Delft University of Technology
Master's Thesis in Embedded Systems

Human Activity Recognition using Channel State Information

Mohammed Nasser Hamdoon Al-Rahbi

embedded software

TU Delft Delft University of Technology

Human Activity Recognition using Channel State Information

Master's Thesis in Embedded Systems

Embedded Software Section
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology
Mekelweg 4, 2628 CD Delft, The Netherlands

Mohammed Nasser Hamdoon Al-Rahbi
M.N.H.Al-Rahbi@student.tudelft.nl

14th August 2019

Author
Mohammed Nasser Hamdoon Al-Rahbi (M.N.H.Al-Rahbi@student.tudelft.nl)
Title
Human Activity Recognition using Channel State Information
MSc presentation
19th August 2019

Graduation Committee

Dr.ir. C.J.M. Verhoeven (chair)	Delft University of Technology
Dr. RangaRao Venkatesha Prasad	Delft University of Technology
Dr.ir. Vijay S. Rao	Delft University of Technology

Abstract

Human Activity Recognition (HAR) is a key enabler of various applications, including smart homes, health care, Internet of Things (IoT), and virtual reality games. A large number of HAR systems are based on wearable sensors and computer vision. However, a challenge that has emerged in the last few years entails recognizing human activities using WiFi Channel State Information (CSI). Exiting state-of-the-art solutions have considered only amplitudes of the CSI to recognize human activities, we explore both amplitudes and phase differences to recognize activities. We utilize Continuous Wavelet Transform (CWT) to generate scalogram images from the CSI measurements. Then, we use these images as input to the pertained Convolution Neural Network (CNN), namely AlexNet to extract features that are resilient to environment changes and classify the activities. The experimental results show that the proposed method achieves an accuracy of 98.18% $\pm$ 1.26% using amplitude and phase difference. We also studied the impact of different environments and people, and the results show its robustness.

Preface

During my master programs in TU Delft, I have been exposed to the internet of things (IoT), smart sensing and wireless network. One of the courses I took was smart phone sensing which involved activity recognition and indoor localization using the sensors from the smart phone. This got me thinking about finding other possibilities without carrying the smart phone. WiFi devices have attracted my attention since they are available almost everywhere. This motivated me to do this thesis which aims to recognize human activities in indoor environments using only commodity WiFi devices.

Firstly, I would like to express my deepest and sincerest gratitude to my supervisor Dr. RangaRao Venkatesha Prasad (VP) for the continuous support of my thesis study and throughout the Embedded Systems master program at TU Delft, for his patience, motivation, and immense knowledge. He helped me developing my critical research skills and testing ideas independently.

Many thanks go to my best friends Musab, Ammar, Hamed, Suhaib, Zakria, and Ziad who have been my source of motivation during my studies. Most importantly, none of this work would have been possible without the support of my family. My mother, brothers and sisters have been with my side no matter what.

Mohammed Al-Rahbi

Delft, The Netherlands
14th August 2019

Contents

Chapter 1

Introduction

Human activity recognition (HAR) has received great attention in recent years and plays an important role in the areas of health care, smart homes, Internet of Things (IoT), security, and virtual reality games [26]. In the health care application, for instance, patients with diabetes, obesity, or heart disease are required to follow a particular exercise routine for their treatment [18, 51]. HAR is considered as a key component of many product consumers. For instance, the exergames on Microsoft Kinect and Nintendo Wii tracks the gestures and full body movements to allow users to interact with the video games and change the game experience [6].

Traditional HAR approaches are based on wearable sensors and computer vision. There have been significant improvements in sensors and mobile devices in terms of price, computational power, and size, leading many researchers to work on sensor-based activity recognition. Sensors used in HAR can be categorized into four groups: motion sensors, environmental sensors, physiological sensors, and position sensors. Accelerometers and gyroscopes are the most common sensors that are used to recognize daily activities. Several papers have been able to recognize daily activities (e.g., walking, running, sitting, lying, etc.), and have reported high accuracies of 95% [23], 97.51% [16], and 97.9% [22] using accelerometers. Others have combined an acceloermeter and a gyroscope to recognize hand gestures [63, 13]. A comprehensive survey of the sensors based on activity recognition can be found in [26, 40]. In spite of the small size and low price sensors, users are required to wear these sensors at all times, which can be inconvenient.

Another promising area used to understand human activities is computer vision. Vision-based activity recognition is the basis of numerous applications, including video surveillance systems, Human-Computer Interaction (HCI), robot learning, and healthcare [69]. With the development of RGB-D cameras that provides both a color image and per-pixel depth estimates, depth-based representations have attracted many researchers. For instance, the Microsoft Xbox Kinect produces real-time skeleton joint positions from

a single depth image. As a result, it allows users to interact with video games and virtual reality using gestures [41]. A comprehensive survey of vision-based activity recognition can be found in [69, 32, 53]. There are a couple disadvantages of using cameras. Users have to be in the line of sight (LOS) with enough lighting in order to be detected. As a result, many cameras are required to increase the coverage area. In addition, the use of cameras inside homes can raise some user privacy concerns [10].

1.1 Related Work

To overcome the limitation of vision and sensors based activity recognition, device-free wireless solutions has been proposed. These solutions can be categorized into three categories: Radio Frequency (RF) based systems, Received Signal Strength Indicator (RSSI), and Channel State Information (CSI) based.

RF based systems such as Software defined radio (SDR) and Universal Software Radio Peripheral (USRP) are used to capture RF signals. In WiSee, USRP has been used to extract the Doppler shifts caused due to hand movement. A positive Doppler shift indicates the body is moving toward the receiver, whereas moving away from the receiver results in a negative Doppler shift. WiSee has been able to identify nine gestures with an average accuracy of 94% [33]. The drawback of RF systems is that these systems require additional hardware which is often expensive.

From the Media Access Control (MAC) layer, RSSI is the most common power feature that is available in wireless radio types, including WiFi, Bluetooth, and ZigBee. Sigg *et al.* were able to recognize 11 gestures using RSSI from a smartphone with a recognition accuracy of 72% [42]. Abdelnasser *et al.* conducted another study in the same matter, but used an access point (AP) and a laptop. They were able to recognize seven gestures with a recognition accuracy of 87.5% [1]. The main drawback of RSSI is that it fluctuates over time and fails to provide unique features in complex indoor environments [65].

Channel State Information (CSI) is considered as a fine-grained measurement from the physical layer that describes both the amplitude and the phase of the wireless signal across the orthogonal frequency-division multiplexing (OFDM) subcarriers and across multi-antennas in multiple-input and multiple-output (MIMO) system. CSI is stable and can capture the multipath effect [65]. Recently, open source drivers for off-the-shelf Network Interface Cards (NIC) such as Intel 5300 [14], Atheros 9390 [61], and BCM4339 network card [36, 35, 34] allow to extract CSI information.

Although several CSI based HAR solutions have achieved good performance and recognition accuracy, prior works on human activity recognition have only considered amplitude as a base signal [56, 66, 62, 9]. Amplitude

is relatively stable compared to the phase as it does not suffer from a lot of random errors. For instance, the authors in [66] uses a deep learning long short-term memory LSTM (LSTM) to extract features and classify six different activities from the raw amplitude. They found that the LSTM performs better than Hidden Markov Model (HMM) and random forest with an average accuracy of 90%. However, the performance degrades when testing in different environment because the extracted features are depended on the trained environment. In CARM [56], a 12-level discrete wavelet transform (DWT) is applied to the denoised signals from which the 27-dimensional feature vector is the extracted and is used to classify different activities. CARM achieves an average accuracy 96% and it is resilience to environment changes. One drawback of using amplitude is that some activities such as sitting down and standing have a similar amplitude pattern change. This can be confusing from the amplitude point of view [66, 54]. Another drawback is that the signal becomes weaker as the distance between the human body and the LOS path increases due to the path loss [54]. A survey of the related work in CSI can be found in the appendix.

1.2 Challenges in CSI based HAR

There exists several challenges when working with CSI signals for HAR.

- How do human activities affect the amplitude, phase and phase difference information of CSI?
- The CSI data suffers from measurement errors due to the imperfect hardware. How to reduce the CSI noises while maintaining the activity signal?
- What methods can be used to extract features that are resilient to environment changes, and how well these methods can generalize across different people?
- What classifier should be used for this problem, and how to evaluate its performance?

1.3 Research Goal & Contributions

All the challenges and limitations in the previous sections 1.1 & 1.2 lead to our research goal. The main goal of this thesis is to develop an activity recognition system that uses commodity WiFi devices which can: (i) recognize activities performed by different people (ii) be applied in complex/untrained environments (iii) perform online classification of the activities which would be suitable for real-time applications. In summary, our main contribution as follows:

- To our best of knowledge, we are the first to utilize both amplitude and phase difference of CSI that successfully recognize human activities in different environments.
- We leverage a Continuous Wavelet Transform (CWT) to generate scalograms of the activity signal, which then we use a pre-trained Convolutional Neural Network (CNN) to classify the different activities.
- Investigate the parameters that effect the performance such as untrained environments, different people, sampling rate and window size. We also compare the impact of having phase difference and having only amplitude.

