published in this field. In hospitals, wireless systems can capture the signal signature of vital signs, such as coughing, shortness of breath, fever, and aches [57]–[59]. Given that these symptoms are closely linked to patient infections, WiFi sensing has the potential to detect illness. The comparison of WiFi sensing and other RF sensing technique. Vital signs estimation belongs to the range of healthcare recognition applications, which is performed by monitoring the motion of the chest and heart. Most papers analyse the respiration rate [50], [60]–[70], some of them detect heartbeats [50],

[62], [63], [65], [67], [70], and another paper demonstrates the biometric estimation [71]. From the perspective of potential in healthcare applications, these systems are useful for instant monitoring of vitals in the non-hospital environments and helpful in the detection of long-term chronic diseases' such as arrhythmia, and some respiratory diseases.

Convincing results with regard to WiFi sensing for biometrics estimation are still lacking in the literature compared to radar based systems [72], [73], which have shown good performance. Therefore, the use of WiFi signals to estimate biometric parameters can be considered a potential application waiting for further development.

In addition to previously reported studies, it is crucial to have reliable localization and tracking systems to complement healthcare monitoring systems. For instance, monitoring vitals during sleeping cannot be performed when the person is not in bed [74], as the position of human is essential to the decision-making process.

Meanwhile, from the timeline shown in the Tab. 2.1, we can conclude the emerging trend is that the researchers are expecting specific WiFi sensing applications like disease's detection, biometric estimation, and recognition of sedentary activity, which have more application value in healthcare. On the other hand, as the types of perceptible activities in the traditional WiFi sensing method are limited, the pose estimation task is proposed to restore the human skeleton in visualization with only WiFi signals. Combined with the state-of-the-art framework for computer vision based recognition of human activity, performance has been expected to improve further [48], [49].

2.3 Review of LoRa sensing technologies

LoRa technology, designed for IoT communication, presents exciting opportunities for long-range, through-wall sensing. It achieves this by employing a low-power wide-area network (LPWAN) approach, utilizing spread spectrum modulation to ensure signal reception even with a very low Signal-to-Noise Ratio (SNR). This feature allows LoRa to have a significantly extended transmission range—up to ten kilometers in rural areas and a few kilometers in urban areas. Meanwhile, LoRa's architecture, utilizing a chirp spread spectrum, gives it a unique edge in maintaining communication over vast distances without draining power, a critical aspect for devices that are on the field for extended periods.

Table 2.1: General WiFi sensing applications

Year	Human Activity Recognition	Vitals Signs Monitoring
2015	gestures recognition [75], [76], human motion [77], localization [78], tracking [79], daily activity [29]	respiration [60]
2016	localization [80], tracking [81], location and velocity [82], smoking [83], gait identification [84], mouth motion [53], gestures recognition[85]	sleep vitals monitoring [61], respiration and heart rate [62]
2017	falling [45], daily activity [86], smoking [46], location and velocity [87],keystroke [52]	respiration and heart rate [63]
2018	falling [71], dynamic velocity [88], daily activity [33], [40], localization and tracking [89], gesture recognition [71], [90], sedentary behavior [47]	respiration [64], respiration and heart rate [65], biometrics estimation [71]
2019	paraparesis detection [55], localization [32], daily activity [41], Parkinson detection [56], 2D pose estimation [48],gesture recognition [91], [92], pedestrian flow estimation [93]	respiration [66], sleep vitals monitoring [67]
2020	pose estimation [50], driver activity & falling [94], 3D pose estimation [49], gesture [95], [96], daily activity [42]	respiration [68], [69], respiration and heart rate[50], [70]
2021	gesture [97], [98], human activity [99], [100] , localization [101], fall detection [102], tracking [103],gait recognition [104], temperature [105]	respiration [106],
2022	fall detection[107], localization[108], gesture [109], human activity [110], gait [111], air-writing[112],	respiration [113],
2023	pain behaviours [114], human activity [115], gesture [116],	sleep vital monitoring [117], [118]

The incorporation of LoRa into human-centric systems brings forth several advantages. Firstly, its long-range communication capability allows for the deployment of monitoring systems over expansive areas, a necessity in large-scale urban settings or in remote health monitoring scenarios. Furthermore, LoRa’s low power requirement ensures longevity in operations, crucial for devices that need to function continuously without frequent maintenance or battery replacements.

LoRa finds its application in a variety of human-centric monitoring systems since it was proposed. For instance, in health monitoring, LoRa enabled wearable devices can continuously transmit patient data to medical professionals, enabling real-time tracking of vital signs in critical care scenarios. In smart cities, LoRa can be utilized for monitoring environmental parameters like air quality or noise levels, directly contributing to improved urban living conditions. An illustrative case study could be the deployment of LoRa in a smart city project, where it facilitated the seamless integration of various IoT devices, contributing to more efficient city management and enhanced citizen experiences. Besides, JCAS based research on perception models using the LoRa protocol, has also received some attention in recent years. The core technique of LoRa sensing is its capability to handle weak target-reflected signals effectively. Traditional wireless sensing technologies such as WiFi and RFID are limited to shorter sensing ranges (3-6 meters for WiFi and even less for acoustic based sensing). LoRa, on the other hand, extends this range considerably. For example, the researches of [23], [24] are in the first step to elucidate the mechanism

by which LoRa signals can be used for perception, and it also explored the feasibility of long-distance breathing detection and activity detection under different environmental backgrounds. It can sense human respiration from as far as 25 meters and detect human walking accurately at a distance of up to 30 meters from the LoRa transceiver pair, shown in Fig. 2.3.



Figure 2.3: Experiment setup in LoRa sensing [23]

Compared to other protocols, LoRa's key differences lie in its extended range, ability to penetrate through walls. Its spread spectrum modulation and use of signal ratio for sensing provide a distinct advantage in long-range sensing scenarios, especially in challenging environments with obstacles or interference.

Despite its advantages, implementing LoRa in human-centric systems presents challenges. In terms of interference, LoRa sensing can be affected by environmental factors. For instance, using an omnidirectional antenna can lead to significant interference from surroundings, while directional antennas with a narrower beamwidth can mitigate this issue to some extent. Besides, scalability is a concern, especially as the number of IoT devices exponentially increases. Security is another critical issue, given the sensitivity of data being transmitted. Moreover, potential interference with other wireless technologies in dense urban environments can affect reliability. Looking ahead, advancements in LoRa technology, such as enhanced encryption for security or integration with emerging IoT standards, can further bolster its application in HCM systems.

2.4 Review of signal processing techniques

Majority of JCAS work of sensing side are mostly considering WiFi sensing, in this case we focus on discussing the techniques of WiFi sensing in this section. The development of WiFi sensing systems involves two stages, the first is applying signal processing techniques, and the second is the algorithm design. The signal processing stage consists of three substages, i.e., denoising, signal transformation, and feature extraction. The algorithm stage explains modelling based and learning based tracks, respectively. A generalized architecture diagram of a typical WiFi sensing system is shown in Fig. 2.4. Firstly, raw WiFi signals are collected by the receiver devices, where they are denoised, transformed, and features are extracted for the data-mining of CSI signals. Secondly, algorithms are applied to classify/recognize/estimate the results. Each of the stages is detailed in the following subsections. In this section, we review various kinds of technologies and classify them in different stages.

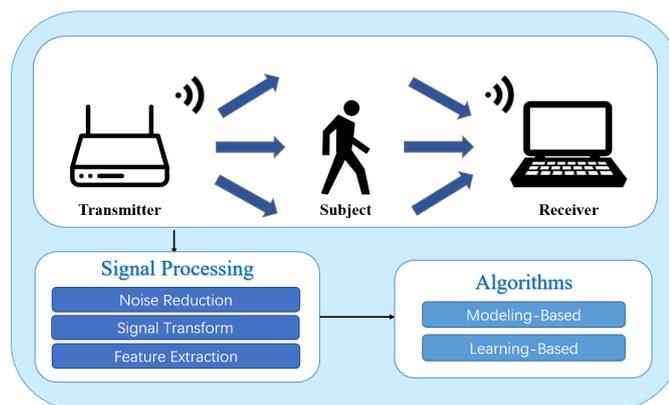


Figure 2.4: WiFi Sensing System Architecture

2.4.1 Signal Processing Techniques

This stage is concerned with the processing of the collected CSI signals captured during the subject's motion. CSI data is processed by different methods to obtain the nature of the information that is required by the system. The signal processing of WiFi signals constitutes three phases: Noise Reduction, Signal Transformation, and Feature Extraction, to feed noise-free information to the algorithms (see Sec. 2.4.1.1, 2.4.1.2 and 2.4.1.3)). Tab. 2.2 lists the various methodologies adopted and applied in the literature to process WiFi signals.

Table 2.2: Signal processing techniques applied in literature for wireless sensing

Reference	Noise Reduction	Signal Transformation	Feature Extraction	Application
WiFall [45]	LOF, MA	N/A	N/A	falling detection
WiHear [53]	N/A	IFFT, DWT	butterworth BPF	mouth motion recognition
Omni-PHD [119]	N/A	N/A	thresholding	human moving detection
Somkey [46]	Hampel filter, interpolation	N/A	thresholding	smoking detection
Widar [82], [87]	N/A	STFT	Butterworth BPF, PCA	localization
PADS [88], [120]	hampel filter	N/A	Combination of Phase Difference and Phase Linear Transform	human moving detection
WiSee [76]	Interpolation	FFT	band pass filter	gesture recognition
Ri-2017 [121]	N/A	N/A	signal separation by ICA	daily objects moving detection
TensorBeat [122]	hampel filter, PBD, SFO, CFO	N/A	thresholding	vitals
CSI-Net [71]	DWT, butterworth LPF	N/A	N/A	human activity recognition
PhaseBeat [70]	DWT, hampel filter	FFT	N/A	vitals

2.4.1.1 Noise Reduction

Noise components, like outliers of CSI data, always exist, which impacts the signal and causes a significant reduction in the recognition accuracy of the overall system. Denoising raw data can reduce the redundant computation of invalid information and improve efficiency and accuracy. De-noising is performed in two stages, the first is the removal of outliers, and the second is performing interpolation.

Outlier is the data that stands out from the rest of the data set, leading to suspicion that no random deviations are resulting from entirely different mechanisms. In a WiFi system, outliers can be caused by hardware or software errors. Moving average (MA) is a primary method to solve the outliers, which uses statistical methods to average the CSI values in a certain period and connect the average values in the time range. A Hampel filter is also used to remove the outliers, where for each sample of the CSI datasets, the median value of the window consisting of the sample and several surrounding samples is calculated, and then the absolute value of the median is used to estimate the standard deviation of the median of each sample pair. Using the median to replace outliers is less sensitive to noise than using mean and standard deviation [36], [62]. The median filter has the same principle as the Hampel filter, which traverses the signal without outlier detection. LOF is used to find abnormal CSI patterns calculating the local density of the points with respect to k-nearest neighbours [123]. The local density of the selected point will be calculated by reach-ability distance to neighbours and compared with other points.

On the other hand, interpolation processing ensures the continuity of the signal in time and reliability of the experimental data, especially when the data packets are collected at a higher frequency. If packets are lost during communication, the interpolation method would take the average of the nearest two points to replace the unperceived data. Meanwhile, to keep the continuity of the signals, linear interpolation is applied in many proposed systems [46], [71], [76].

2.4.1.2 Signal Transformation

The signal transformation method targets the analysis of CSI signals in the time-frequency domain. In the practical environment, the wireless signal will be impacted by high and low-frequency noise. Through frequency domain filtering processing, these noise signals can be effectively reduced. At the same time, the signal components of the frequency band required by the systems can be obtained using a band-pass filter and inverse transformation. The fast Fourier transform (FFT) is a standard method applied in the OFDM systems where the CSI is a sample of FFT of channel impulse response (CIR). Short-time Fourier transform (STFT) frames and windows the original signal first, then performs

FFT on each frame. These characteristic assists researchers in finding the dominant frequency change in the time domain, which is efficient for real-time sensing. However, when the length of the frame is constant, STFT takes a poor balance of signal restoration in the time and frequency domains. Suppose FFT window length (for CSI signals in the time domain) gets extremely short, it will cause inaccurate frequency analysis with inadequate signal information. Inversely, longer window length brings a lower resolution of signal in the time domain. Discrete wavelet transform (DWT) is utilized to decompose signals on different scales to improve the performance compared with the Fourier transform. Meanwhile, DWT is available in time-frequency analysis to judge the signal frequency changes in the time range, the instantaneous frequency, and amplitude at each moment.

2.4.1.3 Feature Extraction

Feature extraction is the process of obtaining information from the signal, which is the basis of a different algorithm for classification and estimation from the CSI data. Phase difference and phase linear transform are used to find the relationship between the changes in the phase and human activities.

Filtering is adequate for detection of the behavior with constant frequency like heartbeat and respiration, even for the detection of walking, which focuses on filtered high frequency CSI signals out to get cleaner human-related signals, such as the respiration and heartbeat rate [70], [122]. Butterworth filter is widely used because the frequency response curve in the pass-band is flat without fluctuations, while it gradually drops to zero in the stop-band.

Threshold is used to distinguish valid signals in the time range based on ToF. As shown in equation 2.1, the ToF value of each path can be estimated by CSI data [81]. Based on the distance of transmission lines, those signals with high ToF value are reflected more times around the environment than others, which is meaningless for systems and can be excluded. Last but not least, signal compression utilizes dimensional decrease methods that generally work in feature extraction, like principal component analysis (PCA) and independent component analysis (ICA) algorithm. PCA is a statistical method, transforming a group of potentially correlated variables into a group of linearly uncorrelated variables through orthogonal transformation. This group of variables after conversion is called the principal component. In WiFi sensing, PCA is mainly adopted to integrate the signals from different subcarriers to extract main components of variance (see Fig. 2.5). ICA is also a method to find the hidden factors of non-Gaussian data, regarded as a powerful method in blind signal analysis. From the previously familiar sample-feature perspective, the prerequisite for using ICA is that implicit factors of independent non-Gaussian distribution generate the sample data.

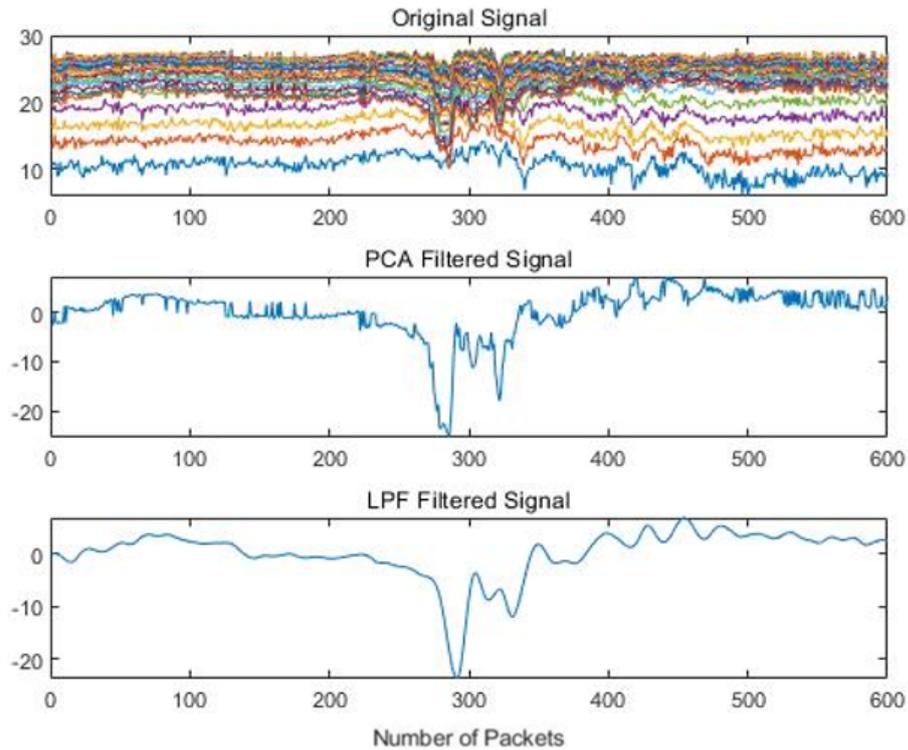


Figure 2.5: Example of PCA & LPF process for CSI amplitude signals

2.4.2 WiFi Sensing Algorithms

Table 2.3: Feature extraction and classification techniques applied in the literature for wireless sensing

Reference	Modelling based	Learning based	Application
WiFall [45]	N/A	KNN, One-Class SVM	falling detection
WiHear [53]	MCFS	DTW	mouth motion recognition
Omni-PHD [119]	EMD	N/A	human moving detection
Smokey [46]	Peak Detection	Autocorrelation	smoking detection
Widar [82], [87]	Doppler Shift, PLCR	N/A	localization
PADS [88], [120]	N/A	One-class SVM	human moving detection
WiSee [76]	Doppler Shift	Pattern Mapping	gesture recognition
CSI-Net [71]	N/A	Deep learning network	human activity recognition
PhaseBeat [70]	Peak Detection, Root-MUSIC	N/A	vitals estimation

The core methodology applied for detection or recognition of activities lies in the algorithm, which is divided into either modelling based or learning based. Some examples from the literature are shown in the Tab. 2.3.

2.4.2.1 Modelling based Algorithm

Modelling based approaches apply statistical or mathematical models to extract specific features, depending on the tasks. These studies are less dependent on training set and have more robustness compared to the learning based methods.

ToF and angle of arrival (AoA) models have been frequently applied for indoor tracking and localization. When receiving signals in the same physical path, the delay should be a constant value. However, due to the multi-path effect, which reflects the transmitted signal, the value of ToF can be influenced. Power delay profile (PDP) is a common approach to get the value of ToF through inverse fast Fourier transform (IFFT), which is popular in tracking and localization [79]. Meanwhile, AoA makes different antennas show different phase observations. Multiple Signal Classification (MUSIC) performs well on the AoA estimation [78], [81], which correlates phase difference with the distance of multiple antennas to estimate the transmission direction.

The concept of Fresnel zones, as illustrated in Fig. 2.6, is fundamental to understanding the propagation of WiFi signals in a wireless communication system. These zones are defined by concentric ellipses with the transmitting and receiving antennas at their foci, lying on a horizontal plane within the WiFi system. While the most efficient path for signal transmission is the direct LOS path, electromagnetic waves inherently spread out and follow a path that forms an ellipsoidal shape, encompassing a much broader area than the narrow LOS. Each Fresnel zone corresponds to a distinct region where the electromagnetic waves, having diverged from the LOS, converge at the receiving antenna with either in-phase or consistently phased differences. This phasing is critical, as it dictates the constructive or destructive interference patterns of the waves. The boundary of each Fresnel zone marks a point where the phase shift is a multiple of half the wavelength, which is significant for signal coherence and strength. [124] designs the related experiments and proves that the motion that happens in the middle of the Fresnel zone is more efficient than happens on the boundary.

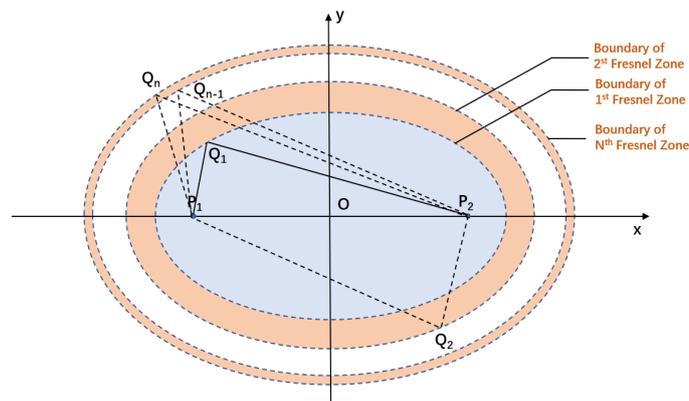


Figure 2.6: Geometry of the Fresnel zone

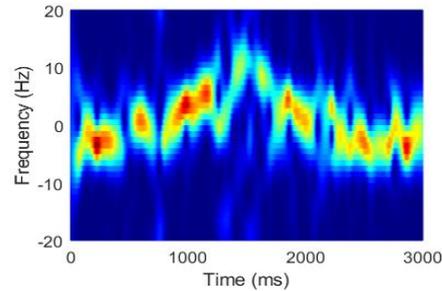


Figure 2.7: DFS representing human activity of leaning forward and back

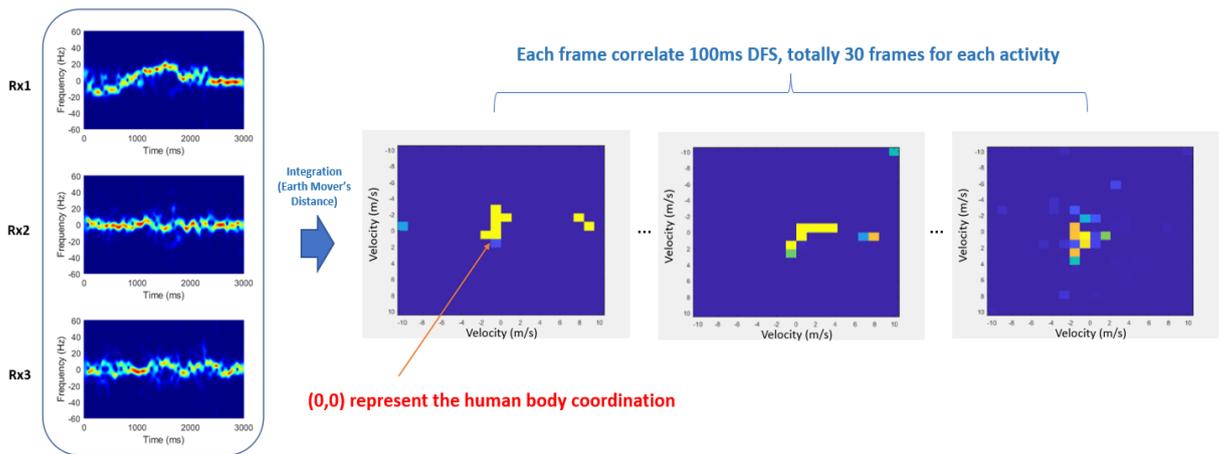


Figure 2.8: BVP methodology proposed in [91]

Doppler frequency spectrum (DFS) represents frequency shift influenced from the active motion, which is feasible to extract the velocity of subjects. CSI itself represents the channel frequency response, so it is convenient to do time-frequency analysis of the WiFi signals (see Fig. 2.7)). The authors of [87] developed an algorithm to correlate static CSI values with active multi-path gradient and utilize the Doppler frequency change to estimate the velocity and location of the humans. Besides, [91] adopts the body velocity profile (BVP) algorithm to apply the earth mover's distance to integrate the multiple spectrum's characteristics to classify the gestures (see Fig. 2.8). In each BVP, the velocity component is projected onto the normal direction of a physical WiFi link and contributes to the power of the corresponding radial velocity component in the DFS profile. Due to the path length change of the Doppler signal, WiFi equipment in a different position collects distinct CSI signals. This Widar3.0 considers the location of devices and maps DFS value into the BVP. This method reduces the negative influence from the environment and has been tested in unknown locations where signals are collected for the training set.

2.4.2.2 Learning based Algorithm

Machine learning based classification algorithms such as the k-Nearest Neighbors (KNN) and support vector machines (SVM) are widely used in detection and recognition tasks [125]. Multi-cluster/class feature selection (MCFS) in the WiHear system [53] sets to extract the optimal feature subset and find the correlation feature between different subsets, using a pattern matching algorithm to avoid over-fitting. On the other hand, with the fixed size of the dataset, the classification process of MCFS on the testing set takes 5 seconds, which is much lower than 3–5 minutes taken by the SVM algorithm. The Dynamic Time Wrapping (DTW) method calculates the similarity between time series data by extending and shortening the sequences widely used in fingerprint based learning methods. Similarly, earth mover’s distance (EMD) defines distance measurement, which can measure the distance between two distributions. The function of EMD and DTW is similar in CSI based classification, which both belongs to the range of linear regression. By contrast, DTW focus on estimation in single dimension CSI sequence, and EMD can be used in higher dimensions, like DFS integration shown in Fig. 2.8.

Besides ML methods, many deep neural network (DNN) frameworks have been widely applied in WiFi sensing. With the rapid development of neural network in recent years in various fields, different DNN structures have been applied in WiFi sensing, for example, convolutional neural network (CNN), long short-term memory (LSTM) and etc. In [41], instead of using long short-term memory (LSTM), an attention based bidirectional long short-term memory (ABLSTM) neural network is proposed to extract 2-dimensional features from WiFi CSI data, representing human activity. In the results, recognition of six different activities in public places, recorded an accuracy of more than 97.5%. In [40] the author applies the CNN classification algorithm, which records a higher accuracy in comparison with the SVM classifier. Similarly, SignFi [90], and 1D-CNN [55] propose the use of different CNN structures for WiFi sensing. The 1D-CNN performs better than the KNN method (average 4% higher in 1D-CNN), and SignFi improved 2% - 4% accuracy compared to the KNN-DTW method proposed in [85], [126]. Instead of using external neural networks (less than five layers) as alternate nonlinear operators, [71] proposed a novel method to extend the size of CSI input from $30 \times 1 \times 1$ to $6 \times 224 \times 224$ with bilinear interpolation, which provided an image-like structure for further deep learning. It offers conditions to apply different backbone networks like AlexNet, visual geometry group (VGG) network, inception cluster, etc. The deep learning network framework proved that body perception of CSI sequences for WiFi could accomplish: two body characterization problems of biometrics estimation (including body fat, muscle, water, and bone rate) and identification, two activity recognition: gesture recognition and fall detection. For both traditional ML and DNN methods, the performance suffers from the distribution shift that arises from different circumstances/locations. To avoid the repeated training of the model and fitting new areas, transfer learning methods can be utilized with lower com-

putation resources [127]. Furthermore, the metric learning approach also helps the model generalize to new environments for applications such as gesture recognition [92]. Although all above methodologies apply distinct network structures, the central task is the same, which adopted DNN to match the CSI signal with the artificial label.

Moreover, pose estimation adopts the label from camera based methods, and proposes a novel DNN structure to match the human skeleton to WiFi CSI data. In the training stage, the skeleton of a human being can be acquired from image processing with cameras. Afterward, the collected WiFi data is labelled and correlated with different patterns of skeleton coordination and trained by a neural network. The authors of [128] proposed a novel network to apply a fully convolutional network (FCN) for estimation of a single person's pose from the collected data and annotations. This work aims to train the specific neural network to map the CSI variance to the human skeletons, and get the fine-grained human skeletons from CSI signals. Furthermore, they developed another structure for multi-persons' pose estimation [48]. Based on a similar theory, [129] proposes an image based preprocess method to get a CSI-image for CNN framework to estimate the pose. However, in the mentioned 2D human skeleton restoration, there are few discussions of the robustness. Due to the sensitivity of CSI signals, environment has the severe impact on channel information, which means the overfitting issue is inevitable. Because 2D pixels obviously can not map all human activities, especially for NLOS side, with CSI variance. To improve the reliability, the study of [49] improves the BVP to 3D velocity profile through changing the antennas' height. These ideas create the condition of more applications' development with pose estimation.

2.5 Review of Radar sensing applications in vital detection

2.5.1 Introduction of Radar based sensing techniques

Radar systems function by estimating the position and/or movement speed of objects, known as targets, using electromagnetic wave reflections. These systems determine the distance of a target by measuring the time delay of the waves that return after hitting the target. The velocity of the target is inferred through changes observed in the frequency of the reflected waves, a phenomenon known as the Doppler effect. Specifically, for vital sign monitoring, these radar systems are based on the concept that the human body's chest movements, caused by heartbeat and breathing, alter the radar's reflected signals. This alteration allows for the indirect measurement of heart and breathing rates, without any physical contact with the individual.

2.5.1.1 IR-UWB radar

Like Wi-Fi and Bluetooth, UWB is a short-range wireless communication protocol. The UWB was defined as the wireless transmission system of which the bandwidth exceeds 500MHz , and each transmit pulse of this communication system can occupy at least 500MHz bandwidth. Instead of modulating with a carrier wave, IR-UWB relies on nano-second (ns) to picosecond (ps) non-sinusoidal narrow impulse radio signals to transmit data. The time based modulation technology increases transmission speed and reduces power consumption. For speech recognition, the UWB system has the following advantages:

1. Strong anti-interference ability: From the RF mechanism, the pulse wave emitted by UWB is more resistant to interference than the continuous electromagnetic waves in short range. Specifically, the permitted work frequency band of UWB is from 3GHz to 10GHz , which suffers less disturbance from the general 2.4GHz WiFi system and other telecommunication signals.
2. The protocol yielded positive results, resulting in a reduction in power consumption for short-range communication applications. The transmit power of the UWB transmitters was found to be typically less than 1 mW, which extended the system's operating time and minimized electromagnetic wave radiation to the human body.

After careful consideration of cost and feasibility, we have selected the XeThru X4M03, an IR-UWB radar system on chip, as our UWB radar. The UWB RF specifications of this radar have been approved by European Telecommunications Standards Institute (ETSI) in Europe, and Federal Communications Commission (FCC) in the USA for commercial use in human living circumstances [130]. This device is a highly reliable sensor that is capable of detecting objects at a range of up to 10 meters. It is also capable of detecting objects in a wide range of angles, up to 180 degrees. This radar system has been adopted in a variety of research projects, ranging from human vital sign detection [131], [132] to activity recognition [133].

$$\mathbf{s}(\tau, t) = \sum_{i=1}^{N_d} a_i(\tau, t) e^{-j2\pi \frac{(d_i(t) + d_a(\tau))}{\lambda}} \quad (2.2)$$

For pulsed radars, the distance between the radar and the target can be determined by $R = \frac{c * \Delta T}{2}$, where c represents microwave speed, ΔT represents the round-trip time of a single pulse, called time of flight (ToF). The signals of IR-UWB can be represented in Eq. 4.1, where the τ represents the ToF of signals impulses in fast-time range, t represents

receiving time of frame in slow-time domain, N_d is the index of the dynamic path, $a_i(\tau, t)$ represents the complex attenuation factor of the i^{th} path; $e^{-j2\pi\frac{d(t)}{\lambda}}$ represents the phase change of i^{th} path; $d_i(t)$ and $d_a(t)$ are the static length and vibrating length of i^{th} path. λ represents the wavelength of the UWB signal.

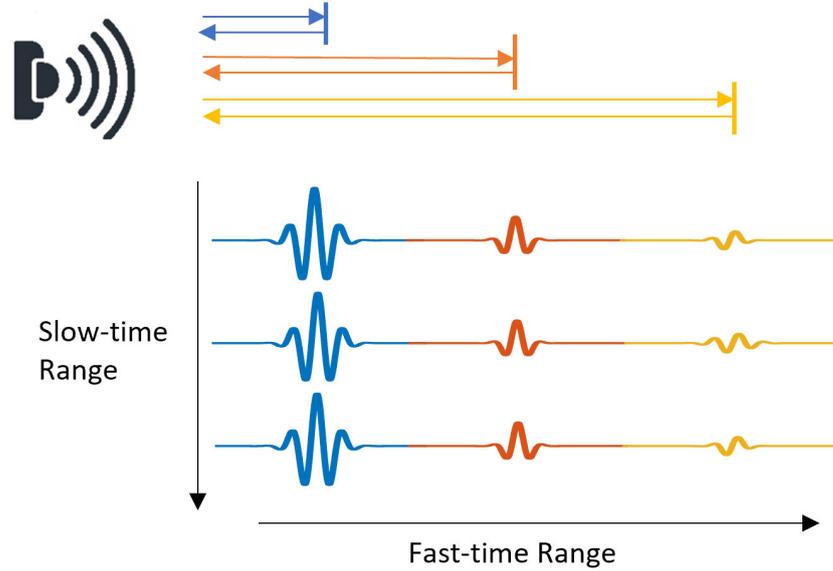


Figure 2.9: IR-UWB impulse signals format

The format of the received signals consists of a fixed set of bins, determined by the timing of the transmitted pulses. These bins are indexed by fast-time and slow-time dimensions, as illustrated in Fig. 3.6. Fast-time and slow-time are two dimensions that are used to describe the format of the received signals in a UWB radar system. Fast-time is the time it takes for the radar to transmit a pulse and receive the reflected signal. Slow-time is the time it takes for the radar to transmit multiple pulses and receive the reflected signals. The fast-time and slow-time dimensions are used to determine the format of the received signals, which is a fixed set of bins. These bins are used to store the information about the objects detected by the radar. The fast-time and slow-time dimensions are also used to determine the range of objects detected by the radar. By using the fast-time and slow-time dimensions, the UWB radar system can accurately detect objects in a wide range of distances. This makes the UWB radar system an ideal choice for a variety of applications, including human vital sign detection and activity recognition.

2.5.1.2 CW radar

Unmodulated continuous wave (CW) radar measures target speed and angular position. CW radar transmits a continuous electromagnetic wave, unlike pulsed radar systems. There is no modulation on the wave in its simplest form. These types of "primary" radars work by transmitting a waveform and then having the radar receiver measure the echoes reflected off the surface of the target. There are two main characters of CW radar:

- **Doppler Shift Detection:** CW radar primarily uses the Doppler effect to detect the velocity of objects. The frequency of the reflected wave changes in relation to the motion of the target relative to the radar, allowing the radar to measure the speed of the target.
- **No Range Information:** Basic CW radar cannot measure the distance to the target because it continuously transmits and receives signals. It lacks the time-of-flight information that pulsed radars use to determine range.

2.5.1.3 FMCW radar

Transmit a continuous sine wave, mainly used to measure the speed of the target (single-frequency continuous wave can only measure speed but not distance). If it's necessary to measure the distance to the target, the transmitted waveform needs to be modulated, such as frequency modulated FM continuous wave. FMCW radars applied in the field of healthcare mainly include Linear Frequency-Modulated Continuous-Wave (LFMCW) radar and Stepped Frequency Continuous-Wave (SFCW) radar. In this section, we take LFMCW as an example to introduce the way that FMCW radar acquires signals at multiple distances and angles, and how to extract physiological signals based on FMCW radar. For LFMCW radar, the transmitting antenna Tx periodically transmits signals whose frequency increases linearly with time.

The radar emits these chirps and receives the echoes reflected from targets (like the human body). By comparing the frequency of the transmitted and received signals, the radar can determine the distance to the target. Compared to CW radar, there are two specific characters of FMCW radar:

- **Distance and Velocity Measurement:** The difference in frequency (beat frequency) between the transmitted and received signals is proportional to the distance to the target. Moreover, by analysing the phase shift between these signals, the radar can also measure the target's velocity.
- **Advantage in Detail Resolution:** Unlike simpler CW radars, FMCW radars can provide more detailed information due to their ability to discern multiple targets at different distances and velocities.

Unlike IR-UWB radars that measure distance using the ToF of instantaneous impulses, FMCW radars rely on the difference in frequency between the transmitted and received signals from a linear variation of signals frequency. In other words, Doppler effect of moving target can be explored from difference between transmission frequency (f_t) and shifted frequency (f_s). The formula shows Doppler velocity of target relative to radar: $v = \frac{c(f_s - f_t)}{2f_t}$.

Although the IR-UWB radar is able to capture the vibration of sub-centimeter motion. The angle resolution is limited by the number of antennas. Meanwhile, we take one commercial off-the-shelf (COTS) 77GHz FMCW radar, AWR2243, for example. This high frequency enables radar signals to capture motion in millimeter, that can be used for both lip motion and vocal folds detection. A FMCW radar chirp signals can be illustrated as Fig. 3.5.

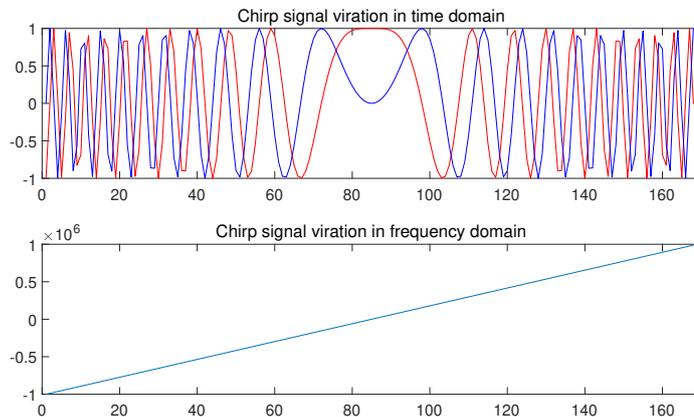


Figure 2.10: The plot of a single mmWave FMCW radar chirp in time and frequency domain.

Except for modulation methods, AWR2243 radar contains 4 receive antennas and 3 transmit antennas, which is possible to adopt AoA for research in horizontal plane. In our experiment, we adopted 4 receive and 1 transmit antennas to increase the sample rate. By adopting 4 channels at a snapshot, we can gain the angular resolution of 30 degrees.

2.5.2 Applications with radar technology

In the realm of smart homes, radar technology monitors the activities of household members, such as walking, sitting, and breathing patterns. This enhances home automation systems like smart lighting, temperature control, and security monitoring. Future applications of radar in smart homes are likely to improve convenience and safety, such as optimizing energy use through more precise behaviour recognition or enhancing the safety of

the elderly and children through abnormal activity monitoring. In the medical field, radar technology is utilized for non-invasive monitoring of patients' vital signs, including heart rate and respiration. This is particularly useful in telemedicine and home care for monitoring the health of the elderly or chronic patients. As radar technology becomes more precise and reliable, it is poised to become an essential tool in routine medical monitoring and emergency response, potentially driving innovation in new medical diagnostics and treatment methods.

These applications not only showcase the multifunctionality and adaptability of radar technology but also hint at its significant role in future technological and societal developments. With ongoing research and development, radar technology's applications in these fields are expected to expand, bringing greater convenience and protection to people's lives and health.

2.5.3 Vital signs detection modelling with Radar

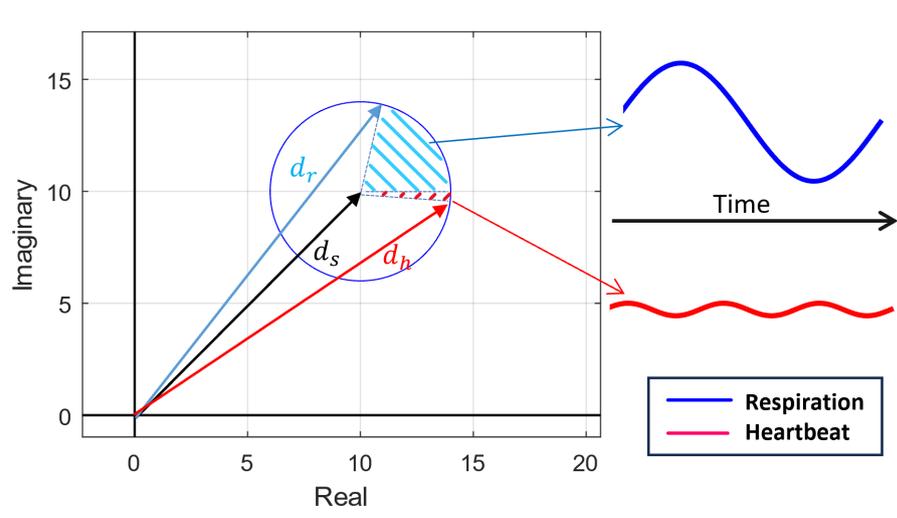


Figure 2.11: Phasor diagram of theoretical plot of respiration and heartbeat reflected by 5.8 GHz WiFi signal

Empirically our heartbeat and respiration are periodical, which reflects the physical motion of chest and ventricle can be modelled by continuous signals, with corresponding motion amplitude of d_{rr} and d_{hr} . If we want to apply wireless sensing methods to get these frequencies from microwave signals, it's necessary to explore the availability of algorithm. Now we model the response of wireless signal as $H(t)$ [134] in below Eq. 4.1.

$$\mathbf{H}(t) = \sum_{i=1}^{N_d} a_i(t) e^{-j2\pi \frac{d_{rr}(t) + d_{hr}(t)}{\lambda}} \quad (2.3)$$

where, N_d is the index of path that signal passes through, $a_i(t)$ represents the complex attenuation factor of the i^{th} path; $e^{-j2\pi\frac{d_i(t)}{\lambda}}$ represents the phase change of i^{th} path, with the changing distance of $d_i(t)$ in i^{th} path. λ represents the wavelength of the wireless signal. For respiratory and cardiac signals, they originate within the human body and influence the propagation characteristics of wireless signals through mechanisms like reflection, absorption, and scattering. These biomechanical and physiological processes are relatively independent. As a result, their effects on wireless signals can be considered independent and hence can be linearly superimposed. This property of linear superposition allows us to extract and analyse these two physiological signals separately from the wireless signals, although effective algorithms might be needed in practical applications to accurately differentiate and extract these signals.

Last but not the least, from the model algorithm it can be discovered that the phase variation is decided by both factors: motion distance and signal wavelength. The normal distance of heartbeat and chest motion of adults are bounded into 0.2-0.5 mm and 4-12 mm respectively [135], [136]. We assume that there is a WiFi signal of 5.8 GHz carrier frequency, of which wavelength is about 5.2 cm, passing through a moving metal surface of 0.5 mm and 12 mm displacement. The phase variation can be illustrated in Fig. 2.11. Compared to the discernible respiration-reflected signal, capturing the heartbeat is challenging, especially when considering the noise from free space and hardware.

where the d_s represents the distance between radar and purely static human position (ideally without any motion including respiration and heartbeats). d_r and d_h represent the distance from radar to surface of chest and Therefore, the hardware that supports RF signal transmission of higher frequency should be adopted for heartbeat detection. The automotive FMCW radar that works at 77 GHz bands offers a superior combination for vehicular sensing including encompassing resolution, potential for miniaturization, designated frequency allocation, and sufficient penetration capability. These factors collectively do not only make this frequency band the ideal choice for automotive radar, but also for vital detection.

2.5.3.1 Literature review of Radar based vital detection

Recent advancements in vital detection have brought forth several validated proposals. Since the phase tracking methods from specific range bin was proposed [138], FMCW radar is beginning to be used to detect physiological signals. Then, [139] introduced a classic scheme utilizing FMCW technology for vital sensing. Their approach incorporates time-frequency analysis and a bandpass filter. By analysing the frequency difference between transmitted and received signals through FFT, they are able to estimate the location of subjects. The periodic response at this distance is assumed to reflect the subject's

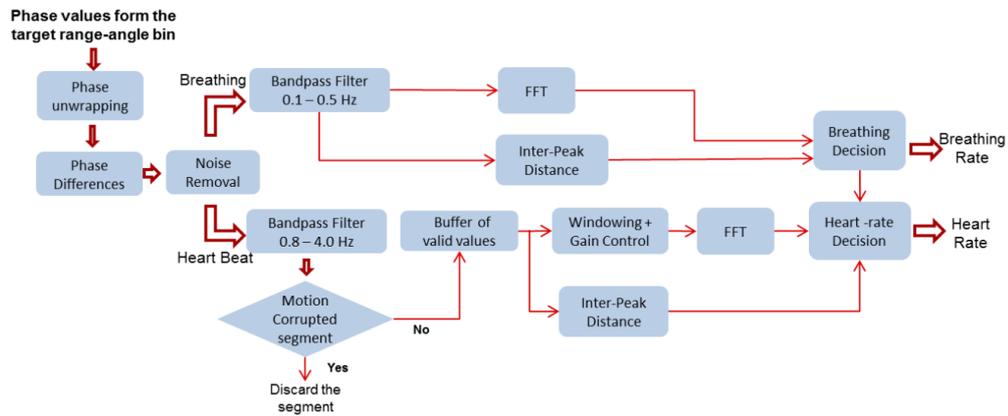


Figure 2.12: Typical vital detection system structure [137]

vital frequency, as it is believed to contain primarily physiological activity information. To detect multiple subjects, [137] introduced the concept of Angle of Arrival (AoA) using linear antennas arrays, building upon a similar scheme, shown in Fig. 2.12. [140] proposed an enhanced method within this framework, incorporating linear demodulation to compensate for phase signal loss. However, these references primarily focused on bandpass filters for extracting and isolating breathing and heartbeat signals from the original phase information, but overlooking the potential of amplitude information. In addition, the papers by [141], [142] suggest that radar signals can capture cardiac vibrations and map Seismocardiography (SCG) to Electrocardiogram (ECG). Nevertheless, these experiments are confined to specific setups mentioned in the papers and face challenges in terms of widespread applicability in daily life. Meanwhile, the preceding article has provided limited discussion on post-phase distortion signal processing methods. In light of this, we conducted a search for approaches aimed at compensating for phase distortion in radar systems.

For the aim to establish a improved sensing scheme for daily usage, it's necessary to capture fine-grained physiological signal to compensate the phase distortion. The author of [143] utilized phase distribution superposition of multiple channel with MIMO radar that have 20×20 arrays to recover the main respiration and heartbeat components. In proposed driver monitoring system [144], an optimization method with smoothing spline was adopted to compensate the distortion of the unwrapped phase. A systematical improved VMD method is proposed by [145] to separate the respiration and heartbeat signals from raw information.

Some papers work on smoothing and fitting the extracted phase signal to a uniform size to eliminate the influence from distortion and subtle motion. The author introduce an fitting method to mapping the raw phase signal to a smooth spline to enhance the vibration tendency that is influenced by human heartbeat and respiration [144]. However, these fitting methods needs longer windows length and higher requirement towards com-

putational capability, which may not be supported by distributed and embedded devices under the daily scenarios of in-vehicle and in-bed. Furthermore, it proposed a joint decomposition method to reconstruct respiratory and heartbeat signals by optimizing important information contained in multiple range-azimuth voxels, similarly as [145]. But both of the mentioned works are limited to driving scenario. For daily usage, it should be compatible within different sensing range. Furthermore, it's hard to find multiple voxels from neonatus due to the smaller body size. In contrast, the proposed system can be flexibly implemented for adults and neonatus.

Originally, the Quadrature Imbalance compensation method was proposed for single-tone continuous waveform (CW) radar without frequency modulated [146], [147]. Meanwhile, few papers notice the availability of IQ signal in vital detection using FMCW radar [148], [149]. The author of [148] provided a deviation based evaluation methods, which firstly checks for IQ imbalance, and only when the IQ signal is relatively balanced does the system record and compute the frequency. The threshold of imbalance checking needs an intuition and initial measurement to fine-tune the system. [149], [150] comprehensively validates the DC compensation of IQ signal can be transferred from CW to single tone FMCW radar system. The improvement make it possible to distinguish different person with their vital signs information in different locations.

2.6 Summary

This chapter provides an extensive review of technological advancements in healthcare monitoring systems, with a special emphasis on RF sensing and its applications in non-hospital settings. It highlights the critical role of RF signals in monitoring vital signs like heart and lung movements, enabling real-time monitoring of respiration and heart rates. The chapter also explores the concept of JCAS in the IoT, underlining the efficiency and cost-effectiveness of integrated sensing and communication systems.

Additionally, the chapter discusses the use of WiFi sensing technologies in healthcare, covering their roles in human activity recognition, vital sign monitoring, and fall detection. It points out the benefits of WiFi sensing in terms of privacy, through-wall sensing capabilities, and cost-effectiveness. The review extends to LoRa sensing, recognized for its long-range communication and low power consumption, suitable for urban health monitoring and remote patient care. Radar sensing is also examined for its precision in detecting the position and movement of targets, contributing to non-invasive patient monitoring.

In conclusion, the chapter reflects on the emerging trends, challenges, and future developments of these technologies, especially their integration into smart homes and telemedicine. It underscores the significant potential of these technologies in enhancing healthcare services in non-hospital environments, and anticipates further advancements in this rapidly evolving field. Integration with IoT and applications in smart homes and telemedicine marks a significant step forward in achieving more convenient, accurate, and non-invasive healthcare services.

WiFi Sensing of Human Activity Recognition

Indoor human activity recognition (HAR) has played a significant role in intelligent internet of things (IoT) controlling and healthcare monitoring. Device-free WiFi sensing in HAR is an emerging research trend for nearly a decade, which spawned various kinds of applications including vitals monitoring, human daily activity recognition, falling detection, and signs recognition [10], [134]. Compared to other equipment of radar based device-free approaches, the WiFi devices objectively own higher cost-effective which is the better choice for general indoor cases implementation.

Since the WiFi signals are approved to be accessible for wireless sensing, there are various kinds of HAR systems have been developed. WiFall system [45] is designed to detect falling activity for health monitoring using the stability of CSI time and diversity of frequency spectrum of WiFi signals. Lip motion detection system [151] is proposed for under-mask speech recognition, which utilizes WiFi signal's variation while human speaking to classify different pronunciation. E-eyes [152] leverages the distribution information of CSI amplitude for HAR tasks. All the above systems are focused on extracting temporal signals influenced by human activity but don't utilize the spatial information brought by the WiFi antenna array. Widar3.0 [153] realizes an overall HAR system that can be transferred to different environments, orientations, and locations. It integrates the Doppler frequency variation from multiple views of WiFi devices in distinct coordinates, with regard to human position. However, the system is only available when the scenario has more than three receivers.

Inspired by the Widar3.0 system, we proposed two feature extraction methods, of which the first is based on Doppler feature and another is based on continuous AoA and ToF maps based method that adopt single pair of commercial WiFi devices to achieve fine-grained human activity recognition task. Our contributions are summarised in three points.

1. We firstly leverage the continuous AoA-ToF information in the HAR task as our best knowledge. Compared to the previous frequency spectrum based classification method that only leverages temporary information, the CATM preserves both temporal and spatial human body information.
2. We design a series of lightweight temporal neural networks for HAR tasks for our CATM based HAR system.
3. We collect and establish a dataset of human activity with WiFi sensing, and human skeleton information for validation.

Besides, the denoise and conventional deep learning based WiFi HAR is discussed ahead of the CATM as comparison studies. In the beginning, we explore the respiration detection to approve the signal sanitation methods for HAR system. In the second section, we explore the SOTA model, transformer model's availability on WiFi HAR system. Next but not the least, we propose our CATM system in detail, which applied the enhanced spatial information of variation to improve the single WiFi device's performance of HAR.

3.1 Preliminaries

Channel state information (CSI) is used to characterize the channel attributes of WiFi links in the physical layer as the sample of channel frequency response, which can be disturbed by multipath fading and shading. In the indoor environment, human movement is one of the significant factors that influence the CSI value. Meanwhile, CSI data contains human-related information, like daily movement and rhythmic physiological phenomena, like respiration and heartbeat. WiFi sensing technique aims to extract and analyse this potential human related information. We could use the equation to describe the CSI:

$$\mathbf{H}(f, t) = e^{-j2\pi\Delta ft} (H_s(f) + H_a(f, t)) + N(t) \quad (3.1)$$

where $e^{-j2\pi\Delta ft}$ is the random phase shift due to the asynchronous sending and receiving process of the WiFi system; H_s and $H_a(f, t)$ represents the CSI signals from all the static paths (including the signals in LOS path and those reflected off the stationary objects) and active paths (including signals reflected from the dynamic objects) respectively. $N(t)$ represents the noise during transmission. CSI signals in active paths can be expressed as:

$$\mathbf{H}_a(f, t) = \sum_{i=1}^{N_d} a_i(f, t) e^{-j2\pi \frac{(d_i(t) + d_a(t))}{\lambda}} \quad (3.2)$$

where N_d is the index of the dynamic path, $a_i(f, t)$ represents the complex attenuation factor of the i^{th} path; $e^{-j2\pi \frac{d_i(t)}{\lambda}}$ represents the phase change of i^{th} path; $d_i(t)$ and $d_a(t)$ are the static length and vibrating length of i^{th} path. λ represents the wavelength of the WiFi signal.

Through coupling the above two equations, it can be found that human related information is reflected from the amplitude and phase shift. However, previous researches found that the raw CSI data is too noisy to extract the feature [128], [154], especially for tiny motion. In this case, we designed a series of experiments of breathing detection to validate the performance of data sanitation techniques. The reason why we choose breathing instead of other activities is the respiration information is periodical and tiny enough, which make it easier to check the correlation between main frequency response and human motion.

3.2 Respiration detection with sanitized CSI phase

Although the captured CSI signals are distorted in distinct paths, the main components of noise can be modeled as a whole considering SFO and CFO effects. First, we apply a proposed calibration method of establishing the solid reference link using a coaxial line to connect transmitter and receiver in [154], shown in Fig. 3.1. The established link is not affected by transmission attenuation in the air. Then calculate the conjugated multiplication of CSI complex signals of reference link and wireless link. Meanwhile, we calculate the IFFT of raw CSI data to get ToF value. By setting the threshold of ToF and filtering long time-delay components, we can reduce the multipath noise.

3.2.1 Experimental Setup

The system is implemented with a pair of WiFi devices equipped with Intel 5300 wireless network interface controller and the open-source CSI driver [155]. The carrier frequency and sample frequency are set to 5.8 GHz and 200 Hz. We adopt one transmit antenna and two receiving antennas, with a coaxial cable and a splitter to connect two devices. The distance of two receive antennas is set to the half-wavelength of 2.6 cm. In each of the experiments, we asked volunteers to breathe in the fixed rhythm of 0.2 Hz, 0.25 Hz and 0.33 Hz respectively, while collecting data.



Figure 3.1: Experiment hardware connection of system

3.2.2 Respiration estimation

After the calibration of CSI data, we acquire the denoised CSI signals which show the lung's periodic movement clearly, shown in the Fig. 3.2.

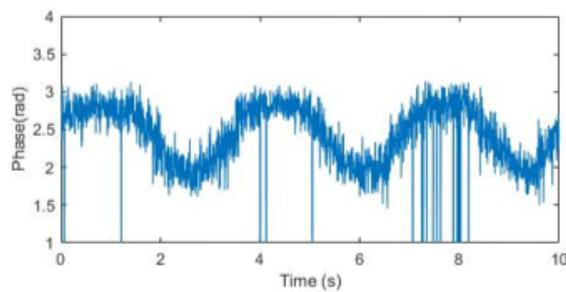


Figure 3.2: Filtered respiration signal.

Then we calculate the ratio of variance and difference of maximum and minimum values to select a single subcarrier that is the most sensitive to lung movement, which can reduce the static components' influence. The respiration rate estimation is performed for one and two human subjects.

3.2.2.1 Respiration rate estimation for one human subject

Firstly, we adopt the peak detection (PD) method on processed signals. However, in some cases we found the vibration range of lungs during breathing can be different, which led to the inconsistent peak in the respiration signals. To improve the performance, we adopt the auto-correlation function (ACF) to enhance the periodicity of signals.

3.2.2.2 Respiration rate estimation for two human subjects

PD cannot work well in the case of the two-person scenario because the components belonging to different identities cannot be separated in the time domain. The accessible solution is to transfer the signal into the frequency domain with FFT.

3.2.3 Evaluation

3.2.3.1 Single person respiration rate detection

In the one-person scenario, we firstly discuss the performance of PD and PD with ACF. The cumulative distribution function (CDF) shows the performance of our system in Fig. 3.9a. The overall median error is 0.7 bpm.

Meanwhile, NLOS signals usually involve rich reflection, diffraction, and refraction, resulting in significant attenuation of transmission power theoretically. We also conduct the experiments under the weak NLOS environment that we placed a cardboard of 2m high and 1m wide, between the transmitter and receiver. The CDF shows the slightly different performance of estimation in LOS and NLOS scenarios in Fig. 3.9b. We can find our system can adapt the NLOS scenarios in the sedentary situation.

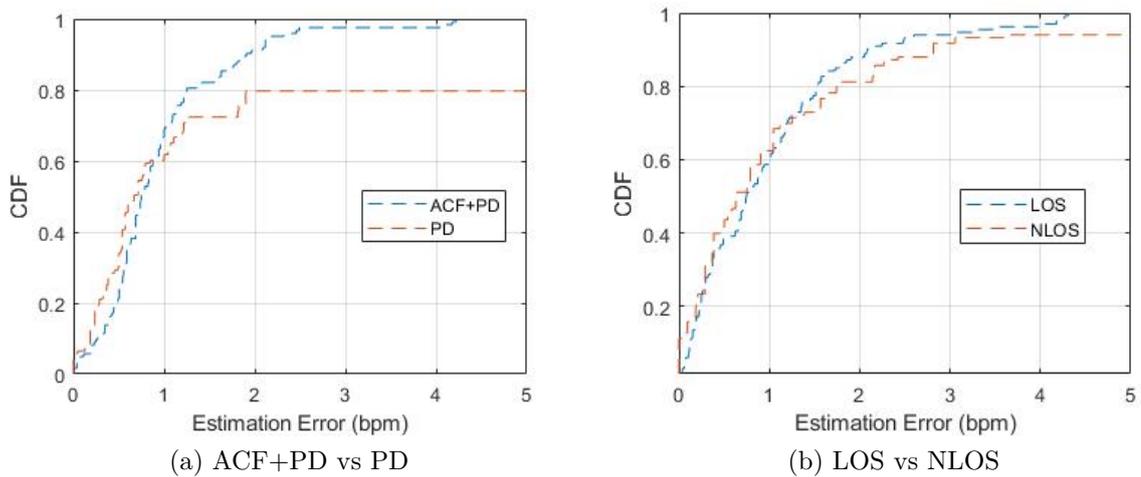


Figure 3.3: CDFs of real-time frequency estimation

3.2.3.2 Two-person respiration rate detection

The image of Fig. 3.4 shows the frequency analysis results in the two-person scenario. The volunteers are asked to breathe in different frequencies (0.2 Hz, 0.25 Hz, and 0.33 Hz). The average error is the account of 1 bpm in total. In this scenario, we didn't consider the spatial information of respiration, so the number of subjects should be known for the system.

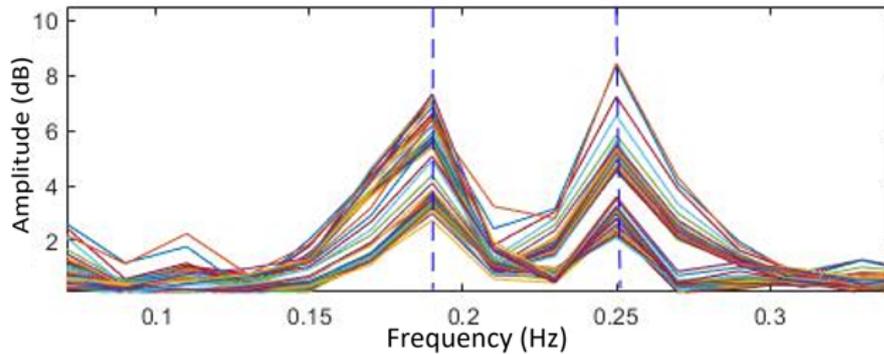


Figure 3.4: Frequency analysis of the two-person scenario

3.3 Methodology and System Design of Doppler WiFi HAR

Previous work has provided a denoising method for subsequent fine-grained HAR. With multiple WiFi devices placed surrounding the monitoring area, wireless CSI data can be collected, but is distorted by human activities. In this section, we introduce the overall processing and classification procedures, including data collection, pre-processing, Doppler based feature extraction and Transformer based recognition. Accordingly, the proposed system can be divided into two modules, the Doppler based feature extraction module and the activity recognition module.

3.3.1 System Setup

System Implementation The proposed system is implemented based on one WiFi transmitter and two WiFi receivers, equipped with Intel 5300 wireless NIC and the relevant CSI driver [155]. Given few interfering radios, the driver is set with monitoring mode on the channel at 5 GHz with 40 MHz bandwidth. The transmitter is set to broadcast the rate of WiFi packets at 1,000 packets per second with the receivers activated in a line, where there is a 5 cm (larger than the half-wave wavelength) distance to avoid the interference of closed antennas.

3.3.1.1 Evaluation Setup

Our system was tested in one indoor environment, illustrated in Fig. 3.5a. The sensing area is with a 2.4 m \times 1.6 m rectangular, with one WiFi transmitter (Tx) and two WiFi receivers (Rx), as shown in Fig. 3.5b. Two volunteers participated in the designed experiments where each performed five different activities for data collection, to increase the inter-class variation and strengthen the Transformer model. In the experiment, all the volunteers were asked to finish one activity in three seconds and each data sample contains 3,000 packets CSI data. The overall dataset contains 500 samples of different activities including walking, jogging, leaning forward and back, putting both hands up and down, and waving the left hand. The choice of activities was in relation to the envisioned applications and use cases of this system, which includes future healthcare sensing for in-home care (walking, jogging, leaning forward and back) and smart-home control (gesture recognition).

3.3.2 Processing of WiFi CSI data

WiFi CSI data is not only influenced by high-frequency noise and phase offset, but also disturbed by the static and dynamic motion from other sources, as indicated in Eq. 4.1. It is necessary to reduce the noise in the data before performing feature extraction. There are three noise components that require attention:

- Static components of CSI data from the environment.
- High-frequency noise from the communication hardware.
- Phase offset from the NIC.

Fortunately, a number of methods have been proposed to solve the above problem. For the phase offset, Widance system [156] provided a reliable phase offset removal methods. Given that the phase offset of three antennas on the same receiver should be the same, which is produced from the same NIC. This method uses the phase information from a

selected antenna with lower sensitivity of dynamic information to unify the phase offset of the rest two antennas. The authors of IndoTrack [157] proposed an amplitude adjustment method to reduce the influence of static components in CSI data. Then, Butterworth band-pass filter and principal component analysis were applied to filter other noise [158]. Our proposed system applies all of the mentioned pre-processing techniques to extract the DFS.

3.3.2.1 Doppler Frequency Shift Extraction

In this section, the Doppler method is used for feature extraction to get the profile of different persons' activities. The DFS generated by different limb movements can be used to identify the specific human activities. As the CSI data has been pre-processed, we applied the short-time Fourier transform on the dataset to get the Doppler spectrum, which represents the velocity variation of torso and limbs. There are two DFS representing two activities as shown in Fig. 3.6.

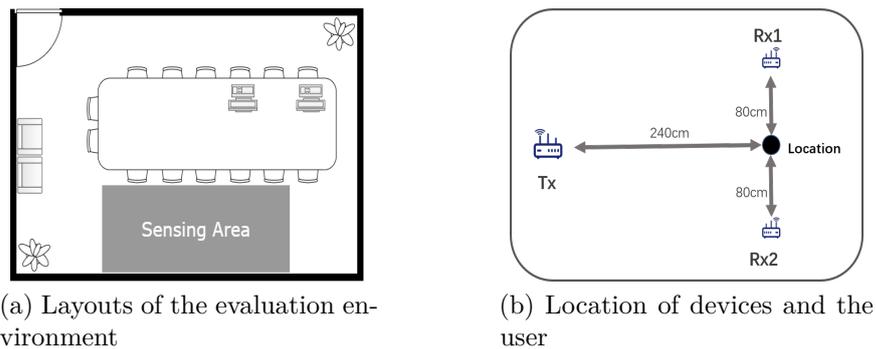


Figure 3.5: Layout of experiment setup

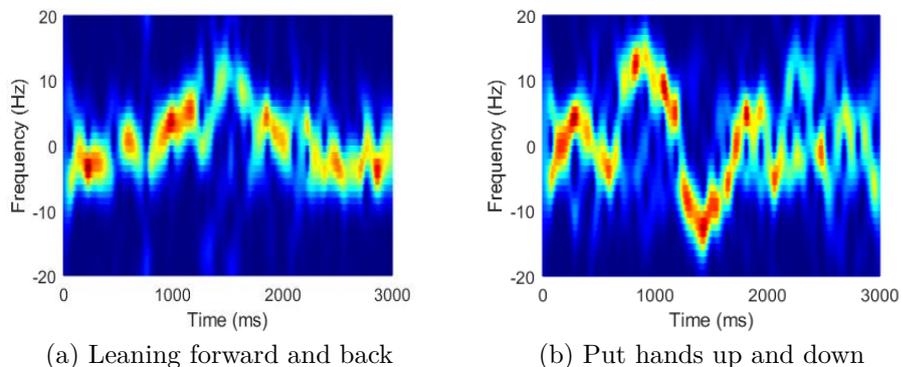


Figure 3.6: Doppler spectrum of indoor activity profile

In Fig. 3.6a, the spectrum illustrates the activity of leaning forward and back, with a peak representing the fastest speed of the torso. In Fig. 3.6b, the abrupt change stands for the fast movement of the arms.

3.3.2.2 Activity Recognition

Our HAR monitoring system is achieved by a two-stream Transformer model, proposed by [159], which performs human activity classification by learning the features of DFS. It involves extracting both the temporal and spatial features which contain the information within the entire sequence. The core of this model is the multiscale convolution augmented transformer. DFS frames are entered into the two-stream self-attention layers to extract features, and finally pass through the feature fusion layer to obtain the prediction of human activity.

3.3.3 Evaluation of Doppler based recognition system

3.3.3.1 Network Training

The input of our network is the Doppler spectrum extracted from raw CSI data, as shown in Fig. 3.6. The input size of each sample is 121×3000 , where the 121 represents the resolution of frequency from -60 Hz to 60 Hz, and 3000 represents the time series of milliseconds, where each sample lasts 3 seconds. To focus on the human activity monitoring, and improve the efficiency of the neural network, we adjusted the human-induced DFS range is within ± 20 Hz, and down sampled 3000 sets of CSI data to 500 packets. Besides, due to the relatively small size of our dataset, we use a single temporal module to train Transformer. At first, the dataset of five activities from two people is training, and then the third people are introduced. Finally an identity classification for the first two users is conducted. The whole dataset is divided into two sets, where 80% data are used as training set and 20% as test set.

3.3.3.2 Doppler based Recognition Results

To quantitatively measure the performance of our system, we applied 5-fold cross validation, which divided sample data randomly into five parts, each time one is selected as the test set and the rest was used for training the system. The results of five activities of two volunteers achieve 87.6% accuracy (on average) in the 5-fold cross validation, with the best result of five achieves 92.7%. Table 3.1 shows the confusion matrix.

The table illustrates similarities between different actions. For example, jogging and walking classes are partially confused because the two activities have similar body movements. To explore the factors that influence activity recognition performance, the identity classification is conducted on a single walking behavior of user-1 and user-2. The dataset is

Table 3.1: Confusion matrix for all five classes with Transformer based activity classification

All Five Classes - Accuracy (87.6%)							
<i>Class</i>			<i>Predict Class</i>				
			1	2	3	4	5
True Class	1	Walking	81	16	2	1	0
	2	Jogging	16	76	4	0	0
	3	Leaning forward and back	2	2	94	0	0
	4	Put the hands up and down	1	0	0	91	8
	5	Wave left hand	0	0	0	4	96

randomly divided into 80% and 20% for training and testing, respectively. Our proposed system performs an overall accuracy of 77%, which shows that different identities influence the Doppler based human activity classification result, and it is possible to recognize the identity of different persons.

The rest parts of the chapter are structured as follows: Section 3.4 demonstrates the methods of signals calibration, feature extraction of CATM, and temporal neural network classification. Section 3.5 analyses the overall performance and give the conclusion respectively. The structure of the proposed system is shown in Fig. 3.7.

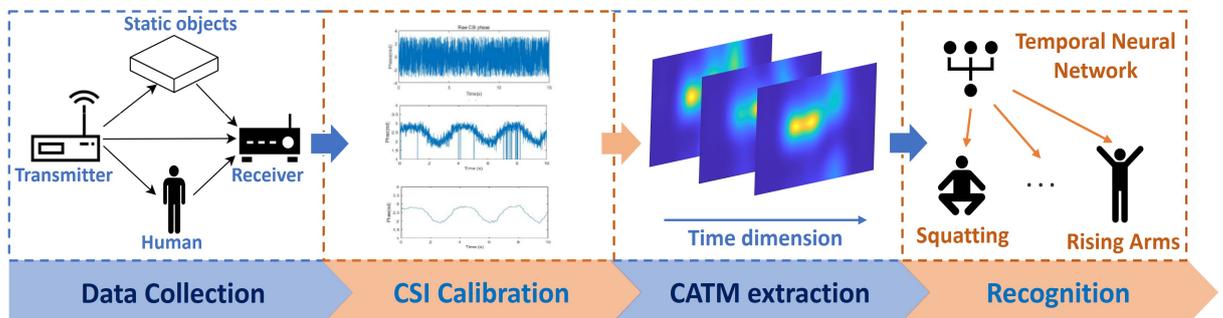


Figure 3.7: CATM based HAR system of WiFi sensing

3.4 Continuous AoA-ToF maps (CATM) Methodology

3.4.1 Calibration technique

At first, we collected CSI data using Intel 5300 NIC and omnidirectional antennas. A volunteer is asked to sit statically between a WiFi receiver and a WiFi transmitter and keep breathing evenly. The plots of raw CSI amplitude and phase are shown in Fig. 3.8. In comparison with CSI amplitude, CSI phase information of human body motion

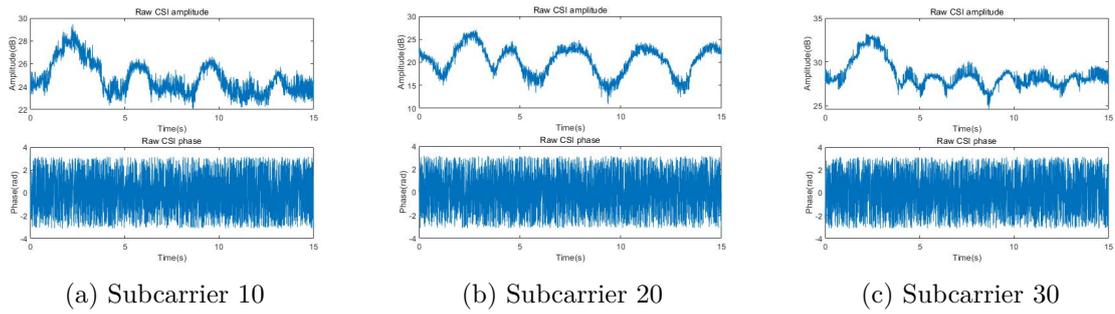


Figure 3.8: Raw CSI data within volunteer keep breathing evenly for 15 seconds of different subcarriers

is more sensitive, especially for respiration detection [154], [160]. However, the random phase shift from asynchronous links pollutes informative phase variation as Fig. 3.8 shows. Phase sanitation is significant for the post-classification task of WiFi sensing. According to the previous analysis of WiFi CSI transmission [161], noise components can be divided into sampling frequency offset (SFO) and packet detection delay (PDD) caused by the receiver:

- SFO: Clocks of transmitter and receiver are out of synchronization, which leads the sampling frequency of the receiver to shift randomly. This will cause the phase of collected signal to be unpredictable shifted.
- PDD: Detection of WiFi signals in receiver side cause the delay.

Next, the conjugation multiplication is adopted in our system to reduce the SFO which has been proved in IndoTrack [162]. It considers the phases of two transmission paths are shifted by the same value from the receiver. In this case, conjugate multiplication can help to reduce SFO and PDD. Secondly, due to the multi-path effect of WiFi signals, receivers can collect multiple impulse signals while the transmitter only sends them once. It shows the time dispersion of wireless transmission, which means the ToF should be various for one transmitted packet. Based on the line-of-sight path, we adopt the ToF based filter to eliminate the high delay components of CSI signals. To explore the influence of ToF corresponding to the respiration signals, we adopt the PDP. The PDP method utilizes the influence of time delay due to the multi-path effect. PDP can be calculated by IFFT of raw CSI data. Due to the multi-path effect, physical WiFi signals are not only transferred by the LOS path but also can be reflected by any objects around the circumstances, which is in NLOS paths. The signals with a long time delay can be regarded as uncertain data that carried more environmental information. Meanwhile, considering the sampling frequency, the time delay should be smaller than the time gap between the two samples. Otherwise, the signals will be attenuated by other samples' information in the hardware or transmission medium. In this case, we calculated the PDP of CSI signals and set the threshold of time to the sampling time to filter the long ToF signals out. The experiment result is shown in Fig. 3.9 tested in distinct subcarriers.

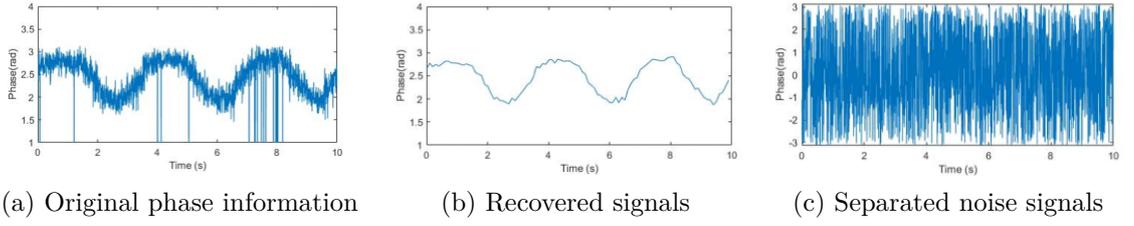


Figure 3.9: The PDP filtered result of CSI.

3.4.2 AoA-ToF Maps Construction

For our activity classification, the directional movement feature of the human body is the main parameter for researchers to establish the mapping relationships from signals to multiple activities class [163]. However, for complex human motions, like waving hands (left and right side), the behavior patterns can be similar to each other. We assume that the human motion is capable to influence the channel parameters of AoA and ToF. AoA indicates the direction of receiving signals with respect to the receiver and ToF indicates the time cost of transmission from the transmitter to receiver, by LOS and NLOS. In this case, we extract CATM, which directly represent the behaviour coordination in physical space.

Firstly, we regard the double antennas as an antenna array and analysis it with a Multiple Signal Classification (MUSIC) signal processing algorithm [164]. However, the MUSIC algorithm requires the number of transmission paths should be smaller than the number of antennas. In our setup, the number of antennas is limited to the number of ports in the commercial WiFi chip. In the equation, we separate the signals received by each antenna into 30 different groups based on the number of subcarriers of the chip. The separated frequency bands of subcarriers lead to frequency-selective fading. For different frequency components of the signal, the wireless transmission channel will show distinct random responses. According to the frequency-selective effect, we simulate a virtual antenna array composed of 30 independent subcarriers from a single antenna. On the other hand, in the linear antenna array, phase differences between two actual antennas are related to the time cost of receiving the data from the same path. Suppose the index of antenna and subcarrier are M and K respectively.

$$\mathbf{X}_{CSI} = \begin{bmatrix} x_{csi}(1,1) & x_{csi}(1,2) & \cdots & x_{csi}(1,K) \\ x_{csi}(2,1) & x_{csi}(2,2) & \cdots & x_{csi}(2,K) \\ \vdots & \vdots & \ddots & \vdots \\ x_{csi}(M,1) & x_{csi}(M,2) & \cdots & x_{csi}(M,K) \end{bmatrix} \quad (3)$$

Next, we analyze the phase shift indicator with the AoA and ToF value of the individual path. We suppose the incident angle of receiving signals towards the linear array is θ .

$$\phi_\theta = \frac{2\pi(m-1)d\sin\theta}{\lambda} \quad (4)$$

We mainly adopt the MUSIC based method which is proposed in [165]. Due to the frequency band f_θ between two adjacent subcarriers, there is a 2.5 radians phase shift. At l_{th} path, the phase shift of k_{th} subcarrier with ToF τ is given by:

$$\phi_\tau = 2\pi(k-1)f_\theta\tau_l \quad (5)$$

Besides, we leveraged the Smoothed-CSI matrix proposed in SpotFi [165] to increase the virtual unit and enhance the linear array association. The Eq. 6 shows the cases of $M = 2$ and $K = 30$. For simplicity, We use \mathbf{X}_{S-CSI} to call the smoothed CSI matrix.

$$\mathbf{X}_{S-CSI} = \begin{bmatrix} x_{csi}(1,1) & x_{csi}(1,2) & \cdots & x_{csi}(1,15) \\ x_{csi}(1,2) & x_{csi}(2,2) & \cdots & x_{csi}(2,16) \\ \vdots & \vdots & \ddots & \vdots \\ x_{csi}(1,15) & x_{csi}(1,16) & \cdots & x_{csi}(1,30) \\ x_{csi}(2,1) & x_{csi}(2,2) & \cdots & x_{csi}(2,16) \\ x_{csi}(2,2) & x_{csi}(2,3) & \cdots & x_{csi}(2,17) \\ \vdots & \vdots & \ddots & \vdots \\ x_{csi}(2,15) & x_{csi}(2,16) & \cdots & x_{csi}(2,30) \end{bmatrix} \quad (6)$$

In our system, we assume the phase difference of different subcarriers is ϕ_τ . Therefore, we replace $x_{csi}(m,k)$ with $e^{-j\phi_\tau}$ and $e^{-j\phi_\theta}$ to explore the phase shift. Meanwhile, we label the transmission path index as L . The steering vector a and steering matrix \mathbf{A} can be written as:

$$a(\theta, \tau) = [1, e^{-j\phi_\tau}, \dots, e^{-j\phi_\tau^{14}}, e^{-j\phi_\theta}, \dots, e^{-j(\phi_\theta + \phi_\tau^{14})}] \quad (7)$$

$$\mathbf{A} = [a(\theta_1, \tau_1), a(\theta_2, \tau_2), \dots, a(\theta_L, \tau_L)]^\top \quad (8)$$

In the MUSIC algorithm, the received data is used for eigen-decomposition to separate the signal subspace and the noise subspace. The orthogonality of the signal direction vector and the noise subspace is used to form a spatially scanned spectrum for a full-domain search of the spectral peaks, which enables parameter estimation of the signal. To apply the algorithm, we firstly establish a linear model of a received signal vector of \mathbf{X} , in the presence of a steering matrix of \mathbf{A} , Gaussian white noise of \mathbf{N} , and Smoothed-CSI matrix of \mathbf{S} :

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{N} \quad (9)$$

The basic idea of the MUSIC algorithm is to analyse the covariance matrix \mathbf{R}_x of received signals:

$$\begin{aligned} \mathbf{R}_x &= \mathbb{E} \{ \mathbf{X}\mathbf{X}^H \} \\ &= \mathbf{A} \mathbb{E} \{ \mathbf{S}\mathbf{S}^H \} \mathbf{A}^H + \mathbb{E} \{ \mathbf{N}\mathbf{N}^H \} \\ &= \mathbf{A}\mathbf{R}_s\mathbf{A}^H + \mathbf{R}_N \end{aligned} \quad (10)$$

where \mathbf{R}_s and \mathbf{R}_N represents the correlation matrix of multi-path incident signals and noise respectively. In the K -size eigenvalues of correlation matrix, the smallest $K - L$ eigenvalues are corresponding to the noise and others eigenvalues correspond to L incident signals. The noise subspace can be constructed: $\mathbf{E}_n = [\vec{e}_1, \vec{e}_2, \dots, \vec{e}_{K-L}]$. Therefore, the power spectrum with respect to the AoA θ and Doppler frequency dop can be expressed as:

$$\mathbf{P}(\theta, \tau)_{MUSIC} = \frac{1}{\mathbf{a}^H(\theta, \tau) \hat{\mathbf{E}}_N \hat{\mathbf{E}}_N^H \mathbf{a}(\theta, \tau)} \quad (11)$$

Besides, attributed to the short time delay of subcarriers on a single antenna and limitation of resolution in MUSIC, there are negative ToF value signal components shown in the matrix. Generally, the MUSIC algorithm only focuses on the peak of the power spectrum which reflects the sources' direction. We suppose the negative parts are produced by which is brought by the operation of conjugation multiplication calibration. These components are significant for classification. These negative ToF values in CATM are artifacts introduced by a conjugation multiplication calibration process, which is likely a step within the signal processing chain designed to improve measurement accuracy or resolve ambiguities in signal phase. Despite their physical implausibility, these negative components are considered significant for the classification tasks performed by the neural network. The authors hypothesize that the pattern or "shape" of the signal's representation in the negative time region could provide useful features for the neural network to learn from. We adopt these parts and suppose the shape of heatmaps in the negative region is helpful for neural network classification. In the discussion of Section. 3.5.2.4 we compare the influence on performance between non-negative ToF CATM (NT-CATM), which only consider the positive ToF features, and full CATM maps. There are two activities, i.e. walking and jogging, that are posted in the time domain as Fig. 3.10. Because the patterns of these human movements are similar, it can be proved that the effectiveness of feature extraction with the similarity of two CATM. The red boxes represent the similarity between the two activities, which shows the robustness via different time slots.