1.4 Organization of the thesis

This thesis is organized as follows. Chapter 2 gives a background on CSI and the sources of measurement errors. Chapter 3 investigates the extracted base signals - amplitude, phase and phase difference, and their effect on the human activities. Chapter 4 describes the preprocessing methods used to remove the noises and improve the extracted features. Chapter 5 shows the extraction of time-frequency features which are essential in classification problem. Chapter 6 discusses the AlexNet Convolutional Neural Network (CNN) architecture used to classify the activities. Chapter 7 presents the experimental results and discussion. Finally, in Chapter 8 draws conclusions and purpose possible future work.

Chapter 2

Preliminaries

This chapter gives a background on Channel State Information (CSI) and how does it recognize activities (Section 2.1). Then, the sources of measurement errors are discussed in Section 2.2.

2.1 Channel State Information (CSI)

Wi-Fi devices with standards such as IEEE 802.11n/ac use multiple transmitters and receiver antenna arrays, which are also known as Multiple Input, Multiple Output (MIMO) systems. MIMO is used in order to increase diversity gain, array gain, spatial multiplexing gain, and interference reduction [30]. It is typically used with Orthogonal Frequency-Division Multiplexing (OFDM) as a physical (PHY) layer [67]. In IEEE 802.11n, OFDM divides a spectrum band (20 MHz/40 MHz) into a set of closely spaced orthogonal (64/128) sub-carriers. The 64 subcarriers in 20 MHz as shown in figure 2.1 are divided into 52 subcarriers for data, 4 for pilot and 8 as null subcarriers. In case of 40 MHz, 128 subcarriers are into 108 subcarriers for data, 6 for pilot and 14 as null subcarriers. The frequency spacing between two consecutive subcarriers is 312.5 KHz (20 MHz or 40 MHz / 64 or 128) [12, 10, 31]. The modified firmware of Intel 5300 provides CSI values for 30 subcarriers only per transmit-receive (TR) link which corresponds to the grouping Ng = 2 for 20 MHz and Ng = 4 for 40 MHz as shown in Table 2.1 [14].

In a narrowband flat-fading channel with MIMO, the Channel Frequency Response (CFR) can be modeled as:

$$y = Hx + n \tag{2.1}$$

where y and x are the received and transmitted vectors, respectively, H is the channel matrix, and n is an additive white Gaussian noise vector. The noise component is often modeled as a circular symmetric complex normal with n $\sim \mathcal{CN}(0, S)$, where the mean value is zero and the noise co-variance

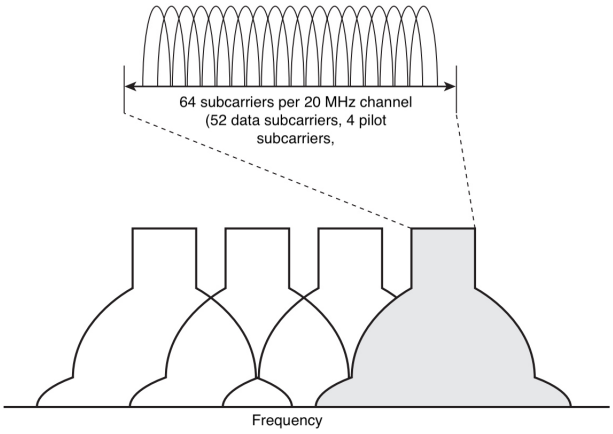


Figure 2.1: OFDM Subcarriers in 802.11n Wi-Fi

BW	**Grouping *Ng***	***Ns***	**Carriers for which matrices are sent**
20 MHz	1	56	All data and pilot carriers: –28, –27,…–2, –1, 1, 2,…27, 28
	2	30	–28,–26,–24,–22,–20,–18,–16,–14,–12,–10,–8,–6,–4,–2,–1, 1,3,5,7,9,11,13,15,17,19,21,23,25,27,28
	4	16	–28,–24,–20,–16,–12,–8,–4,–1,1,5,9,13,17,21,25,28
40 MHz	1	114	All data and pilot carriers: –58, –57, …, –3, –2, 2, 3,…, 57, 58
	2	58	–58,–56,–54,–52,–50,–48,–46,–44,–42,–40,–38,–36,–34,–32,–30, –28,–26,–24,–22,–20,–18,–16,–14,–12,–10,–8,–6,–4,–2, 2,4,6,8,10,12,14,16,18,20,22,24,26,28, 30,32,34,36,38,40,42,44,46,48,50,52,54,56,58
	4	30	–58,–54,–50,–46,–42,–38,–34,–30,–26,–22,–18,–14,–10,–6, –2, 2,6,10,14,18,22,26,30,34,38,42,46,50,54,58

Table 2.1: IEEE 802.11n Standards

matrix S is known. Therefore, an estimation of H, also known as Channel State Information (CSI), is:

$$\hat{H} = \frac{y}{x}$$

In OFDM, the CSI of a single subcarrier can be expressed as:

$$h = |h|e^{jsin\theta} \tag{2.2}$$

, where $|h|$ and θ represent amplitude and phase, respectively [59].

As shown in Figure 2.2, the received signal in an indoor environment is a sum of LOS path, static reflected paths from static objects such as ceiling or floor, and dynamic reflected paths caused by human movement. In a static environment where there is no motion in the room, the channel

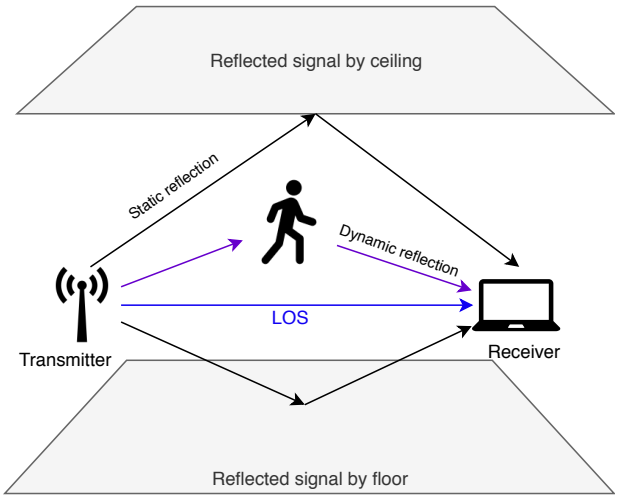


Figure 2.2: Multipaths caused by human movement in an indoor environment

is relatively stable. However, when a person is moving in the room, the channel fluctuates, leading to amplitude attenuation and phase distortion.

2.2 Wi-Fi CSI Error Sources

Due to the hardware imperfection in the wireless signal processing, the CSI suffers from measurement errors. To understand the sources of CSI error, one must first understand the overall signal processing in IEEE 802.11n. Figure 2.3 illustrates a typical signal processing architecture at the receiver. First, the Automatic Gain Controller (AGC) compensates the incoming signal amplitude attenuation s(t) to the transmitted power level. Then, the signal is sampled as a discrete digital signal s(n) by Analog-to-Digital converter (ADC). When there is a correlation between s(n) and the pre-defined 802.11 preamble pattern, a packet is detected and, then, the signal central frequency is estimated and corrected by the Central Frequency Offset (CFO) corrector. In order to extract the data correctly, the receiver estimates the instantaneous CSI and, then, the channel equalizer compensates amplitude attenuation and phase errors [61, 72, 47].

As shown in Figure 2.3, a number of possible errors have occurred while processing the signal. These errors include power amplifier uncertainty (PAU), sampling frequency offset (SFO), carrier frequency offset (CFO), packet detection delay (PDD), phase-locked-loop phase offset (PLO), phase ambiguity (PA), and system nonlinearity (SN). The impact of these errors can cause CSI amplitude offset, CSI phase rotation, and offset [61, 71, 46, 46, 52, 50].

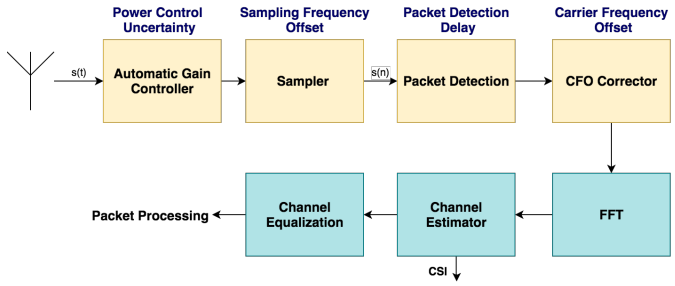


Figure 2.3: Overview of Signal Processing in 802.11n

2.2.1 Power Amplifier Uncertainty (PAU)

The AGC cannot perfectly calibrate the signal amplitude attenuation to the transmitted power level; so, the measured CSI amplitude is a sum of the calibrated power level and the power control uncertainty error. As a result, high amplitude impulse and burst noises are introduced in the CSI stream [56]. Averaging individual bands can mitigate the CSI amplitude offset. However, this is not always the case when the number of CSI measurements on each band is insufficient [19, 61].

2.2.2 Sampling Frequency Offset (SFO)

Because of non-synchronized ADC clocks, the sampling signal s(n) at the receiver is shifted by an offset $\phi_S(t)$ relative to the transmitter. However, the clock offset is almost constant over a short time and, therefore, the phase rotation errors are constant.