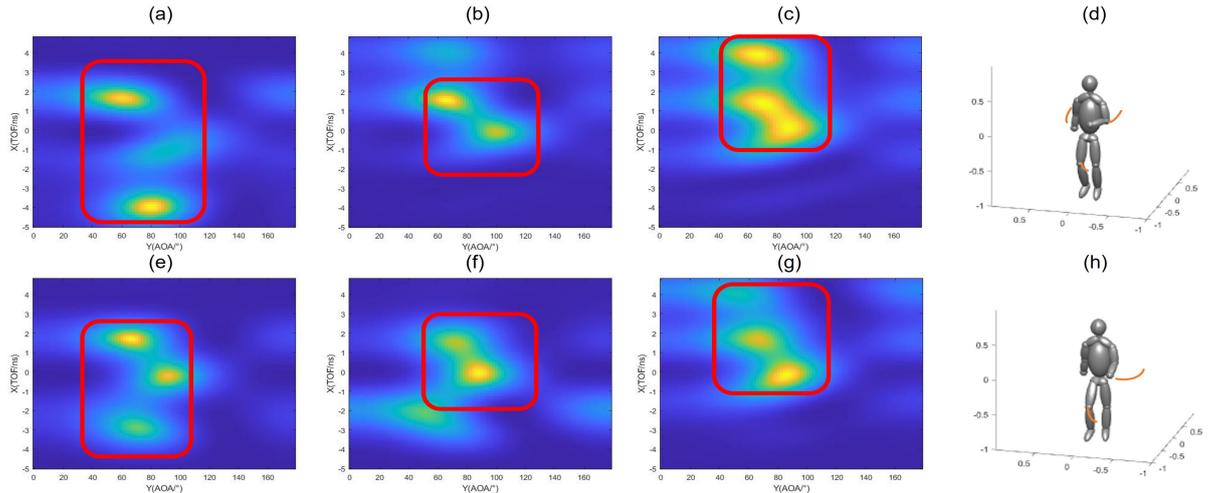


Figure 3.10: ATM result of CSI while human jogging (a,b,c) and walking (e,f,g).

3.4.3 Temporal classification neural network structure

3.4.3.1 1D LSTM-CNN structure

The structure of our proposed neural network is illustrated in Fig. 3.11. It contains 3 components: frame embedding, sequential fusion, and classification. The frame embedding consists of one convolutional layer, one residual block, and one pooling layer. We implement the residual block following the function in [166], where an identity mapping is introduced for faster back propagation of gradient. The following is the sequential fusion, which intends to collect the information from the temporal sequence. In the module, we adopt an LSTM structure, which has a robust ability to retain the long-term information in time sequence. In the final step, a simple linear projection with SoftMax is utilized as the classification module that generates the probability distribution of predicted motions. The model is trained by CrossEntropy loss and Adam optimizer.

3.4.3.2 Fusion LSTM-CNN structure

Although the WiFi signals from single direction is sufficient to recognize the human activity, we further improve the performance by combining the information from different routers.

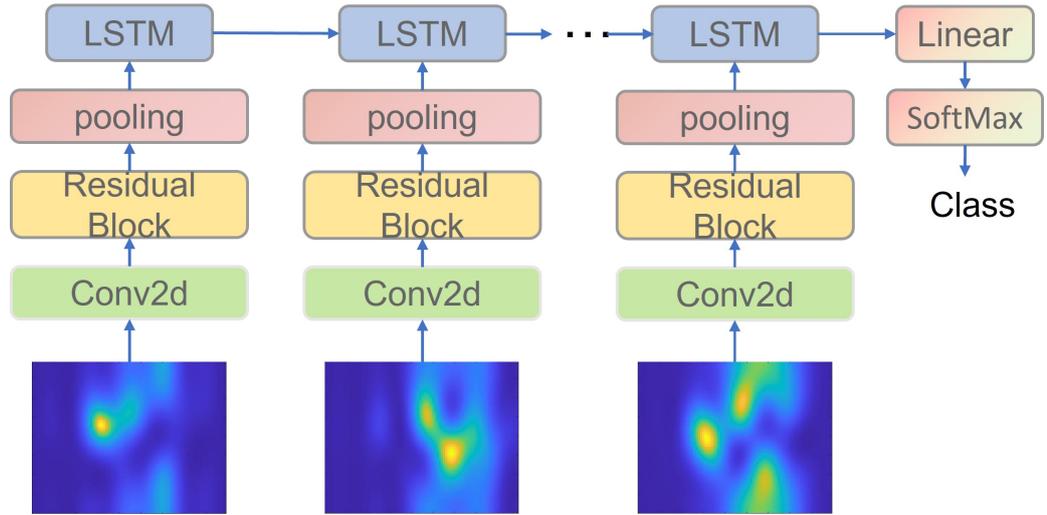


Figure 3.11: Res-LSTM Temporal neural network structure

3.5 Evaluation of CATM system

3.5.1 Experiment Setup

We adopted four PCs equipped with Intel 5300 COTS WiFi devices to sense 8 different activities of 5 volunteers. As Fig. 3.12(d),(h) shows, we set up 1 transmitter and 3 receivers to collect CSI data. Meanwhile, we equipped a Kinect V2 depth camera by the side of the transmitter to capture the human skeleton coordinates for labelling the activity types, as Fig. 3.10 shows. For CSI data collection, each profile contains 3-second human activity with a sampling frequency of 1000 Hz.

After CSI calibration and CATM extraction, the dimension of each profile is set to $S \times H \times W$, where S represents the length of the sequence in one packet, H and W both refer to the dimension of a single ATM. We set the size of CATM profile to $100 \times 90 \times 90$. In total, we collected 7998 profiles. To standardize the training step, we set 100 epochs with 32 batch sizes for our system and other comparison tests. The whole dataset is randomly picked 20% of profiles as test set.

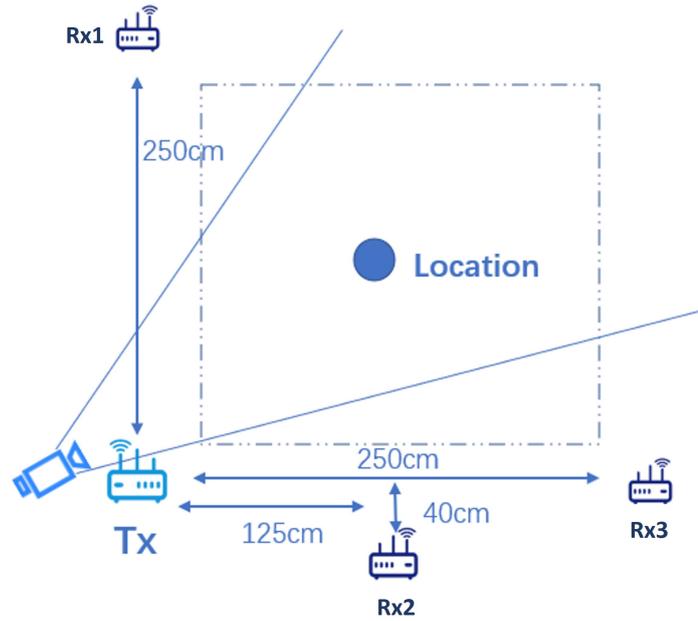


Figure 3.12: Experimental setup of data collection, the signal of "Rx1", "Rx2", "Rx3" collected we labelled it as "D1", "D2", "D3"

Table 3.2: Recognition accuracy vs Activity type

	Accuracy (%)		
	D1	D2	D3
Walking	90.3	88.2	87.1
Jogging	78.6	72.4	79.3
Squatting	91.4	82.9	84.4
Chest expansion	83.9	82.4	96.2
Rising arms	84.4	89.5	83.3
Lunge press (Right)	85	82.8	91.2
High-step stretch	86.1	88.2	86.8
Lunge press (Left)	85.3	94.1	80
Average	85.8	85.4	86.8

3.5.2 Overall evaluation and Discussion

The overall classification results are composed of 3 pairs of Tx and Rx, which are calculated to be 85.8%, 85.4%, and 86.8% respectively. For simplicity, we classify the data collected from the different paths into 3 directions. For example, D1 represents the pair of Tx and Rx1, with the same for others. To evaluate the factors that influence system performance, we analyze the system with a control variable method of different direction paths and user identity.

Table 3.3: Recognition accuracy vs User identity

	Accuracy (%)				
	U1	U2	U3	U4	U5
Walking	85.7	95.2	77.8	95.5	87.5
Jogging	85	83.3	77.8	66.7	70.6
Squatting	83.3	95	73.7	95	84
Chest expansion	91.3	94.1	78.9	85	83.3
Rising arms	90.1	92	90	91.3	63.6
Lunge press (Right)	90.5	95	79.2	81.3	86.4
High-step stretch	92.6	100	83.3	93.3	70.4
Lunge press (Left)	80	77.8	93.8	92.3	87
Average	87.8	91.9	81.6	88.5	79.1

3.5.2.1 Influence from direction

From the accuracy comparison among the 3 directions, we observe that the performance of the classification system of D1 and D3 works slightly better than D2. In WiFi sensing, the proximity to the first Fresnel zone corresponds to higher sensing precision compared to areas farther away. It is the area theoretically closest to the unobstructed LOS path. Any objects or movements within this zone have a more significant impact on the signal, causing diffraction, reflection, or obstruction, thereby affecting the phase and strength of the signal that reaches the receiver. These changes are detected by the WiFi sensing system to identify the location, movement, and other environmental variables of objects.

Since WiFi sensing systems typically rely on minor signal variation to detect movements or changes in the environment, movements occurring within the first Fresnel zone result in more pronounced signal changes. These changes can be translated into more accurate sensing data. Therefore, movements near the first Fresnel zone are more critical for improving sensing accuracy. Conversely, movements occurring in farther Fresnel zones have a lesser impact on the signal and can lead to decreased precision in the sensing system. In this case, it is intuitive that the distance between D2's Fresnel zone and user location is larger than others, which is the cause of performance reduce.

3.5.2.2 Influence from user identity

In common sense, behaviour habits are diverse in distinct identities. To validate the generalization performance of our system among different persons, we invited 5 volunteers to participate in our experiments for data collection. The performance is shown in the Table. 4.1. Identities indeed influence the performance with a maximum gap of 12.8%.

3.5.2.3 Comparison of other HAR systems

We replicate and compare other three different WiFi HAR systems to validate the performance of our work: Widar3.0 [153], THAT [167], Dop_HAR [168]. The comparison results are illustrated in Fig. 3.13. The Widar3.0 system gets the highest HAR accuracy of 91.4% compared to ours CATM of 86.8% in D3. However, to adopt Widar3.0 system, there should be at least 3 WiFi receivers for establishment. In our system, only one receiver is needed. From the comparison results with other two systems, our CATM system gains the highest classification accuracy.

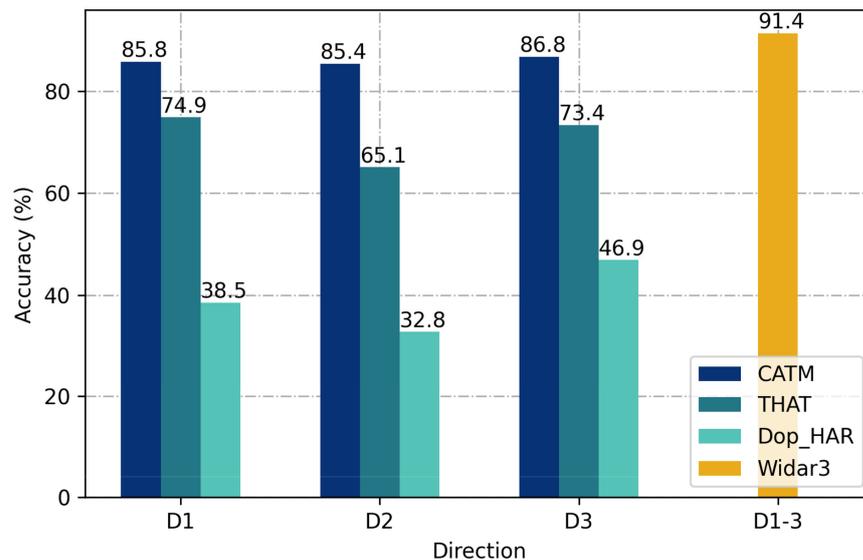


Figure 3.13: Accuracy comparison of different WiFi HAR systems in 3 directions

3.5.2.4 Comparison of CATM and NT-CATM

In our comparison test, we reproduce the NT-CATM described in Section 3.4.2. The results in Fig. 3.14 validate our assumption that negative ToF values in CATM contain the effective information related the human behaviours.

3.5.3 Section Conclusion

This paper proposes a CTAMs based system for human activity monitoring in the indoor environment, which adopts spatial phase variation information of CSI. In this system, we firstly analysed the noise components in raw CSI data and leverage various techniques to filter the phase offset of CSI information for data sanitation. Then, we extracted the features with continuous AoA-ToF maps. Finally, we established a temporal neural network for the HAR task and achieved 85.6% accuracy among 5 volunteers and 8 activities.

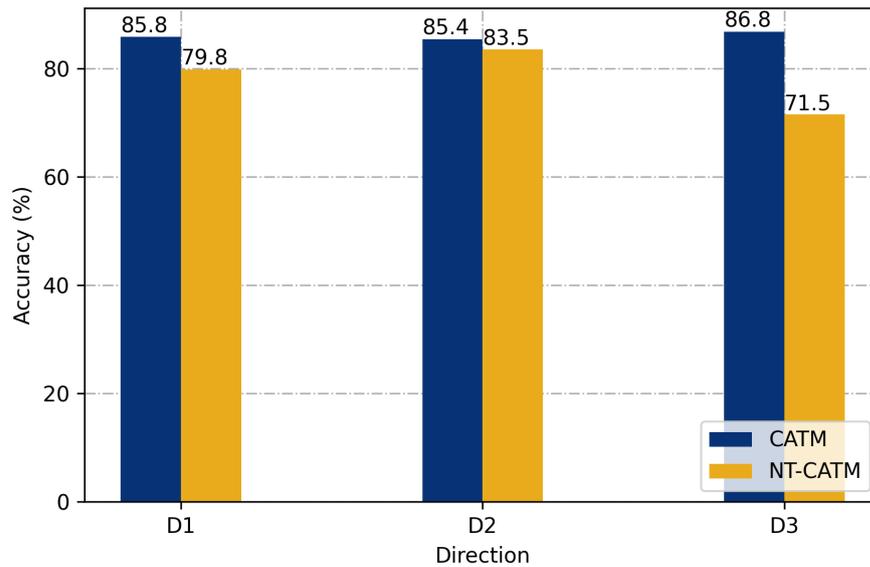


Figure 3.14: Accuracy comparison of CATM and NT-CATM in 3 directions

The comparison results demonstrate the ubiquitous performance under the conditions of different directions and user identities. Meanwhile, from the comparison tests with other WiFi HAR systems, we get the highest accuracy in the systems considering single WiFi receiver.

3.6 Summary

This chapter delves into the use of WiFi sensing for HAR within intelligent IoT and health-care monitoring systems, emphasizing the importance of indoor activity recognition. It reviews the evolution of device-free WiFi sensing technologies, focusing on methods to improve accuracy and efficiency in single-device activity recognition. Innovations like the Continuous Angle of Arrival-Time of Flight maps, which preserve temporal and spatial information of human bodies through WiFi signals, are explored alongside the development of specialized temporal neural networks for these systems.

The chapter meticulously outlines experimental setups, calibration techniques, and system designs. It details the use of sanitized CSI phase for respiration estimation, evaluates Doppler based recognition systems, and discusses the construction and assessment of AoA-ToF maps. Results show the effectiveness of these methods in accurately detecting and recognizing human activities, with comparisons drawn to other HAR systems to highlight advantages and areas for improvement.

Concluding, the chapter emphasizes advancements in HAR using single-end WiFi, particularly the integration of spatial phase variation from CSI data. This progress in single terminal contexts, enriched with spatiotemporal data, has significantly enhanced HAR's effectiveness, paving the way for future applications in smart-home design and healthcare monitoring.

LoRa based Remote Sensing of Human Gait Identification

4.1 Introduction

Gait recognition utilizing communication signals traditionally grapples with limitations in detection range, a challenge particularly pronounced in WiFi-based sensing systems. This chapter focus on the introduction part of the proposed LoGait system, which push the gait recognition to the range of $20m$ compared to the previous WiFi scheme of WiDIGR [16] of $5m$ and CAUTION [17] of $7m$. Furthermore, the LoRa signal utilized in our system demonstrates enhanced penetration capabilities through obstacles, making it particularly suitable for indoor through-wall sensing, which can be challenging to achieve with WiFi signals. For UWB technique, the work of [19] proposes a detailed analysis of gait training with depth camera. However, the detected space is also limited to a $3m$ size chamber, which is challenging to demonstrate the applicability of this system to general scenarios. [169] presents a gait recognition system based on RFID technology, utilizing 8 RFID antennas and 6 tags for short-range human gait detection within distances ranging from 0.9m to 2.2m. While their RFID sensing system has demonstrated effectiveness in various scenarios, its deployment method incurs higher costs compared to our LoGait system, which utilizes a single transmitter and two receiver antennas to cover a range of $20m$ in long-distance environments. Meanwhile, the bandwidth of our LoRa methods only take up $125kHz$, which is significantly lower the UWB of $500MHz$ in minimum and WiFi of $300MHz$ in common 802.11n protocol that CSI-tool [170] adopted. Conserving bandwidth resources allows for reliable connectivity and a better network experience for large-scale deployments of IoT applications. The previous works of gait recognition in JCAS field are listed in Table 4.1 respectively.

Table 4.1: Review of RF sensing works with comparison of our proposed work

Reference	Protocol	Carrier Frequency	Bandwidth	Application	Experimental Setup Range / Subjects	Performance
CARM [14]	WiFi	5 GHz	20 MHz* 30 subcarriers	human activity recognition	7.7*6.5M / 25 subjects	8 activities recognition with over 96.5% accuracy in average
HARRNN [15]	WiFi	5 GHz	20 MHz* 30 subcarriers	human activity recognition	indoor 5*6m, 8*6m/10 subjects	6 activities recognition with over 95% accuracy in average.
WIDIGR [16]	WiFi	5.825 GHz	20MHz* 30 subcarriers	gait recognition	5*5m / 60 subjects	78.28% accuracy for 6 subjects recognition
CAUTTON [17]	WiFi	5 GHz	40MHz* 114 subcarriers	gait recognition	5.8*6.3 m 7.2*5.2 m/ 20 subjects	88% accuracy in 15 subjects identification
[19]	UWB	4.3 GHz	2 GHz	gait recognition	3m in chamber / Not mentioned	normal and spastic gait recognition with 94.9% accuracy
[20]	FMCW and UWB radar	25 GHz FMCW / 7.5 GHz UWB	2GHz FMCW / 1.5GHz UWB	gait recognition	2.7*1.8m / 14 subjects	Gait recognition in 14 subjects with 84% accuracy in average
[21]	UWB	5.2 GHz	8.7 GHz	human activity recognition	2.5m / 13 subjects	12 non-in-situ motions recognition with an average of 88.9% accuracy, in-situ motions 89.7% average accuracy.
[22]	UHF RFID	865 MHz	3 MHz	gestures recognition	2.5m / 15 subjects	87% accuracy for recognition of drinking episodes for young volunteers and 79% for older volunteers
[23]	LoRa	915 MHz	125 kHz	respiration sensing / human tracking	25m(respiration) / 35m(tracking)	Achieve long-range through-wall respiration sensing with 0.25bpm mean absolute error, human tracking with 4.27cm average absolute error.
LoGait (Ours)	LoRa	868.1 MHz	125 kHz	gait recognition	5m, 20m (LOS) & 6m(NLOS) /13 subjects	Adopt LoRa signals to extend the gait recognition in various environment including 20m corridor, with 82.8% accuracy.

4.1.1 Contributions

The problem of recognising humans from their walking patterns is known as gait recognition. It has many potential applications in surveillance, healthcare, and human-computer interaction. In the literature, WiFi based sensing has provided various solutions for gait recognition [171], [172]. However, the nature of the WiFi signal restricts the range of detection [23], [173]. Therefore, WiFi based systems cannot be adopted in long-range spaces, e.g., corridors. Inspired by the previous LoRa based sensing work [24], [174]–[176], we attempt to analyse the feasibility of adopting LoRa sensing in long and narrow environments. We propose LoGait system to push the range of gait recognition to multiple indoor scenarios including a corridor of $20m$ length. The proposed LoGait system is shown in Fig. 4.1. In summary, the following are the major contributions of this paper.

1. To the best of our knowledge, this study is the first attempt toward using LoRa signals for gait recognition in a 20 meters range.
2. We propose a pipeline for performing preprocessing of LoRa signals for gait feature extraction and classification.
3. We collect LoRa signals containing the gait patterns of different subjects in various scenarios. Our experimental results in different settings validate the effectiveness of our proposed LoGait system in performing gait recognition. Our work fills an absence in this field.

The rest parts of the paper are structured as follows: Section 4.2.1 introduces the preliminary analysis of the LoRa signal. Then Section 4.2.2 describes the detailed methods for LoRa gait feature extraction, and dynamic time warping (DTW) based classification. Section 4.3 presents the evaluation of the proposed LoGait system in different scenarios.

4.2 Methodology

4.2.1 Preliminaries

Unlike WiFi which applies OFDM to divide channel bandwidth into different subcarriers, LoRa adopts full bandwidth for Chirp Spread Spectrum technology, which encodes information on radio waves using chirp pulses. The technology operates in a fixed-bandwidth channel of 125 kHz for up-link. The frequency of linear chirps increases from $f_c - \frac{B}{2}$ to $f_c + \frac{B}{2}$ over the sampling period of $-\frac{T_s}{2} < t \leq \frac{T_s}{2}$, where the f_c and T_s represent carrier frequency and sampling time, respectively. The exponential representation of LoRa transmitting signal is composed of two elements, chirp signal and carrier frequency modulation:

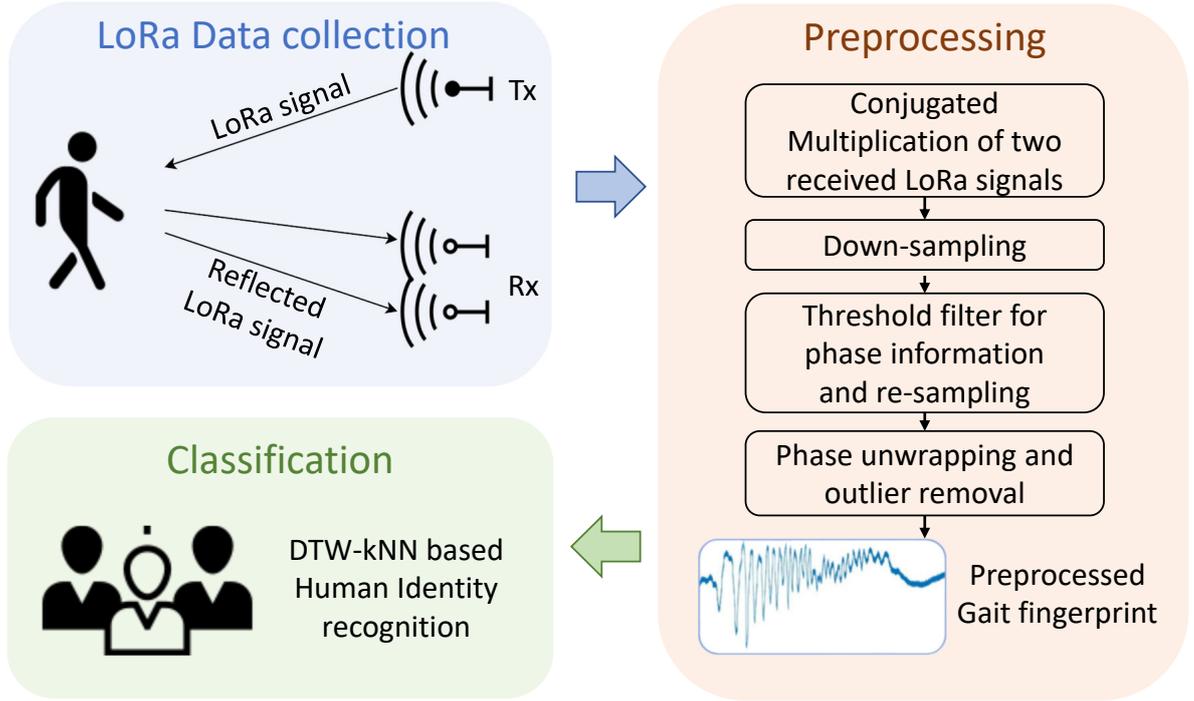


Figure 4.1: Overview of proposed LoGait system that consists of three components: (1) LoRa Data Collection; (2) Preprocessing; and (3) Classification using DTW and kNN.

$$\mathbf{T}\mathbf{x}(t) = e^{j\pi f_s(t) + j2\pi f_c t}, \text{ with } f_s(t) = \pm \frac{B}{T_s} t \quad (4.1)$$

where, $f_s(t)$ represents the chirp signal with sweep rate. Existing literature on LoRa based sensing suggests that the channel response at the receiver end can be represented without considering the chirp signal [23]:

$$H(t) = \frac{R\mathbf{x}(t)}{T\mathbf{x}(t)} = e^{-j2\pi\Delta t} (H_s + H_a(t)) + N(t) \quad (4.2)$$

where, $e^{-j2\pi\Delta t}$ is due to the SFO and CFO; H_s and $H_a(t)$ represents the LoRa signals from the time-invariant static paths (including the signals in LOS path and those reflected off the stationary objects) and time-variant dynamic paths (including signals reflected from the dynamic objects). $N(t)$ represents the free space transmission noise. LoRa signals in active paths can be expressed as:

$$H_a(t) = \sum_{i=1}^{N_d} a_i(t) e^{-j2\pi \frac{d_i(t)}{\lambda}} \quad (4.3)$$

where, N_d is the index of path that signal passes through, $a_i(t)$ represents the complex attenuation factor of the i^{th} path; $e^{-j2\pi\frac{d_i(t)}{\lambda}}$ represents the phase change of i^{th} path, with the changing distance of $d_i(t)$ in i^{th} path. λ represents the wavelength of the LoRa signal.

However, the channel response cannot be calculated directly with reference data. In this case, we replicate the setup from previous work with two receiver antennas to get the conjugate multiplication (CM) signal [23], [177]. There are various parameters that correlate with signal ratio. However, to find the dominant dynamic path for estimation, we select two directional antennas which perform better in the reduction of the multipath effect ($N_d = 1$). So we assume in the ideal situation, that there is a single path with relatively less noise ($d_i(t) = d(t)$).

$$\begin{aligned}
\mathbf{R}_{\text{CM}}(t) &= \mathbf{R}_{x_1}(t)\overline{\mathbf{R}_{x_2}(t)} = \mathbf{T}_x(t)\overline{\mathbf{T}_x(t)}H_1(t)\overline{H_2(t)} \\
&= \mathbf{T}_x(t)\overline{\mathbf{T}_x(t)}(e^{-j2\pi\Delta ft}(H_{s1} + H_{a1}(t)))(e^{j2\pi\Delta ft}(\overline{H_{s2}} + \overline{H_{a2}(t)})) \\
&= \|\mathbf{T}_x\|^2(H_{s1} + H_{a1}(t))(\overline{H_{s2}} + \overline{H_{a2}(t)}) \\
&= \underbrace{\|\mathbf{T}_x\|^2}_{(1)} \underbrace{(H_{s1}\overline{H_{s2}})}_{(2)} + \underbrace{H_{s1}\overline{H_{a2}(t)}}_{(3)} + \underbrace{\overline{H_{s2}}H_{a1}(t)}_{(4)} + \underbrace{H_{a1}(t)\overline{H_{a2}(t)}}_{(5)}
\end{aligned} \tag{4.4}$$

From the representation, the components of chirp signal, CFO and SFO are removed. On the other hand, the equation is divided into five parts for analysis: the transmission part of (1) and product of static components of (2) can be regarded as constant value, and the product of active components of (5) is small that can be ignored. Meanwhile, we consider the extended changing path, ΔS , which is caused from the different physical locations of two receiver antennas. Meanwhile, for a short time duration the path attenuation factor can be regarded as static value. This value is assumed as the constant value due to the setup receiver antennas are close to each other. Next, we can rewrite the superposition of rest components in Eq. 4.5.

$$\begin{aligned}
(3) + (4) &= H_{s1}\overline{H_{a2}(t)} + \overline{H_{s2}}H_{a1}(t) \\
&= H_{s1}(a_2e^{j2\pi\frac{d(t)+\Delta S}{\lambda}}) + \overline{H_{s2}}(a_1e^{-j2\pi\frac{d(t)}{\lambda}}) \\
&= (H_{s1}a_2e^{j2\pi\frac{\Delta S}{\lambda}})e^{j2\pi\frac{d(t)}{\lambda}} + \overline{H_{s2}}a_1e^{-j2\pi\frac{d(t)}{\lambda}}
\end{aligned} \tag{4.5}$$

The exponential form can be converted to trigonometric form using Euler's formula and then they can be added together. To summary all, the approximate CM result can be represented as Eq. 4.6. Meanwhile, we replace the representation of $A = H_{s1}a_2e^{j2\pi\frac{\Delta S}{\lambda}}$, and $B = \overline{H_{s2}}a_1$.

$$\begin{aligned}
\mathbf{R}_{\text{CM}}(t) &\approx \|Tx\|^2 (Ae^{j2\pi\frac{d(t)}{\lambda}} + Be^{-j2\pi\frac{d(t)}{\lambda}}) \\
&\approx \|Tx\|^2 ((A+B) \cos(2\pi\frac{d(t)}{\lambda}) + j\|Tx\|^2 ((A-B) \sin(2\pi\frac{d(t)}{\lambda}))
\end{aligned} \tag{4.6}$$

In this case, we can conclude that both amplitude and phase variation of CM result can be influenced by components from the dynamic path of $d(t)$, which is available to be adopted for gait feature extraction. Experimental validation in Section 4.2.4 proves the above conclusion.

4.2.2 Preprocessing methods of LoGait system

The inference provided in Section 2.5.3 shows that human motion can be indicated by the phase variance of CM results. However, to extract any relevant data related to the motion, it is necessary to go through a series of preprocessing steps before using the Dynamic Time Warping (DTW) recognition. This section outlines all of the steps that we proposed which need to be taken prior to the DTW recognition. All experiment implementation is detailed, introduced in Section III.A.(1).

4.2.2.1 Conjugated multiplication of two antennas signal

The Fig. 4.2 shows the different amplitude of raw LoRa signals with/without dynamic physical interference. Although the envelope shape of LoRa amplitude is explicit, it is required to transform these perturbations into measurable values. Compared to visible variation from amplitude information, phase information shown in Fig. 4.3c provides random information that cannot be intuitively observed.

In the next step, we observed a considerable number of blank interpolations inside the received envelope, which represents the receiver end collected noise during packet duration. Removal of this blank information is important to extract accurate gait features. Firstly, we calculated the CM result from raw LoRa signals of dual antennas, which is shown in Equation. 4.4. Demonstrated by the Equation, the gain of $\|Tx\|^2$ can determinately increase the amplitude of the received signal, which differentiates the meaningful LoRa signal from noise. From this point of view, the noise duration can be removed by setting the low amplitude threshold of CM result. In Fig. 4.3a, the red dashed line represents signal components from noise and the green one from the LoRa chirp signal. Meanwhile, we assume the transmitted power of LoRa signals and free space attenuation is stable. In this case, the threshold was set to the mean value of the first second's receiving signals.

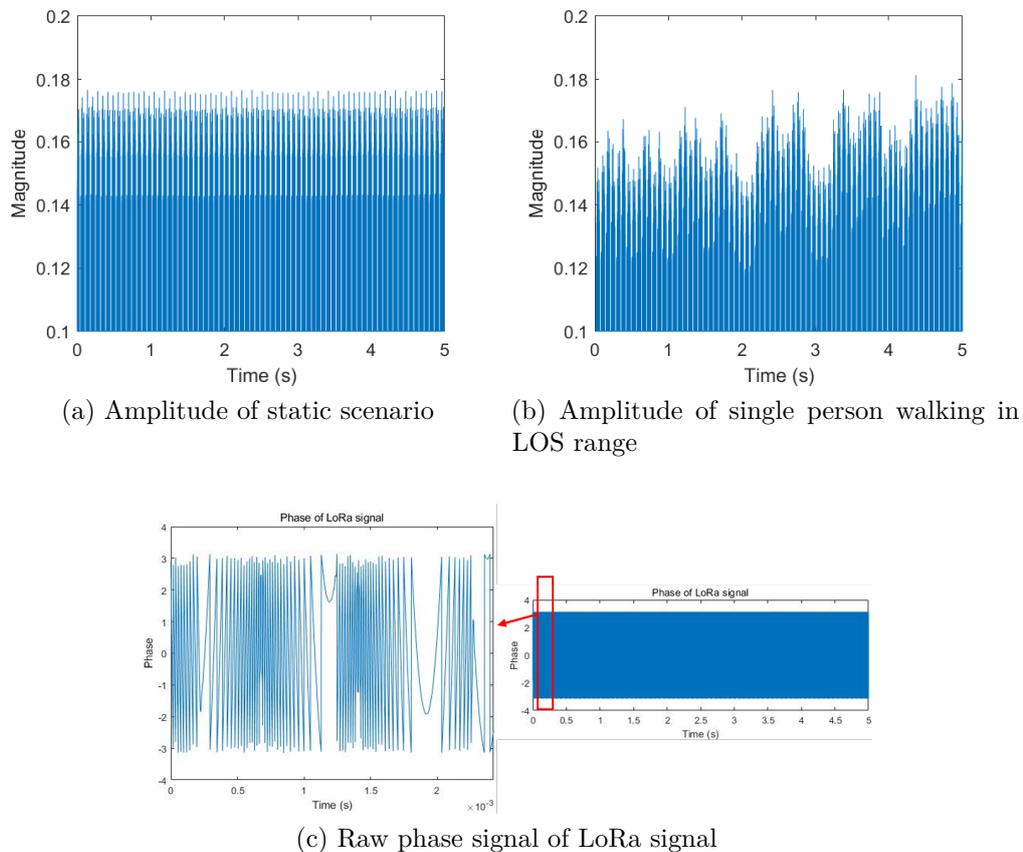


Figure 4.2: Amplitude and phase plots of raw LoRa signals.

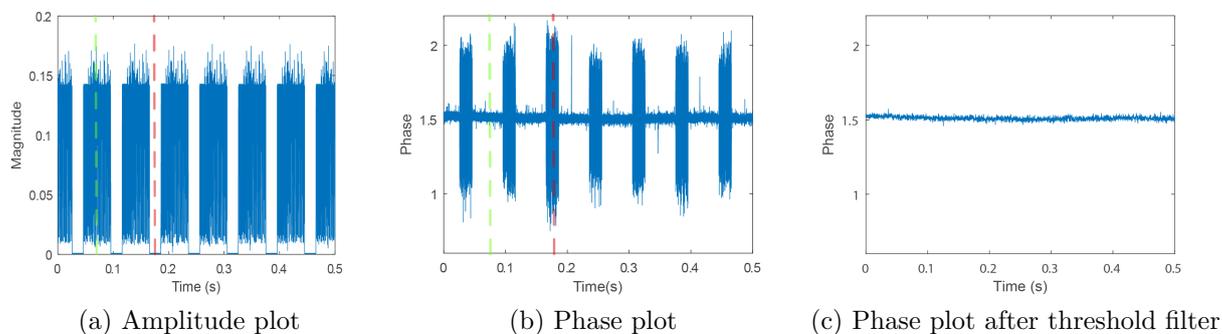


Figure 4.3: Plot of CM results under static environment, with the red dashed line labelled for noise components and green one for LoRa chirp signals in (a) and (b).

4.2.2.2 Threshold filter for phase information and down-sampling

Meanwhile, there are two downsample operations executed before and after threshold filtering. For amplitude threshold, it is a waste of computational resources for searching and comparing all $800k$ samples per second. On the other hand, the Doppler shift frequency range that human activity can generate is limited to $60Hz$ [178], $800kHz$ sampling rate is highly redundant. Therefore, we set the first downsample of $1k$ sample rate ahead of threshold operation. In practical scenario, there must be a silent duration while transmitting chirp signals to reduce the power consumption and avoid interference among different packages by multipath effect, which is called Inter-Symbol Spacing (ISS). However, the receiver side is not able to differentiate if received samples belongs to ISS. In this case, we adopt an adaptive thresholding calculation method to filter the ISS information out of gait profile. At first we statically sorted the non-zero value of CM results and pick the smaller components of whole CM sequence based on the percentage that ISS occupies. The priority assumption is that under silent conditions, the receiver can only get a noisy signal with a stable amplitude and a spurious phase, which are shown in Fig. 4.3a and Fig. 4.3b. By calculating the mean and standardized variation of estimated ISS information, we can determine the value of threshold. To make it more generalised, the value is set to the mean plus double standardised value. After performing thresholds filtering, we discovered that the length of the LoRa chirp signal is not constant. Then, we resample the filtered data to $1kHz$ for unifying constant sampling frequency among different profiles. Fig. 4.3c shows the LoRa signal after threshold filtering with twice downsampling.

4.2.2.3 Phase unwrapping and outlier removal

In this stage, we acquire meaningful phase information that can reflect the channel environment. However, the outliers and mismatched phase data appear. We adopt the Hampel filter and unwrap operation to denoise the signals, with the shown comparison figures shown in Fig. 4.4. The signal components that are framed out by a red box represent the recovered parts by the methods.

4.2.3 Gait analysis

Considering the ISS, we only adopt the phase of CM sequence for gait analysis. According to the Eq. 4.6, the phase is given as follows:

$$\Delta\phi_{RCM} = \|Tx\|^2(A - B)\sin(2\pi\frac{d(\Delta t)}{\lambda}) \quad (4.7)$$

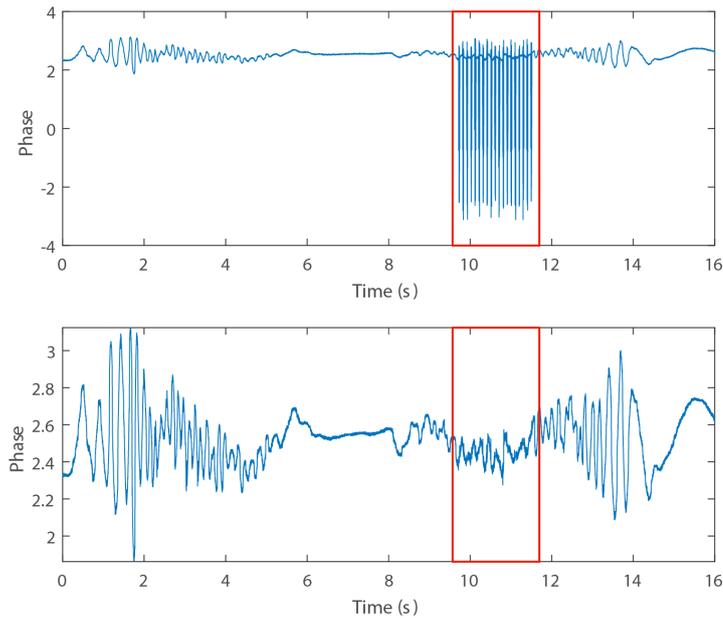


Figure 4.4: Comparison plots with/without Hampel filter and phase unwrapping, with circled outliers in red boxes. The first graph shows original gait signals and the second shows Hampel filtered and phase-unwrapped gait signals.

where $\|Tx\|^2(A-B)$ can be treated as constant. Then we take the derivative of this equation, the period of the phase change is related to the dynamic path $d(t)$. In this case, the system can be considered as a monocular radar system, which represents the path length of moving target should be considered twice of varying distance: $d(\Delta t) \approx 2\Delta d$. In this case, we can get the $\Delta\phi_{RCM} \propto \sin(4\pi\frac{\Delta d}{\lambda})$. By leveraging the periodic nature of the sinusoidal signal, we can deduce the distance the monitored object has traversed by inverting the observed phase period. We name the periodicity of ϕ_{RCM} as T , and get $T = 2\frac{\Delta d}{\lambda}$. Meanwhile, the wavelength of LoRa signal is fixed to $\lambda = c/f = \frac{3 \times 10^8}{868.1 \times 10^6} = 0.346m$. From the equation, we can get the moving distance of target is related to the periodicity of phase variation, which is $\Delta d = 0.173T$. Then, to verify the above derived equation formula, we collected moving data from a humanoid robot and a human volunteer moving $2m$ in fixed area respectively, shown in Fig. 4.6. All the peaks are labelled with dotted line. In the robot moving profile as Fig. 4.5 shows, we observed 11 peaks and 10 periods of signals, which represents the minimum of moving distance of robot reaches $1.73m$. Considering the start and end period are not properly counted, the result is close to $2m$. However, instead of robot profile, the human profile illustrates 14 peaks, which represents the distinct motion beyond $2m$. Meanwhile, we discover that there are some fluctuated peaks having less amplitude variation than torso motion, which is labelled in red block. Excluding the possibility of environmental influences, this part of the correspondence could only be generated by the movement of the limbs. Therefore, in the robot scenario, the phase change only reflects by $2m$ displacement. In the human scenario, the phase change is caused mainly by both displacement of body torso motion and waving limbs.

Although the processed signal contains all gait information of subject, this information is difficult to quantified and modelled accurately due to the complex dynamical structure of realistic human gait. However, there is a prerequisite that different human's gait schemes are different [179]. By constructing a database of LoRa signals generated by the gait of different volunteers, we can use statistical based machine learning methods to find differences in the gait of identity.

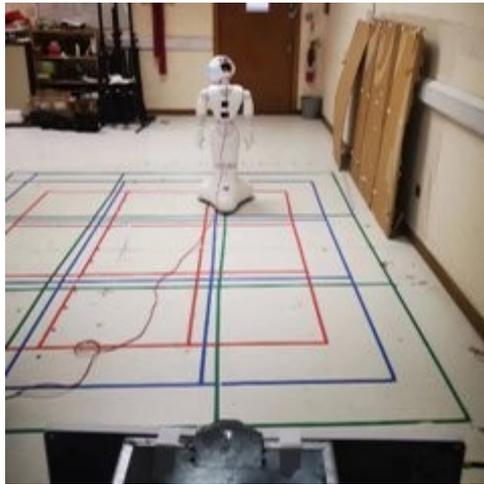


Figure 4.5: Picture of humanoid robot motion.

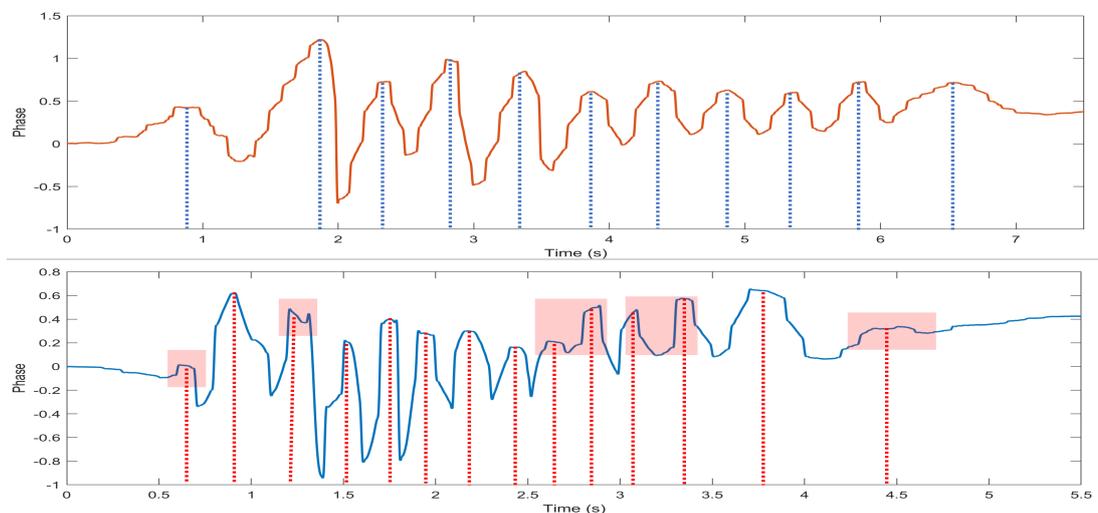


Figure 4.6: Gait profiles of humanoid robot motion (first graph) and human volunteer (second graph).

4.2.4 DTW based Gait Recognition

After preprocessing of LoRa signals, we compared the signals collected from different activities and gait signals, shown in Fig. 4.7. The gait experiment setup is the same as LOS experiment that mentions in Sec. 4.3.1.

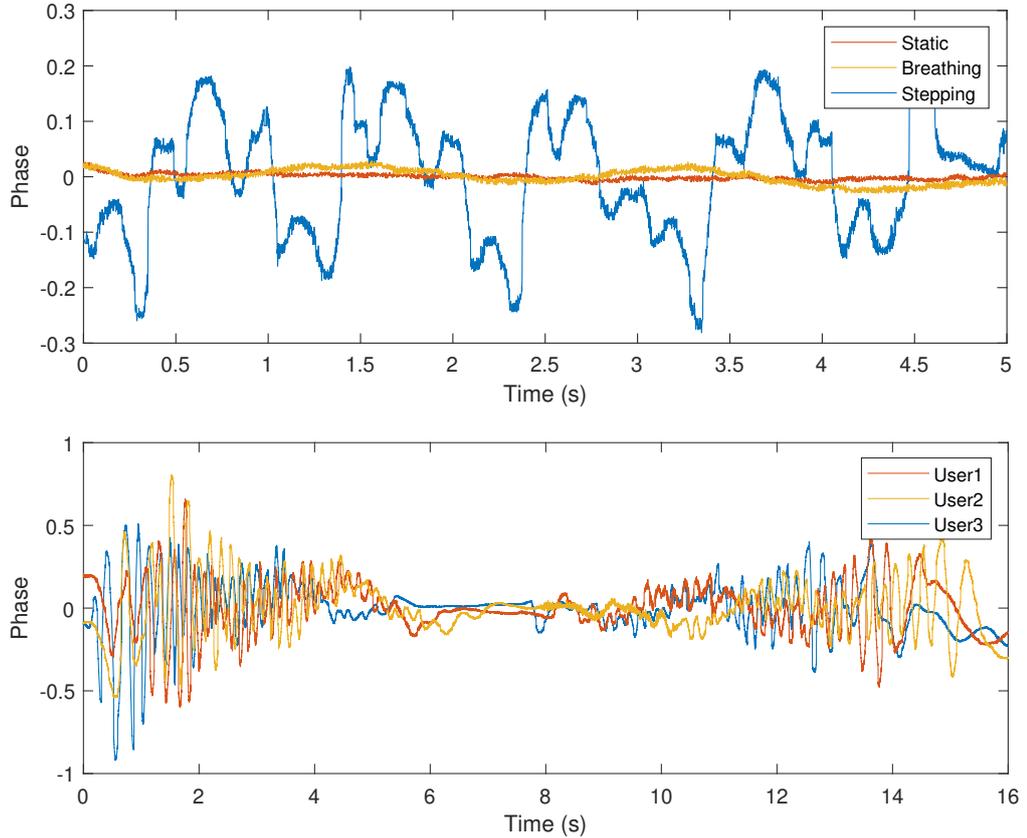


Figure 4.7: Comparison of LoRa signals under different scenarios of no-person, a person standing still, a person stepping shown in the first figure; and gait profiles of three users in the second figure.

From the intuitive view, we observed the collected activity signals under three different scenarios that match our normal experiences: the human presence scene only contains chest motion of respiration, and stepping signals contains multiple signal peaks from human skeleton motion. From the gait profiles of three users, we found that the human behaviour patterns of different identities are highly overlapping compared to human activity recognition. Besides, there are two main challenges that were observed in the comparison:

1. Variation of motion speed can result in various lengths of gait signals from a single subject.
2. Temporal gait signal can not be completely aligned while the data collection, which causes the distortion of information.

The general method of measuring the similarity of two-time series signals is to calculate the Euclidean distance. However, lock-step Euclidean distance measurement refers to those distances that compare the *i*th point of one series to the *i*th point of another, which is significantly influenced by incomplete alignment [180]. To solve the alignment problem and improve the recognition performance, DTW based method was adopted.

DTW is a similarity measurement method, which exhausts all the correspondences with restrictions and finds the one with the smallest distance. Then the cumulative distance of the selected path is used for their similarity judgement. The Equation. 4.8 describes the algorithm of DTW distance.

$$D_{min}(i, j) = M(i, j) + \min \begin{cases} D(i-1, j-1) \\ D(i-1, j) \\ D(i, j-1) \end{cases} \quad i, j \geq 1 \quad (4.8)$$

To align the two sequences, a matrix with two dimensions of sequences' length is required. The matrix element $M(i, j)$ denotes the Euclidean distance $d(x_i, y_j)$ between the two points x_i and y_j . The shortest distance of the current element $D_{min}(i, j)$ is necessarily the length of the shortest path of the previous element plus the value of the current element. There are three possible directions for the previous element, so we take the minimum value of three possibilities into DTW distance.

For initial validation of DTW, we calculated the DTW distance of gait profiles of different identities and the same identity respectively to verify the algorithm's availability. The comparison graph is shown in Fig. 4.8 which illustrated the distance between the same user and different users. The larger DTW distance verified our assumption that the gait profiles of different identities have mismatched information, and the data from a single identity has similar features. Therefore, we adopted a KNN cluster based algorithm to classify different identities of gait signals.

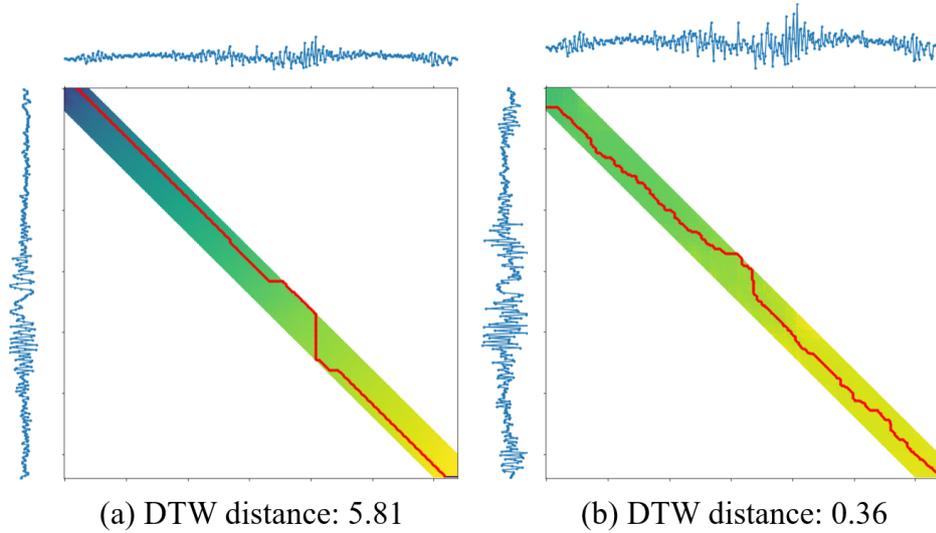


Figure 4.8: Comparison of DTW distance matrix of (a) gait profiles from the first user and second user; (b) gait profiles from the first user.

4.3 Evaluation

4.3.1 Experiment Setup

4.3.1.1 Devices

Our implementation considers one pair of devices to imitate the general LoRa link. We select one USRP b205mini and one USRP x310 as transmitter and receiver, respectively, shown in Fig. 4.11. On the transmitting side, we have opted for a directional linear polarized antenna of Aaronia Ag. This type of antenna emits signals in a specific direction, allowing us to focus the transmission towards the target area, namely the long corridor under observation. The directional nature of this antenna ensures that the transmitted signals are concentrated on a desired path, enhancing the accuracy of gait information collection. On the receiving side, we have implemented two antennas of SlimLine A5010 Circular Polarized Antenna with 8.5 dBi gain. Circular polarized antennas are known for their ability to capture signals from various polarization angles, making them suitable for scenarios where the incoming signals may have different polarization orientations. By utilizing a circular polarized antenna, we can effectively receive the transmitted signals, regardless of their polarization alignment. Moreover, our choice of directional and circular polarized antennas offers an advantage in mitigating the impact of multipath propagation. With an omni-directional antenna, signals can bounce off obstacles and create interference due to multipath reflections. However, by utilizing directional antennas, we can minimize the reception of reflected signals and primarily capture signals from the intended pointing direction. This helps to reduce the effects of multipath interference, improving the reliability and accuracy of the gait information obtained. The LoRa signal is generated by an

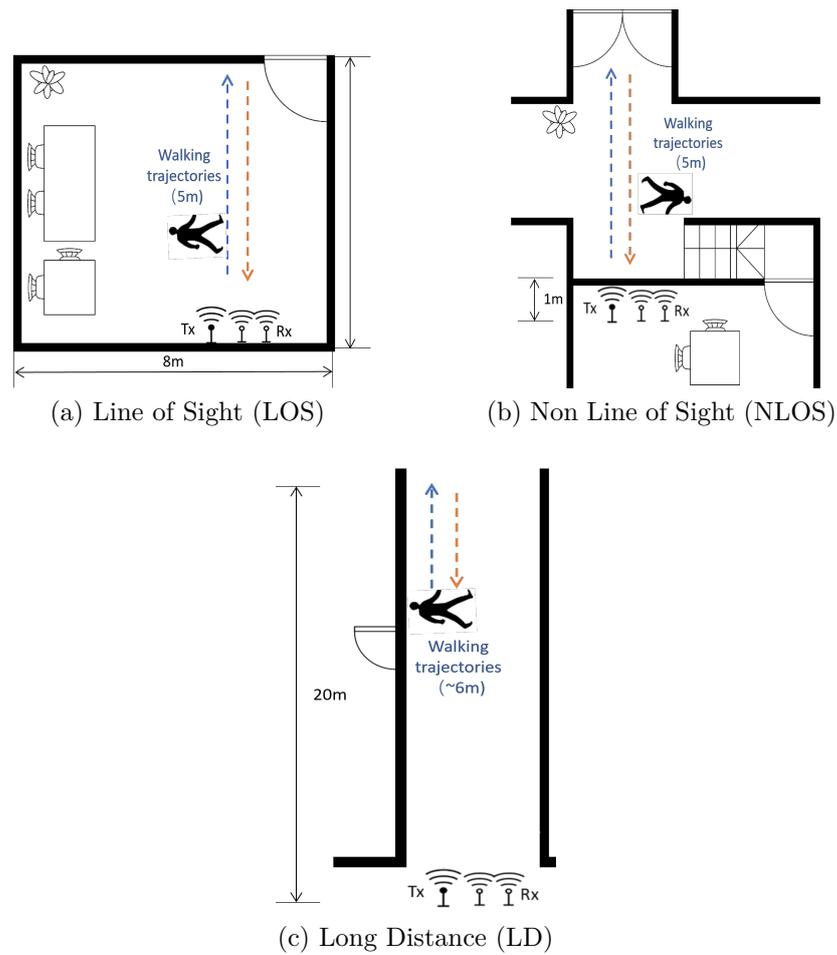
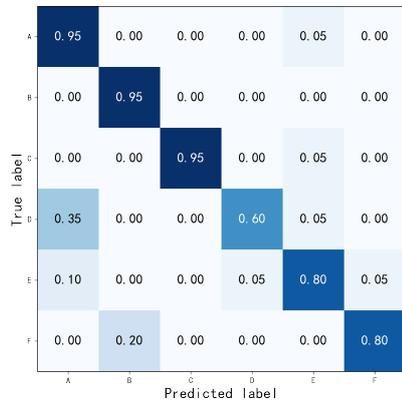


Figure 4.9: Experimental setup of 3 scenarios.

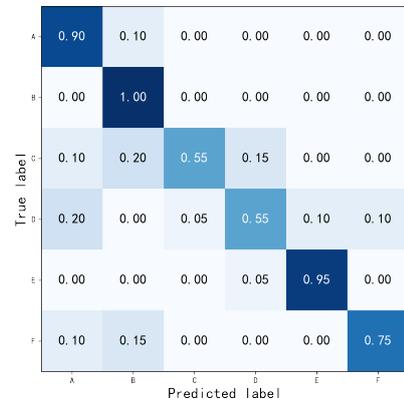
open-source project of LoRa communication in the physical layer [181]. On the receiver side, we configured Labview based system to collect LoRa signal. The experimental setup can be viewed in Fig. 4.12. The sampling rate and packet duration are set to $800kHz$ and $20ms$, respectively.

4.3.1.2 Application scenarios

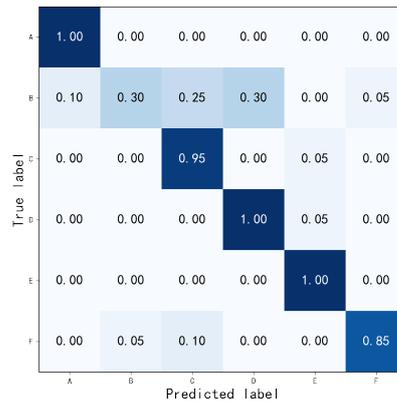
The experiments were conducted in 3 different scenarios: line of sight (LOS), NLOS and long distance (LD) range, with the top-view structure graph shown in Fig. 4.9. In the LOS scenario, the room area occupied for the activity experiment is $6m$ in length and $5m$ in width. One line has marked a $5m$ distance on which people walk to and from the front of the transmitter and receiver. In NLOS scenario, all apparatus containing the transmitter and receiver are in one room, and activity is being monitored outside the room, shown as walking trajectories. The space between devices and humans is separated by a brick wall. In the LD scenario, the implementation is setup in a corridor of $20m$ length. Volunteers were arranged to walk along the trajectory at the end of the corridor, shown in Fig. 4.12.



(a) LOS



(b) NLOS



(c) LD

Figure 4.10: Confusion matrix of gait recognition in 3 scenarios



(a) USRP b025mini



(b) USRP x310

Figure 4.11: Devices of LoRa signal modulation & demodulation sources

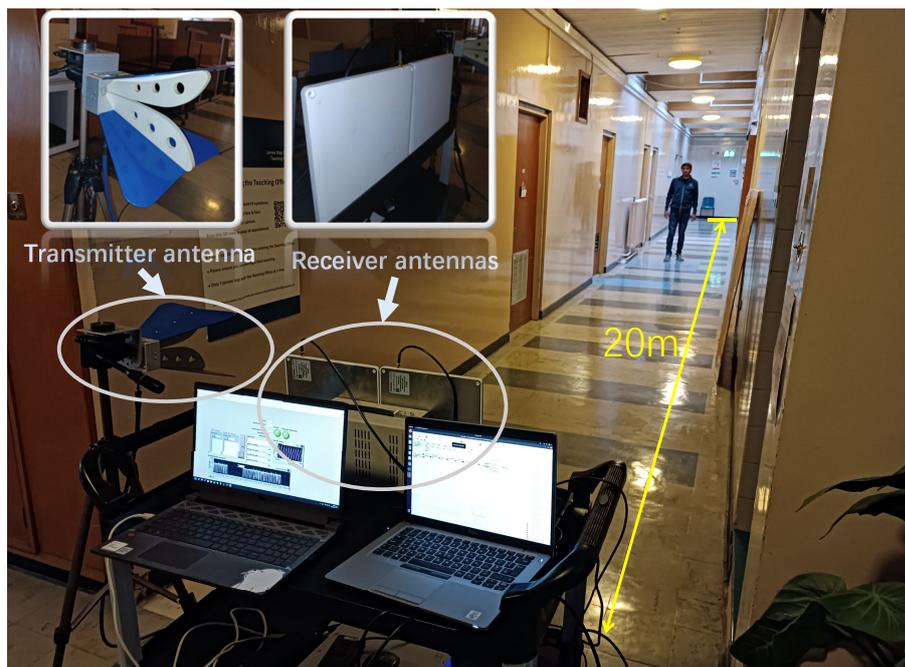


Figure 4.12: Experimental setup in Long distance scenario

4.3.1.3 Gait profiles

The basic human gait in our experiment contains three phases: rotating, walking and standing. In this study, our focus was primarily on the identification of gait patterns in single-person walking scenarios. Meanwhile we discuss the challenges posed by multiple walking scenarios in Sec. 4.3.2.6 of the paper. During one data collection, one person was asked to turn back, and walk along the trajectory and stand still for 8 seconds. Then, it took another 8 seconds for subjects to walk back to the starting point. The gait signal in each profile lasts for 16 seconds in total. Besides, we downsample the signals to 200Hz for speeding up the machine learning algorithm.

We recruited 13 volunteers for data collection of human gait including 4 females and 9 males. This human involved research has got the ethic approval from College of science and engineering from University of Glasgow, approval no: 300210309. In each scenario, we ask 6 subjects to conduct the experiments and data was collected for 20 rounds for each person that provides 16s gait data. In total, we have collected 5744 seconds of gait signals for experimental validation.

4.3.2 Overall performance and Discussion

The recognition performance of 3 scenarios with confusion matrices is shown in Fig. 4.10. We perform a 5-fold cross-validation on collected data with an overall accuracy of 85.13% in LOS range, 79.13% in NLOS and 84.14% in LD respectively. The average accuracy of 82.8% validates the effectiveness of our system. To study the performance of the LoGait system affected by different factors, we design the comparison tests and analyse the influence of Sample rate, Classification distance algorithm. Meanwhile, we explain the difference in Data collection methods of gait signal between LOS/NLOS and LD scenarios.

4.3.2.1 Sample Rate

The sample rate of gait profiles is the significant parameter that balance of recognition performance and processing speed. We resample the gait profiles from 10Hz sample rate to 500Hz to test the performance of the LoGait system. The recognition performance is illustrated in Fig. 4.13, which validates that a higher sample rate preserves more gait information.

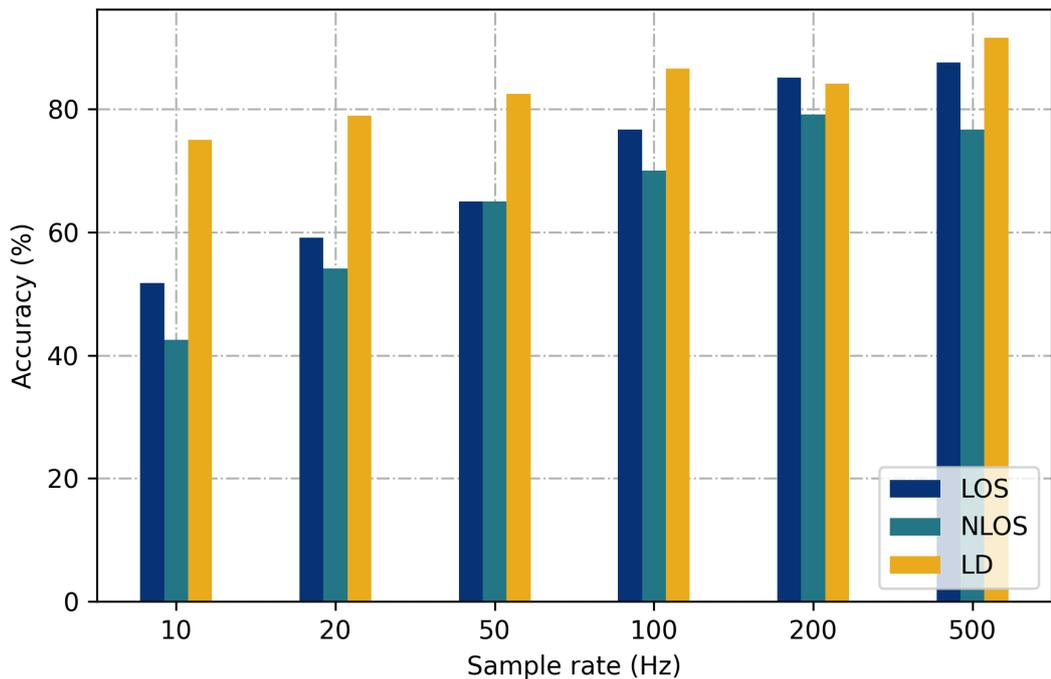


Figure 4.13: Recognition accuracy via sample rate in 3 scenarios

4.3.2.2 Vector distance algorithm

To approve the assumption in Sec. 3.4.2, we compared the classification results using the other three distance algorithms with DTW distance, shown in Fig. 4.14. It illustrated that the DTW based classification method acquires the best performance among the traditional distance estimation algorithms.

4.3.2.3 Walking directions

In our experiment, we divided the data collection of one profile into two phases, moving forward and back, as shown in Fig. 4.9 (depicted by blue and red dashes). To effectively evaluate the performance of our system, we separated the profile into three parts: forward, back, and integration.

4.3.2.4 Data collection methods

In both LOS and NLOS scenarios, we asked volunteers to do free walking in a restricted space. In this case, the users are easy to control their speed without following usual habits. To study the robustness of the gait recognising system, in the LD scenario, the volunteers were asked to walk freely in a given time slot instead of limiting the moving area. From the results, the accuracy under the LD environment approves the popularization potential of the LoGait system.

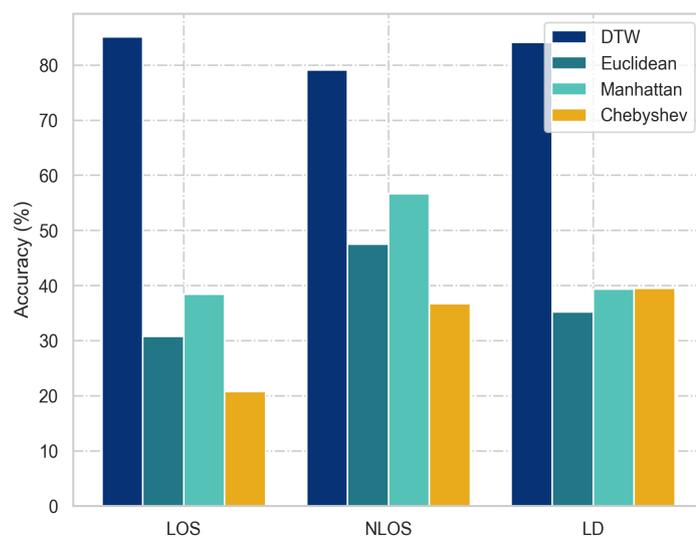


Figure 4.14: Recognition accuracy of different distance algorithms for classification in 3 scenarios: LOS, NLOS, LD

4.3.2.5 Strength influence

In this section we complement the LoGait identification performance on different signal strength condition by analysing human and moving robot gait profiles under the same environment. From the Eq. 4.6, we know the CM result is significantly influenced by transmission gain of communication system. In the absence of environmental changes, the noise in the received signal will remain stable, but resulting in varying gain of human related moving signals due to changes in transmission gain. Intuitively, we assume that lower signal strength will diminish the stylized components of the gait, leading to increasingly noise which damages the similarity between gait signals. In this case, we set the fixed trajectory of single person, and recollect LoRa signals in different transmitter gain. By manipulating the transmitted signal strength, we compute the euclidean distance after DTW operation between the two motion profiles of human subject himself/herself. As we decrease the signal strength, we expect to observe a increase of the DTW distance, indicating the noise interferes with the gait signals between gait patterns from same person. Larger DTW distance represents the system prefers to split two profiles of the same category into two different parts, which can be predicted to reduce the accuracy of the system for identification. The result of mentioned experiments from three volunteers is shown in Fig. 4.15. At gains higher than $40dB$, we observed random variations in the DTW distance of the three subjects, which can be attributed to the natural variation in their respective gaits. However, beyond the $40dB$ gain threshold, all three trends exhibited an upward trend, indicating that the noise components were increasingly dominating the gait profiles and diminishing the relevance of individual gait patterns. This outcome aligns with our assumption that lower signal strength significantly hampers the identification recognition capabilities of our system.

Consequently, distinguishing between different individuals becomes more challenging under the scenario of lower transmission gain for the models employed in this study.

4.3.2.6 Multiple scenarios

Recently, we only utilise single transmitter and phase difference of a pair of LoRa channels. The prerequisite for identification in a multi-person environment is to obtain the complete gait signal of each person. Typically, both the extraction of beamforming signals in a specific azimuth direction and independent component analysis have the potential to perform blind source separation of gait signals. However, due to the limitations of the experimental apparatus, the only signals we obtained were a set of phase differences computed from the two channels. Therefore, the received signals can not provide independent signals for multiple gait identification recognition. Separating and analysing distinct gait signals will be our next stage target.

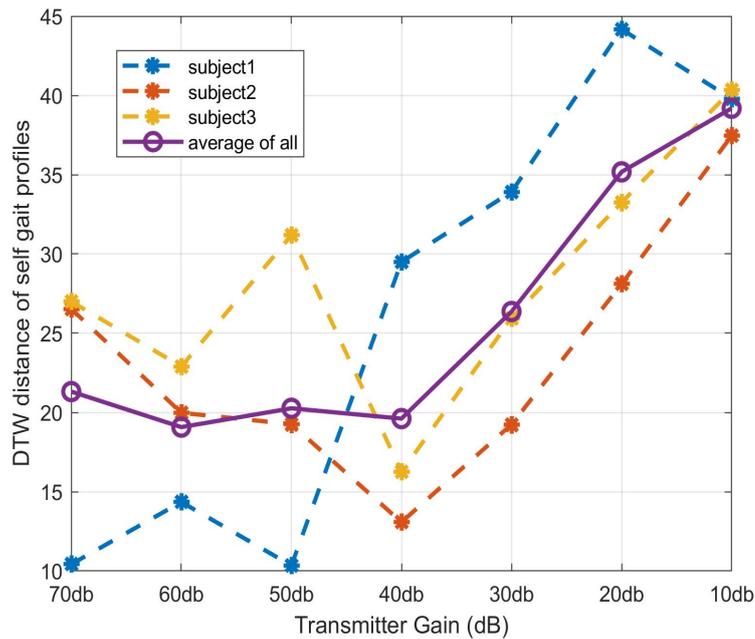


Figure 4.15: DTW distance between two profiles from single subject via Transmitter Gain

Although the recent LoGait system is not able to complete identification recognition in multiple targets scenario, it has the ability to distinguish the number of targets walking in sensing area. We collected data from five different classes, involving the empty environment and individuals walking. The number of participants ranged from 1 to 4. Considering the size of walking area, we selected LD scenario for experiments. And collect 10 minutes for all participants' random walking which is the same as the single LD experiment setup. To implement as the realtime setup, we totally collect 472 profiles for by extracting the profile of 50 minutes data. By using proposed classification scheme, the average accuracy that detect scenarios arrives at 86.9%, shown in Fig. 4.16. The result proves that the system has limited ability to capture response from multiple human walking pattern. In the next step, we will work on solving the multiple gait separation and recognition with multiple-channel device.

4.4 Summary

This chapter introduces the LoGait system, an innovative approach for recognizing human gait across various indoor settings, including living rooms, corridors, and through-wall scenarios. The system leverages LoRa sensing technology, applying conjugated multiplication and several preprocessing techniques to extract distinct gait profiles. These techniques focus on filtering out communication symbols and isolating the physical variations characteristic of human gait. For identifying users, the system uses a DTW based

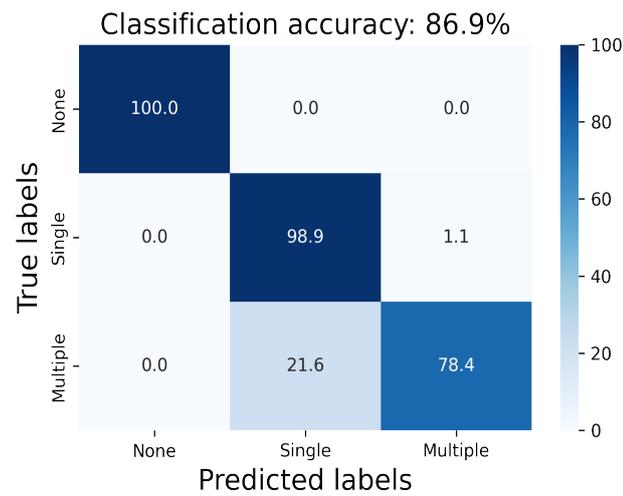


Figure 4.16: Recognition accuracy of LD scenario with different amount of walking subjects

machine learning algorithm. Experimental evaluations conducted in three different scenarios show solid results, with the system achieving identification accuracy of 85.13% in living rooms, and 79.13% in through-wall scenarios, and 84.14% in corridors measuring 20 meters. The LoGait system's capability to operate effectively in complex indoor environments, including challenging through-wall settings and extended corridors, represents a notable development in gait recognition technology with communication based signals,

Radar based Respiration detection with Ellipse Normalization

5.1 Introduction

In the landscape of contemporary healthcare, monitoring vital signs is a cornerstone practice, enabling the assessment of patient health indicators such as heart rate, respiratory rate, and blood pressure from a distance. The rise of wireless and non-invasive detection technologies has notably enhanced the feasibility and effectiveness of remote vital sign monitoring. At the forefront of this innovation is FMCW radar [182]. This method utilizes modulated waves to not only gauge the distance and velocity of objects but also to detect them without physical contact. This feature is particularly beneficial in situations where avoiding contact is crucial. FMCW radar's high accuracy, precision, and energy efficiency make it a key element in wireless monitoring of vital signs. Its integration with wireless technology simplifies the setup and upkeep, revolutionizing convenience in healthcare.

Current healthcare research underscores the need for tailored monitoring solutions, especially for at-risk groups like the elderly and children, due to their heightened risk of sudden health changes. Elderly patients often face a variety of health issues, from heart irregularities to complications from falls or respiratory issues. Incorporating FMCW radar into healthcare systems marks a shift towards non-invasive monitoring, reducing the need for frequent and sometimes intrusive physical check-ups. These innovative systems are promising for early detection of abnormalities, paving the way for pre-emptive medical actions.

Radar technology is promising in respiratory monitoring, offering non-contact, constant observation, crucial for vulnerable preterm infants prone to infections. However, radar systems face challenges in detecting subtle chest movements and quick changes in breathing patterns. The sensitivity and adaptability of the radar system need to be high to accommodate the different size and variable breathing patterns of human beings.

When considering the application of radar in everyday scenarios, particularly within indoor settings, several challenges become salient:

- **Dynamic and Unpredictable Movement:** Within indoor environments, the target subjects, be it humans or otherwise, are seldom static. Their movements can be arbitrary, not adhering to any set pattern or trajectory. This unpredictability complicates the task for the radar, which must constantly adjust and recalibrate to ensure accurate tracking and data acquisition.
- **Subtle Body Movements:** Even when a person might seem relatively stationary, the human body is never truly static. The natural rhythms of breathing, heartbeats, or even minute involuntary twitches introduce a degree of motion. Such subtle dynamics, while seemingly trivial, can present nuances in the interpretation of radar data, demanding greater adaptation to the detection mechanisms.
- **Interference from Other Dynamic Entities:** An indoor setting is seldom devoid of other dynamic entities. The presence of other moving objects, be they other people, pets, or even mechanical entities like fans or robotic appliances, can introduce interference. These entities might produce reflective signals, which could potentially confound the primary readings from the intended target, thereby challenging the radar's ability to discern and compensate relevant data.

Inspired by the literature of Doppler radar signal imbalance compensation with ellipse fitting in algebra [146], we propose a series of smoothing methods to remove the distortion and jumps in phase due to subtle motion based on FMCW radar. In this chapter, we design a vital detection system including the modules of subject localization, vital information extraction, subtle moving elimination and frequency estimation.

5.2 Methodology

5.2.1 Preprocessing methods

The preprocessing methods are illustrated in Fig. 5.1, including the steps of subject localization with Range-Angle maps (RAMs) and Constant False Alarm Rate (CFAR), vital feature extraction from RAMs with ellipse normalization.

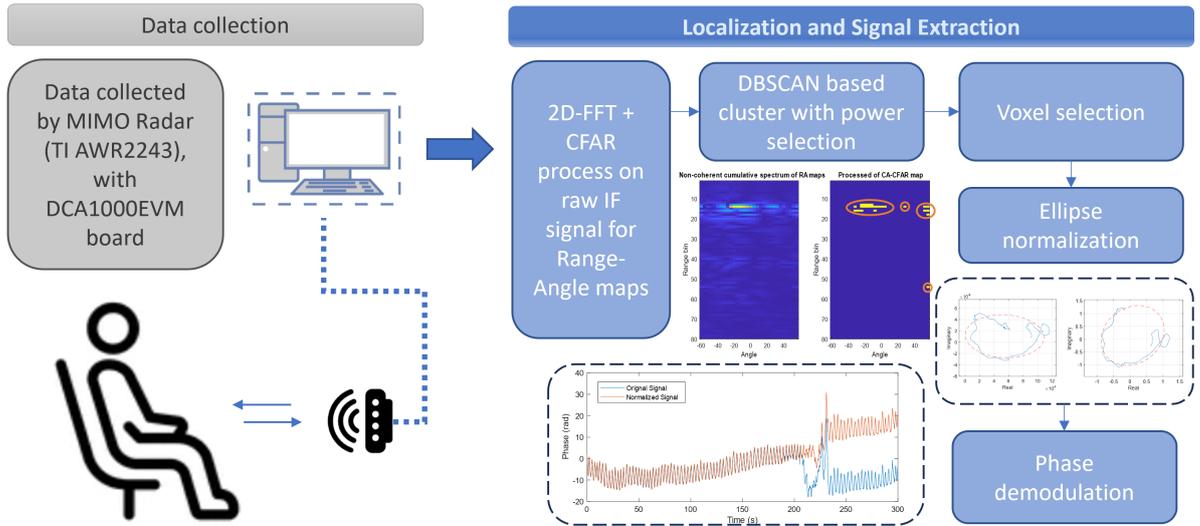


Figure 5.1: Overview of preprocess methods of vital detection.

5.2.1.1 Range-Angle maps

The radar transmits a continuous wave signal with a frequency that varies linearly over time, often referred to as the 'chirp' in FMCW radar systems. As this wave interacts with objects, such as the human body, a segment of it is reflected back. The duration it takes for this reflected wave to return indicates the object's distance from the radar. By mixing the echoed signal with the initial transmitted wave, a resultant beat frequency is achieved. Analysing this frequency via a FFT provides a Range spectrum, with distinct peaks representing objects at varying distances.

For determining the direction or angle from which the signal arrives, the phase disparities among signals received across multiple antenna elements are measured. This variation in phase is indicative of the AoA. By processing the FFT of range data from various virtual channels originating from paired antennas, a comprehensive RAM can be constructed. Within this visualization, the horizontal axis denotes range, and the vertical axis signifies angle. Distinct peaks or intensities on this map depict the spatial orientation and distance of detected entities.

In instances where the target exhibits movement beyond the range resolution, there is an adjustment to the frequency bin corresponding to the target. Conversely, if the target's motion is below the range resolution, it results in a modification of the phase within a solitary frequency bin. To judge if the motion of chest can exceed the range resolution, we can theoretically calculate the resolution at first. The general equation of FMCW radar's range resolution is $\Delta R = c/(2 * BW)$, where the ΔR is the range resolution in meters. c is the speed of light, approximately 3×10^8 meters per second. BW is the bandwidth of the chirp signal in hertz. Here we take an example of TI mmWave radar AWR2243, the

extremity parameter of bandwidth is 4 GHz. It obtains the range resolution of 4 cm, of which is larger than the maximum of chest movement of 1.2 cm. Therefore, when the radar wave interacts with a human, the subtle motions caused by respiration and cardiac rhythms induce minute alterations in the phase of the reflected wave without exceeding the range of bin. Studying the amplitude and phase of these spectrum peaks allows for the identification of consistent patterns stemming from chest movements, a phenomenon rooted in micro-doppler effects.