2.2.3 Carrier Frequency Offset (CFO)

Similarly, the central frequencies of the transmitter and receiver cannot synchronize perfectly due to a mismatch between the transmitter's and the receiver's local oscillator. Even though there exists a module "CFO corrector", which calibrates the central frequency, the signal s(n) still carries residual errors. As a result, it causes phase offsets $\phi_C(t)$ for all CSI sub-carriers. Wang et al. indicated that a CFO of 5 GHz band can reach up to 80 KHz, leading to a phase shift of 8π. However, the body movement generally leads to a phase shift of 0.5π, which is not observable from the raw CSI phase [56].

2.2.4 Packet Detection Delay (PDD)

One of the simplest ways to detect a packet while requiring a minimum amount of signal processing is to detect its energy based on the few first samples [19]. If a number of samples cross its energy detection threshold,

then the packet is detected. However, the number of required samples required to detect the energy varies based on the received signal power as well as noise. As a result, the packet detection module produces a time shift signal $\phi_P(t)$ due to the sensitivity of the packet detector [61, 52, 72].

2.2.5 Phase-Locked Loop Offset (PLO)

The phase-locked loop is responsible for generating a centre frequency for both the transmitter and the receiver [3]. Due to hardware imperfection, both the transmitter and the receiver have some initial random phase, resulting in a phase difference in the received OFDM symbol. Therefore, a relatively constant phase offset ϕ_{PLL} is added to the CSI phase [52, 72, 21].

2.2.6 Phase Ambiguity (PA)

When analyzing the phase difference between two receiving antennas, Tzur et al. pointed that there exists another source of error known as "phase ambiguity", which can be found in Intel 5300 NICs when working on 2.4 GHz. This error exists because of the firmware implementation. There are two sources of ambiguity: symmetric ambiguity and phase-periodic ambiguity.

Symmetric ambiguity occurs when the system cannot distinguish between the physical AoA θ and its symmetric $\pi - \theta$

The phase-periodic ambiguity occurs when $d \geq \lambda/2$, which is the physical phase difference in the range $[-\pi, \pi]$, and is given by:

$$\Delta\phi = \Delta\hat{\phi} + 2\pi k \tag{2.3}$$

Where k is an integer. If $d < \lambda/2$, then $k = 0$, and the ambiguity is eliminated [50].

2.2.7 System Nonlinearity (SN)

The authors in [61, 72, 73] identified a non-negligible non-linear source of phase errors that exists in the IQ imbalance module, as shown in Figure 2.3. The non-linear errors exist in both NICs: Intel 5300 and Atheors AR9380. These errors are stable over time but sensitive to the used frequency band.

Chapter 3

Amplitude, Phase & Phase Difference Analysis

This chapter investigates the extracted base signals: the amplitude (Section 3.1), phase (Section 3.2) and phase difference (Section 3.3) variations across the different subcarriers, and their effects on the human activities. Each base signal has its own pros and cons.

3.1 Amplitude Analysis

Since there are a total of 90 subcarriers for the three received antennas from a single transmitter, we study the correlation amplitudes of subcarriers when there is a body movement. The reason for this study is that we are looking to reduce dimensions instead of processing all 90 subcarriers and hence good computational complexity. Also, to investigate whether averaging all 30 subcarriers [54, 58] into one signal per each antenna is effective.

To measure the linear correlation between subcarriers, the Pearson Correlation Coefficient (PCC) [15] is calculated using Equation 3.1:

$$r_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(\frac{x_i - \bar{x}}{s_x})(\frac{y_i - \bar{y}}{s_y}) \tag{3.1}$$

where x and y represents two subcarrier signals, n is the length of the signal. The term $(\frac{x_i-\bar{x}}{s_x})$ and $(\frac{y_i-\bar{y}}{s_y})$ is the z-score of x and y where $\bar{x}$, $\bar{y}$ and s_x, s_y is the mean and standard deviation of the subcarrier x and y respectively. The r_{xy} values ranges from -1 to +1. A value of +1 indicates that the two subcarriers have a strong positive linear relationship, 0 is no linear relationship and -1 indicates for a strong negative linear relationship. Figure 3.1 shows the linear correlation between subcarriers. We observe that the adjacent subcarriers from the same antenna are more correlated than the subcarriers from different antenna. It can be seen that the first and second

subcarrier for a body movement is strongly correlated than the first and 30th subcarrier. Figure 3.2 shows that the first and second subcarrier have identical amplitude changes whereas the 23th and 83th subcarrier have non identical amplitude changes. Also, the adjacent subcarriers from the second antenna (31-60) are strongly correlated compared to first and third antenna.

Because the further subcarriers are less correlated than the adjacent subcarriers, averaging of all the 30 subcarriers [54, 58] into one signal per link is ineffective because the filtered signal will lose some of its important high-frequency components and sharp transitions, which are essential in recognizing an activity. Hence, a principle component analysis (PCA) is utilized and discussed in section 4.3 to reduce the dimension and extract the most important information from data.

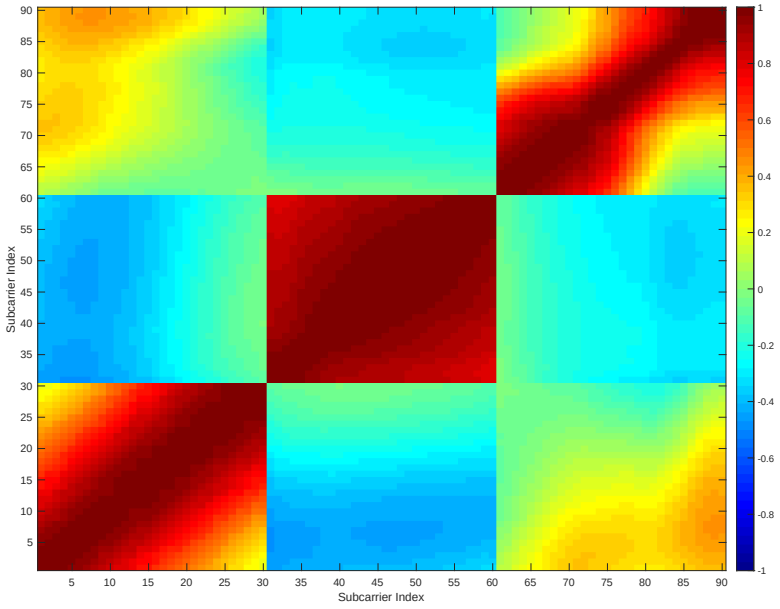


Figure 3.1: Correlation between subcarriers of a body movement across 3 received antennas. Subcarrier index from [1-30, 31-60, 61-90] represents the first, second and third received antenna respectively

3.2 Phase Analysis

Due to the hardware imperfection discussed in section 2.2, the raw phase information cannot distinguish human activities. Figure 3.3 shows the polar histogram of the phase information (1st subcarrier) of a walking activity. As it is shown, the raw phase information is distributed randomly and it is unstable.

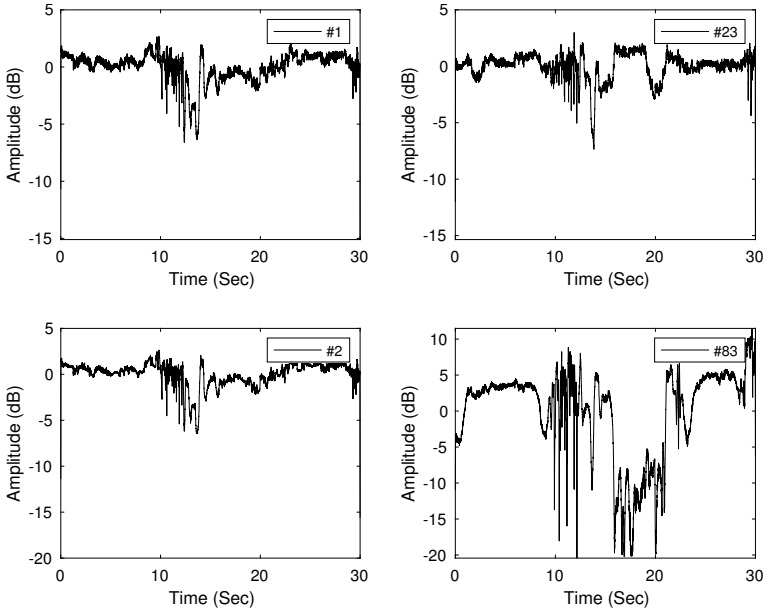


Figure 3.2: Amplitude variations of a body movement across different subcarriers

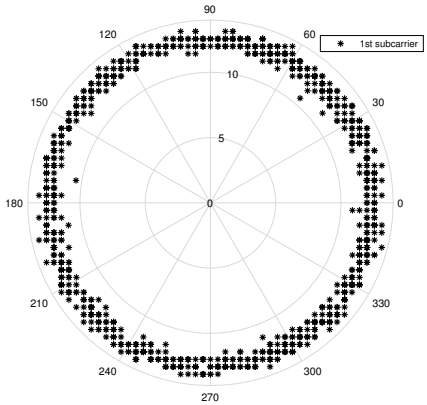


Figure 3.3: The raw phase information of a walking activity

One of the common techniques used to remove the phase error is linear transformation [39, 60]. Let the measured phase $\hat{\phi}_i$ for the subcarrier i be:

$$\hat{\phi}_i = \phi_i + 2\pi \frac{s_i}{N} \Delta t + \beta + Z \tag{3.2}$$

where ϕ_i is the true phase, s_i is the subcarrier index in Table 2.1, N is the FFT size (64 for 20 MHz or 128 for 40 MHz), Δt is the timing offset due to SFO, β is the unknown phase offset due to CFO, and Z is a measurement noise. The key idea of linear transformation is to eliminate the timing Δt and the unknown phase offset β by considering phase across the entire

measured subcarriers (data and pilot subcarriers) which is 56 subcarriers for the 20 MHz or 114 subcarriers for the 40 MHz. Since only 30 subcarriers are measured, the subcarrier indices from the Table 2.1 are asymmetric. In other words, the $\sum_{i=1}^{30} s_i \neq 0$. Assuming they are symmetric $\sum_{i=1}^{30} s_i = 0$ ie ([-15,-14,...,14,15]), the phase slope a between the first and last subcarrier and the mean of the phases of all the subcarriers b can be calculated with the following:

$$a = \frac{\hat{\phi_{30}} - \hat{\phi_1}}{s_{30} - s_1} = \frac{\hat{\phi_{30}} - \hat{\phi_1}}{15 - (-15)} = \frac{\hat{\phi_{30}} - \hat{\phi_1}}{30} \tag{3.3}$$

$$b = \frac{1}{30} \sum_{r=1}^{30} \hat{\phi}_r \tag{3.4}$$

By subtracting the linear equation: $as_i + b$ from the raw phase $\hat{\phi}_i$ equation 3.2, the random phase offsets are removed. The calibrated phase $\widetilde{\phi}_i$ is given by:

$$\widetilde{\phi}_i = \hat{\phi}_i - as_i - b$$

$$\widetilde{\phi}_i = \hat{\phi}_i - (\frac{\hat{\phi_{30}} - \hat{\phi_1}}{30})s_i - \frac{1}{30} \sum_{r=1}^{30} \hat{\phi}_r \tag{3.5}$$

Figure 3.4 shows the linear transformed phase compared to the raw phase. As it can be seen that linear transformed phase is concentrated between [0°-15°]. Figure 3.5 shows the phase variation of walking activity with respect to time.

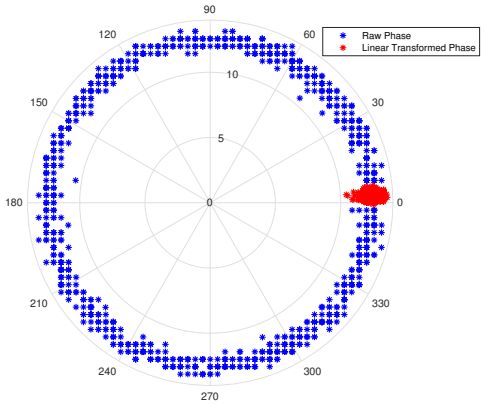


Figure 3.4: The calibrated phase information of the walking activity

Even though the randomness is removed to some extent, the linear transformation does not capture activities that have small phase changes. In our case, it is found that phase changes of a clapping activity is buried in

phase noise whereas activities that have large phase changes such walking is observed as shown in Figures 3.5, 3.6.

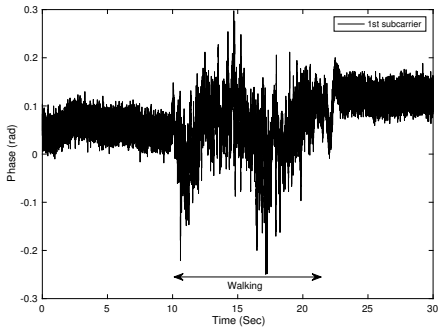


Figure 3.5: The phase changes of a walking activity

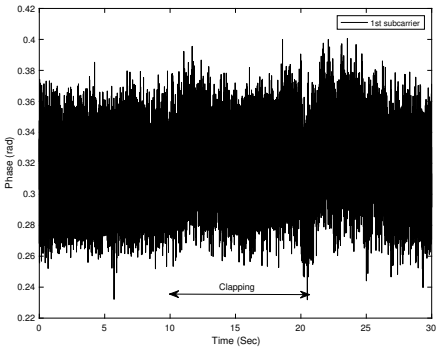


Figure 3.6: The phase changes of a clapping activity

3.3 Phase Difference Analysis

Due to the limitation of linear transformation, the next approach is to exploit the phase difference between two antennas. As shown in Figure 3.7, a phase difference between two subsequent antennas exists because a signal has to travel further to reach each subsequent antenna [37]. The phase difference between two subsequent antennas $\Delta\phi$ is given by:

$$\Delta\phi = 2\pi \frac{d \sin\theta}{\lambda} \tag{3.6}$$

Where d is the distance between two antennas, λ is the wavelength of the signal, and θ is the angle of arrival (AoA) [50].

Due to a shared clock between antennas and having the same SFO, CFO, PDD, and PLO offsets, the offsets are cancelled out. Hence, the phase difference between two antennas is stable [57]. We examined the stability of phase difference between the antennas, the polar histogram in Figure 3.8 shows that the phase difference between two subsequent antennas is being shifted by multiple of 90° in the 2.4 GHz band. The phase difference between first and second antenna, and between second and third antenna, clusters around [0°, 90°, 180°, 270°], [70°, 160°, 250°, 340°] respectively. This is referred as a "four-way ambiguity" which has been observed by Tzur et al [50]. It occurs because of the firmware implementation which measures the phase in modulo $\pi/2$ rather than modulo 2π. As a result, the measured phase difference is unstable, and hard to determine the variation of the phase difference of a clapping activity as shown in Figure 3.9.

We found that the ambiguity issue exists only when using an access point (AP) as a transmitter and a receiver that has Intel 5300 NIC (network

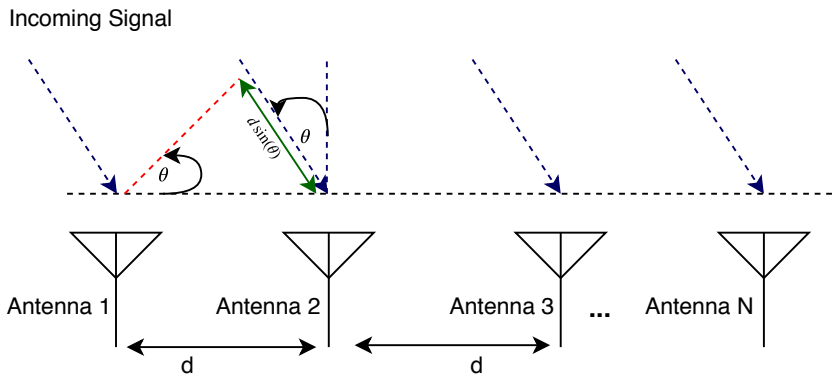


Figure 3.7: A signal hits an antenna array. There are N antennas spaced d apart. As a result, an additional path of $dsin(\theta)$ is added at each subsequent antenna.

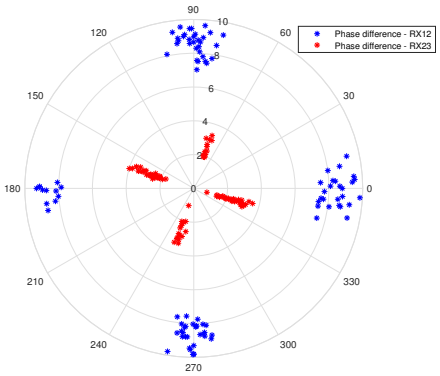


Figure 3.8: Polar histogram of the phase difference between two subsequent antennas in 2.4 GHz band using AP mode

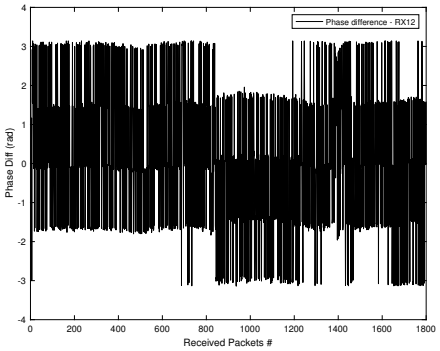


Figure 3.9: The phase difference of a clapping activity in 2.4 GHz using AP mode

interface card). However, it does not exit when working on a monitoring mode where packets are captured without associating to an access point. The monitoring mode works when both WiFi devices have Intel 5300 NIC. Figure 3.10 shows the polar histogram of the phase difference between two subsequent antennas when working on monitor mode. The phase difference values between the first and second antenna is concentrated between 340°and 15°. The second and third antenna's phase difference is clustered from 15°to 30°. Unlike linear transformation, the phase difference data captures small body movement such as clapping as shown in Figure 3.11.