5.2.1.2 Range of interest (ROI) selection

The RAM displays the intensity of received signals across different ranges and angles, accentuating potential targets. Initially, we employ CFAR detection to search the range of interest (ROI), which scans all the bins and detect targets by dynamically adjusting the threshold based on surrounding clutter. Then, the density based clustering non-parametric algorithm (DBSCAN) method is adopted to identify clusters representing detected objects, as illustrated in Fig. 5.1. However, multiple clusters can correspond to various static objects such as tables, chairs, walls, and the like. To distinguish humans from these static objects, it's pivotal to factor in the influence of human physiological activity on the reflected signals. Since our system is expected to monitor relatively stationary human subjects, the voxels representing the location of the subject would reflect the respiratory activity of the object, as well as the micro-motions produced by the body. Based on the previous discussions, we can identify two methods to lock onto the ROI. Firstly, within a time slice, the ROI should exhibit stronger signal fluctuations representing physiological activities of the human body compared to the static regions. Secondly, these signal fluctuations should be continuous rather than instantaneous. In previous work [144], the authors proposed a non-coherent accumulation of the Moving Target Indicator (MTI) method to recognize the voxels that represent subject's location. However, only considering the accumulation can introduce the fake ROI of those stationary objects that near the target (phase offset was included to enhance the accumulation response), and the range within extensive dynamic motion. Such instantaneous signals often represent moving targets active within the scanning area. Based on the mentioned two principles, we referred to the previous MTI formula [144] and proposed a method that, in addition to calculating the cumulative MTI, also integrates an average difference heatmap by performing a Hadamard product calculation on the two sets of RAM spectrograms, as denoted in Eq. 5.1.

$$\begin{aligned}
 MTI_{sum}(r, \theta) = & \sum_{i=1}^{N-1} |M_{i+1}(r, \theta) - M_i(r, \theta)| \\
 & \odot |M_N(r, \theta) - \frac{1}{N-1} \sum_{i=1}^{N-1} M_{N-i}(r, \theta)|
 \end{aligned} \tag{5.1}$$

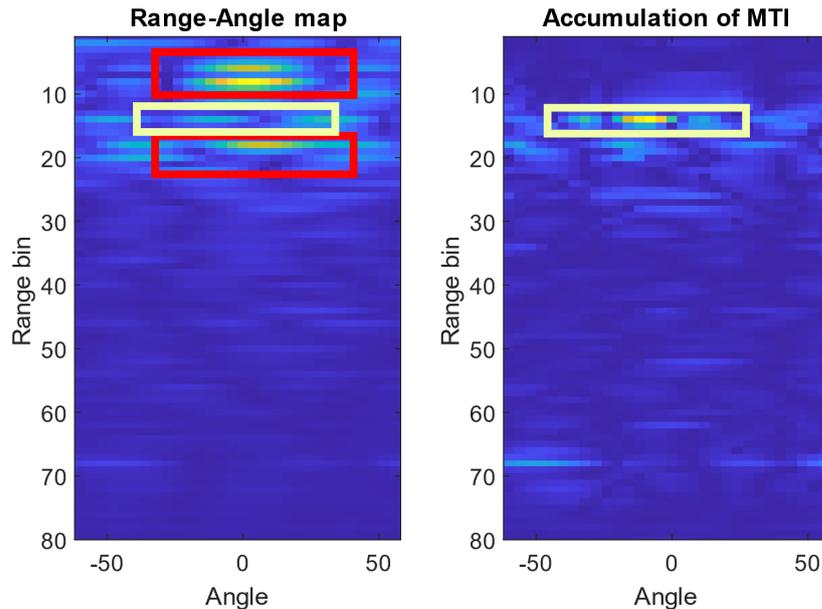


Figure 5.2: Comparison of RAM and accumulation MTI map. Red boxes indicate the static object, yellow boxes indicate the human location.

where the $MTI_{sum}(r, \theta)$ and $M_i(r, \theta)$ represent the MTI result and RAM at i th frame respectively; the detection window length is shown as N . After removal of static cluster, we are able to adopt CFAR to capture the voxels of active bins from the accumulated spectrum, shown in Fig. 5.2.

Due to the leakage and incomplete elimination of echoes by the MTI, and taking into account the multipath effects of radar signals, such false detection points are inevitable. Since the voxels where the human body is located produce periodic signals, we believe that the signal fluctuations caused by human physiological activities are stronger than those caused by static noise. We postulate that signal variations resulting from human physiological processes surpass those arising from static noise. Therefore, time-sequence variance can provide a metric for the sensitivity and data richness of each voxel. In addressing these clusters, we first utilize DBSCAN for spatial categorization. We then ascertain the variance of each cluster's central voxel over a predefined time span and select the cluster which shows largest value of variance. Furthermore, a uniform voxel count is pivotal for signal extraction from the cluster. In the presence of noise interference, our approach extracts the average time-sequence information from the voxels of top three most significant variances.

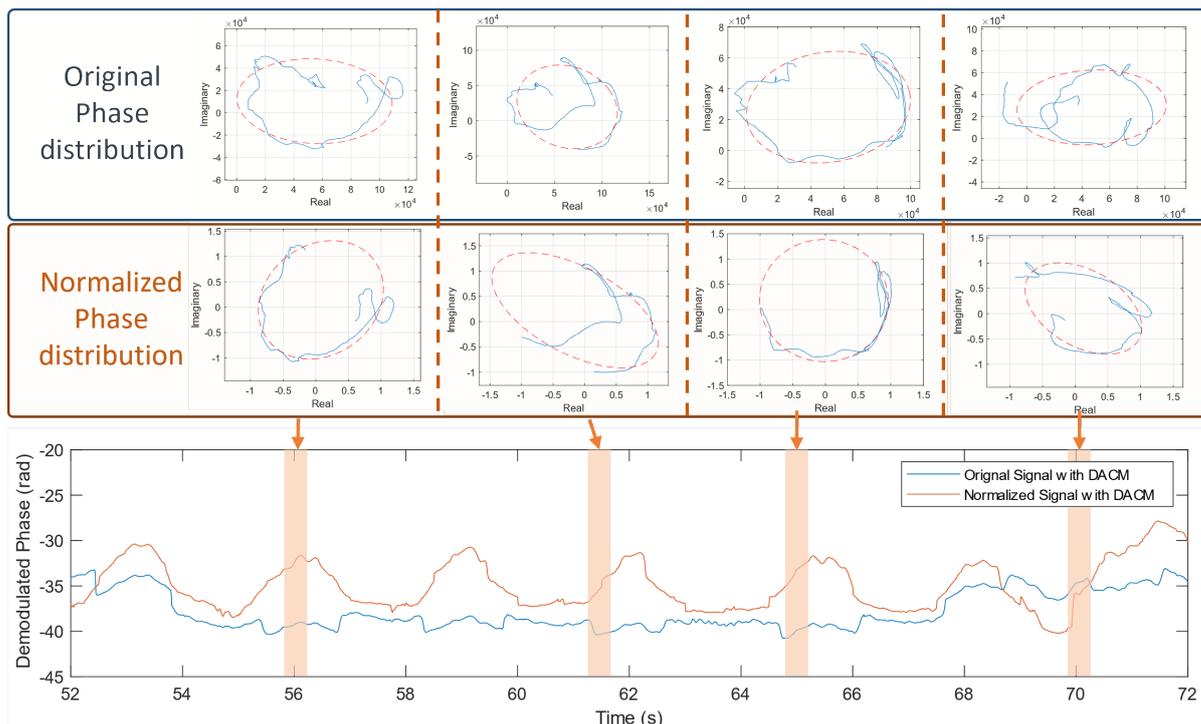


Figure 5.3: Normalization steps of phase signal in slices

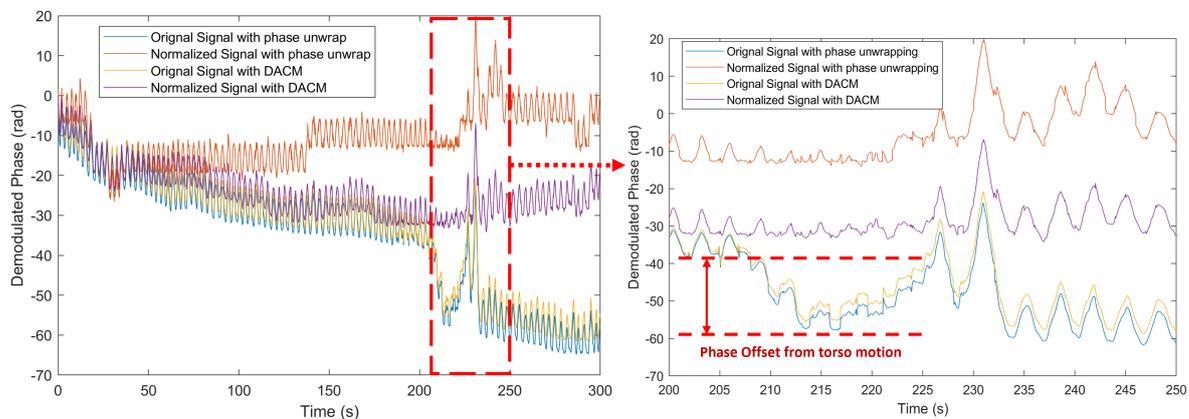


Figure 5.4: Demodulated phase signal in time dimension

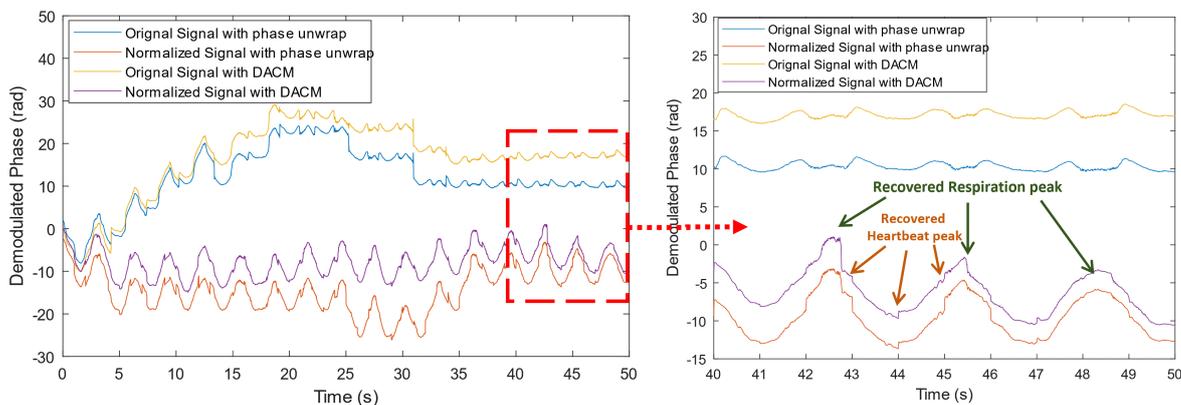


Figure 5.5: Demodulated phase signal in time dimension

5.2.1.3 Subtle moving elimination with Ellipse normalization

When we get the specific address of ROI that represents human body, it is feasible to extract amplitude and phase information of corresponding voxels. Nevertheless, the slight body movement can cause significant offset on extracted signals, especially phase information. Although the band pass filter is able to eliminate the phase variation trend, the low-frequency components of motion, such as body movement, which are similar to the variation mode of respiration, can be reserved as part of filtered signal that distorts the estimated rate. Toward this issue, we proposed an ellipse normalization method without any filtering, which contains two steps: amplitude transformation and phase rotation.

From the basic sensing model in Sec. 2.5.3, we conclude that the signal response induced by respiration can be regarded as a superposition of two unit circles with different frequencies and sliding phases. However, the equalization extend of the IQ signal in actual received signal is distorted. Inspired by the problems, we propose a two-step ellipse normalization method of which both operations optimize the quadrature signals within a time window, thereby mapping the distribution of the original phase variation caused by breathing to the positions with the maximum phase response, shown in Fig. 5.3.

1. **DC cancellation and Amplitude transformation.** Most disturbance of phase in this section comes from the imbalance of IQ signal. Given that we know human respiration can be simplified as harmonic motions with specific frequencies, and their echo responses can be simplified as phase rotations on a unit circle, it is important to note that practical signal acquisitions often suffer from IQ imbalance. Therefore, our initial step involves fitting the uniformly sampled points using an ellipse, followed by operations such as translation and stretching of the sampling distribution based on the ellipse's center and major/minor axes. These operations are aimed at restoring the distribution of IQ signals to the distribution of an approximate unit circle. Details are shown in the Algorithm 1.
2. **Phase rotation.** After compensating for the amplitude of the circularly mapped IQ signals, we can enhance the phase information. This is achieved by adjusting for the additional phase shift, as our interest lies in the variance of the phase rather than its precise value. Illustrated by Fig. 5.5. The details of the process are listed in Algorithm 2.

In our ellipse normalization method, we naturally mapped the distorted complex signal to the unit circle distribution of the equalized IQ signal, and shift the phase distribution to the place that induces largest phase variation. A series of operations neither increased nor decreased the amount of information inherent in the data. Instead, they enhanced changes induced by uniform signals and suppressed distortions caused by irregular variations in the signal.

vlined 1 Ellipse Normalization Algorithm

Data: x, y : Lists of x and y coordinates of the data points**Result:** Parameters of the fitted ellipse ($a, b, c, d, e, X0, Y0, \text{orientation}$)**procedure** Normalization

Compute the mean values:

$mean_x = \text{mean}(x)$

$mean_y = \text{mean}(y)$

Center the data points:

$x = x - mean_x$

$y = y - mean_y$

 Construct the matrix X based on the centered data points:

$X = [x^2, x \cdot y, y^2, x, y]$

Solve the linear system using least squares to get the parameters:

$a = \text{SolveLeastSquares}(X)$

 Extract ellipse parameters from the vector a :

$a, b, c, d, e = a[1], a[2], a[3], a[4], a[5]$

 Check the orientation of the ellipse using b, a , and c :**if** $\left| \frac{b}{a} \right|$ or $\left| \frac{b}{c} \right|$ is greater than a threshold **then**

$orientation = 0.5 \times \arctan\left(\frac{b}{c-a}\right)$

Correct the coefficients for the orientation

Compute the center of the ellipse:

$X0 = mean_x - \frac{d}{2a}$

$Y0 = mean_y - \frac{e}{2c}$

return $a, b, c, d, e, X0, Y0, \text{orientation}$

vlined 2 Process of Maximizing Phase Variance

Data: Complex sequence z , the calculated phase shift of previous clip θ_{i-1} , steps τ **Result:** Sequence $z_{\text{optimized}}$ **procedure** MaximizePhaseVariance $N \leftarrow \text{length of } z$ $\text{shift_angle} \leftarrow 0$ $\text{best_variance} \leftarrow 0$ **for** θ from 0 to 2π in steps of τ **do**

$z_{\text{rotated}} \leftarrow z \times \exp(i \times \theta)$

 $\phi \leftarrow \text{angle of } z_{\text{rotated}}$ $\text{var_current} \leftarrow \text{variance of } \phi$ **if** $\text{var_current} > \text{best_variance}$ **then** $\text{best_variance} \leftarrow \text{var_current}$ $\text{shift_angle} \leftarrow \theta$

$z_{\text{optimized}} \leftarrow z \times \exp(i \times \text{shift_angle} \times \theta_{i-1})$

return $z_{\text{optimized}}$

Next, to demodulate the normalized complex series, we utilized the differentiate and cross-multiply method as described in [183] rather than general phase jump method which is mentioned in [184]. DACM method addresses the codomain limitations inherent in the arc-tangent demodulation approach. The comparison of different combination of demodulated signals are within the Fig. 5.4.

5.2.2 Estimation of respiration rate

5.2.2.1 Filter design

To reduce the high-frequency noise in the original demodulated phase signal, a Savitzky-Golay filter of order 5 with a window length of 11 is employed for filtering. Besides, we used Variational Mode Decomposition (VMD) to decompose and simplify complicated signals into 5 intrinsic mode functions (IMFs), and a clean phase signal was constructed by summing all the VMD modes except for the first, which contains most of the high-frequency noise, and the last, which contains the low-frequency baseline oscillation. Finally, the phase signals are separated into the breath waveforms by Butterworth filters in frequency ranges from 0.1 to 1.2 Hz.

5.2.2.2 Rate estimation

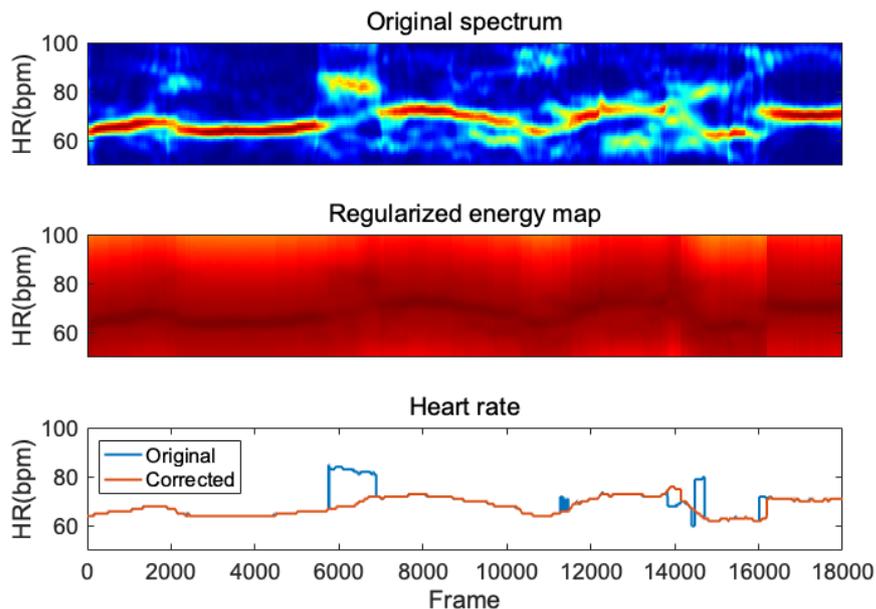


Figure 5.6: Comparison of original rate and corrected rate from spectrum

For quantitative comparison with the reference of the monitor, the respiratory rate is calculated from built-in algorithm. The rate of each time bin was analyzed from the signal fragment within a sliding window of length L (typically 15 seconds). The frequency trace of each time bin can be defined as

$$\mathbf{f} = \{(f(n), n)\}_{n=1}^L \quad (5.2)$$

where $f : [1, L] \rightarrow [1, M]$ is the FFT. The frequency trace based on each fragment are combined into the final time-frequency spectrum. The time-frequency spectrum can be expressed as $\mathbf{Z} \in \mathbb{R}_+^{M \times N}$, where N is the length of the spectrum along the time axis and M is the length of the spectrum along the frequency axis. Based on the spectrum \mathbf{Z} and a candidate trace \mathbf{f} , the energy function for this trace is defined as:

$$E(\mathbf{f}) = \sum_{n=1}^N \mathbf{Z}(f(n), n) \quad (5.3)$$

which has a reasonable estimate $\hat{\mathbf{f}}$ of frequency trace for the input signal. The $\hat{\mathbf{f}}$ is the maximum of the energy function and is expressed as

$$\hat{\mathbf{f}} = \underset{\mathbf{f}}{\operatorname{argmax}} E(\mathbf{f}) \quad (5.4)$$

Due to the superior breath waveforms provided by ellipse normalization, we can directly select the frequency with the maximum of \mathbf{Z} at each moment to construct the final RR trace using Eq. 5.4. For the extraction of weak RR, the movement of the human body and the harmonic noise generated by breathing tend to result in a low signal-to-noise ratio (SNR), thus, we proposed to consider both energy as well as the smoothness of the RR for a more accurate estimate. Considering the inherent continuity of RR, a regularization term was added to penalize jumps in frequency. We model the jump between two consecutive frames as a one-step Markov chain, and here the distribution of all frequencies in frame n_{th} is considered a normal distribution with the output frequency of the $(n-1)_{th}$ frame as the mean and the standard deviation of $\sigma = 3$ (The standard deviation should be close to the range at which the heart rate fluctuates). Specifically, when $n = 1$, we assume that all frequencies follow a uniform distribution. Thus, the Eq. 5.4 is rewritten as

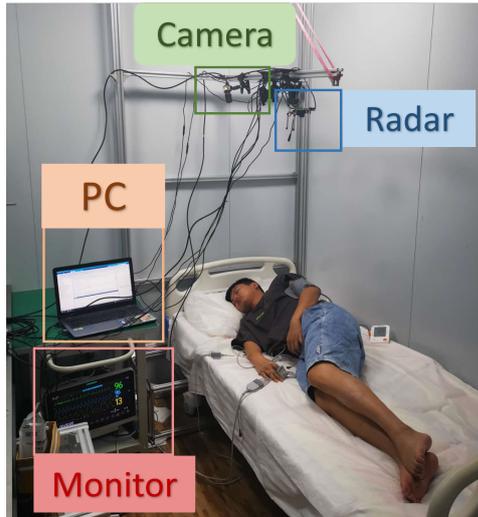
$$\hat{\mathbf{f}} = \underset{\mathbf{f}}{\operatorname{argmax}} E(\mathbf{f}) + \lambda P(\mathbf{f}) \quad (5.5)$$

where $P(f) \triangleq \log P(f(1)) + \sum_{n=2}^N \log P(f(n) | f(n-1))$, and $\lambda > 0$ is a parameter, along with the σ to control the smoothness of the resulting RR. Eq. 5.5 produces a regularized energy map $G \in \mathbb{R}_+^{M \times N}$ based on the recursion on Z in the time direction, and the operation is expressed as

$$G(m,n) = \begin{cases} Z(m,n) + \lambda \log P(f(1)) & n = 1; \\ \max_{m'} G(m',n-1) + \\ Z(m,n) + \lambda \log P(f(n)|f(n-1)) & n > 1. \end{cases} \quad (5.6)$$

where $\forall n \in [1, N]$, $\forall m, m' \in [1, M]$ and m' is the frequency of the $(n-1)_{th}$ frame. The maximum of the $G(m,n)$ at each time bin is output to construct final RR trace.

5.3 Evaluation



(a) Lab room setup

Figure 5.7: Experiment setup in lab room

5.3.1 Data collection

In this section, we first introduce the experimental settings and the protocol. Then the referred baseline methods are described. In summary, we collected 10 human subjects including 6 males and 4 females in lab room, shown in Fig. 5.7. In the laboratory setting, we had subjects lie still on a bed and connected them to the medical monitor produced by Mindray, to collect real-time ECG and fingertip PPG signals. Simultaneously, a radar and a camera turned on, and they recorded timestamps calibrated to a well-synchronized

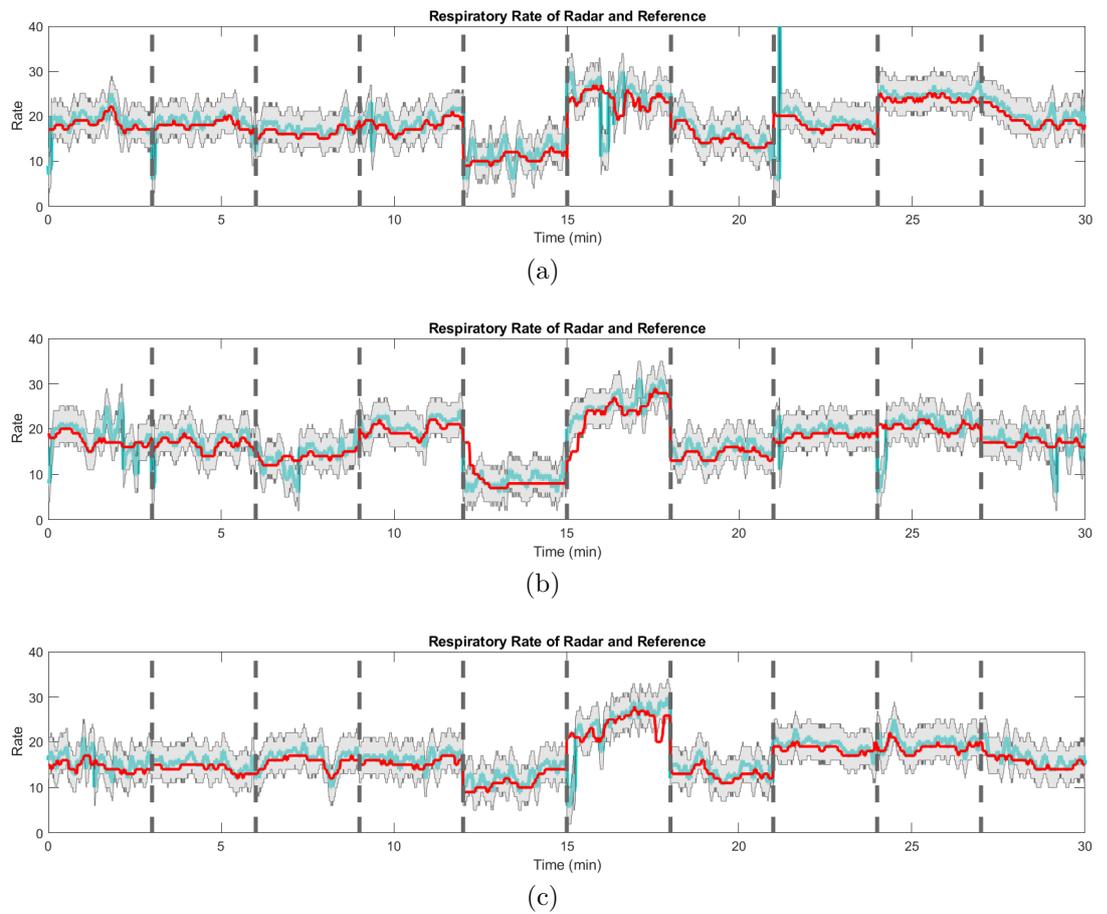
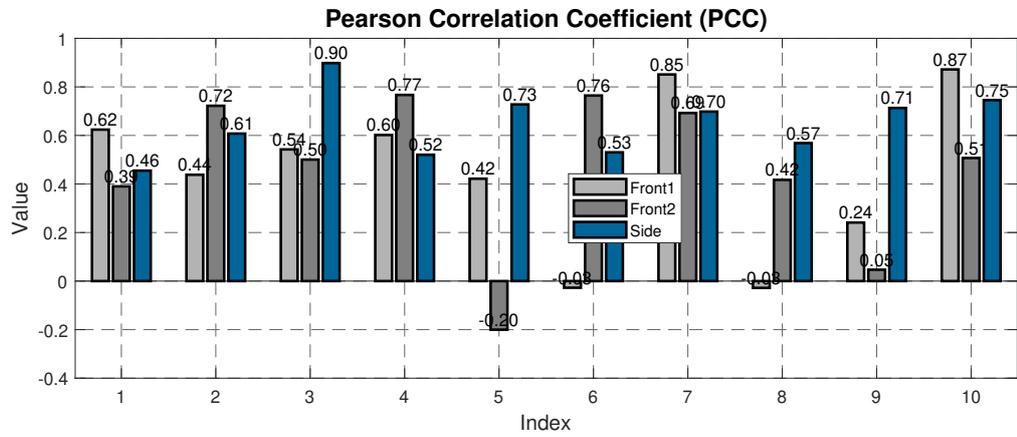
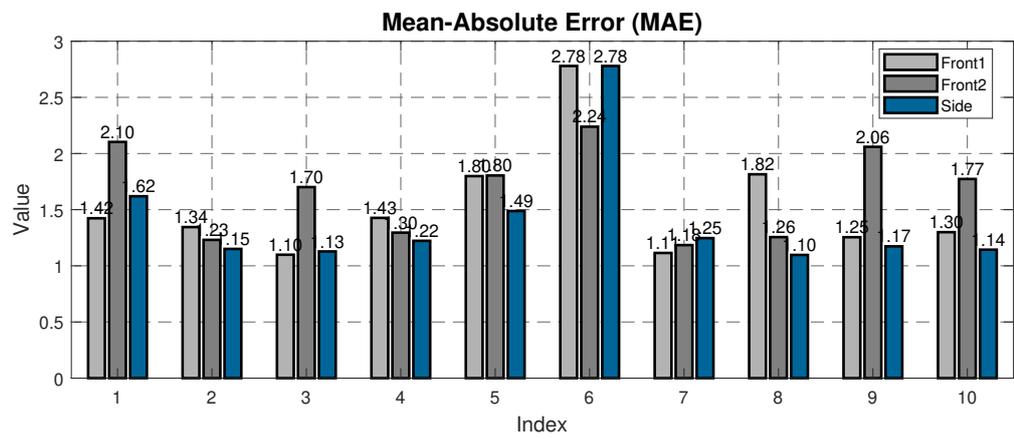


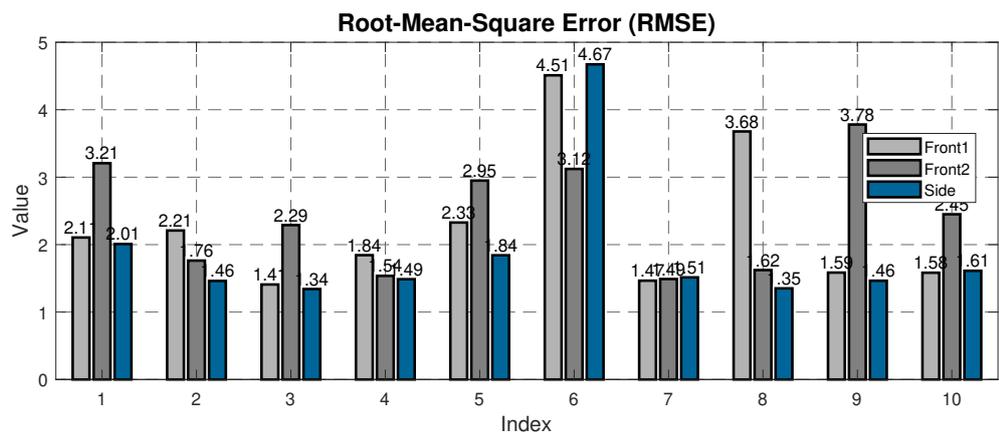
Figure 5.8: Experiment setup in lab: (a) and (b) are from front body, (c) is from body side. Red line represents the radar estimation result and blue link represents the ground truth. Gray part covers the range of ± 5 bpm.



(a)



(b)



(c)

Figure 5.9: Performance of PCC, MAE, RMSE in lab scenario with 10 subjects.

Table 5.1: Parameter set of FMCW radar

Parameter	Value
Center frequency	79 GHz
ADC Sampling frequency	10 msp/s
ADC samples	512
Frame rate	100 Hz
Ramp End Time	40 μ s
Bandwidth	3.8 GHz
Range resolution	3.95 cm
Number of Rx	4
Number of Tx	3

NTP server. Each sample lasted for 3 minutes. The radar settings are as shown in the Tab. 5.1. Meanwhile, Fig. 5.8 illustrates the performance of the proposed system with adult subjects in the lab. We collected 90 minutes respiration data in total of all subjects, including different posture which demonstrates in Sec. 5.3.2.4.

5.3.2 Overall Performance

We introduce three parameters to evaluate the respiration detection performance: mean absolute error (MAE), root mean square error (RMSE) and Pearson correlation coefficient (PCC): MAE measures the average magnitude of errors in a set of predictions without considering their direction. It is the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight. The formula for MAE is: $MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$. RMSE is a quadratic scoring rule that measures the average magnitude of the error. It's the square root of the average of squared differences between prediction and actual observation. The formula for RMSE is: $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$, where y_i is the true value and \hat{y}_i is the predicted value. PCC measures the linear correlation between two variables X and Y . It has a value between +1 and -1, where 1 represents completely positive linear correlation, 0 is no linear correlation, and -1 is total negative linear correlation. The PCC formula is: $\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$, where $cov(X,Y)$ is the covariance of the two variables and σ_X and σ_Y are the standard deviations of X and Y respectively. Fig. 5.9 illustrates the performance of our proposed system.

5.3.2.1 Baseline Approaches

To fairly compare our work with the existing technology, we selected the method that was implemented on same mmWave MIMO radar from Texas Instrument [137]. In their proposed system, the respiration rate is output from bandpass filter with peak detection. However, the performance of rate estimation from peak detection method can be easily varied from auto-correlation window size. Therefore, we apply our estimation algorithm on baseline method to evenly compare the performance from multiple signal processing methods, of which represents baseline method and our proposed system.

5.3.2.2 Comparison and Ablation Study

Table 5.2: Pearson correlation coefficient (PCC) vs User identity

	PCC										
	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	Mean
Baseline	0.49	0.45	0.58	0.63	0.3	0.46	0.53	0.29	0.27	0.67	0.47
VMD	0.48	0.41	0.62	0.61	0.41	0.51	0.57	0.31	0.14	0.72	0.48
PCA+VMD	0.45	0.52	0.63	0.63	0.29	0.56	0.57	0.24	0.33	0.65	0.49
EN+VMD +PCA	0.49	0.59	0.65	0.63	0.32	0.42	0.75	0.32	0.33	0.71	0.52

Table 5.3: Mean absolute error (MAE) vs User identity

	MAE (bpm)										
	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	Mean
Baseline	1.79	2.03	1.75	2.11	1.62	6.44	1.58	1.71	2.72	1.45	2.32
VMD	1.75	1.7	1.37	1.99	1.52	5.51	1.44	1.43	2.88	1.36	2.1
PCA+VMD	1.96	1.37	1.36	1.32	1.83	2.95	1.35	1.45	1.57	1.54	1.67
EN+VMD +PCA	1.72	1.24	1.31	1.31	1.7	2.6	1.18	1.39	1.5	1.41	1.54

Table 5.4: Root mean square error (RMSE) vs User identity

	RMSE (bpm)										
	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	Mean
Baseline	2.49	2.85	2.23	2.9	2.29	8.18	2.2	2.57	3.68	1.94	3.13
VMD	2.66	2.43	1.75	2.7	2.12	7.09	1.95	1.98	3.87	1.7	2.83
PCA+VMD	3.02	2.06	1.77	1.62	2.52	4.33	1.84	2.5	2.42	2.23	2.43
EN+VMD +PCA	2.44	1.81	1.68	1.62	2.37	4.1	1.49	2.22	2.28	1.88	2.19

We conducted a comprehensive ablation study to assess the efficacy of each component in our respiration detection system, as detailed in Tab. 5.2, Tab. 5.3, and Tab. 5.4. These studies, conducted in a laboratory setting, demonstrate a significant enhancement in performance, evidenced by a notable decrease in RMSE from 3.13 bpm to 2.19 bpm.

Additionally, our method achieved a modest improvement in PCC, aligning closely with the trend of reference rate variation. Notably, our approach effectively addresses outlier values, such as those observed in sample *U6*, where the RMSE was reduced from 8.18 bpm to 4.1 bpm, a reduction of nearly 50%.

5.3.2.3 Gaussian frequency estimation

Table 5.5: RMSE vs User identity without Gaussian estimation

	RMSE (bpm)										
	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	Mean
Baseline		3.27	2.44	3.66	2.38	9.54	2.31	3.19	4.99	2.34	3.74
VMD	2.97	2.76	2.35	3.28	2.04	8.9	1.89	2.24	3.74	1.78	3.19
PCA+VMD	3.02	1.94	1.74	1.63	2.52	4.71	1.84	1.95	2.49	3.45	2.53
EN+VMD +PCA	2.38	1.73	1.68	1.63	2.38	5.04	1.51	1.69	2.07	1.94	2.21

Meanwhile we compare the estimation algorithm with simple Maximum algorithm in lab scenario, shown as Tab. 5.5. We discover the RMSE results are higher than the system with proposed Gaussian estimation, which approves the availability of our methods.

5.3.2.4 Position towards target’s chest

Table 5.6: RMSE vs radar placement direction

	PCC										
	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	Mean
Front	0.51	0.58	0.52	0.68	0.11	0.37	0.77	0.19	0.14	0.69	0.46
Side	0.45	0.61	0.9	0.52	0.73	0.53	0.7	0.57	0.71	0.75	0.65
	MAE (bpm)										
	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	Mean
Front	1.76	1.29	1.4	1.36	1.8	2.51	1.15	1.53	1.66	1.54	1.6
Side	1.62	1.15	1.13	1.22	1.49	2.78	1.25	1.1	1.17	1.14	1.41
	RMSE (bpm)										
	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	Mean
Front	2.65	1.99	1.85	1.69	2.64	3.82	1.48	2.65	2.68	2.02	2.35
Side	2.01	1.46	1.34	1.49	1.84	4.67	1.51	1.35	1.46	1.61	1.88

Majority of works only consider the single direction from radar. However, generally the movement of the human lungs, as they function, is primarily in the anterior-posterior direction, relative to the body. This means that the motion is predominantly towards and away from the chest or frontal plane of the body. However, it’s inevitable that the sensing target turns its direction, e.g. rolling over in bed. Therefore, we performed breathing monitoring on subjects while lying on their backs and on their sides, the two most common

sleeping positions. From the result shown in Tab. 5.6, PCC, MAE and RMSE show the wireless detection performance of side lying is better than normally lying on back. This can be the reference of displacement of radar sensor for next generation healthcare detection system.

5.4 Summary

This chapter introduces our proposed method of radar based respiration detection. It employs multiple signal processing techniques for improved sensing accuracy. These innovations include the application of Ellipse Normalization to address data variations and the use of autocorrelation and Gaussian frequency estimation for signal analysis. The robustness of the methodology is validated through extensive evaluations, demonstrating its efficacy in various real-world scenarios and settings.

Meanwhile, it presents a comprehensive evaluation of a radar based respiration detection system, which includes a lab experiment with 10 human subjects and focuses on MAE, RMSE, and PCC. The evaluation demonstrates significant performance improvements, notably a reduction in RMSE from 3.13 bpm to 2.19 bpm. Additionally, the study addresses outlier values, reducing the RMSE for specific cases by nearly 50%. This extensive evaluation underlines the system's effectiveness in accurately detecting breath rate. This improved approach marks a potential advancement in the field of respiratory monitoring.

Data collection of multimodal RF sensing of human speech recognition

6.1 Background & Summary

In general speech recognition tasks, acoustic information from microphones is the main source for analyzing the verbal communication of humans [185]. The speech process is not just a means of conveying linguistic information, which can also provide valuable insight into the speaker's characteristics such as gender, age, social and regional origin, health, emotional state, and in some cases even their identity. Recently, the automatic speech recognition (ASR) technique has already matured and been marketed [186]. In addition to sound signals, the series of physiological processes that produce sound, such as lip movement, vocal cord vibration, and head movement, also retain semantic and speaker information to some extent. On the other hand, there are two main limitations in specific environments that only audio information can not perfectly work for ASR: silent speech recognition (SSR) and multiple speakers environments. Both issues can be solved if considering the speaker physics properties and will be explained in following paragraphs.

First, SSR can be considered a significant branch of speech recognition that provides understandable and enhancing communication methods to assist patients with severe speech disorders. In recent years, research in silent speech recognition has explored a variety of approaches, including wearable sensors, radar based systems, and other non-invasive techniques, to address the challenges of capturing and processing speech-related information. The contactable methods mainly focus on detecting brain and muscle activity with electroencephalogram (EEG) sensor, articulator movements headset and other types of implantable sensors [187]. However, contact based methods are highly dependent on wearable and implant sensors, which is dedicated to patients but does not collect a large

dataset from a normal person. Meanwhile, users should consider the potential health risk of contactable devices. For voice disorder and other patients who maintain the capability to control the vibration of vocal folds and face muscles, non-invasive SSR has the potential to improve their quality of life compared to electronic sensors.

In addition, in scenarios with multiple speakers, the microphone captures the sounds from the surroundings without distinguishing the person's identity, which seriously lowers the accuracy of speech recognition. This issue is similar to the cocktail party effect [188], which is a phenomenon in which an individual can focus on one conversation despite being surrounded by several other simultaneous conversations. The effect is mainly attributed to the brain's ability to process auditory frequency and highlight certain sounds, allowing the individual to focus in on the source of interest without being easily distracted. However, it is a challenge to separate different sources only using acoustic data. In this case, additional radar or laser devices can assist the model in distinguishing the audio according to the physical information. For example, the proposed work [189] combined the audio and radar signals to filter after added noise. And Secondly, the voice information including tone and speaking habits of individual contains a variety of personal data that can be used to create a unique voice fingerprint, such as speaking habits and intonation. This will cause a risk of sensitive data leakage, as the voice fingerprint could be used for identification. For the wireless sensing side based algorithm, vocal folds vibration focuses only on the tone of speech, which does not include privacy information. Third, previous speech recognition research has focused mainly on visual based mouth movements, posing a risk of lack of privacy and overlooking internal mouth movements.

In this chapter, we proposed a dataset of human speech by collecting data from multiple sensors information while people are speaking specific corpus. The contributions of our dataset are concluded in following points:

1. In this work, we present a novel dataset that incorporates multiple modalities for silent speech recognition, including UWB radars, mmWave radar, and depth camera data, which we believe will be a valuable resource for researchers in the field. The dataset is expected to reduce the labour for researchers who expect to work on SSR from wireless signals or enhancing audio signals.
2. Our system takes into account the physical movements of all parts of the head during human speech, including mouth movements and vibrations of the vocal cord, which is illustrated in Fig. 6.1.
3. The diverse range of modalities in our dataset offers ample opportunities for conducting research in the field of speech recognition. The range contains but is not limited to the following application: radar based vowels and words classification, speaker identification, speech enhancement in noisy environment, radar based lip reconstruction, etc.

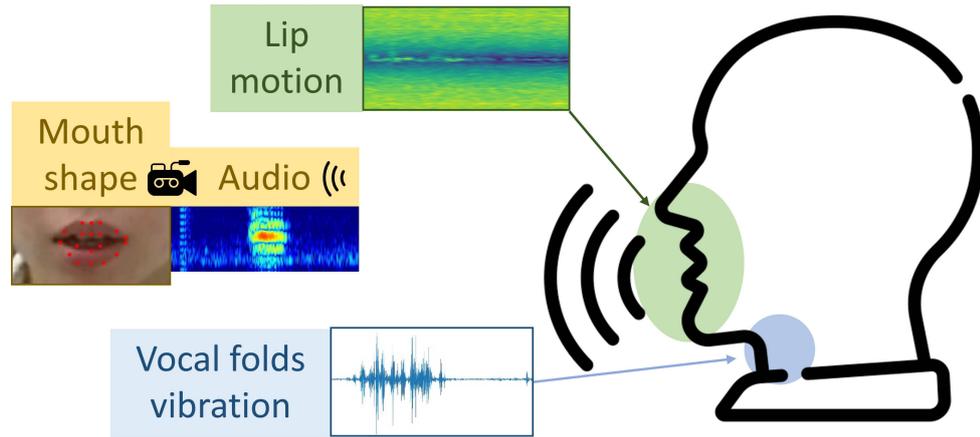


Figure 6.1: Schematic diagram of the Multimodal signals

6.1.1 Literature survey of radar-enabled speech recognition

There are various kinds of sensors has been adopted for speech research: UWB, mmWave radar and laser speckle detector [190]. For SSR task, the work of UWB demonstrated the lip reading work with the vowels of $[\text{æ}]$, $[\text{i}]$, $[\text{ə}]$, $[\text{ɔ:}]$, $[\text{u:}]$ and static scenario, with a face mask. The result of 95% approves that the mouth motion produces informative signals for UWB sensing [191]. FMCW radar is also an optional choice which has been proven in the result of the paper [192]. The mentioned work adopts point clouds of human mouth while speaking as data feature for classification work of 13 words with 4 speakers. It gains 88% accuracy using Linear Regression classifiers. To expand the work and exploit more possibilities, we added sentences for data collection regarding the reference. Besides, mmWave FMCW radar has been used for speech enhancement in the published works [189], [190]. These two researches have distinct focus directions: the paper [189] considered the distance coefficient for radar signals and successfully make speech enhancement system implementation in 7 meters; and works [190], [193] on audio separation of multiple speakers with radar based spatial information. For laser-related information, paper [194] proposed a remote measurement technique for healthy individuals that involves capturing the reflected laser speckle from the surface of the neck skin. This system is capable of capturing the micro-vibration on the surface of the neck produced by blood pressure, which can also be adopted for the extraction of voice signals without audio signals through detecting the vibration caused by throat. Inspired by mentioned works, we decided to adopt radar sensors of FMCW radar and UWB radar, laser speckle detection system, and Kinect camera for mouth skeleton stream and speech voice as the source of our multimodal dataset.

Furthermore, we conclude the dataset of those multimodal based speech recognition works in Table 6.1. To the best of our knowledge, most open-access datasets in speech recognition focus on audiovisual topics instead of considering radio signals. Although there has been research talking about wireless signals based speech processing, it is difficult to get the dataset from authors. Therefore, our core contribution is establishing a contactless based speech dataset for research on combination of audio signals and physically vibration from wireless signal.

	Dataset size	Dataset usage	Multimodal type	Open access of dataset provided
RAVDESS [195]	7356 samples	emotion recognition	visual video stream, voice	Yes
SpeakingFaces [196]	13000 samples	biometric authentication, speech recognition	visual video stream, thermal video stream, voice	Yes
UltraSE [197]	8000 samples of 5 seconds speech	speech enhancement	ultrasound signal, voice	Not found
RadioSES [198]	5700 sentences	speech separation and enhancement	FMCW radar signal, voice	Not found
Speckle detection [199]	1000 sample of speech	speckle noise removal	laser signal, voice	Not found
Electromyogram [200]	660 words	silent speech recognition	high-density surface electromyogram	Not found
SSR [201]	few samples for validation	speech enhancement	laser signals	Not found
AV-corpus [202]	34000 sentences	speech recognition	audio and video	Yes
IEMOCAP [203]	10039 samples of 4.5 seconds speech	emotion recognition	audio, video, face and hand markers	Yes
TaL [204]	18221 samples	speech recognition	audio, video stream, ultra sound image	Yes
Ours	6000 samples (including vowels, words and sentences)	silent speech recognition, speech separation, speech enhancement	FMCW radar signal, UWB radar signal, mouth skeleton, laser signal, and voice	Yes

Table 6.1: Dataset review of multimodal based speech recognition works

6.2 Methods

Firstly, we conducted a literature review to establish the necessary sensors and experimental setup for radar based speech recognition, given the absence of a standard and corpus. Meanwhile, we demonstrate the availability of all the sensors we adopted and then establish our data collection approach, referencing previous work.

6.2.1 Data acquisition scheme

The overall data collection system was organized by four laptops and four types of sensors: Microsoft Kinect V2 for audio and video including mouth landmark, X4M03 UWB radar from NOVELDA, AWR2243 mmWave radar from Texas Instrument, and laser measurement system for physical vibration of human speech. The selection of devices is referred from mentioned previous research. To keep the data synchronised with different sensors, we used the TCP / IP connection to control the distinct host laptops with the same network time protocol (NTP) for recording the time stamp while data collection. A multi-threaded control script has been developed and employed that automatically initiated and terminated data recording scripts, minimizing data recording latency to the greatest extent. Once we run the script on a master laptop, the master will send the commands to the other three sockets in series. The mean delay from the master to sockets of other devices is around 80 ms, which is considered in our post-synchronization processing. Furthermore, we employed expert supervision and manual calibration to ensure time synchronization across different sensors. We calibrated the devices, monitored the data collection process, and made necessary adjustments throughout the entire data collection. Considering the potential research for speech recognition, we designed three data collection schemes shown in the following. The adopted corpus is recorded in an additional folder in our dataset.

- Single person speech of vowels, words and sentences.
- Dual-person speech simultaneously of complex sentences.
- Single person speech of vowels, words, and sentences with different distance from radar to speakers.

The details of data collection from specific sensor are demonstrated below, with experiment setup shown in Fig. 6.2.

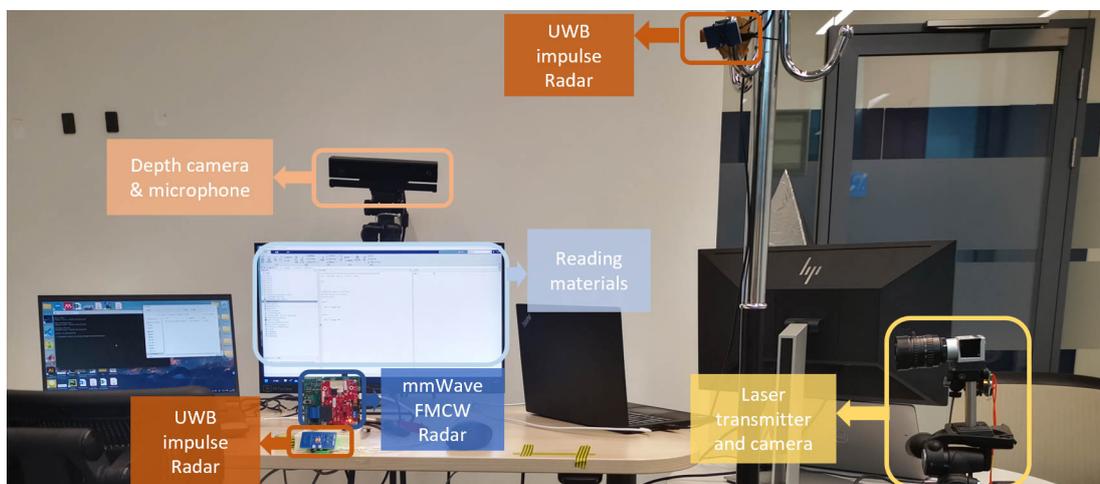


Figure 6.2: Data collection setup with device label in the real scenario

6.2.2 Speech voice

We used Kinect v2 for collecting vocalised speech. With the enable Kinect v2 to collect the accurate acoustic information. The sample rate of audio data is $16kHz$, and bit depth is 16-bit. The frequency range of recording audio is up to $8kHz$, that can cover the frequency range of human voice.

6.2.3 Mouth skeletal points

The Kinect v2 is also used in collecting the facial landmark information. An RGB camera and an infrared camera are intergrated in kinect v2. By measuring the ToF using the IR camera Kinect can get the depth image. Meanwhile, we use the lip recognition method proposed in paper [205] for extraction of the lip skeleton, which is provided as part of our dataset.

For data collection, we adopted Moduleconnector API which is supported by the IR-UWB radar on Windows MATLAB 2021b. The detailed parameters of radar are listed in the Table 6.2. The detailed technique principle is mentioned in Section 2.

6.2.4 mmWave FMCW radar

In data collection, we utilised mmWaveStudio API on Windows with the configuration file. The parameters were set to the value shown in Table 6.2. The frequency response reflected by the radar signals was of paramount interest, particularly given the average fundamental frequency range of the vocal folds, which spanned from $85Hz$ to $255Hz$ [206]. Aligning with the Nyquist criterion, we adjusted our radar frame rate to $1018Hz$, ensuring the sampling rate exceeds twice the highest frequency present in the signal.

6.2.5 Laser-speckle system

The laser measurement system consists of a 532nm green laser diode (DJ532-40, Thorlabs) as transmitter and a high-speed CMOS camera from Basler as receiver, where the laser diode emits a laser beam pointing to the face outline of the testing subject and the camera captures the reflected laser speckle patterns. Both transmitter and receiver are fixed on a 1.2m tripod, and the camera is connected with a laptop via an USB 3.0 cable for powering and data transferring. The green laser diode has a distance of approximately 1 m to the participants, it will produce an illumination spot of around 5mm diameter on the human skin by considering the beam divergence. For the laser safety, laser power exposed on

Parameter	Value
Center frequency	8.745 GHz
Sampling frequency	23.328 GHz
Frame rate	300 Hz
Bandwidth	1.5 GHz
Number of antennas	1 Tx and 1 Rx

(a) UWB radar setup

Parameter	Value
Center frequency	79 GHz
ADC Sampling frequency	10 msp/s
ADC samples	512
Frame rate	1018 Hz
Ramp End Time	60 μ s
Bandwidth	3.8 GHz
Range resolution	3.95 cm

(b) mmWave radar setup

Table 6.2: Radar parameter setup of UWB X4M03 and TI AWR2243

human skin is controlled to be less than 0.5 mW (CLASS 1), therefore it is safe for long-term eye and skin exposure. The focal length and f-stop of the camera objective are set as 25 mm and 0.95, respectively, allowing the camera system to detect the laser speckle from a very close range (0.1 m) to a relatively far range (up to 3 m). Furthermore, the size of the ROI window is chosen as 128x128 pixels, and the camera exposure time is set as 600 μ s, shown in Table 6.3. The laser and camera are carefully aligned before the experiments to ensure that the selected ROI includes the movements of speckles. For each measurement, the collected data is in a format of $W \times H \times N$, where W and H represent the width and height of the ROI, respectively, and the N equals to the number of frames within the measurement period. In our case, N is correlated with the sampling frequency of the CMOS camera, which is set as 1.47kHz.

Parameter	Value
Power	0.5 mW
Wavelength	532 nm
Working Mode	continuous wave
Beam Divergence	12 mrad
Operating Current	330 mA
Operating Voltage	1.9 V

(a) Laser transmitter setup

Parameter	Value
FPS	1470
Gain	12
Exposure Time	600 μ s
ROI Size	128x128 pixel

(b) Camera setup

Table 6.3: Parameter setup of Laser speckle detection system

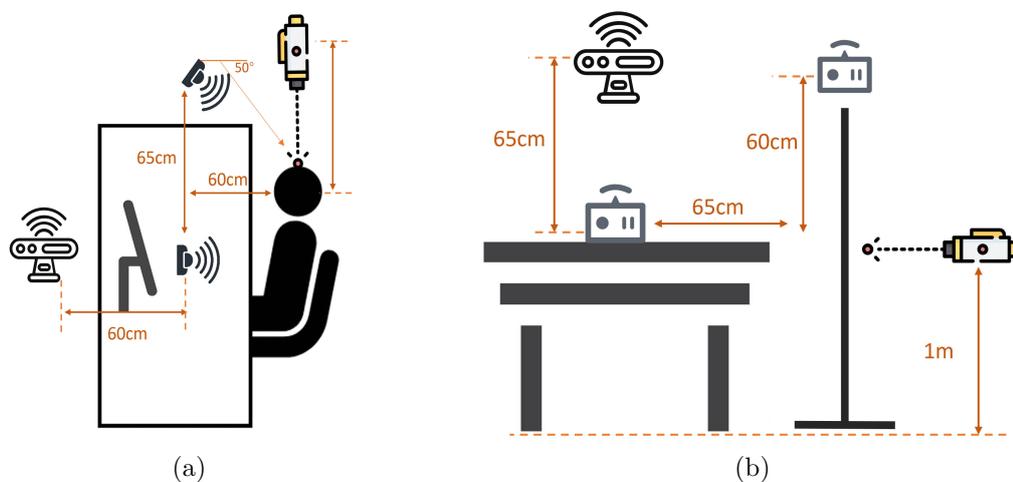


Figure 6.3: Detailed setup schematic diagram for single person scenario from (a) Top view and (b) Front view (Laser's location is not fixed due to the camera based signals process only require the laser directly point to skin of subjects. The UWB radar facing to subject directly was called 'xe2' in dataset folder, another is called 'xe1').

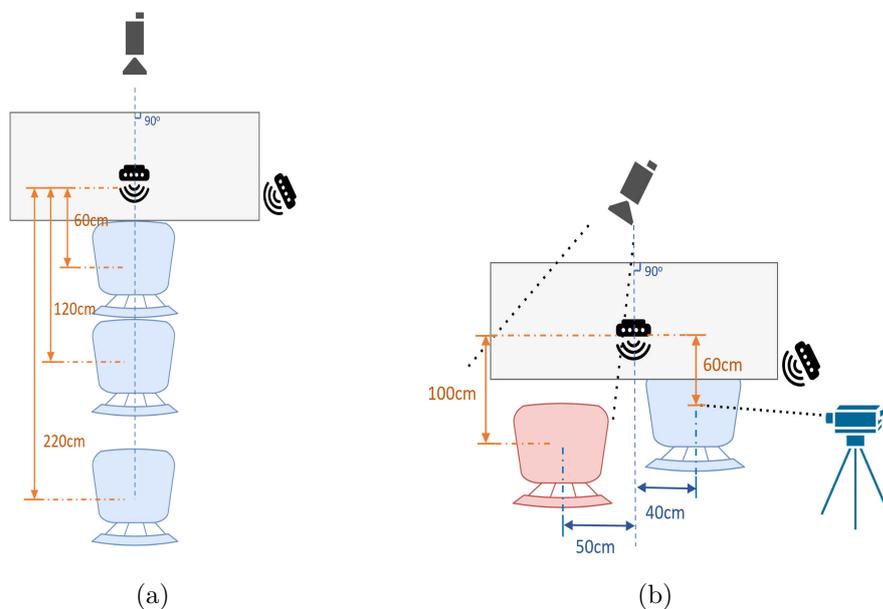


Figure 6.4: Detailed setup schematic diagram from top view of (a) second scenario which considers the distance of 120cm and 220cm between sensor and speaker, and (b) third scenarios which considers speech data collected under multiple speakers. Camera was set to collect left speaker and laser to right speaker.

Volunteer index	Age	Gender	Native language	Education
1	25	Male	Chinese	BEng
2	25	Female	Chinese	BEng
3	24	Male	Chinese	MSc
4	25	Male	Chinese	BEng
5	22	Male	Chinese	BEng
6	25	Male	Chinese	BEng
7	31	Male	Arabic	MSc
8	24	Female	Chinese	BEng
9	30	Male	English	PhD
10	25	Male	Chinese	MSc

Volunteer index	Age	Gender	Native language	Education
11	24	Male	Chinese	BEng
12	23	Male	Chinese	BEng
13	24	Female	Chinese	BEng
14	34	Female	Arabic	MSc
15	25	Female	English	BEng
16	30	Female	Arabic	MSc
17	25	Male	Chinese	BEng
18	25	Female	Chinese	MSc
19	24	Male	English	PhD
20	31	Male	Arabic	PhD

Table 6.4: Volunteer information sheet

6.2.6 Participant

There are 20 volunteers who contribute to our experiment, who come from different country regions including Europe, China, and Pakistan. The volunteer information sheet is listed in Table 6.4. Our dataset for speech recognition presents both opportunities for generalization and challenges due to the volunteers’ diverse backgrounds, resulting in distinct accents. To the issue bring from body size, we adopted an adjustable table for subjects that can keep the relative distance same between the head of speaker and different sensors. The characteristic of accents and speech habits can be extracted from our dataset with lip motion, vocal folds vibration, and audio, which has potential for related multimodal ASR research.

Meanwhile, all participants were informed about the purpose of the study, the implications of identifying information, and what was expected from them. They agreed to the open publication of identifiable information including voice signal in released dataset. Experiment consent forms were obtained from each participant prior to experiments.

6.2.7 Setup of data collection

This section provides the data collection protocol including the introduction of corpus, experiment setup, and data formats. The corpus of the single person scenario is listed in Table 6.5, and specific setup is illustrated in Fig. 6.3. During data collection, we asked volunteers to pronounce a specific vowel / word / sentence with timestamps on laptops. All laptops were synchronized using the same NTP server. During collection, volunteers were guided by automatic voice instruction to read the corpus and relax. The timestamps of audio instruction were instantaneously recorded in kinect/timestamp. However, there is a few seconds of uncontrolled latency in activating all radars and laser equipment, which

Type	Corpus	Index	Participants
vowel	[æ], [i], [ə], [ɔ:], [u:]	1-5 in sequence	User 1 - 20
word	order, assist, help, ambulance, bleed, fall, shock, medical, sanitize, doctor, accident, rescue, emergency, heart, break	1-15 in sequence	User 1 - 20
sentences	I need help.	1	User 1 - 20
	Call for an ambulance.	2	
	The building's on fire.	3	
	Can you smell smoke.	4	
	Where's the fire escape.	5	
	There's been an accident.	6	
	Is there a doctor here?	8	User 1,2,4,6,7
	The staff sanitized the sickroom.	7	
	Medical care is important.	9	
	Don't worry about bleeding.	10	User 3, 8 - 13
	I am having trouble breathing.	7	
	I think I'm having a heart attack.	9	
	My heart is failling.	10	User 5, 14 - 20
	Need emergency treatment at shock stage.	7	
	He need a rescue for a heart attack.	9	
Don't worry about falling.	10		

Table 6.5: Corpus list for single subject experiment. The index of each participant is identical to the user label.

Type	Corpus	Label	Participants
vowels	[æ], [i], [ə], [ɔ:], [u:]	v1-5 in sequence	User 4 of 1.2m and 2.2m (Index No. 24 and 25 in dataset), User 5 of 1.2m and 2.2m (Index No. 26 and 27 in dataset)
words	order, ambulance, medical, sanitize, accident	w1-5 in sequence	
sentences	Call for an ambulance	s1	
	There's been an accident	s2	
	The staff sanitized the sickroom	s3	
	Is there a doctor here?	s4	
	Medical care is important.	s5	

Table 6.6: Corpus list for supplementary experiments of changing the distance.

Type	Corpus	Label	Participants
article	From view of Kinect, volunteer on the left side read 'Mr Sticky', on the right side read 'The king of the birds'.	b1-11	User 6 (Left) and User 4 (Right), recorded in Index 21
		b12-22	User 4 (Left) and User 6 (Right), recorded in Index 21
		b1-11	User 4 (Left) and User 5 (Right), recorded in Index 22
		b12-23	User 5 (Left) and User 4 (Right), recorded in Index 22
		b1-11	User 5 (Left) and User 1 (Right), recorded in Index 23
		b12-23	User 1 (Left) and User 5 (Right), recorded in Index 23
static	sitting without speaking	b24-26	User 5 (Left) and User 4 (Right), recorded in Index 22

Table 6.7: Corpus list for supplementary experiments of two-person scenario. The reading materials are referred from corpus publication [207].

disrupted the devices' ability to synchronize acquisition data. In this case, we decided to keep these devices recording for one minute and write timestamp while data collection is activated so that the signals can be cropped according to kinect timestamps. All were recorded alongside the data, which reduces the effort of manually separating the data.

Meanwhile, we also collected data via different distances in the single-person scenario and the two-person scenario, which is shown in Fig. 6.4, with corpus listed in Table 6.6. Instead of original $60cm$, we asked volunteers to sit $1.2m$ and $2.2m$ away from radar equipment, respectively, which is a potential for researchers to explore the relationship of radar based audio detection with distance. In addition, in the two-person experiment, we kept one volunteer sitting in the same place as the single-person scenario, and then let another speaker sit on the left side of the first-mentioned volunteer. The two subjects were asked to normally read different corpus shown in Table 6.7, which was shown on the screen in one minute, without repeating words. The laser equipment was pointed to the first volunteer, and the kinect camera only took information from another subject. This kind of dataset will contribute to multiple audio source separation.

6.3 Data Record

The multimodal speech detection dataset is accessible with the doi link of the dataset [208], designed for ease of access and reproducibility. Detailed in this Fig. 6.5 is the file structure and content description of the dataset.

6.3.1 Data storage structure

Processed data files, such as *land_proc_<Repetitive_index>.csv*, include columns for coordinates of detected landmarks by Kinect V2.0, among other variables. The methodology for data collection is expounded in the Section 6.2 of this document, providing the necessary context for dataset curation. After saving the data, all files were integrated into specific folders according to the data class, of which structure is illustrated in Fig. 6.5. The entire dataset was divided into raw data and processed data to match the limitation of file size. Firstly, due to the data size limitation of file, we put mmWave FMCW radar data and laser data in a separate folder and other sensors in another. The radar signals files were kept in binary format with radar timestamps in text format (FMCW radar signals of subject 12 were missing).

Meanwhile, information from kinect and two UWB radars was kept in same folder as the similar storage structure, which contains timestamps in JSON format, audio in WAV, landmarks of user's head in BVH, and UWB radar signals in MAT format. Additionally, to ensure a license-free distribution of the dataset, the preprocessed data was converted from MAT to NPY and BVH to CSV files regarding the usages. The timestamps reader and conversion scripts are all provided with our dataset. The description below completely introduces the statement of data files recorded in the proposed dataset.

6.3.2 Raw Data Statement

Comprising raw data from all sensors, segregated by subject IDs. Each subject's data is further segmented by type of modality, including laser data, FMCW radar signal, UWB radar signal and kinect based audio, facial skeleton and timestamps:

- *< SubjectID > _ < CorpusType > _ < CorpusIndex > .bin*: Raw FMCW radar signals in binary format.
- *radarlog < SubjectID > _ < CorpusType > _ < CorpusIndex > .txt*: Raw radar logs with timestamps.
- *laser_data_index_ < SubjectID > .mat*: Laser data matrices with timestamps.

- *timestamp_ < Repetitive_index > .json*: Kinect timestamp records including audio and landmark data.
- *audio_ < Repetitive_index > .wav*: Original audio recordings.
- *audio_ < Repetitive_index > .bvh*: Landmarks of facial expression.
- *xethru_datafloat.dat*: Raw UWB radar signal.
- *timestamp_ < Repetitive_index > .mat*: UWB radar timestamps.

6.3.3 Processed Data Statement

Each subject's folder lies processed data in NPY format, with the same name of data files of *sample_ < Repetitive_index > .npy* in separate folder, detailed as follows:

- *kinect_processed*: Data from Kinect processed to obtain landmarks of *land_proc_ < Repetitive_index > .csv*, vocal features of *audio_proc_ < Repetitive_index > .wav* and video frame features of boxes and landmarks of *< frames_index > .npy*, and mouth frame of *< frames_index > .png*.
- *radar_processed*: FMCW radar data tailored for specific signal characteristics.
- *laser_processed*: Laser data processed pertinent to facial recognition tasks.
- *uwb_processed*: UWB radar data tailored for specific signal characteristics.

6.4 Technical validation

Effectiveness of the collected data is validated in three parts for validation and benchmark: signal processing and analysis, laser signals based speech separation, and multimodal based speech recognition.

6.4.1 Signals analysis

In this section, we analyse the entire process of lip motion and vibration of the vocal fold combined with video frames of the skeletal mouth and information on the voice, as shown in Fig. 6.6.

For UWB and FMCW radar signals, we transferred the raw data to the Doppler spectrum, shown with the speech spectrum and skeleton motion. The Fig. 6.6. shows all synchronized data types that were collected in the dataset. For UWB data, to sanitize the stationary object, the raw signals were first multiplied by a MTI filter, which is a radar process

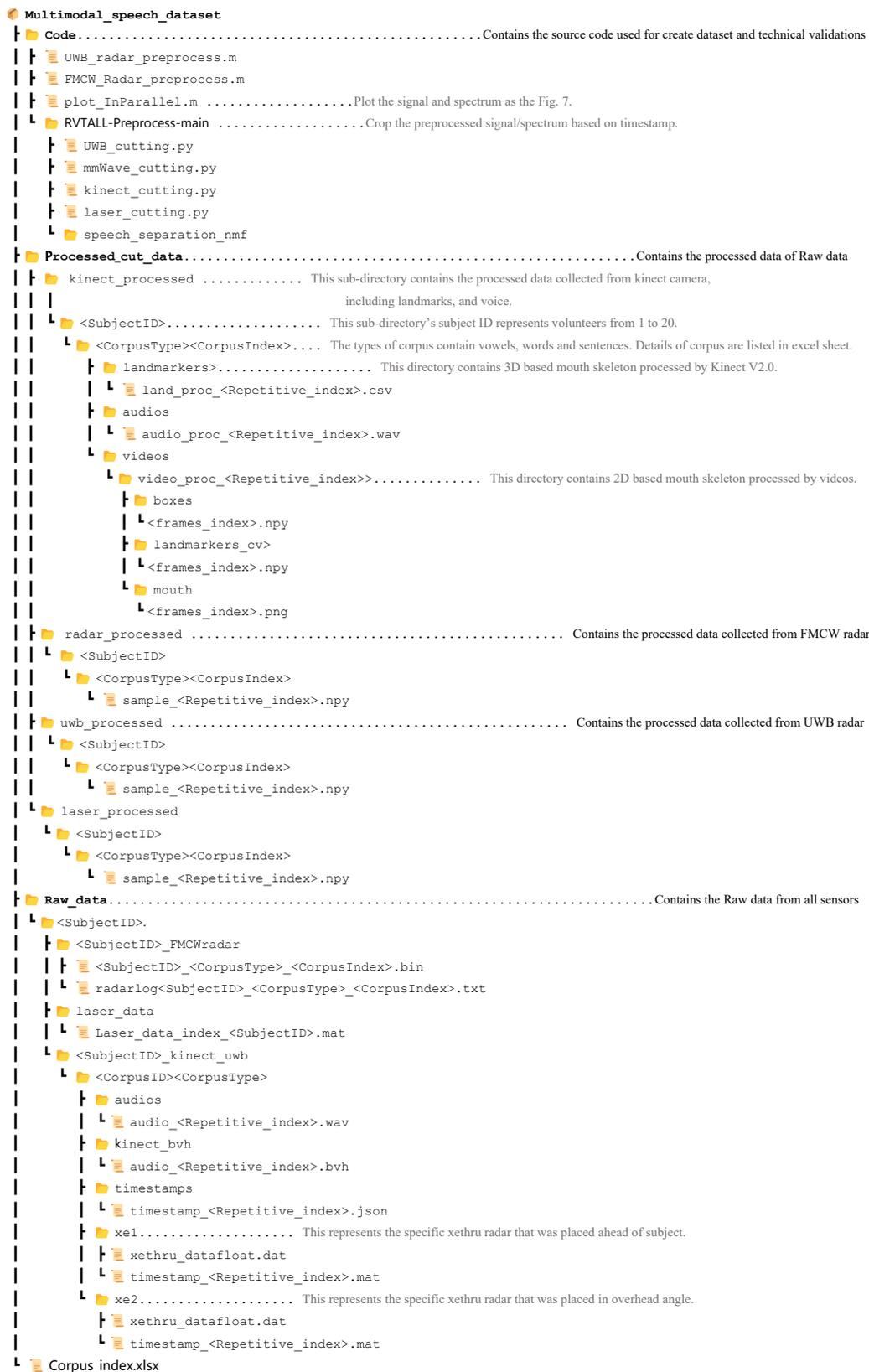


Figure 6.5: The structure of the multimodal speech dataset

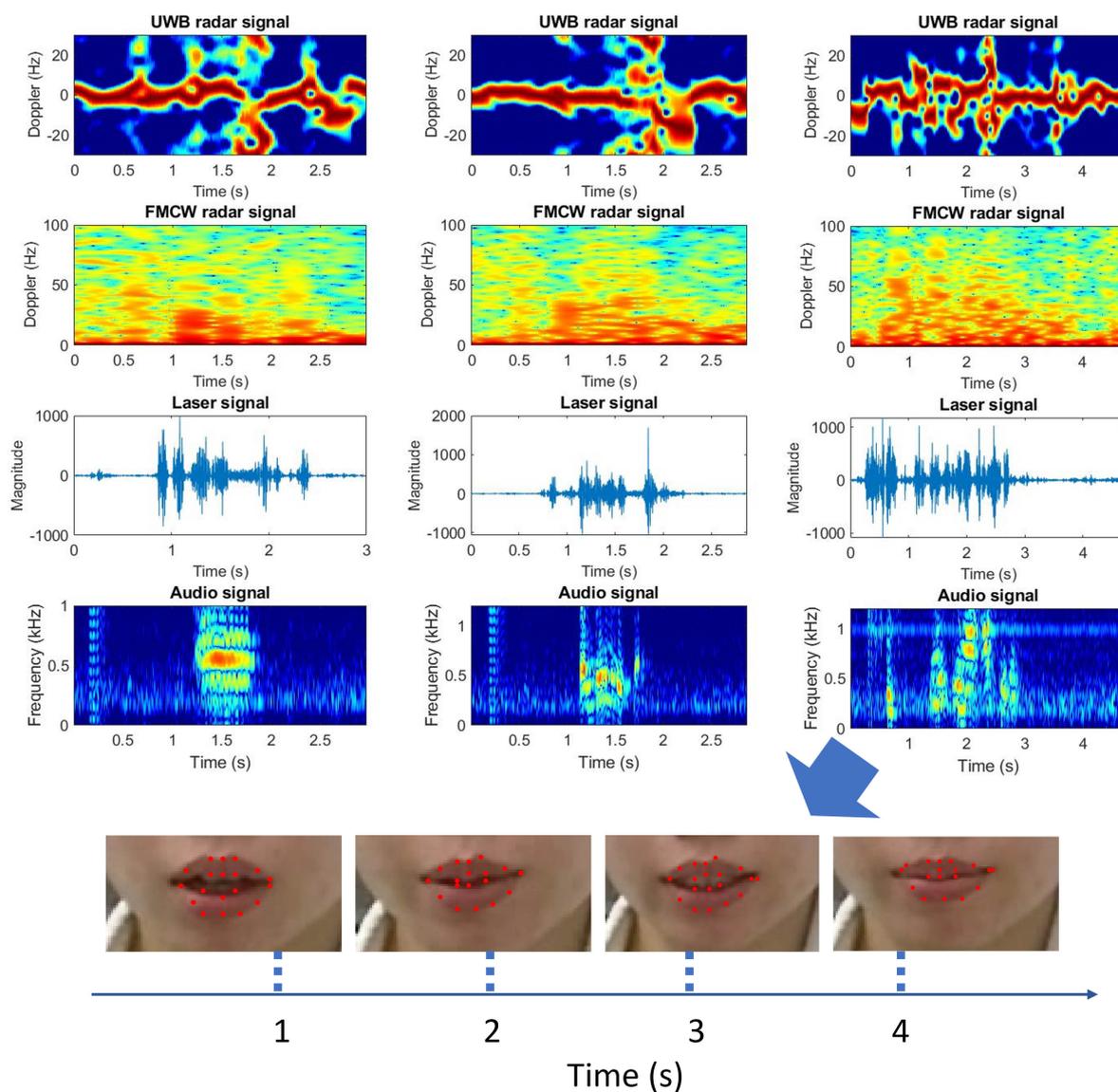


Figure 6.6: Multimodal data illustration including UWB and mmWave radar signal, laser, audio, image, and mouth skeleton points. From left to right columns, the first represents the volunteer is speaking of vowel '[ɔ:]', second is speaking of word 'bleed', and the third is speaking of sentence 'There's been an accident'. The last row illustrates the camera vision of the volunteer's mouth with the processed skeleton.

method that allows the UWB radar to detect and track targets that are moving in relation to the radar devices. From Eq. 4.1, we know that the channel impulses indicate different ranges. To consider all channel vibrations, we calculate the STFT result on each channel and then add all channels together, which is shown in Fig. 6.6.

For the FMCW radar, the primary step involves converting the IQ data to range-bin data via 1D-FFT. Notably, our approach emphasizes a distinct form of beamforming using FFT in the AoA dimension. This FFT based beamforming is designed to obtain precise angle information, capturing the location of radar signals and thus assisting in the accurate detection of movements, especially in the mouth and vocal folds. Instead of conventional methods that harness the velocity dimension with multiple chirps, our strategy elevates the frame rate to yield refined and continuous radar signals. With the vast capabilities of radars, it's essential to filter out superfluous data. We employed MTI to sieve out noise echoed by static objects. Subtracting range-angle spectrograms at certain intervals aids in diminishing false alarms due to substantial indoor clutter, optimizing the clarity of our results. Consequently, by focusing on the radar strength within our range of interest—which corresponds to the human's location, we could discern detailed information about mouth movement and vocal fold vibrations [209]. Given the predetermined speaker location, the IQ signal regarding that specific range-angle bin can be extracted and analyzed [210]. As the benchmark validation, we only extract the amplitude variation of radar signal.

Furthermore, we transferred the video to images with 30 frames per second and voice signals in spectrum, shown in Fig. 6.6 together with Doppler spectrograms of UWB and FMCW radar signals. To retrieve the sound signals from raw laser speckle data, an optical flow based method, notably the Farneback algorithm, is utilized to estimate the displacement of laser speckles on participants' faces. The input of this algorithm is every frame, denoted as a 2D function ($f(x, y)$), whereas a quadratic polynomial expansion is adopted to approximate the gray value of each pixel and its neighbours. The signals model based on the local coordinates of the selected pixel could be written as the Eq. 6.1.

$$f(x) = x^T M x + n^T + q \quad (6.1)$$

where x is the local coordinate (x, y) , M is a symmetric matrix equal to $\begin{bmatrix} C_4 & C_6/2 \\ C_6/2 & C_5 \end{bmatrix}$, n is a vector equal to $\begin{bmatrix} C_2 \\ C_3 \end{bmatrix}$ and q is a scalar equal to C_1 , C_1 to C_6 are the coefficients of the quadratic polynomial expansion. The new signals could be expressed using a displacement index Δd as the Eq. 6.2 indicates.

$$f(x - \Delta d) = (x - \Delta d)^T M_1 (x - \Delta d) + n_1^T (x - \Delta d) + q_1 g(x) = x^T M_2 x + n_2^T + q_2 \quad (6.2)$$

Simply let $f(x - \Delta d) = g(x)$, then we can get $n_2 = n_1 - 2M_1\Delta d$, leading to the solution of displacement index. Then the computed optical flow needs to be filtered with a band-pass filter. The cut-off frequency of the filter is chosen to be $80Hz$ and $255Hz$ for removing the frequency components caused by non-speaking activities such as head and skin movement. Then we integrate all sorts of cropped data that were mentioned above and show the matched samples in Fig. 6.6.

6.4.2 Speech separation task assisted by laser signal

Speech separation is used in a variety of applications, including telecommunications, hearing aids, speech recognition systems, and audio and video conferencing systems. These techniques typically involve analysing speech signals from different sources and then using filtering, spectral shaping, or other signals processing methods to remove or reduce unwanted components while preserving or enhancing the speech signal. Since the components of the human voice are mainly in the same frequency band, the frequencies of the different components are mixed together. Therefore, identification in the frequency domain alone is almost impossible.

Laser speckle signals are able to observe the vibration of skin covering cheek through capturing the laser speckle motion mentioned in Section 6.4.1. We empirically observe that there is a relationship between laser signals and audio signals. Previous work also approved that laser speckle signals can be used for audio denoising [201], which can be replicated with our dataset. Meanwhile, in our dataset, we introduce the laser speckle signals as the source of speech enhancement / separation in single channel case. As the validation example of speech enhancement application with laser speckle signal, we will introduce non-negative matrix factorization (NMF) based work which utilizes the processed laser signals to enhance the voice made from the speaker. Then the rest of the audio components of another speaker can be revealed. Validation experiments follow the scheme shown in Fig. 6.7. The input contains audio and laser signals from subject A, and delayed audio signals from subject B. First, we add up two audio input to simulate the complex environment of two volunteers speaking simultaneously. Then the laser and audio signals are synchronized with timestamp and frequency. Meanwhile, the audio signals were down-sampled to match the vibration frequency limit of the vocal folds. After that, the applied STFT with hanning window to get spectrogram of laser and audio signal. Then we applied NMF to the magnitude spectrogram to obtain a set of basis matrix and corresponding activation coefficients for each source. The decomposed basis matrices reflect the speech characteristic in the time domain. Two clusters of matrices can be separated depending on whether they are correlated with reference laser signals. Finally, we reconstruct the audio with time-frequency mask that was clustered. The detailed steps can be referred to the code with the dataset.

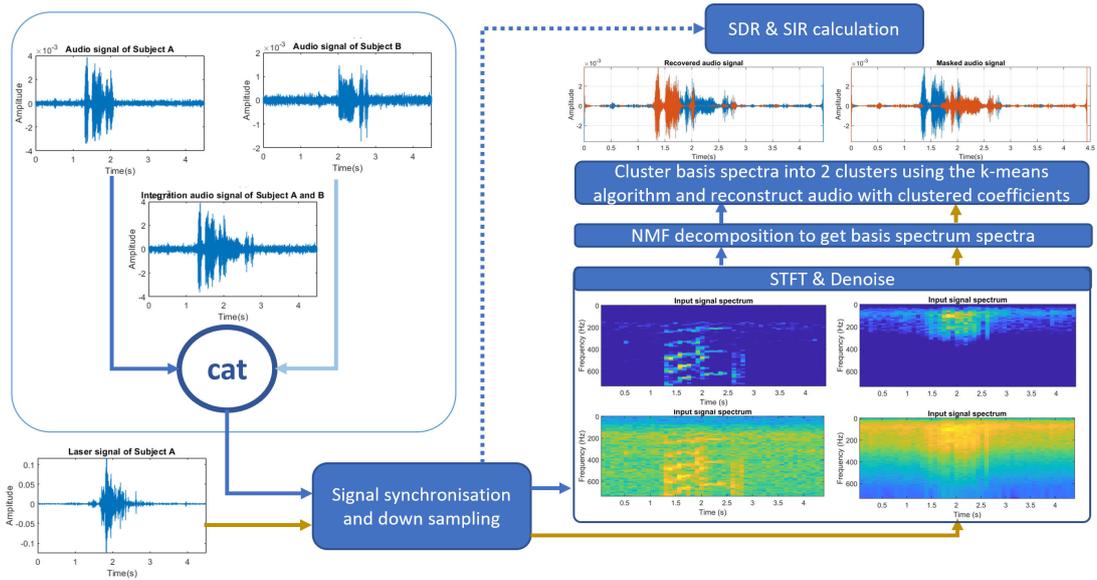


Figure 6.7: Speech separation scheme of mixed audio signals and laser-speckle signals using a NMF based method

To explore the performance of our scheme, we selected the collected laser-audio dataset of sentences from No.17 and 18 subjects. We named the audio signals of laser target as recovered audio signals, and another as masked signals. The average signal-to-distortion ratio (SDR) and the signal-to-interference ratio (SIR) of recovered signals are 1.04 dB and 35.97 dB respectively. Meanwhile, the performance of masked signals that were not correlated with laser signals, gains 0.0002 dB of SDR and 35.97 dB of SIR. Overall, this work provides another view of single channel based speech enhancement and separation tasks. Meanwhile, the data collected from third scenario provides raw signals under two volunteers' speech, which is a good source for users to test their methods.