Furthermore, we observed that the ambiguity issue does not exist when working on a 5 GHz band for both monitor and AP mode. As it can be seen from Figure 3.12, the phase difference between two antennas is concentrated

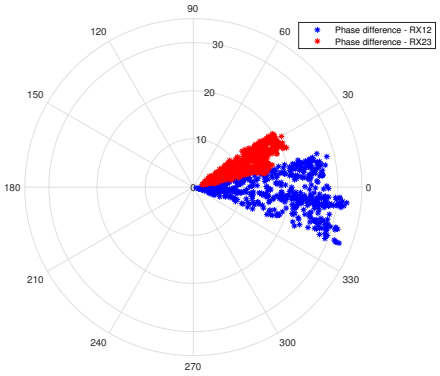


Figure 3.10: Polar histogram of the phase difference between two subsequent antennas in 2.4 GHz band using monitor mode

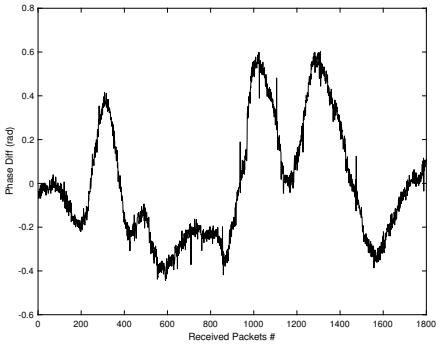


Figure 3.11: The phase difference of a clapping activity in 2.4 GHz using monitor mode

in 255°and 228°. The ambiguity of the clapping activity is eliminated in 5 GHz band as shown in Figure 3.13. The 5 GHz band is chosen through out our experiment because the wavelength is shorter compared to 2.4 GHz which leads to a better movement speed resoultion as shown in Figure 3.13. Also, the 5 GHz is a better option for less inter-channel interference and more reliable communication.

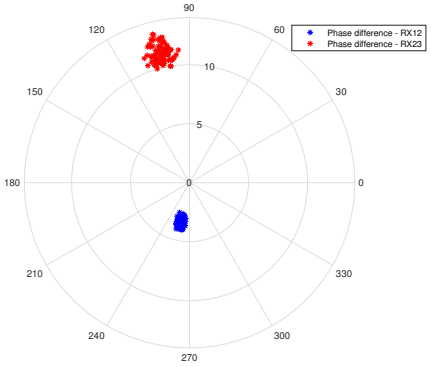


Figure 3.12: Polar histogram of the phase difference between two subsequent antennas in 5 GHz band

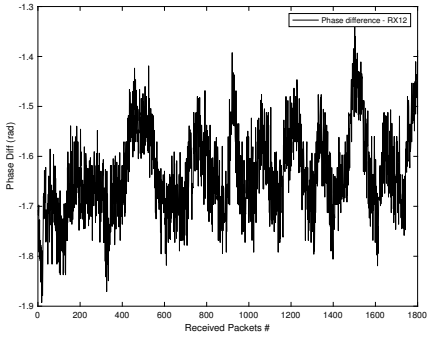


Figure 3.13: The phase difference of a clapping activity in 5 GHz

We also found that the correlation of phase differences between 60 subcarriers (ie phase difference between first and second antenna, and second and third antenna) behave the same as amplitudes correlation between subcarriers (discussed in section 3.1). The subsequent subcarriers are more correlated than the further subcarriers. Hence, PCA is used and discussed

in section 4.3 to reduce the dimensions and extract important information from the data.

3.4 Summary

- The subsequent subcarriers are more correlated than the further subcarriers from the same and different antennas. Hence, averaging all the 30 subcarriers would not be an optimum solution.
- Linear transformation of the phase purposed in [39, 60] removes the random phase error to some extent. We observed that the method only captures large activity movement that have large phase changes such as walking. However, activities with the small phase changes such as clapping is often buried in the noise, and the phase changes cannot be captured when using linear transformation.
- Phase difference between subsequent antennas is more stable than than the phase information. Due to the shared clock between antennas, the offsets are cancelled. However, phase ambiguity is existed in the 2.4 GHz band because of the firmware implementation of Intel 5300. We found that the phase ambiguity is eliminated when working on a monitor mode 2.4 GHz or also working in the 5 GHz band.
- The 5 GHz band is used through out the experiments because it provides a better movement resolution, and has less inter-channel interference.

Chapter 4

CSI Preprocessing

This chapter describes the preprocessing methods used to remove the environmental and internal hardware noises, and improve extracted features. Section 4.1 discusses the linear interpolation used to construct missing data. Then, Section 4.2 describes the windowing approach which enables an online classification. Section 4.3 shows how noise and dimensions of the subcarriers are reduced.

4.1 1D Linear Interpolation

Although the transmitter is configured to send packets at a fixed rate, the collected CSI data are non-uniformly sampled in the time domain because of the packet loss and transmission delay. This raises an issue when extracting features of a signal that is discontinuous. Also evenly spaced data is required to perform time-frequency analysis and obtain spectrogram [70]. A 1-D linear interpolation method [10, 27] is adopted to get an evenly spaced samples with the spacing of 1 ms since our sampling rate of the collected dataset is 1000 Hz. Given two data points (x_1, y_1) and (x_2, y_2) where x_1, x_2 are the packet arrival time, y_1, y_2 are the signal value (amplitude or phase), x is the new equal spaced time, the interpolant is a straight line segment given by:

$$y = y_1 + (x - x_1)\frac{y_2 - y_1}{x_2 - x_1} \tag{4.1}$$

Although 1D linear interpolation is simple and quick to compute, it results in a discontinuous at each point and sharp edges between two adjacent segments especially for non-linear functions [4]. Hence, other interpolation methods such as polynomial or spline would result in a "smooth" interpolating function. Because of the nature of CSI signal in which the amplitude is fluctuated and has abrupt changes as shown in Figure 4.1, a 1D linear interpolation is utilized in the experiment.

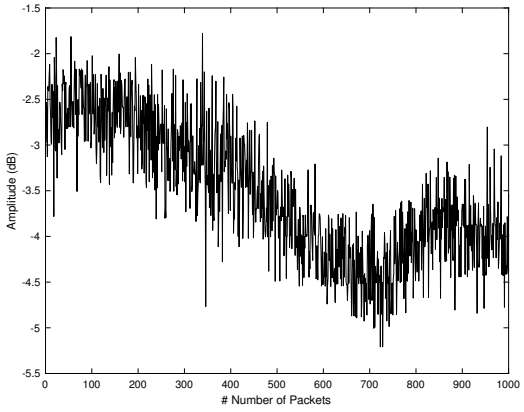


Figure 4.1: Fluctuated Raw CSI amplitude

4.2 Sliding Window

Before extracting the features, the CSI signals are segmented into windows. This will reduce the computation, and enable an online classification of the activity. Selection of the window size is important because short windows may not provide sufficient information of the performed activity, and large windows may have more than one activity. Also, the windows are calculated with the certain overlap to handle the transition more accurately and reduce the misclassification [26]. In this study, a fixed sliding window approach of 4 seconds is used with a 50% overlap. Different window sizes are tested in Section 7.4.2 to evaluate the performance of the classifier.

4.3 Noise & Dimension Reduction

From section 3.1 and 3.3, we showed that the adjacent subcarriers from the same antenna are more correlated than the further subcarriers for both amplitudes and phase differences, and averaging the 30 subcarriers is ineffective due to loss of activity information. Also, we discussed in section 2.2 that burst noise and high amplitude impulse exists for every subcarrier due to power amplifier uncertainty (PAU). The question is how to ensure that the noise is removed and at the same time the dimension is reduced?

We firstly utilize principle component analysis (PCA) [20] to reduce the dimension and noise. This is achieved by transforming a set of possibly linearly correlated variables into linearly uncorrelated variables called principle components (PCs) in such a way that the first principle component has the highest possible variance which retains most of the data variability. In our experiment, the PCA is applied for each antenna (ie 30 subcarriers) separately for both amplitudes and phase differences because the results from section 3.1 show that subcarriers from different antennas tend to be

uncorrelated.

There are four main steps of applying PCA to the CSI matrix H which is $(n \times m)$ where n is the signal length, and $m = 30$ the total number of subcarriers so that we have:

$$H = [H_1, H_2, H_3, ..., H_m] \tag{4.2}$$

1. *Normalizing H matrix.* We normalize the H matrix using z-score such that each column i=1,...,m will have a zero mean and unit variance. The normalized version of H is denoted as Z. The effect of normalization is studied in section 4.3.1.

2. The *covariance matrix* Q of Z is calculated such that $Q = Z^T Z$. The covariance matrix is a $m \times m$ symmetric matrix.

3. *Eigen-decomposition* of the covariance matrix is performed to calculate the eigenvectors v_i, i = 1,2,...m such that eigenvectors are ranked in order of their eigenvalues from highest to lowest.

4. *Reconstruction* of the signal is performed such that $z_i = Z \times v_i$ where i is the ith principle component.

5. *The optimum number of PCs* that describes more than 90% of the variance of the data is the first 2 PCs. Based in our experiment with the training dataset in Table 7.1, we observe that only the top two PCs that describes more than 90% of the information. Hence, the **total number of PCs = 10 PCs** (2 PCs amplitudes x 3 antennas + 4 PCs phase differences between two subsequent antennas: RX12 & RX23).