6.4.3 Multimodal speech recognition

For the benchmark of speech contents classification, we selected mmWave radar, UWB radar, laser, audio and video data of 5 subjects and established a CNN based ResNet classification network. The batch size is 16 and the primary learning rate is 0.01. Each model uses 80% data for training and 20% for validation and is trained for 50 to 100 epochs. The classification performance of multiple sources is shown in Fig. 6.8. It is evident that there is potential to improve speech recognition performance based on non-invasive UWB and laser technology.

Besides, we further considered a sensor fusion scheme that combines data from UWB radar and multiple sources for word recognition. We employed a multi-input ResNet18 for this task, which includes two input blocks consisting of a convolutional layer, a batch normalization layer, and a ReLU activation, as shown in Fig. 6.9. The initial feature extraction

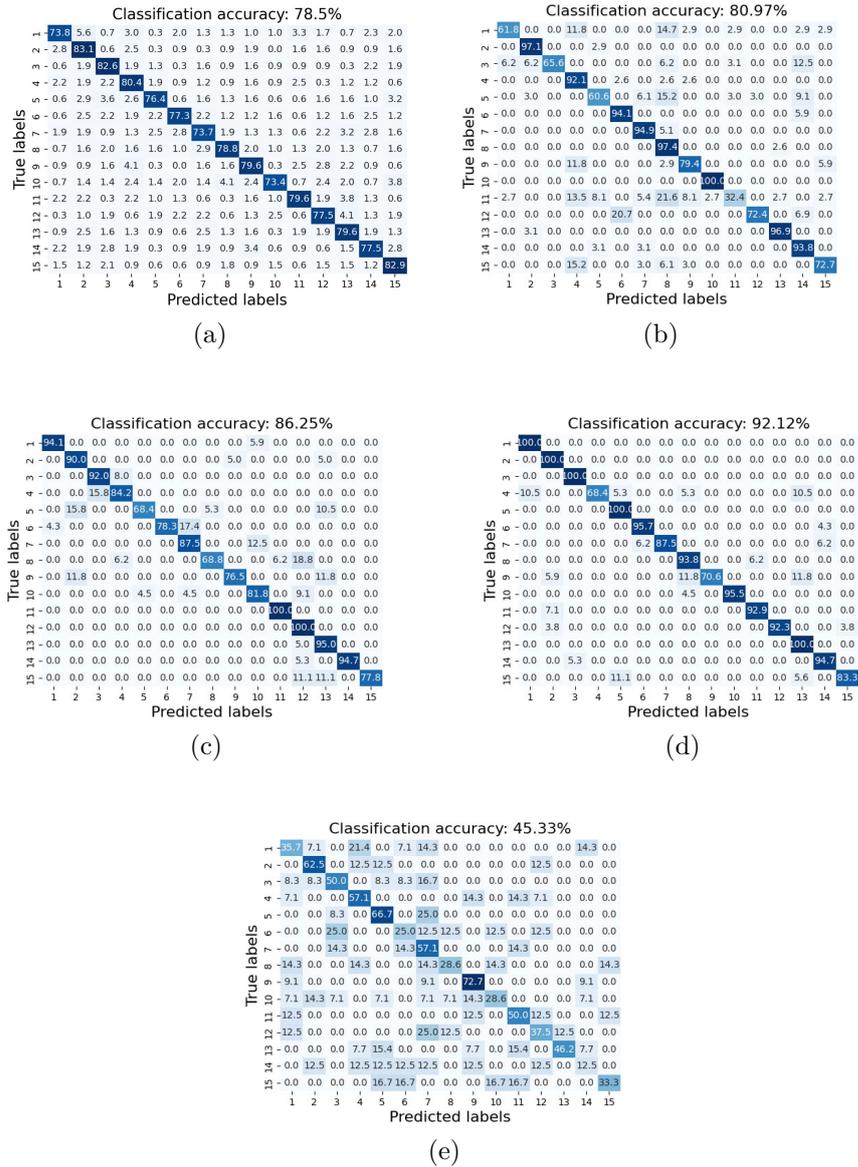


Figure 6.8: Classification performance of human speech across 15 Words with confusion matrix of (a) mmWave radar (b) UWB radar, (c) video stream, (d) audio signal, (e) laser speckle signals.

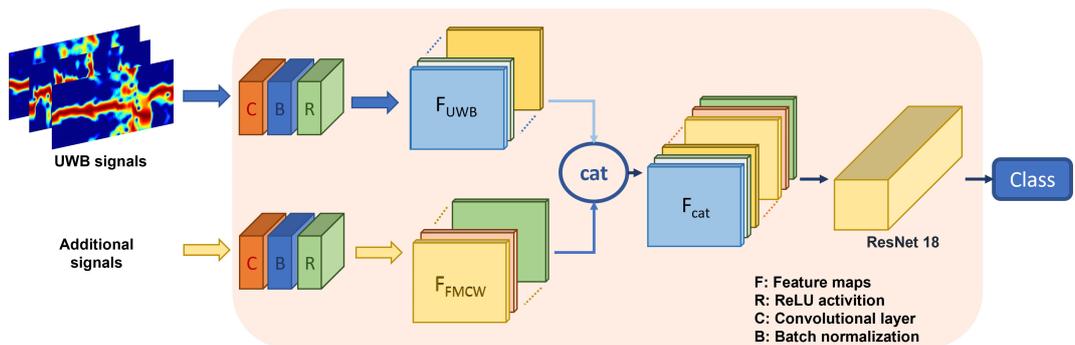


Figure 6.9: Multi-modal sensor fusion scheme for word speech classification.

was completed by feeding spectrograms from the UWB radar and additional information into their respective input blocks. The resulting feature maps were then stacked along the channel axis and processed by ResNet18 for final analysis. Then, we also applied audio data on word recognition as comparison with radio based methods. The performance of the multimodal based recognition systems is shown in Fig. 6.10. Meanwhile, we adopt the methods on sentences classification with the result shown in Fig. 6.11.

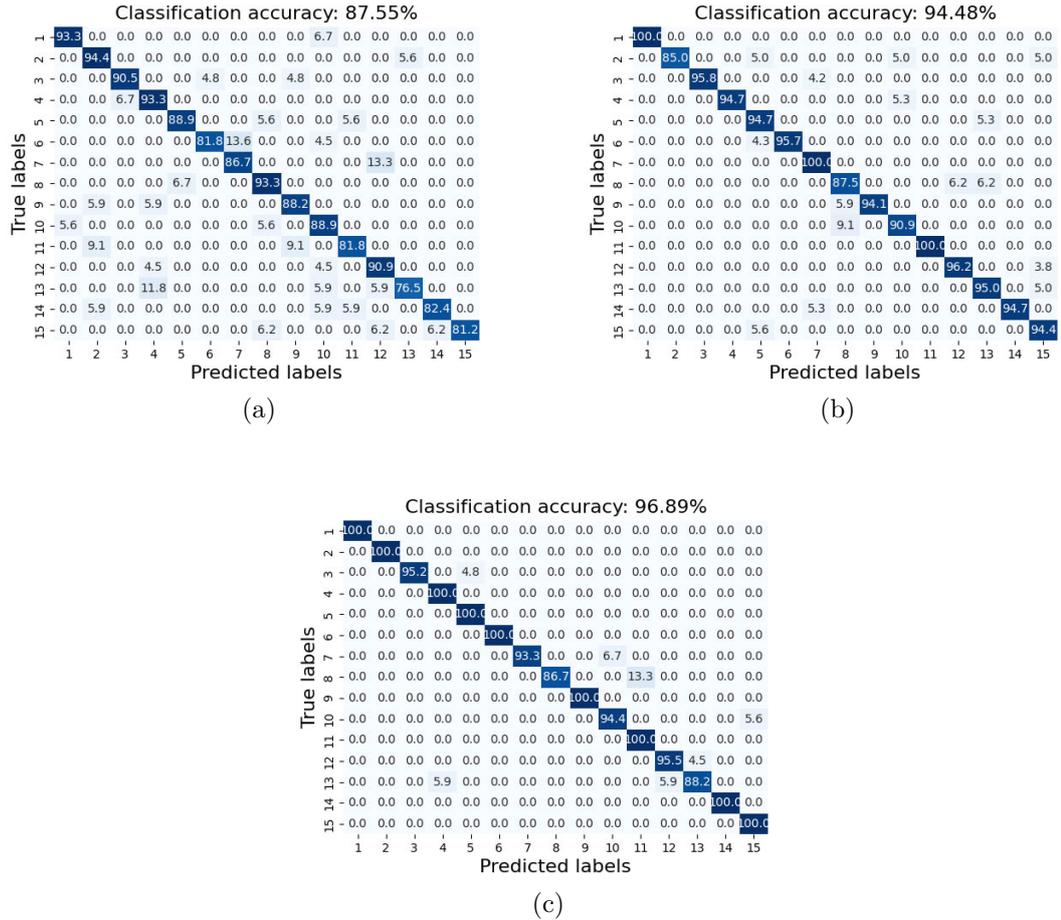


Figure 6.10: Classification performance of human speech across 15 Words with confusion matrix of (a) the fusion of video and UWB, (b) the fusion of audio and UWB, (c) the fusion of video, audio and UWB.

Another challenge encountered in real-world applications is the varied "context". For instance, in our scenarios, the dynamic distance between a target and the receiver may cause a substantial performance decline for radar based models trained on a fixed-distance dataset. However, obtaining new data samples to account for different "contexts" is often impractical. Consequently, we further explore the use of our dataset in transfer learning tasks, with the goal of enhancing the scalability of radar based lip-reading systems. By fine-tuning a pre-trained model with minimal additional data, we strive to address the challenges posed by diverse contexts. We carried out experiments using UWB signals gathered from volunteer 4 at three distinct distances. The pre-training phase was conducted on data from the first distance, employing a batch size of 16, a learning rate of 0.01,

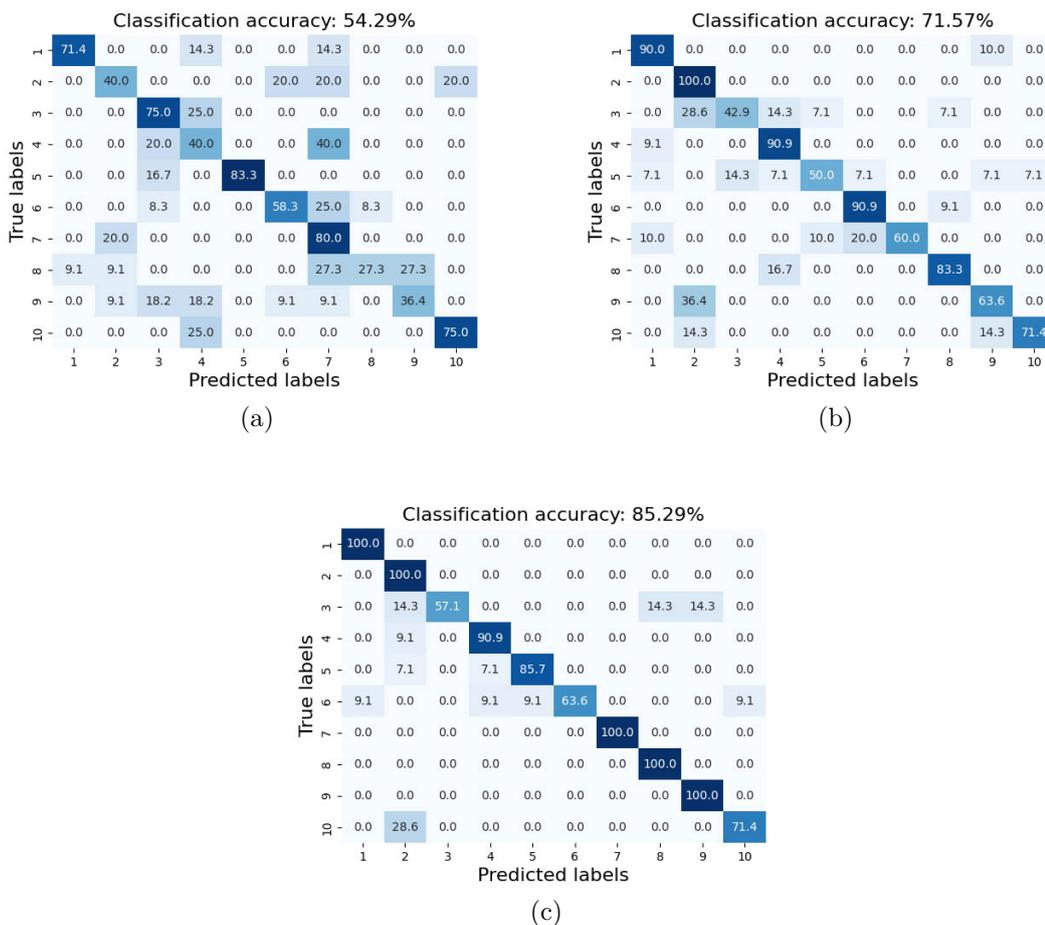


Figure 6.11: Classification performance of human speech across 10 Sentences with confusion matrix of (a) UWB signals, (b) the fusion of video and UWB, (c) the fusion of audio and UWB.

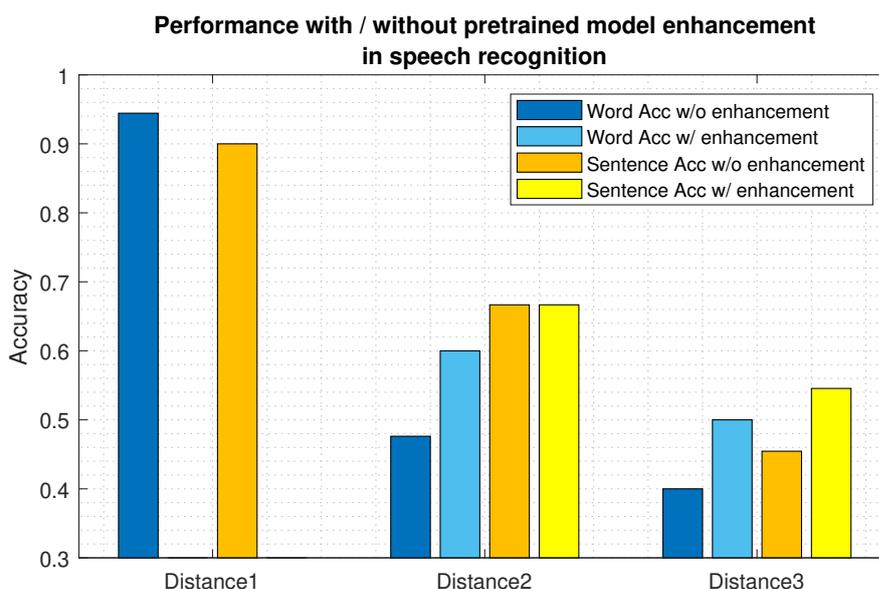


Figure 6.12: Comparative performance with or without pretrained model enhancement in speech recognition

and 50 to 100 training epochs. Subsequently, we assessed the model's performance at the second and third distances, both with and without fine-tuning (utilizing the same training settings as the pre-training phase), focusing on its SSR performance. A comparison of sentence and word performance revealed that complex sentences presented greater challenges for SSR with radar. Additionally, we examined the impact of varying distances between the user and sensors in Fig. 6.12, which corresponds to the second scenario discussed in Section 6.3.

6.5 Summary

As a significant part of small-scale motion activity, speech related lip and vocal fold reading is crucial for understanding the how wireless sensing mechanism work in speech recognition. Our research introduces RVTALL, a unique multimodal dataset designed to advance speech recognition through noninvasive remote sensing techniques. Integrates a diverse range of data sources, including Radio Frequency, visual, text, audio, laser, and lip landmark information. The dataset features multiple data types, such as *7.5GHz* CIR from UWB radars, *77GHz* radar data from mmWave FMCW radar, visual and audio information, lip landmarks, and laser data. Utilizing a depth camera, we recorded detailed lip landmarks and voices from 20 participants, resulting in 400 minutes of annotated speech profiles encompassing vowels, words, and sentences. This comprehensive dataset has been rigorously validated and shows great promise for lip reading and multimodal speech recognition studies, offering new perspectives in the field.

Conclusions and Future Expectations

7.1 Conclusions

The thesis firstly presents a literature review of the advancements and applications of various technologies in healthcare monitoring systems, with a particular focus on RF sensing. The discussion spans across multiple aspects of this field, starting with the role of RF sensing in non-hospital environments. Here, the Chapter 2 sheds light on how RF signals are instrumental in monitoring vital signs such as heart and lung movements, thereby facilitating real-time detection of respiration and heartbeat.

The narrative then transitions to the concept of JCAS, highlighting its significance in the realm of the IoT. This convergence of sensing and communication in a single system underscores the efficiency, cost-effectiveness, and enhanced performance that these integrated devices offer. Furthermore, the chapter delves into the applications of WiFi sensing technologies in healthcare. It elaborates on their role in human activity recognition, vital sign monitoring, and fall detection, while also emphasizing their advantages in maintaining privacy, enabling through-wall sensing, and their cost-effectiveness. The exploration of technologies extends to LoRa sensing, an emerging technology noted for its long-range communication capabilities and low power requirements. This makes it particularly suitable for applications in large-scale urban health monitoring or in scenarios requiring remote monitoring. Additionally, the chapter discusses the use of radar sensing in vital detection, particularly its effectiveness in estimating the position and movement speed of targets, which includes non-invasive monitoring of patients' vital signs.

In the third chapter, a thorough exploration of WiFi sensing for HAR within the context of intelligent IoT and healthcare monitoring systems is undertaken. The importance of indoor human activity recognition in IoT and healthcare is initially highlighted, setting the stage for a detailed review of the evolution and advancements in device-free WiFi sensing technologies.

Focus is then shifted to the development of various methods and systems aimed at enhancing the accuracy and efficiency of WiFi based sensing for activity recognition from single device. The introduction of the Continuous Angle of Arrival-Time of Flight maps methodology, an innovation in preserving both temporal and spatial human body information through WiFi signals, is discussed. Furthermore, the design of specialized lightweight temporal neural networks, tailored for these sensing systems, is elaborated upon.

The experimental setups, including calibration techniques and system designs, are meticulously outlined. The processes involved in respiration estimation using sanitized CSI phase, the evaluation of Doppler based recognition systems, and the construction and assessment of AoA-ToF maps are described in depth. In the results and discussion section, the outcomes of various experiments are presented. The efficacy of the proposed methods in accurately detecting and recognizing human activities is demonstrated. Comparisons are made with other HAR systems to underscore the advantages and identify areas for potential improvement. It highlights the progress in HAR using single-end WiFi, focusing on the incorporation of spatial phase variation from CSI data. This exploration into single terminal contexts, enriched with spatiotemporal data, has enhanced HAR's effectiveness, building upon previous research.

Although WiFi based system is potential to most of activity detection, due to the power limitation of devices and protocols, it cannot be used to detect long range scenario. In the next chapter, thesis presents the LoGait system, an LoRa protocol based approach for human gait recognition in various indoor scenarios, including living rooms, through-wall scenarios, and corridors. The system analyses the feasibility of LoRa sensing and employs conjugated multiplication, along with multiple preprocessing methods, to extract LoRa gait profiles. These methods focus on filtering communication symbols and extracting physical variation information from human gait. User identity recognition is achieved using a DTW based machine learning algorithm. Experimental evaluations in three scenarios demonstrate promising results, with an identification accuracy of 85.13% in living rooms, 79.13% in through-wall scenarios, and 84.14% in 20m corridors. By addressing complex indoor environments, including through-wall scenarios and long-distance corridors, the LoGait system represents a significant advancement in gait recognition prototypes with great potential.

On the other hand, instead of only focusing on JCAS device, radar device has ability to locate the target with specific RF signal modulation. Meanwhile, like WiFi signal, radar is able to detect the tiny motion from human being like vital signs and speech. Chapter 5 of the thesis presents a detailed exploration of radar based respiration detection, emphasizing the application of ellipse normalization for enhanced accuracy. It begins with an overview of the significance of remote vital sign monitoring in healthcare, highlighting the potential of FMCW radar in this domain. The chapter identifies challenges in implementing radar technology for detecting subtle body movements and adjustments in dynamic environments. It introduces preprocessing methods such as Range-Angle maps and Constant False Alarm Rate techniques. The ellipse normalization method is proposed to counteract subtle motion distortions. The chapter also describes methodologies for accurate respiration detection system using advanced filtering techniques and rate estimation algorithms. The performance of the system is evaluated through experimental setups, benchmark comparisons, and an ablation study, demonstrating the effectiveness of each system component in the respiration detection process. Like respiration detection, radar based speech recognition has shown a growing interest in small-scale motion detection through non-invasive remote sensing techniques.

In Chapter 6, we aim to contribute to this area through our dataset paper, which is designed to enhance and restore speech information from a variety of data sources. This chapter introduces RVTALL, a novel multimodal dataset that integrates Radio Frequency, visual, text, audio, laser, and lip landmark information, offering a comprehensive approach to speech recognition research. Our dataset comprises a rich amalgamation of data types: $7.5GHz$ CIR from UWB radars, $77GHz$ radar data from mmWave FMCW radar, alongside visual and audio information, lip landmarks, and laser data. This unique multimodal composition paves the way for advanced research in speech recognition. Furthermore, we have employed a depth camera to meticulously record the lip landmarks and voice of the subjects. Our dataset encompasses approximately 400 minutes of annotated speech profiles, meticulously collected from 20 participants. These profiles include the articulation of 5 vowels, 15 words, and 16 sentences, providing a comprehensive range for analysis. The dataset has undergone rigorous validation and shows immense potential for use in studies focusing on lip reading and multimodal speech recognition. This work not only contributes a novel dataset to the field, but also opens new avenues for exploring speech recognition through a multimodal lens.

7.2 Future Research

The direction of author's future research will focus on enhancing multimodal RF sensing technologies for more robust and diverse human-centric applications in IoT systems. Specifically, there are four main points of topic for author's future research.

- **Advancing RF Sensing Technologies in Systems:**

The future research will concentrate on enhancing the capabilities of multimodal RF sensing technologies. This advancement aims to bolster robust and diverse healthcare applications, particularly within IoT systems. The focus will be on refining these technologies to make them more efficient, versatile, and suited to the dynamic needs of human-computer interaction, ensuring seamless integration and functionality within IoT frameworks.

Meanwhile, the essential work for this target will be the continuous expansion and diversification of multimodal datasets. This involves integrating various data types, such as RF, visual, and audio information, to create comprehensive datasets. These datasets will support diverse types of recognition research, opening new avenues in non-invasive, remote healthcare monitoring and human activity recognition. The goal is to leverage these datasets to revolutionize healthcare monitoring and diagnosis, providing innovative solutions for modern healthcare challenges.

- **Precision Enhancement in WiFi and LoRa Technologies for Activity and Vital Sign Monitoring:**

A significant part of future research will involve advancing WiFi and LoRa technologies to achieve greater precision in human activity recognition and vital sign monitoring. This entails improving the accuracy and reliability of these technologies, especially in varying environmental conditions. The goal is to adapt these technologies for more effective indoor and remote monitoring, catering to the diverse requirements of modern healthcare systems.

- **Improving Radar based Small Scale Detection in Dynamic Environments:**

Research will also focus on enhancing radar based detection methods, especially for small-scale applications and in dynamic settings. This includes developing new technologies and refining existing methods to increase the radar's sensitivity and precision. The aim is to effectively detect subtle movements and physiological changes in various dynamic environments, thereby improving the overall accuracy and reliability of radar based healthcare monitoring systems.

- **Integration of Digital Twin and Sensor Technologies for Quality of Life Enhancement:**

Future research will explore the integration of digital twin and sensor technologies, aiming to meet the growing demands for improved quality of life. This direction involves harnessing the advancements in these technologies to capture comprehensive data on daily life and health parameters. The objective is to use this data to enhance living standards, facilitate better health monitoring, and provide more personalized healthcare solutions.

Bibliography

- [1] P.-Y. Hsu, P.-H. Hsu, T.-H. Lee and H.-L. Liu, ‘Heart rate and respiratory rate monitoring using seismocardiography’, in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, 2021, pp. 6876–6879. doi: 10.1109/EMBC46164.2021.9630298.
- [2] S. Patel, H. Park, P. Bonato, L. Chan and M. Rodgers, *A review of wearable sensors and systems with application in rehabilitation*, 2012. doi: 10.1186/1743-0003-9-21.
- [3] L. P. Malasinghe, N. Ramzan and K. Dahal, ‘Remote patient monitoring: a comprehensive study’, *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 1, pp. 57–76, 2019, issn: 18685145. doi: 10.1007/s12652-017-0598-x. [Online]. Available: <http://dx.doi.org/10.1007/s12652-017-0598-x>.
- [4] Q. Huang, H. Chen and Q. Zhang, ‘Joint Design of Sensing and Communication Systems for Smart Homes’, *IEEE Network*, vol. 34, no. 6, pp. 191–197, 2020, issn: 1558156X. doi: 10.1109/MNET.011.2000107.
- [5] S. Majumder, T. Mondal and M. J. Deen, ‘Wearable sensors for remote health monitoring’, *Sensors (Switzerland)*, vol. 17, no. 1, 2017, issn: 14248220. doi: 10.3390/s17010130.
- [6] D. Xiong, D. Zhang, X. Zhao and Y. Zhao, ‘Deep learning for emg-based human-machine interaction: A review’, *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 3, pp. 512–533, 2021.
- [7] M. Andriluka, L. Pishchulin, P. Gehler and B. Schiele, ‘2d human pose estimation: New benchmark and state of the art analysis’, in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3686–3693. doi: 10.1109/CVPR.2014.471.
- [8] W. Taylor, S. A. Shah, K. Dashtipour, A. Zahid, Q. H. Abbasi and M. A. Imran, ‘An intelligent non-invasive real-time human activity recognition system for next-generation healthcare’, *Sensors*, vol. 20, no. 9, p. 2653, 2020.
- [9] Y. Ge, W. Li, M. Farooq *et al.*, ‘Logait: Lora sensing system of human gait recognition using dynamic time wrapping’, *IEEE Sensors Journal*, 2023.
- [10] M. Usman, J. Rains, T. J. Cui, M. Z. Khan, M. A. Imran, Q. H. Abbasi *et al.*, ‘Intelligent wireless walls for contactless in-home monitoring’, *Light: Science & Applications*, vol. 11, pp. 212–224, 2022.

- [11] W. Li, B. Tan and R. Piechocki, ‘Passive radar for opportunistic monitoring in e-health applications’, *IEEE journal of translational engineering in health and medicine*, vol. 6, pp. 1–10, 2018.
- [12] Y. Ge, A. Taha, S. Shah *et al.*, ‘Contactless wifi sensing and monitoring for future healthcare - emerging trends, challenges and opportunities’, *IEEE Reviews in Biomedical Engineering*, pp. 1–1, 2022. doi: 10.1109/RBME.2022.3156810.
- [13] Y. Ma, G. Zhou and S. Wang, ‘Wifi sensing with channel state information: A survey’, *ACM Computing Surveys (CSUR)*, vol. 52, no. 3, pp. 1–36, 2019.
- [14] W. Wang, A. X. Liu, M. Shahzad, K. Ling and S. Lu, ‘Device-free human activity recognition using commercial wifi devices’, *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1118–1131, 2017.
- [15] J. Ding and Y. Wang, ‘Wifi csi-based human activity recognition using deep recurrent neural network’, *IEEE Access*, vol. 7, pp. 174 257–174 269, 2019. doi: 10.1109/ACCESS.2019.2956952.
- [16] L. Zhang, C. Wang, M. Ma and D. Zhang, ‘Widigr: Direction-independent gait recognition system using commercial wi-fi devices’, *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 1178–1191, 2020. doi: 10.1109/JIOT.2019.2953488.
- [17] D. Wang, J. Yang, W. Cui, L. Xie and S. Sun, ‘Caution: A robust wifi-based human authentication system via few-shot open-set recognition’, *IEEE Internet of Things Journal*, vol. 9, no. 18, pp. 17 323–17 333, 2022. doi: 10.1109/JIOT.2022.3156099.
- [18] Z. Zhang, Z. Tian and M. Zhou, ‘Latern: Dynamic continuous hand gesture recognition using fmcw radar sensor’, *IEEE Sensors Journal*, vol. 18, no. 8, pp. 3278–3289, 2018. doi: 10.1109/JSEN.2018.2808688.
- [19] S. P. Rana, M. Dey, M. Ghavami and S. Dudley, ‘Markerless gait classification employing 3d ir-uwb physiological motion sensing’, *IEEE Sensors Journal*, vol. 22, no. 7, pp. 6931–6941, 2022. doi: 10.1109/JSEN.2022.3154092.
- [20] H. Li, J. Le Kernec, A. Mehul, S. Z. Gurbuz and F. Fioranelli, ‘Distributed radar information fusion for gait recognition and fall detection’, in *2020 IEEE Radar Conference (RadarConf20)*, IEEE, 2020, pp. 1–6.
- [21] C. Ding, L. Zhang, C. Gu *et al.*, ‘Non-contact human motion recognition based on uwb radar’, *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 8, no. 2, pp. 306–315, 2018. doi: 10.1109/JETCAS.2018.2797313.
- [22] A. Jayatilaka and D. C. Ranasinghe, ‘Real-time fluid intake gesture recognition based on batteryless uhf rfid technology’, *Pervasive and Mobile Computing*, vol. 34, pp. 146–156, 2017.
- [23] F. Zhang, Z. Chang, K. Niu *et al.*, ‘Exploring lora for long-range through-wall sensing’, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 2, pp. 1–27, 2020.
- [24] B. Xie and J. Xiong, ‘Combating interference for long range lora sensing’, in *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, 2020, pp. 69–81.

- [25] J. Qi, P. Yang, L. Newcombe, X. Peng, Y. Yang and Z. Zhao, ‘An overview of data fusion techniques for internet of things enabled physical activity recognition and measure’, *Information Fusion*, vol. 55, pp. 269–280, 2020.
- [26] A. Rashed, A. Ibrahim, A. Adel *et al.*, ‘Integrated iot medical platform for remote healthcare and assisted living’, in *2017 Japan-Africa Conference on Electronics, Communications and Computers (JAC-ECC)*, 2017, pp. 160–163. doi: 10.1109/JAC-ECC.2017.8305801.
- [27] Y. Ge, S. Ansari, A. Abdulghani, M. A. Imran and Q. H. Abbasi, ‘Intelligent instruction-based iot framework for smart home applications using speech recognition’, in *2020 IEEE International Conference on Smart Internet of Things (SmartIoT)*, 2020, pp. 197–204. doi: 10.1109/SmartIoT49966.2020.00037.
- [28] Y. Wang, M. Li and M. Li, ‘The statistical analysis of iee 802.11 wireless local area network–based received signal strength indicator in indoor location sensing systems’, *International Journal of Distributed Sensor Networks*, vol. 13, no. 12, p. 1550147717747858, 2017. doi: 10.1177/1550147717747858.
- [29] W. Wang, A. X. Liu, M. Shahzad, K. Ling and S. Lu, ‘Understanding and modeling of wifi signal based human activity recognition’, in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, ser. MobiCom ’15, Paris, France: Association for Computing Machinery, 2015, pp. 65–76, isbn: 9781450336192. doi: 10.1145/2789168.2790093. [Online]. Available: <https://doi.org/10.1145/2789168.2790093>.
- [30] Z. Wang, K. Jiang, Y. Hou *et al.*, ‘A Survey on CSI-Based Human Behavior Recognition in Through-The-Wall Scenario’, *IEEE Access*, vol. 7, pp. 78772–78793, 2019, issn: 21693536. doi: 10.1109/ACCESS.2019.2922244.
- [31] D. Halperirr, W. Hu, A. Sheth and D. Wetherall, ‘Tool release: Gathering 802.11n traces with channel state information’, *Acm Sigcomm Computer Communication Review*, vol. 41, no. 1, pp. 53–53, 2011.
- [32] Y. Xie, Z. Li and M. Li, ‘Precise power delay profiling with commodity wi-fi’, *IEEE Transactions on Mobile Computing*, vol. 18, no. 6, pp. 1342–1355, 2019. doi: 10.1109/TMC.2018.2860991.
- [33] J. Yang, H. Zou, H. Jiang and L. Xie, ‘Device-Free Occupant Activity Sensing Using WiFi-Enabled IoT Devices for Smart Homes’, *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3991–4002, 2018, issn: 23274662. doi: 10.1109/JIOT.2018.2849655.
- [34] Y. Xie, Z. Li and M. Li, ‘Precise power delay profiling with commodity wifi’, in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, ser. MobiCom ’15, Paris, France: ACM, 2015, pp. 53–64, isbn: 978-1-4503-3619-2. doi: 10.1145/2789168.2790124. [Online]. Available: <http://doi.acm.org/10.1145/2789168.2790124>.

- [35] M. Atif, S. Muralidharan, B. Yoo and H. Ko, ‘Wi-ESP — A tool for CSI-based Device-Free Wi-Fi Sensing (DFWS)’, vol. 7, no. May, pp. 644–656, 2020. doi: 10.1093/jcde/qwaa048.
- [36] Y. Ma, G. Zhou and S. Wang, ‘WiFi sensing with channel state information: A survey’, *ACM Computing Surveys*, vol. 52, no. 3, 2019, issn: 15577341. doi: 10.1145/3310194.
- [37] M. Rehman, R. A. Shah, M. B. Khan *et al.*, ‘Contactless small-scale movement monitoring system using software defined radio for early diagnosis of covid-19’, *IEEE Sensors Journal*, vol. 21, no. 15, pp. 17180–17188, 2021. doi: 10.1109/JSEN.2021.3077530.
- [38] S. A. K. Tanoli, M. Rehman, M. B. Khan *et al.*, ‘An experimental channel capacity analysis of cooperative networks using universal software radio peripheral (usrp)’, *Sustainability*, vol. 10, no. 6, p. 1983, 2018.
- [39] B. Tan, Q. Chen, K. Chetty, K. Woodbridge, W. Li and R. Piechocki, ‘Exploiting wifi channel state information for residential healthcare informatics’, *IEEE Communications Magazine*, vol. 56, no. 5, pp. 130–137, 2018. doi: 10.1109/MCOM.2018.1700064.
- [40] T. Z. Chowdhury, ‘Using Wi-Fi channel state information (CSI) for human activity recognition and fall detection’, pp. 24–50, 2018. [Online]. Available: <https://open.library.ubc.ca/cIRcle/collections/ubctheses/24/items/1.0365967>.
- [41] Z. Chen, L. Zhang, C. Jiang, Z. Cao and W. Cui, ‘WiFi CSI Based Passive Human Activity Recognition Using Attention Based BLSTM’, *IEEE Transactions on Mobile Computing*, vol. 18, no. 11, pp. 2714–2724, 2019, issn: 15580660. doi: 10.1109/TMC.2018.2878233.
- [42] S. Zhou, W. Zhang, D. Peng, Y. Liu, X. Liao and H. Jiang, ‘Adversarial wifi sensing for privacy preservation of human behaviors’, *IEEE Communications Letters*, vol. 24, no. 2, pp. 259–263, 2020. doi: 10.1109/LCOMM.2019.2952844.
- [43] Q. Chen, Y. Liu, B. Tan, K. Woodbridge and K. Chetty, ‘Respiration and Activity Detection Based on Passive Radio Sensing in Home Environments’, *IEEE Access*, vol. 8, pp. 12426–12437, 2020, issn: 21693536.
- [44] Y. Ge, S. Li, M. Shentu *et al.*, ‘A doppler-based human activity recognition system using wifi signals’, in *2021 IEEE Sensors*, 2021, pp. 1–4. doi: 10.1109/SENSORS47087.2021.9639680.
- [45] Y. Wang, K. Wu and L. M. Ni, ‘WiFall: Device-Free Fall Detection by Wireless Networks’, *IEEE Transactions on Mobile Computing*, vol. 16, no. 2, pp. 581–594, 2017, issn: 15580660. doi: 10.1109/TMC.2016.2557792.
- [46] X. Zheng, J. Wang, L. Shangguan, Z. Zhou and Y. Liu, ‘Design and implementation of a csi-based ubiquitous smoking detection system’, *IEEE/ACM Transactions on Networking*, vol. 25, no. 6, pp. 3781–3793, 2017. doi: 10.1109/TNET.2017.2752367.

- [47] J. Yang, H. Zou, H. Jiang and L. Xie, ‘CareFi: Sedentary Behavior Monitoring System via Commodity WiFi Infrastructures’, *IEEE Transactions on Vehicular Technology*, vol. 67, no. 8, pp. 7620–7629, 2018, issn: 00189545. doi: 10.1109/TVT.2018.2833388.
- [48] F. Wang, S. Zhou, S. Panev, J. Han and D. Huang, ‘Person-in-wifi: Fine-grained person perception using wifi’, in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.
- [49] W. Jiang, H. Xue, C. Miao *et al.*, ‘Towards 3D human pose construction using wifi’, *Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM*, pp. 295–308, 2020. doi: 10.1145/3372224.3380900.
- [50] L. Guo, Z. Lu, S. Zhou, X. Wen and Z. He, ‘When healthcare meets off-the-shelf wifi: A non-wearable and low-Costs approach for in-home monitoring’, *arXiv*, 2020, issn: 23318422. arXiv: 2009.09715.
- [51] S. Zhou, L. Guo, Z. Lu, X. Wen, W. Zheng and Y. Wang, ‘Subject-independent human pose image construction with commodity wi-fi’, in *ICC 2021 - IEEE International Conference on Communications*, 2021, pp. 1–6. doi: 10.1109/ICC42927.2021.9500994.
- [52] K. Ali, A. X. Liu, W. Wang and M. Shahzad, ‘Recognizing Keystrokes Using WiFi Devices’, *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1175–1190, 2017, issn: 07338716. doi: 10.1109/JSAC.2017.2680998.
- [53] G. Wang, Y. Zou, Z. Zhou, K. Wu and L. M. Ni, ‘We Can Hear You with Wi-Fi!’, *IEEE Transactions on Mobile Computing*, vol. 15, no. 11, pp. 2907–2920, 2016, issn: 15361233. doi: 10.1109/TMC.2016.2517630.
- [54] M. M. Hassan, A. Gumaei, G. Aloï, G. Fortino and M. Zhou, ‘A smartphone-enabled fall detection framework for elderly people in connected home healthcare’, *IEEE Network*, vol. 33, no. 6, pp. 58–63, 2019. doi: 10.1109/MNET.001.1900100.
- [55] L. Guan, F. Hu, F. Al-Turjman, M. B. Khan and X. Yang, ‘A Non-Contact Paraparesis Detection Technique Based on 1D-CNN’, *IEEE Access*, vol. 7, pp. 182 280–182 288, 2019, issn: 21693536. doi: 10.1109/ACCESS.2019.2959023.
- [56] A. Tahir, J. Ahmad, S. A. Shah *et al.*, ‘WiFreeze : Multiresolution Scalograms for Freezing of Gait Detection in Parkinson ’ s Leveraging 5G Spectrum with Deep Learning’, pp. 1–18, 2019. doi: 10.3390/electronics8121433.
- [57] M. B. Khan, Z. Zhang, L. Li *et al.*, ‘A systematic review of non-contact sensing for developing a platform to contain covid-19’, *Micromachines*, vol. 11, no. 10, pp. 1–23, 2020, issn: 2072666X. doi: 10.3390/mi11100912.
- [58] M. Alizadeh, G. Shaker and S. Safavi-Naeini, ‘Remote heart rate sensing with mm-wave radar’, in *2018 18th International Symposium on Antenna Technology and Applied Electromagnetics (ANTEM)*, 2018, pp. 1–2. doi: 10.1109/ANTEM.2018.8572982.