Not only does PCA reduce the dimension, but also reduce the burst noise. Figure 4.3 shows the the top two PCs of the walking activity. As it can be seen that the noise has been reduced significantly in the second PC which represents 8% of the activity information compared to the raw CSI amplitude in Figure 4.2. However, there is still some correlated noises in the first PC - 83% of the activity information that PCA could not remove it. Hence, we utilize a moving average to smooth the signal and reduce the burst noise. We choose the window size to be 50 ms so that it is large enough to smooth the signal and at the same time not to distort the activity signal. Figure 4.4 shows the smoothed version of the walking activity signal. We can see that the burst noises has been removed as well as the large impulse that exist at time 15 second.

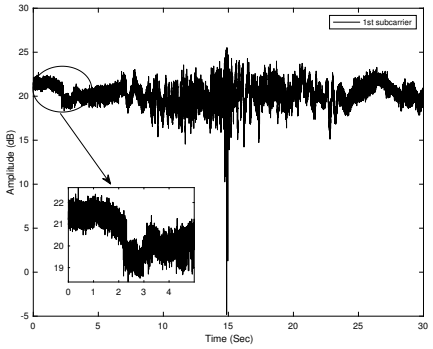


Figure 4.2: Raw CSI amplitude of a walking activity

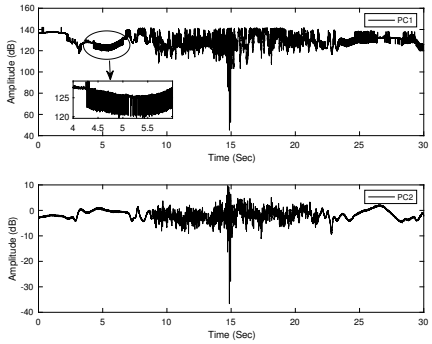


Figure 4.3: After PCA. PC1 represents 83% of the total variance of data. PC2 represents 8%

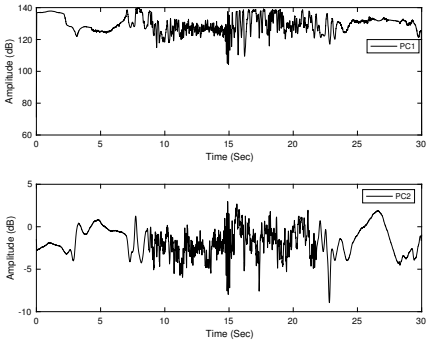


Figure 4.4: After moving mean

4.3.1 Impact of Standardization in PCA

Standardization is important in PCA because subcarriers with high amplitudes will not dominate subcarriers with smaller domains, and hence preventing from having a biased result. For instance, we performed a PCA on the first antenna (30 subcarriers) without normalization for a sitting activity, and the results in Figure 4.5 shows that the first two components explain 95% of the total variance and the first PC explains about 78% of the total variance. However, the results in Figure 4.6 shows that the the first three components explain more than 95% and the first PC explains only 50% when applying a z-score normalization. It is clear that the second and third PC also contribute to the overall variance of the data.

The reason why standardization does effect the outcome of PCA is that is the covariance is scaled by the product of standard division as derived in Equation 4.5. We know that the standardization of a variable can be written as:

$$z = \frac{x_i - \bar{x}}{\sigma} \tag{4.3}$$

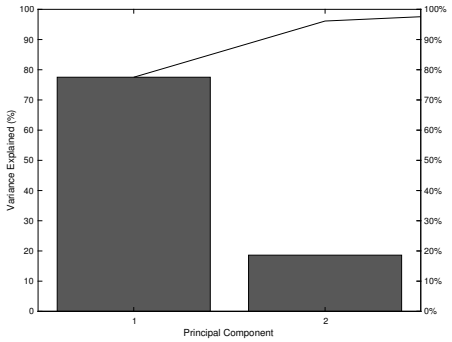


Figure 4.5: Before normalization: the variance accounted for by each component for a sitting activity

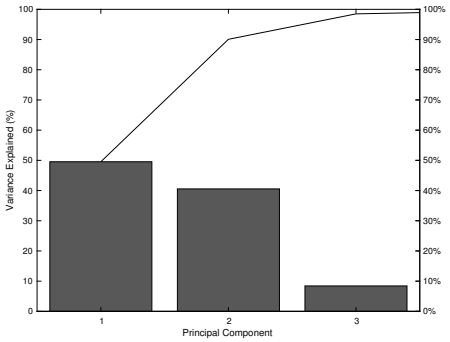


Figure 4.6: After normalization: the variance accounted for by each component for a sitting activity

and the covariance between subcarrier x and y can be expressed as:

$$cov(x,y) = \frac{1}{n-1}\sum_{i}^{n}(x_i - \bar{x})(y_i - \bar{y}) \tag{4.4}$$

After standardizing both subcarriers x and y:

$$x_i^{'} = \frac{x_i - \bar{x}}{\sigma_x} \text{ and } y_i^{'} = \frac{y_i - \bar{y}}{\sigma_y}$$

The new covariance be written as:

$$cov(x,y)' = \frac{1}{n-1}\sum_{i}^{n}(x_i^{'} - \bar{x})(y_i^{'} - \bar{y})$$

Since standardized subcarrier x, y has a zero mean, then:

$$= \frac{1}{n-1}\sum_{i}^{n}(x_i^{'} - 0)(y_i^{'} - 0)$$

$$= \frac{1}{n-1}\sum_{i}^{n}(\frac{x_i - \bar{x}}{\sigma_x})(\frac{y_i - \bar{y}}{\sigma_y})$$

$$= \frac{1}{(n-1)(\sigma_x\sigma_y)}\sum_{i}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$cov(x,y)' = \frac{cov(x,y)}{\sigma_x\sigma_y} \tag{4.5}$$

4.4 Summary

- Performing time-frequency analysis and obtaining spectrogram requires data to be evenly spaced. Due to the transmission delay, packet loss, the collected CSI data are non-uniformly sampled in the time domain. Hence, 1D linear interpolation is adopted to have an evenly spaced of 1 ms since the sampling rate of our CSI data is 1000 Hz.
- A fixed sliding window with 50% is applied to ensure the transition is smooth and reduce the misclassfication.
- PCA is used to remove the burst noise that exists for each subcarrier. Only the first and second principle component are utilized which represents 90% of the variance of the activity information. Since there are correlated noises that mostly exists in the first principle component, a moving average is used to smooth the signal.
- Standardization of the CSI data affects the outcome of PCA because the covariance is scaled by the product of standarnd division.

Chapter 5

Time-Frequency Analysis

This chapter describes extraction of time-frequency features which are essential in classification problem. Section 5.1 explains why time-frequency analysis is used. Section 5.2 gives an insight into Continuous Wavelet Transform (CWT), and how it is different from Short-Time Fourier Transform (STFT). Section 5.3 discusses the selection of CWT parameters: the scale and wavelet family. Lastly, Section 5.4 shows how to apply CWT to the principle components.

5.1 Why Time-Frequency Analysis?

The most important part in identifying different activities from the signal is feature extraction. If the selected features are unique and well defined, then any simple classification method would be able to recognize and classify the different activities. The aim of our work is classify different activities in different environments performed by different people. Hence, the selected features need to be:

- Representative: The ability of the selected features to represent the activity that is being done regardless the different environment and people.
- Discriminative: Not only does the selected features need to represent the activity, but also discriminative among other activities.

Statistical time domain features such as mean, standard deviation, median absolute deviation...etc have been commonly used in activity recognition. For instance, Wifall [59] uses seven features to detect a fall. These features are normalized standard deviation of CSI, offset of signal strength, period of the motion, median absolute deviation, inter-quartile range, signal entropy, and velocity of signal change. Wifall can achieve a detection precision of 87% in average. However, the results from Wifall show that the performance degrades when the environment changes.

Due to the multipath, the same activities performed in different environments result in different amplitude changes and different statistical properties. However, representing the CSI signal in frequency domain which is related to the speed of the change in multipaths will result in similar features obtained from different environments [56]. Because the CSI signal is non stationary where frequency changes over time, time frequency domain features are extracted in our experiment.

5.2 Continuous Wavelet Transform (CWT)

The Continuous Wavelet Transform (CWT) provides a high localization in time and frequency by continuously varying the scale a and translation (shifting) b parameter of the wavelets. The wavelet transform of a continuous time signal $x(t)$ can be defined as:

$$X(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t)\psi^* \left(\frac{t-b}{a}\right) dt \tag{5.1}$$

where $\psi^*(t)$ is the complex conjugate of the analyzing *mother wavelet* function $\psi(t)$. The a parameter causes the dilation of the wavelet, while b causes shifting of the wavelet. The output $X(a,b)$ is wavelet coefficients X which are a function of scale and position. The magnitude spectrum of CWT is represented by $|X(a,b)|$. The $\frac{1}{\sqrt{|a|}}$ factor ensures energy normalization so that the energy of the scaled mother wavelet equal to energy of the original mother wavelet. When the wavelet is compressed (ie the the scale $0 < a1$), abrupt changes and high frequency components are captured while stretched wavelet $a > 1$ helps capturing low frequency components. The CWT of the signal is represented using a *scalogram* which is a time frequency representation of the signal by wavelet transform.