- [59] F. Khan, A. Ghaffar, N. Khan and S. H. Cho, ‘An overview of signal processing techniques for remote health monitoring using impulse radio UWB transceiver’, *Sensors (Switzerland)*, vol. 20, no. 9, 2020, issn: 14248220. doi: 10.3390/s20092479.
- [60] H. Abdelnasser, K. A. Harras and M. Youssef, ‘Ubibreathe: A ubiquitous non-invasive wifi-based breathing estimator’, in *ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 2015, pp. 277–286.
- [61] P. Nguyen, X. Zhang, A. Halbower and T. Vu, ‘Continuous and fine-grained breathing volume monitoring from afar using wireless signals’, in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, 2016, pp. 1–9. doi: 10.1109/INFOCOM.2016.7524402.
- [62] J. Shang and J. Wu, ‘Fine-grained vital signs estimation using commercial wi-fi devices’, in *Proceedings of the Eighth Wireless of the Students, by the Students, and for the Students Workshop*, ser. S3, New York City, New York: Association for Computing Machinery, 2016, pp. 30–32, isbn: 9781450342551. doi: 10.1145/2987354.2987360. [Online]. Available: <https://doi.org/10.1145/2987354.2987360>.
- [63] X. Wang, C. Yang and S. Mao, ‘Phasebeat: Exploiting csi phase data for vital sign monitoring with commodity wifi devices’, in *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, 2017, pp. 1230–1239. doi: 10.1109/ICDCS.2017.206.
- [64] C. Chen, Y. Han, Y. Chen *et al.*, ‘Tr-breath: Time-reversal breathing rate estimation and detection’, *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 3, pp. 489–501, 2018. doi: 10.1109/TBME.2017.2699422.
- [65] S. Lee, Y.-D. Park, Y.-J. Suh and S. Jeon, ‘Design and implementation of monitoring system for breathing and heart rate pattern using wifi signals’, in *2018 15th IEEE Annual Consumer Communications Networking Conference (CCNC)*, 2018, pp. 1–7. doi: 10.1109/CCNC.2018.8319181.
- [66] Y. Zeng, D. Wu, J. Xiong, E. Yi, R. Gao and D. Zhang, ‘Farsense: Pushing the range limit of wifi-based respiration sensing with csi ratio of two antennas’, 121, 2019.
- [67] Y. Gu, X. Zhang, Z. Liu and F. Ren, ‘Wifi-based real-time breathing and heart rate monitoring during sleep’, in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6. doi: 10.1109/GLOBECOM38437.2019.9014297.
- [68] M. Ibrahim and K. N. Brown, ‘Vehicle in-cabin contactless wifi human sensing’, in *2021 18th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, 2021, pp. 1–2. doi: 10.1109/SECON52354.2021.9491580.
- [69] W. Li, R. J. Piechocki, K. Woodbridge, C. Tang and K. Chetty, ‘Passive wifi radar for human sensing using a stand-alone access point’, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 1986–1998, 2021. doi: 10.1109/TGRS.2020.3006387.

- [70] X. Wang, C. Yang and S. Mao, ‘On csi-based vital sign monitoring using commodity wifi’, *ACM Trans. Comput. Healthcare*, vol. 1, no. 3, May 2020, issn: 2691-1957. doi: 10.1145/3377165. [Online]. Available: <https://doi.org/10.1145/3377165>.
- [71] F. Wang, J. Han, S. Zhang, X. He and D. Huang, ‘CSI-Net: Unified Human Body Characterization and Pose Recognition’, *arXiv 1810.03064*, no. 2, 2018. arXiv: 1810.03064.
- [72] D. Rissacher and D. Galy, ‘Cardiac radar for biometric identification using nearest neighbour of continuous wavelet transform peaks’, in *IEEE International Conference on Identity, Security and Behavior Analysis (ISBA 2015)*, 2015, pp. 1–6. doi: 10.1109/ISBA.2015.7126356.
- [73] K. Diederichs, A. Qiu and G. Shaker, ‘Wireless biometric individual identification utilizing millimeter waves’, *IEEE Sensors Letters*, vol. 1, no. 1, pp. 1–4, 2017. doi: 10.1109/LSENS.2017.2673551.
- [74] S. A. Shah and F. Fioranelli, ‘Rf sensing technologies for assisted daily living in healthcare: A comprehensive review’, *IEEE Aerospace and Electronic Systems Magazine*, vol. 34, no. 11, pp. 26–44, 2019.
- [75] H. Abdelnasser, M. Youssef and K. A. Harras, ‘Wigest: A ubiquitous wifi-based gesture recognition system’, in *2015 IEEE Conference on Computer Communications (INFOCOM)*, 2015, pp. 1472–1480. doi: 10.1109/INFOCOM.2015.7218525.
- [76] Q. Pu, S. Gupta, S. Gollakota and S. Patel, ‘Gesture Recognition Using Wireless Signals’, *GetMobile: Mobile Computing and Communications*, vol. 18, no. 4, pp. 15–18, 2015, issn: 2375-0529. doi: 10.1145/2721914.2721919.
- [77] C. Wu, Z. Yang, Z. Zhou, X. Liu, Y. Liu and J. Cao, ‘Non-invasive detection of moving and stationary human with wifi’, *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 11, pp. 2329–2342, 2015. doi: 10.1109/JSAC.2015.2430294.
- [78] M. Kotaru, K. Joshi, D. Bharadia and S. Katti, ‘Spotfi: Decimeter level localization using wifi’, *SIGCOMM Comput. Commun. Rev.*, vol. 45, no. 4, pp. 269–282, Aug. 2015, issn: 0146-4833. doi: 10.1145/2829988.2787487. [Online]. Available: <https://doi.org/10.1145/2829988.2787487>.
- [79] K. Joshi, D. Bharadia, M. Kotaru and S. Katti, ‘Wideo: Fine-grained device-free motion tracing using {rf} backscatter’, in *12th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 15)*, 2015, pp. 189–204.
- [80] X. Li, S. Li, D. Zhang, J. Xiong, Y. Wang and H. Mei, ‘Dynamic-music: Accurate device-free indoor localization’, in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp ’16, Heidelberg, Germany: Association for Computing Machinery, 2016, pp. 196–207, isbn: 9781450344616. doi: 10.1145/2971648.2971665. [Online]. Available: <https://doi.org/10.1145/2971648.2971665>.

- [81] Y. Xie, J. Xiong, M. Li and K. Jamieson, ‘xD-Track: Leveraging multi-dimensional information for passive Wi-Fi tracking’, *Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM*, pp. 39–43, 2016. doi: 10.1145/2980115.2980127.
- [82] K. Qian, C. Wu, Z. Yang, C. Yang and Y. Liu, ‘Decimeter level passive tracking with WiFi’, in *Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM*, 2016, pp. 44–48, isbn: 9781450342513. doi: 10.1145/2980115.2980131.
- [83] X. Zheng, J. Wang, L. Shangguan, Z. Zhou and Y. Liu, ‘Smokey: Ubiquitous smoking detection with commercial WiFi infrastructures’, in *Proceedings - IEEE INFOCOM*, vol. 2016-July, IEEE, 2016, isbn: 9781467399531. doi: 10.1109/INFOCOM.2016.7524399.
- [84] W. Wang, A. X. Liu and M. Shahzad, ‘Gait recognition using WiFi signals’, *UbiComp 2016 - Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 363–373, 2016. doi: 10.1145/2971648.2971670.
- [85] H. Li, W. Yang, J. Wang, Y. Xu and L. Huang, ‘WiFinger: Talk to your smart devices with finger-grained gesture’, *UbiComp 2016 - Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 250–261, 2016. doi: 10.1145/2971648.2971738.
- [86] W. Wang, A. X. Liu, M. Shahzad, K. Ling and S. Lu, ‘Device-Free Human Activity Recognition Using Commercial WiFi Devices’, *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1118–1131, 2017, issn: 07338716. doi: 10.1109/JSAC.2017.2679658.
- [87] K. Qian, C. Wu, Z. Yang, Y. Liu and K. Jamieson, ‘Widar: Decimeter-level passive tracking via velocity monitoring with commodity Wi-Fi’, *Proceedings of the International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, vol. Part F1291, 2017. doi: 10.1145/3084041.3084067.
- [88] K. Qian, C. Wu, Z. Yang, Y. Liu, H. E. Fugui and T. Xing, ‘Enabling contactless detection of moving humans with dynamic speeds using CSI’, *ACM Transactions on Embedded Computing Systems*, vol. 17, no. 2, pp. 1–18, 2018, issn: 15583465. doi: 10.1145/3157677.
- [89] S. Shi, S. Sigg, L. Chen and Y. Ji, ‘Accurate Location Tracking from CSI-Based Passive Device-Free Probabilistic Fingerprinting’, *IEEE Transactions on Vehicular Technology*, vol. 67, no. 6, pp. 5217–5230, 2018, issn: 00189545. doi: 10.1109/TVT.2018.2810307.
- [90] Y. Ma, G. Zhou, S. Wang, H. Zhao and W. Jung, ‘SignFi: Sign Language Recognition Using WiFi’, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, pp. 1–21, 2018, issn: 2474-9567. doi: 10.1145/3191755.

- [91] Y. Zheng, Y. Zhang, K. Qian *et al.*, ‘Zero-effort cross-domain gesture recognition with wi-fi’, in *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys ’19, Seoul, Republic of Korea: Association for Computing Machinery, 2019, pp. 313–325, isbn: 9781450366618. doi: 10.1145/3307334.3326081. [Online]. Available: <https://doi.org/10.1145/3307334.3326081>.
- [92] J. Yang, H. Zou, Y. Zhou and L. Xie, ‘Learning Gestures From WiFi: A Siamese Recurrent Convolutional Architecture’, *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10 763–10 772, 2019, issn: 23274662. doi: 10.1109/JIOT.2019.2941527.
- [93] B. Huang, G. Mao, Y. Qin and Y. Wei, ‘Pedestrian flow estimation through passive wifi sensing’, *IEEE Transactions on Mobile Computing*, vol. 20, no. 4, pp. 1529–1542, 2021. doi: 10.1109/TMC.2019.2959610.
- [94] Y. Bai and X. Wang, ‘CARIN: Wireless CSI-based driver activity recognition under the interference of passengers’, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 1, 2020, issn: 24749567. doi: 10.1145/3380992.
- [95] J. Yang, H. Zou, S. Cao, Z. Chen and L. Xie, ‘MobileDA : Toward Edge-Domain Adaptation’, vol. 7, no. 8, pp. 6909–6918, 2020.
- [96] H. Fei, F. Xiao, J. Han, H. Huang and L. Sun, ‘Multi-variations activity based gaits recognition using commodity wifi’, *IEEE Transactions on Vehicular Technology*, vol. 69, no. 2, pp. 2263–2273, 2020. doi: 10.1109/TVT.2019.2962803.
- [97] K. Niu, F. Zhang, X. Wang, Q. Lv, H. Luo and D. Zhang, ‘Understanding wifi signal frequency features for position-independent gesture sensing’, *IEEE Transactions on Mobile Computing*, vol. 21, no. 11, pp. 4156–4171, 2022. doi: 10.1109/TMC.2021.3063135.
- [98] R. Gao, M. Zhang, J. Zhang *et al.*, ‘Towards position-independent sensing for gesture recognition with wi-fi’, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 2, pp. 1–28, 2021.
- [99] X. Ding, T. Jiang, Y. Zhong, Y. Huang and Z. Li, ‘Wi-fi-based location-independent human activity recognition via meta learning’, *Sensors*, vol. 21, no. 8, p. 2654, 2021.
- [100] Y. Ma, S. Arshad, S. Muniraju *et al.*, ‘Location-and person-independent activity recognition with wifi, deep neural networks, and reinforcement learning’, *ACM Transactions on Internet of Things*, vol. 2, no. 1, pp. 1–25, 2021.
- [101] H. Li, X. Chen, J. Wang, D. Wu and X. Liu, ‘Dafi: Wifi-based device-free indoor localization via domain adaptation’, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 4, pp. 1–21, 2021.
- [102] Y. Wang, S. Yang, F. Li, Y. Wu and Y. Wang, ‘Fallviewer: A fine-grained indoor fall detection system with ubiquitous wi-fi devices’, *IEEE Internet of Things Journal*, vol. 8, no. 15, pp. 12 455–12 466, 2021.

- [103] D. Wu, Y. Zeng, R. Gao *et al.*, ‘Witraj: Robust indoor motion tracking with wifi signals’, *IEEE Transactions on Mobile Computing*, vol. 22, no. 5, pp. 3062–3078, 2021.
- [104] L. Zhang, C. Wang and D. Zhang, ‘Wi-pigr: Path independent gait recognition with commodity wi-fi’, *IEEE Transactions on Mobile Computing*, vol. 21, no. 9, pp. 3414–3427, 2021.
- [105] J. Li, A. Sharma, D. Mishra and A. Seneviratne, ‘Thermal profiling by wifi sensing in iot networks’, in *2021 IEEE Global Communications Conference (GLOBECOM)*, IEEE, 2021, pp. 1–6.
- [106] J. Liu, Y. Zeng, T. Gu, L. Wang and D. Zhang, ‘Wiphone: Smartphone-based respiration monitoring using ambient reflected wifi signals’, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 1, pp. 1–19, 2021.
- [107] T. Nakamura, M. Bouazizi, K. Yamamoto and T. Ohtsuki, ‘Wi-fi-based fall detection using spectrogram image of channel state information’, *IEEE Internet of Things Journal*, vol. 9, no. 18, pp. 17 220–17 234, 2022.
- [108] X. Chen, H. Li, C. Zhou, X. Liu, D. Wu and G. Dudek, ‘Fidora: Robust wifi-based indoor localization via unsupervised domain adaptation’, *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 9872–9888, 2022.
- [109] R. Gao, W. Li, Y. Xie *et al.*, ‘Towards robust gesture recognition by characterizing the sensing quality of wifi signals’, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 1, pp. 1–26, 2022.
- [110] Z. Shi, Q. Cheng, J. A. Zhang and R. Y. Da Xu, ‘Environment-robust wifi-based human activity recognition using enhanced csi and deep learning’, *IEEE Internet of Things Journal*, vol. 9, no. 24, pp. 24 643–24 654, 2022.
- [111] D. Wang, J. Yang, W. Cui, L. Xie and S. Sun, ‘Caution: A robust wifi-based human authentication system via few-shot open-set recognition’, *IEEE Internet of Things Journal*, vol. 9, no. 18, pp. 17 323–17 333, 2022.
- [112] J. Zhang, Y. Li, H. Xiong, D. Dou, C. Miao and D. Zhang, ‘Handgest: Hierarchical sensing for robust-in-the-air handwriting recognition with commodity wifi devices’, *IEEE Internet of Things Journal*, vol. 9, no. 19, pp. 19 529–19 544, 2022.
- [113] J. Qiu, P. Zheng, K. Chi, R. Xu and J. Liu, ‘Respiration monitoring in high-dynamic environments via combining multiple wifi channels based on wire direct connection between rx/tx’, *IEEE Internet of Things Journal*, vol. 10, no. 2, pp. 1558–1573, 2022.
- [114] X. Wei, T. Olugbade, F. Shi *et al.*, ‘Leveraging wifi sensing toward automatic recognition of pain behaviors’, in *2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, IEEE, 2023, pp. 1–8.

- [115] J. Liu, W. Li, T. Gu *et al.*, ‘Towards a dynamic fresnel zone model to wifi-based human activity recognition’, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 7, no. 2, pp. 1–24, 2023.
- [116] M. Yang, H. Zhu, R. Zhu, F. Wu, L. Yin and Y. Yang, ‘Witransformer: A novel robust gesture recognition sensing model with wifi’, *Sensors*, vol. 23, no. 5, p. 2612, 2023.
- [117] C. Zakaria, G. Yilmaz, P. M. Mammen, M. Chee, P. Shenoy and R. Balan, ‘Sleep-more: Inferring sleep duration at scale via multi-device wifi sensing’, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 6, no. 4, 2023. doi: 10.1145/3569489. [Online]. Available: <https://doi.org/10.1145/3569489>.
- [118] Y. Gu, X. Zhang, H. Yan, Z. Liu and F. Ren, ‘Wital: Wifi-based real-time vital signs monitoring during sleep’, *Authorea Preprints*, 2023.
- [119] Z. Zhou, Z. Yang, C. Wu, L. Shangguan and Y. Liu, ‘Omnidirectional coverage for device-free passive human detection’, *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 7, pp. 1819–1829, 2014, issn: 10459219. doi: 10.1109/TPDS.2013.274.
- [120] K. Qian, C. Wu, Z. Yang, Y. Liu and Z. Zhou, ‘PADS: Passive detection of moving targets with dynamic speed using PHY layer information’, *Proceedings of the International Conference on Parallel and Distributed Systems - ICPADS*, vol. 2015-April, pp. 1–8, 2014, issn: 15219097. doi: 10.1109/PADSW.2014.7097784.
- [121] K. Ohara, T. Maekawa and Y. Matsushita, ‘Detecting state changes of indoor everyday objects using wi-fi channel state information’, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 3, Sep. 2017. doi: 10.1145/3131898. [Online]. Available: <https://doi.org/10.1145/3131898>.
- [122] X. Wang, C. Yang and S. Mao, ‘TensorBeat: Tensor decomposition for monitoring multi-person breathing beats with commodity WiFi’, *arXiv*, vol. V, pp. 1–28, 2017. arXiv: 1702.02046.
- [123] M. M. Breunig, H. P. Kriegel, R. T. Ng and J. Sander, ‘Lof: Identifying density-based local outliers’, *Proc.acm Sigmod Int.conf.on Management of Data*, 2000.
- [124] H. Wang, D. Zhang, J. Ma *et al.*, ‘Human respiration detection with commodity wifi devices: Do user location and body orientation matter?’, in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 25–36.
- [125] C. Ieracitano, A. Adeel, M. Gogate *et al.*, ‘Statistical analysis driven optimized deep learning system for intrusion detection’, in *International Conference on Brain Inspired Cognitive Systems*, Springer, 2018, pp. 759–769.
- [126] P. Melgarejo, X. Zhang, P. Ramanathan and D. Chu, ‘Leveraging directional antenna capabilities for fine-grained gesture recognition’, *UbiComp 2014 - Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 541–551, 2014. doi: 10.1145/2632048.2632095.

- [127] H. Zou, J. Yang, Y. Zhou, L. Xie and C. J. Spanos, ‘Robust WiFi-enabled device-free gesture recognition via unsupervised adversarial domain adaptation’, *Proceedings - International Conference on Computer Communications and Networks, ICCCN*, vol. 2018-July, 2018, issn: 10952055. doi: 10.1109/ICCCN.2018.8487345.
- [128] F. Wang, S. Panev, Z. Dai, J. Han and D. Huang, ‘Can WiFi estimate person pose?’, *arXiv:1904.00277*, no. March, 2019. arXiv: 1904.00277.
- [129] L. Guo, Z. Lu, X. Wen, S. Zhou and Z. Han, ‘From signal to image: Capturing fine-grained human poses with commodity wi-fi’, *IEEE Communications Letters*, vol. 24, no. 4, pp. 802–806, 2020. doi: 10.1109/LCOMM.2019.2961890.
- [130] F. Thullier, A. Beaulieu, J. Maître, S. Gaboury and K. Bouchard, ‘A systematic evaluation of the XeThru X4 Ultra-Wideband radar behavior’, *Procedia Computer Science*, vol. 198, pp. 148–155, 2022, 12th International Conference on Emerging Ubiquitous Systems and Pervasive Networks / 11th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare, issn: 1877-0509. doi: <https://doi.org/10.1016/j.procs.2021.12.222>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877050921024613>.
- [131] Z. Yang, M. Bocca, V. Jain and P. Mohapatra, ‘Contactless breathing rate monitoring in vehicle using UWB radar’, in *Proceedings of the 7th International Workshop on Real-World Embedded Wireless Systems and Networks*, ser. RealWSN’18, Shenzhen, China: Association for Computing Machinery, 2018, pp. 13–18, isbn: 9781450360487. doi: 10.1145/3277883.3277884. [Online]. Available: <https://doi.org/10.1145/3277883.3277884>.
- [132] T. Zheng, Z. Chen, C. Cai, J. Luo and X. Zhang, ‘V2iFi: In-vehicle vital sign monitoring via compact RF sensing’, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 4, no. 2, 2020. doi: 10.1145/3397321. [Online]. Available: <https://doi.org/10.1145/3397321>.
- [133] H. Li, A. Shrestha, H. Heidari, J. Le Kerneec and F. Fioranelli, ‘Bi-LSTM network for multimodal continuous human activity recognition and fall detection’, *IEEE Sensors Journal*, vol. 20, no. 3, pp. 1191–1201, 2020. doi: 10.1109/JSEN.2019.2946095.
- [134] Y. Ge, A. Taha, S. A. Shah *et al.*, ‘Contactless wifi sensing and monitoring for future healthcare-emerging trends, challenges and opportunities’, *IEEE Reviews in Biomedical Engineering*, 2022.
- [135] A. De Groote, M. Wantier, G. Chéron, M. Estenne and M. Paiva, ‘Chest wall motion during tidal breathing’, *Journal of Applied Physiology*, vol. 83, no. 5, pp. 1531–1537, 1997.
- [136] G Ramachandran and M Singh, ‘Three-dimensional reconstruction of cardiac displacement patterns on the chest wall during the p, qrs and t-segments of the ecg by laser speckle inteferometry’, *Medical and Biological Engineering and Computing*, vol. 27, pp. 525–530, 1989.

- [137] A. Ahmad, J. C. Roh, D. Wang and A. Dubey, ‘Vital signs monitoring of multiple people using a fmcw millimeter-wave sensor’, in *2018 IEEE Radar Conference (RadarConf18)*, IEEE, 2018, pp. 1450–1455.
- [138] G. Wang, J.-M. Muñoz-Ferrerias, C. Gu, C. Li and R. Gómez-García, ‘Application of linear-frequency-modulated continuous-wave (lfmcw) radars for tracking of vital signs’, *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 6, pp. 1387–1399, 2014. doi: 10.1109/TMTT.2014.2320464.
- [139] F. Adib, H. Mao, Z. Kabelac, D. Katabi and R. C. Miller, ‘Smart homes that monitor breathing and heart rate’, in *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 2015, pp. 837–846.
- [140] M. Mercuri, I. R. Lorato, Y.-H. Liu, F. Wieringa, C. V. Hoof and T. Torfs, ‘Vital-sign monitoring and spatial tracking of multiple people using a contactless radar-based sensor’, *Nature Electronics*, vol. 2, no. 6, pp. 252–262, 2019.
- [141] R. Grisot, P. Laurent, C. Migliaccio *et al.*, ‘Monitoring of heart movements using an fmcw radar and correlation with an ecg’, *IEEE Transactions on Radar Systems*, vol. 1, pp. 423–434, 2023. doi: 10.1109/TRS.2023.3298348.
- [142] J. Chen, D. Zhang, Z. Wu, F. Zhou, Q. Sun and Y. Chen, ‘Contactless electrocardiogram monitoring with millimeter wave radar’, *IEEE Transactions on Mobile Computing*, pp. 1–17, 2022. doi: 10.1109/TMC.2022.3214721.
- [143] Y. Rong, I. Lenz and D. W. Bliss, ‘Vital signs detection based on high-resolution 3-d mmwave radar imaging’, in *2022 IEEE International Symposium on Phased Array Systems & Technology (PAST)*, IEEE, 2022, pp. 1–6.
- [144] F. Wang, X. Zeng, C. Wu, B. Wang and K. J. R. Liu, ‘Driver vital signs monitoring using millimeter wave radio’, *IEEE Internet of Things Journal*, vol. 9, no. 13, pp. 11 283–11 298, 2022. doi: 10.1109/JIOT.2021.3128548.
- [145] T. Zheng, Z. Chen, C. Cai, J. Luo and X. Zhang, ‘V2ifi: In-vehicle vital sign monitoring via compact rf sensing’, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 4, no. 2, 2020. doi: 10.1145/3397321. [Online]. Available: <https://doi.org/10.1145/3397321>.
- [146] A. Singh, X. Gao, E. Yavari *et al.*, ‘Data-based quadrature imbalance compensation for a cw doppler radar system’, *IEEE Transactions on Microwave Theory Techniques*, vol. 61, pp. 1718–1724, Apr. 2013. doi: 10.1109/TMTT.2013.2249525.
- [147] M. P. Ebrahim, N. Tom, J.-M. Redoute and M. R. Yuce, ‘A low-frequency portable continuous wave radar system for vital signs monitoring’, *IEEE Sensors Journal*, vol. 23, no. 8, pp. 8876–8886, 2023. doi: 10.1109/JSEN.2023.3251978.
- [148] M. Arsalan, A. Santra and C. Will, ‘Improved contactless heartbeat estimation in fmcw radar via kalman filter tracking’, *IEEE Sensors Letters*, vol. 4, no. 5, pp. 1–4, 2020.

- [149] J. Liu, Y. Li, C. Li, C. Gu and J.-F. Mao, ‘Accurate measurement of human vital signs with linear fmcw radars under proximity stationary clutters’, *IEEE Transactions on Biomedical Circuits and Systems*, vol. 15, no. 6, pp. 1393–1404, 2021.
- [150] T. K. V. Dai, K. Oleksak, T. Kvelashvili *et al.*, ‘Enhancement of remote vital sign monitoring detection accuracy using multiple-input multiple-output 77 ghz fmcw radar’, *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. 6, no. 1, pp. 111–122, 2022. doi: 10.1109/JERM.2021.3082807.
- [151] H. Hameed, M. Usman, A. Tahir *et al.*, ‘Pushing the limits of remote rf sensing by reading lips under the face mask’, *Nature Communications*, vol. 13, pp. 5168–5176, doi: 10.1038/s41467-022-32231-1.
- [152] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang and H. Liu, ‘E-eyes: Device-free location-oriented activity identification using fine-grained wifi signatures’, in *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom ’14, Maui, Hawaii, USA, 2014, pp. 617–628, isbn: 9781450327831.
- [153] Y. Zhang, Y. Zheng, K. Qian *et al.*, ‘Widar3. 0: Zero-effort cross-domain gesture recognition with wi-fi’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [154] D. Zhang, Y. Hu, Y. Chen and B. Zeng, ‘Breathtrack: Tracking indoor human breath status via commodity wifi’, *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 3899–3911, 2019. doi: 10.1109/JIOT.2019.2893330.
- [155] D. Halperin, W. Hu, A. Sheth and D. Wetherall, ‘Tool release: Gathering 802.11n traces with channel state information’, *ACM SIGCOMM on Computer Communication Review*, vol. 41, no. 1, p. 53, 2011.
- [156] K. Qian, C. Wu, Z. Zhou, Y. Zheng, Z. Yang and Y. Liu, ‘Inferring motion direction using commodity Wi-Fi for interactive exergames’, *Conference on Human Factors in Computing Systems - Proceedings*, vol. 2017-May, pp. 1961–1972, 2017. doi: 10.1145/3025453.3025678.
- [157] X. Li, D. Zhang, Q. Lv *et al.*, ‘IndoTrack: Device-Free Indoor Human Tracking with Commodity Wi-Fi’, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 1–22, 2017, issn: 2474-9567. doi: 10.1145/3130940.
- [158] B. Zheng and R. Zhang, ‘Intelligent reflecting surface-enhanced ofdm: Channel estimation and reflection optimization’, *IEEE Wireless Communications Letters*, vol. 9, no. 4, pp. 518–522, 2020. doi: 10.1109/LWC.2019.2961357.
- [159] B. Li, W. Cui, W. Wang, L. Zhang, Z. Chen and M. Wu, ‘Two-stream convolution augmented transformer for human activity recognition’, in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, 2021, pp. 286–293.

- [160] X. Wang, C. Yang and S. Mao, ‘Tensorbeat: Tensor decomposition for monitoring multiperson breathing beats with commodity wifi’, *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 9, no. 1, pp. 1–27, 2017.
- [161] C. Wu, Z. Yang, Z. Zhou, X. Liu, Y. Liu and J. Cao, ‘Non-invasive detection of moving and stationary human with wifi’, *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 11, pp. 2329–2342, 2015.
- [162] X. Li, D. Zhang, Q. Lv *et al.*, ‘Indotrack: Device-free indoor human tracking with commodity wi-fi’, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 1–22, 2017.
- [163] M. Z. Khan, A. Taha, W. Taylor, M. A. Imran and Q. H. Abbasi, ‘Non-invasive localization using software-defined radios’, *IEEE Sensors Journal*, vol. 22, no. 9, pp. 9018–9026, 2022. doi: 10.1109/JSEN.2022.3160796.
- [164] R. Schmidt, ‘Multiple emitter location and signal parameter estimation’, *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986. doi: 10.1109/TAP.1986.1143830.
- [165] M. Kotaru, K. Joshi, D. Bharadia and S. Katti, ‘Spotfi: Decimeter level localization using wifi’, *SIGCOMM Comput. Commun. Rev.*, vol. 45, no. 4, pp. 269–282, 2015, issn: 0146-4833. doi: 10.1145/2829988.2787487. [Online]. Available: <https://doi.org/10.1145/2829988.2787487>.
- [166] K. He, X. Zhang, S. Ren and J. Sun, ‘Identity mappings in deep residual networks’, in *European conference on computer vision*, Springer, 2016, pp. 630–645.
- [167] W. W. L. Z. Z. C. Bing Li Wei Cui and M. Wu, ‘Two-stream convolution augmented transformer for human activity recognition’, in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021.
- [168] Y. Ge, S. Li, M. Shentu *et al.*, ‘A doppler-based human activity recognition system using wifi signals’, in *2021 IEEE Sensors*, IEEE, 2021, pp. 1–4.
- [169] Y. Chen, J. Yu, L. Kong, Y. Zhu and F. Tang, ‘Rfpass: Towards environment-independent gait-based user authentication leveraging rfid’, in *2022 19th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, 2022, pp. 289–297. doi: 10.1109/SECON55815.2022.9918573.
- [170] D. Halperin, W. Hu, A. Sheth and D. Wetherall, ‘Tool release: Gathering 802.11n traces with channel state information’, *SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, p. 53, 2011, issn: 0146-4833. doi: 10.1145/1925861.1925870. [Online]. Available: <https://doi.org/10.1145/1925861.1925870>.
- [171] W. Wang, A. X. Liu and M. Shahzad, ‘Gait recognition using wifi signals’, in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 363–373.
- [172] Y. Zhang, Y. Zheng, G. Zhang, K. Qian, C. Qian and Z. Yang, ‘Gaitid: Robust wi-fi based gait recognition’, in *International Conference on Wireless Algorithms, Systems, and Applications*, Springer, 2020, pp. 730–742.

- [173] Y. Zeng, D. Wu, J. Xiong, E. Yi, R. Gao and D. Zhang, ‘Farsense: Pushing the range limit of wifi-based respiration sensing with csi ratio of two antennas’, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 3, no. 3, 2019. doi: 10.1145/3351279. [Online]. Available: <https://doi.org/10.1145/3351279>.
- [174] B. Xie, Y. Yin and J. Xiong, ‘Pushing the limits of long range wireless sensing with lora’, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 3, pp. 1–21, 2021.
- [175] F. Zhang, Z. Chang, J. Xiong and D. Zhang, ‘Exploring lora for sensing’, *GetMobile: Mobile Computing and Communications*, vol. 25, no. 2, pp. 33–37, 2021.
- [176] F. Zhang, Z. Chang, J. Xiong *et al.*, ‘Unlocking the beamforming potential of lora for long-range multi-target respiration sensing’, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 2, pp. 1–25, 2021.
- [177] Y. Zeng, D. Wu, R. Gao, T. Gu and D. Zhang, ‘Fullbreathe: Full human respiration detection exploiting complementarity of csi phase and amplitude of wifi signals’, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 3, 2018. doi: 10.1145/3264958. [Online]. Available: <https://doi.org/10.1145/3264958>.
- [178] K. Qian, C. Wu, Y. Zhang, G. Zhang, Z. Yang and Y. Liu, ‘Widar2.0: Passive human tracking with a single Wi-Fi link’, *MobiSys 2018 - Proceedings of the 16th ACM International Conference on Mobile Systems, Applications, and Services*, pp. 350–361, 2018. doi: 10.1145/3210240.3210314.
- [179] C. Wan, L. Wang and V. V. Phoha, ‘A survey on gait recognition’, *ACM Comput. Surv.*, vol. 51, no. 5, 2018, issn: 0360-0300. doi: 10.1145/3230633. [Online]. Available: <https://doi.org/10.1145/3230633>.
- [180] A. Abanda, U. Mori and J. A. Lozano, ‘A review on distance based time series classification’, *Data Mining and Knowledge Discovery*, vol. 33, no. 2, pp. 378–412, 2019.
- [181] J. Tapparel, O. Afisiadis, P. Mayoraz, A. Balatsoukas-Stimming and A. Burg, ‘An open-source lora physical layer prototype on gnu radio’, in *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, IEEE, 2020, pp. 1–5.
- [182] J. Zhang, R. Xi, Y. He *et al.*, ‘A survey of mmwave-based human sensing: Technology, platforms and applications’, *IEEE Communications Surveys Tutorials*, pp. 1–1, 2023. doi: 10.1109/COMST.2023.3298300.
- [183] J. Wang, X. Wang, L. Chen, J. Huangfu, C. Li and L. Ran, ‘Noncontact distance and amplitude-independent vibration measurement based on an extended dacm algorithm’, *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 1, pp. 145–153, 2014. doi: 10.1109/TIM.2013.2277530.
- [184] M. Shin, Y. Jung, J. Kim *et al.*, ‘Fmcw radar-based vital signal monitoring technique using adaptive range-bin selection’, in *2023 IEEE Radar Conference (RadarConf23)*, 2023, pp. 1–6. doi: 10.1109/RadarConf2351548.2023.10149752.

- [185] C. Cai, R. Zheng and J. Luo, ‘Ubiquitous acoustic sensing on commodity IoT devices: A Survey’, *IEEE Communications Surveys Tutorials*, vol. 24, no. 1, pp. 432–454, 2022. doi: 10.1109/COMST.2022.3145856.
- [186] M. Benzeghiba, R. De Mori, O. Deroo *et al.*, ‘Automatic speech recognition and speech variability: A review’, *Speech Communication*, vol. 49, no. 10, pp. 763–786, 2007, Intrinsic Speech Variations, issn: 0167-6393. doi: <https://doi.org/10.1016/j.specom.2007.02.006>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167639307000404>.
- [187] J. A. Gonzalez-Lopez, A. Gomez-Alanis, J. M. Martín Doñas, J. L. Pérez-Córdoba and A. M. Gomez, ‘Silent speech interfaces for speech restoration: A Review’, *IEEE Access*, vol. 8, pp. 177 995–178 021, 2020. doi: 10.1109/ACCESS.2020.3026579.
- [188] A. Bednar and E. C. Lalor, ‘Where is the cocktail party? Decoding locations of attended and unattended moving sound sources using EEG’, *NeuroImage*, vol. 205, p. 116 283, 2020, issn: 1053-8119. doi: <https://doi.org/10.1016/j.neuroimage.2019.116283>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1053811919308742>.
- [189] T. Liu, M. Gao, F. Lin *et al.*, ‘Wavevoice: A noise-resistant multi-modal speech recognition system fusing mmWave and audio signals’, in *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, ser. SenSys ’21, Coimbra, Portugal: Association for Computing Machinery, 2021, pp. 97–110, isbn: 9781450390972. doi: 10.1145/3485730.3485945. [Online]. Available: <https://doi.org/10.1145/3485730.3485945>.
- [190] M. Z. Ozturk, C. Wu, B. Wang and K. J. R. Liu, ‘RadioMic: Sound sensing via radio signals’, *IEEE Internet of Things Journal*, vol. 10, no. 5, pp. 4431–4448, 2023. doi: 10.1109/JIOT.2022.3217968.
- [191] H. Hameed, M. Usman, A. Tahir *et al.*, ‘Pushing the limits of remote RF sensing by reading lips under the face mask’, *Nature communications*, vol. 13, no. 1, pp. 1–9, 2022.
- [192] D. Ferreira, S. Silva, F. Curado and A. Teixeira, ‘Exploring silent speech interfaces based on frequency-modulated continuous-wave radar’, *Sensors*, vol. 22, no. 2, p. 649, 2022.
- [193] M. Z. Ozturk, C. Wu, B. Wang, M. Wu and K. R. Liu, ‘Beyond Microphone: MmWave-based interference-resilient voice activity detection’, in *Proceedings of the 1st ACM International Workshop on Intelligent Acoustic Systems and Applications*, 2022, pp. 7–12.
- [194] L. Cester, I. Starshynov, Y. Jones, P. Pellicori, J. G. Cleland and D. Faccio, ‘Remote laser-speckle sensing of heart sounds for health assessment and biometric identification’, *Biomedical Optics Express*, vol. 13, no. 7, pp. 3743–3750, 2022.
- [195] S. Livingstone and F. Russo, ‘Ryerson audiovisual database of emotional speeches and songs (ravdess): A dynamic, multimodal set of north american english face and voice expressions’, *Plos One*, vol. 13, no. 5, e0196391, 2018.

- [196] M. Abdrakhmanova, A. Kuzdeuov, S. Jarju, Y. Khassanov, M. Lewis and H. A. Varol, ‘Speakingfaces: A large-scale multimodal dataset of voice commands with visual and thermal video streams’, *Sensors*, vol. 21, no. 10, p. 3465, 2021.
- [197] K. Sun and X. Zhang, ‘Ultrase: Single-channel speech enhancement using ultrasound’, in *Proceedings of the 27th annual international conference on mobile computing and networking*, 2021, pp. 160–173.
- [198] M. Z. Ozturk, C. Wu, B. Wang, M. Wu and K. J. R. Liu, ‘Radioses: MmWave-based audioradio speech enhancement and separation system’, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 31, pp. 1333–1347, 2023. doi: 10.1109/TASLP.2023.3250846.
- [199] Y. Wang, W. Zhang, Z. Wu, X. Kong and H. Zhang, ‘Speckle noise detection and removal for laser speech measurement systems’, *Applied Sciences*, vol. 11, no. 21, p. 9870, 2021.
- [200] X. Chen, Y. Xia, Y. Sun *et al.*, ‘Silent speech recognition based on high-density surface electromyogram using hybrid neural networks’, *IEEE Transactions on Human-Machine Systems*, 2023.
- [201] X. Hao, D. Zhu, X. Wang, L. Yang and H. Zeng, ‘A speech enhancement algorithm for speech reconstruction based on laser speckle images’, *Sensors*, vol. 23, no. 1, p. 330, 2023.
- [202] M. Cooke, J. Barker, S. Cunningham and X. Shao, ‘An audio-visual corpus for speech perception and automatic speech recognition’, *The Journal of the Acoustical Society of America*, vol. 120, no. 5, pp. 2421–2424, 2006.
- [203] C. Busso, M. Bulut, C.-C. Lee *et al.*, ‘Iemocap: Interactive emotional dyadic motion capture database’, *Language resources and evaluation*, vol. 42, pp. 335–359, 2008.
- [204] M. S. Ribeiro, J. Sanger, J.-X. Zhang *et al.*, ‘Tal: A synchronised multi-speaker corpus of ultrasound tongue imaging, audio, and lip videos’, in *2021 IEEE Spoken Language Technology Workshop (SLT)*, IEEE, 2021, pp. 1109–1116.
- [205] A. Haliassos, K. Vougioukas, S. Petridis and M. Pantic, ‘Lips don’t lie: A generalisable and robust approach to face forgery detection’, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 5039–5049.
- [206] I. R. Titze, ‘Physiologic and acoustic differences between male and female voices’, *The Journal of the Acoustical Society of America*, vol. 85, no. 4, pp. 1699–1707, 1989.
- [207] R. Futrell, E. Gibson, H. J. Tily *et al.*, ‘The natural stories corpus: A reading-time corpus of english texts containing rare syntactic constructions’, *Language Resources and Evaluation*, vol. 55, no. 1, pp. 63–77, 2021.
- [208] Y. Ge, T. Chong, L. Haobo *et al.*, ‘A large-scale multimodal dataset of non-invasive human speech recognition’, *University of Glasgow*, 2023. doi: 10.5525/gla.researchdata.1408.

- [209] S. Kim and K.-K. Lee, ‘Low-Complexity joint extrapolation-MUSIC-based 2-D parameter estimator for vital FMCW radar’, *IEEE Sensors Journal*, vol. 19, no. 6, pp. 2205–2216, 2019. doi: 10.1109/JSEN.2018.2877043.
- [210] G. Wang, J.-M. Muñoz-Ferreras, C. Gu, C. Li and R. Gómez-García, ‘Application of linear-frequency-modulated continuous-wave (lfmcw) radars for tracking of vital signs’, *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 6, pp. 1387–1399, 2014. doi: 10.1109/TMTT.2014.2320464.