The scale parameter in CWT is described in terms of octaves and voices. The number of octaves determine the frequency range to be analyzed. For instance, if the wavelet is dilated by a factor of 2 for a sampling frequency of 1000 Hz, then it results in an equivalent frequency by an octave - 500 Hz. The number of octaves can be determined by:

$$\text{Number of octaves} = log_2(\frac{f_{max}}{f_{min}}) \tag{5.2}$$

where f_{max} and f_{min} is the maximum and minimum frequency respectively. The number of voices per octave determine the number of scales between successive octaves. For instance, a 10 voices per octave results in ratio of $2^{1/10}$ between voices. Given the number of octaves N and number of voices per octave L, the total number of scales would equal to $N \times L$. The scale value for a given octave n_0 and voice l_0 is:

$$\text{Scale value} = 2^{(n_0 + \frac{l_0}{L})} \tag{5.3}$$

where $n_0 = [0, 1, 2, ..., N-1]$ and $l_0 = [0, 1, 2, ..., L-1]$

5.2.1 Difference to STFT

We utilize CWT because it provides better time localization for short duration, high frequency events, and better frequency resolution for long duration, low frequency events. Hence, CWT would be able to localize activities such as fall and sit which causes abrupt changes in high frequency components for a short period of time. Unlike Short-Time Fourier Transform (STFT), the analysis window is fixed. Large window size will lead to high frequency resolution but bad time resolution while small window size will lead to better time resolution but bad frequency resolution. Choosing the right window size is a trade-off between time and frequency [38]. The time-frequency plane tiling of the STFT and CWT is shown in Figure 5.1.

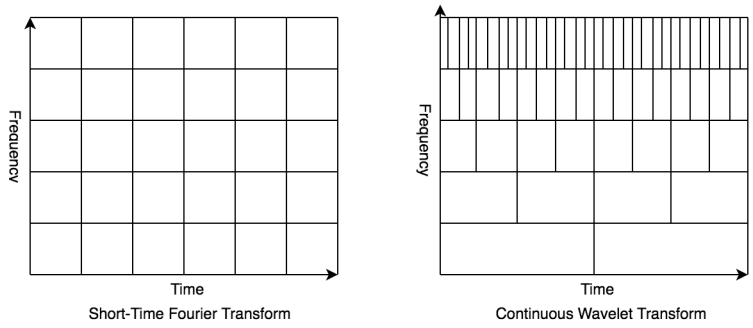


Figure 5.1

We have applied STFT and CWT to the first principle component derived from the 30 subcarriers amplitudes of the first antenna in order to investigate the time-frequency trade-off. Our sampling frequency is 1000 Hz and the window size is set to 500 samples and 100 samples so that it gives a time resolution of (500/1000) = 500 ms and (100/1000) = 100 ms. Figure 5.2 shows the spectograms and scalogram of the fall activity that occur at $t = 1.5s$. We can see that with the window size of 500 ms in Figure 5.2(b), the fall activity occurs precisely at the frequency of 22 Hz but it is not possible to locate the end of activity; hence bad time resolution. With the window size of 100 ms, the fall activity in Figure 5.2(c) can be located in time but imprecise estimation of the frequency (from 15 Hz to 30 Hz). However, the CWT in Figure 5.2(d) achieves good time localization at the high frequency components (20 Hz to 30 Hz), and good frequency resolution at the low frequency components (1 to 10 Hz).

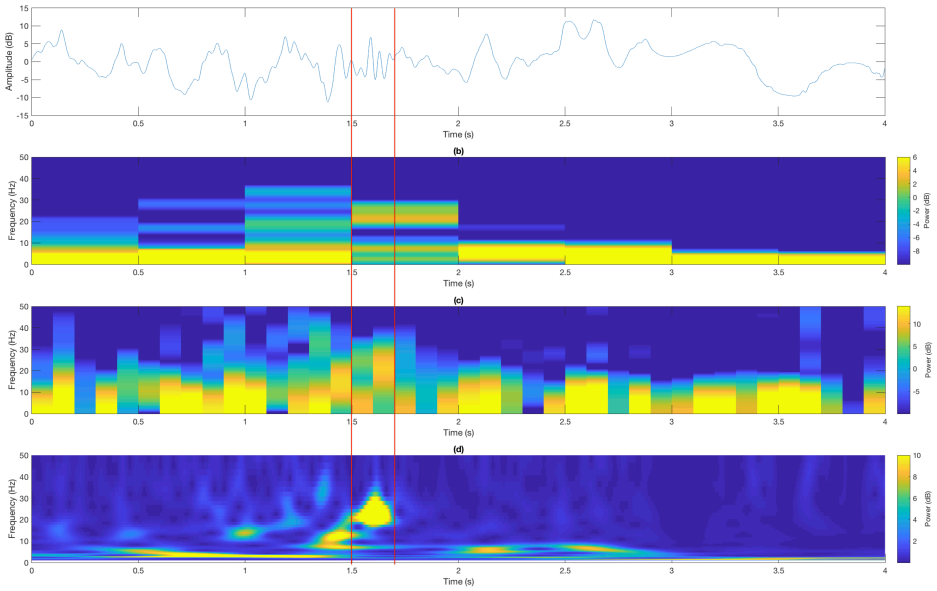


Figure 5.2: Illustration of the properties of STFT and CWT. (a) A fall activity at time $t = 1.5s$. (b) STFT with window length of 500 ms. (c) STFT with window length of 100 ms. (c) CWT with the Morlet wavelet

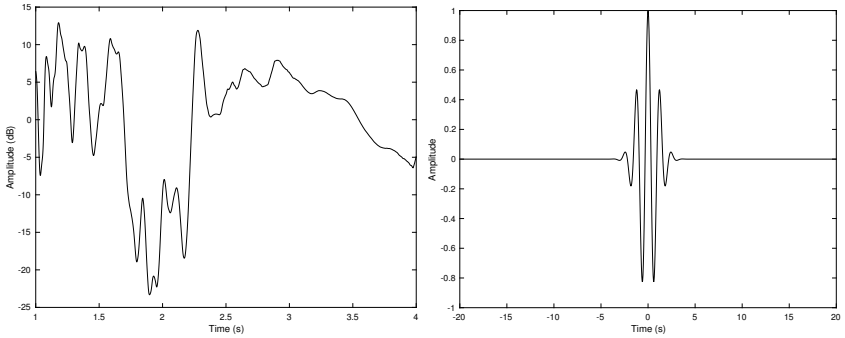
5.3 Selection of Scale and Wavelet Family

As seen from Equation 5.1, the scale and wavelet family should be defined to analyze the signal. The scale parameter is determined by the frequency range we interested in analyzing the human activity. In this study, we analyze four different activities: walk, clap, fall and sit. Walk and fall will have the highest average velocity compared to sit and clap which causes to have the highest frequency. According to [5], the average indoor walk is about 1.2 ± 0.3 m/s, and a typical human fall can reach to a maximum speed of 5 m/s [45]. Hence, the maximum frequency will be obtained from the fall activity. Using the Doppler Frequency formula, the maximum Doppler frequency can be calculated:

$$\Delta f_{max} = \frac{2v_{max}}{c} F_0 \tag{5.4}$$

where v_{max} is the maximum velocity, c is the speed of light, and F_0 is the center frequency. For a WiFi at 5.7 GHz, the maximum Doppler frequency for the fall activity is 190 Hz. To ensure the activity coverage, we set the maximum frequency threshold to 200 Hz and the minimum frequency is 1 Hz in our study. Hence, the number of octaves according to Equation 5.2 would be around 7 octaves. The number of voices per octave is set to 15 voices so that redundancy and computation are compromised.

There many different wavelets that can be used in CWT, and each wavelet has its own properties depending on the signal being analyzed. These properties such as orthogonality, compact support, symmetry and vanishing moment. Some wavelets have the same properties, so the shape of the wavelet is also considered [28, 49]. In our study, a Morlet wavelet is chosen because it has the same structure and shape with CSI signal. The Morlet wavelet is a product of sinusoidal and gaussian function as shown in Figure 5.3b. Figure 5.3a shows the fall activity of the first principle component from the first antenna.



(a) A fall activity from the first PC (b) A Morlet Wavelet

5.4 Applying CWT to the Principle Components

Given there are a total of 10 principle components (2 PCs amplitudes x 3 antennas + 4 PCs phase difference between two subsequent antennas: RX12 & RX23) as seen from section 4.3, now the CWT is applied for each principle component individually.

Because the principle components are orthogonal to each other and each one of them has unique frequency components, a higher amount of activity information can be obtained by summing the magnitude of CWT coefficients from the same antenna. Therefore, the total number of scalograms for each activity will be **5 scalograms** (1 scalogram/amplitude x 3 antennas + 2 scalograms/phase differences between two subsequent antennas (RX1 & RX2)).

The summed magnitude scalograms of the four different activities from PCs of the first antenna amplitudes are illustrated in Figure 5.4. It can be seen that the energy in the walking activity is spreading across different frequencies from 1 Hz to 50 Hz, while the clapping activity has short peak of 50 Hz. In fact, the number of claps can be determined by counting the number of peaks. Looking at the image, the number of claps is 3 claps per second in average. The fall activity is characterized by high peak of about 60 Hz and after that the energy is spreading in lower frequencies for a period

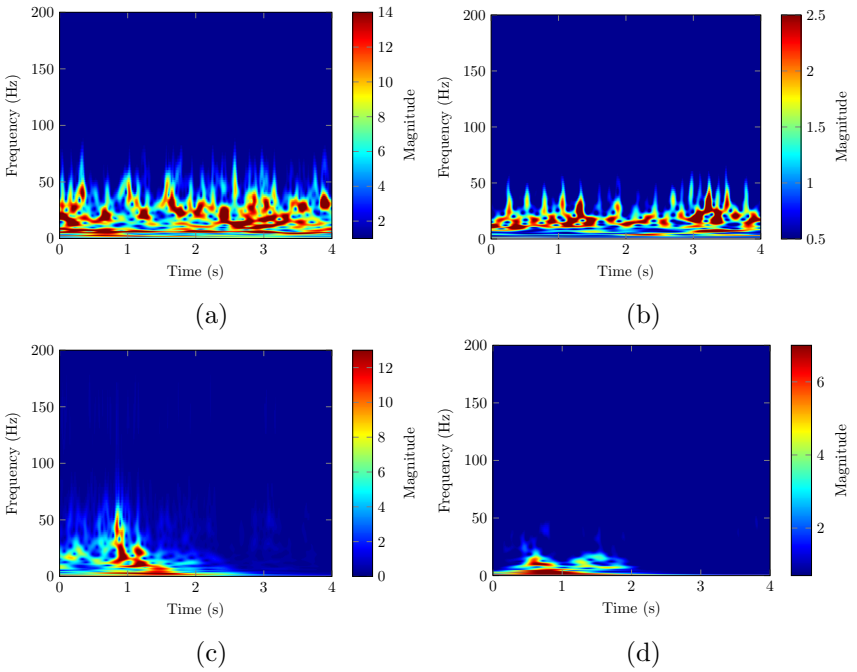


Figure 5.4: Comparison of the CWT scalograms of different activities. (a) Walking, (b) Clapping, (c) Falling, (d) Sitting.

of 1.5 seconds, but the sit activity has only a short peak of 25 Hz. Algorithm 1 summarizes the procedure of getting the five scalograms.

Algorithm 1: Pseudo Code of generating the five scalograms

Result: Five generated scalograms
input: CSI_RAW: CSI measurements from 3 received antennas n
1 Get the phase difference between subsequent antennas
2 Interpolate both amplitude and phase difference
3 Apply PCA and moving average
4 Apply CWT to only the first and second principle component (PC)
5 Sum the magnitude of CWT coefficients of both PC1 & PC2
6 Generate the scalograms from the summed magnitude

5.5 Summary

- Time-Frequency analysis is applied to extract features that are resilience to environment changes. These features represent the speed of change in multipath with respect to time.
- Continuous Wavelet Transform is chosen due to high resolution time-frequency representations. Unlike STFT, CWT provides better time resolution at high frequency and better frequency resolution at low frequency.
- A typical human fall can reach up to a maximum speed of 5 m/s which results to maximum frequency of 190 Hz with the 5.7 GHz WiFi band. Hence, the maximum number of octaves according to Equation 5.2 is 7 octaves. The number of voices per octave is set to 15 voices so that redundancy and computation are compromised.
- The Morlet wavelet is selected based on the similarity between the CSI signal and wavelet.
- The CWT is applied individually to the principle components of amplitudes and phase difference between antennas. There are total of 10 principle components (2 PCs amplitudes x 3 antennas + 4 PCs phase difference between two subsequent antennas: RX12 & RX23).
- Because the principle components are orthogonal, a higher amount of activity information is obtained by summing the magnitude of CWT coefficients from the same antenna.

Chapter 6

Deep Learning Based Activity Recongiton

This chapter shows that Convolution Neural Network can be used to train the obtained scalograms of CSI amplitude and phase difference to extract features automatically. Section 6.1 discusses why CNN is used and its basic structure. Then, Section 6.2 describes the architecture of AlexNet CNN and how the scalograms are feed into the architecture.

6.1 Why Convolution Neural Network (CNN)?

Existing WiFi based activity recognition uses K-Nearest Neighbors (KNN) [2], Hidden Markov Model (HMM) [56] and Support Vector Machine (SVM) [59, 8, 11, 54, 68, 29, 55, 58] to process the extracted features and classify different activities. However, these models experience issues when dealing with CWT scalograms.

The CWT scalograms of the different activities have different features in both time and frequency dimensions. The KNN model does not consider time, so activities that have similar frequencies but different in time cannot be distinguished. HMM considers time but it is based on the Markov assumption which states that the probability of a given state at time t only depends on the state at time $t-1$ but it is not always the case with activities where dependencies extend to sequence of states. Also, it is based on a strong assumption that the successive observations are independent but in reality they rarely independent [7]. The SVM is simple and effective but it requires features engineering, fine tuning a number of parameters and the optimality of the solution is depended on the kernel [43].

Convolution Neural Network (CNN or ConvNet) is one of the common deep neural networks which is used especially in images, and has many applications in computer vision. The strength of CNN lies on automating the feature extraction from raw inputs and ability to learn localized patterns

through weights sharing and pooling. Furthermore, the extracted features of CNN are shift-invariance which is the ability to recognize the patterns at different locations of the input [64]. This is an important property in classifying the CWT scalograms because the activity signal can happen at any period of time. CNN consists of input, hidden and output layers. There are three main types of the hidden layers: convolution layer, pooling layer and fully connected layer [17].

The convolution layer is the main component of CNN which consists of several filters or kernels that are small in size. These filters are used to extract particular features of the image such as edges, curves, lines,...etc. They have the shape of $(size \times size \times depth)$ which is basically an array of numbers also referred as weights, and each point in the filter is a neuron. The filters are applied to the input image of shape $(height \times width \times depth)$, where depth is the number of channels of the image. The output of this layer is called *feature maps* with the shape of $size \times size \times$ number of filters.

The pooling or down-sampling layer is responsible for reducing the spatial dimension of the input and hence reducing the number of neurons and computation in the network. Another advantage is that it extracts high level features regardless of the positional information so that scale changes, rotation, and occlusion of the image are handled. Most common types of pooling layer are max and average pooling. An example of max and average pooling is shown in Figure 6.1.

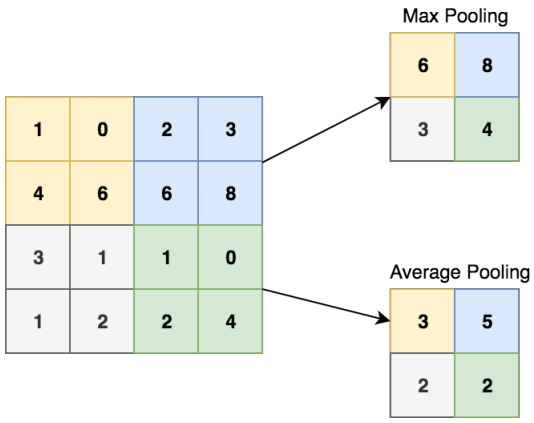


Figure 6.1: Max and average pooling 2×2 filter with 2 strides

The activation layer takes a value and passes into non linear functions which squashes the value into range. The output of the activation layer is also called *activation maps*. The most common activation function is ReLU (Rectified Linear Unit) which is defined by: $f(x) = max(0, x)$ where x is the input to a neuron.

It is common with the large network size, dropout layer is applied to a certain layer in order avoid the problem of overfitting [44]. Overfitting

occurs when the model fits the data too well which results in high accuracy in training set but bad accuracy in test sets. The dropout layer randomly drops certain neurons during the training, and this significantly improves the performance of the network. After several convolution, activation and polling layers, the fully connected layer maps the extracted visual features to the desired output. Typically two or three fully connected layers are used and the size of last layer of the network is the same as the number of predicated classes.

One limitation of using CNN is that training is expensive and it requires large amount of labeled data to train and achieve a good convergence. However, there is a concept called *knowledge transfer* or *transfer learning* where a model is trained in one problem domain and applied to another domain but related problem. This will speedup the training process and overcome the issue of data shortage.

6.2 AlexNet

Since our dataset is limited, we utilize AlexNet pre-trained model to fine tune and classify the CWT scalograms. AlexNet is one of the famous CNN architectures that won the 2012 ImageNet LSVRC (Large Scale Visual Recognition Challenge) with the top five errors of 15.3% [24]. The ImageNet LSVRC project contains a database of 14 millions images and more than 20,000 categories. AlexNet contains eight layers with weights: five convolution layers and three fully connected layers. The dropout layer is used in the first two fully connected layers. The output layer has four connected neurons which represent the four activities. The ReLU layer is applied after every convolution. One of the main challenges in using AlexNet is that the input layer accepts an RGB image size of $224 \times 224 \times 3$ but there are five scalogram images that represent the activity signal obtained from section 5.4.

The question is how do we feed the five scalograms of the activity into AlexNet CNN? There are several options that we could follow. The first option would be to train each scalogram separately and then combine the results but this will result in a poor performance because the five scalograms are dependent with each other. Another option would be to place an image on top of each other on a single image but with 5 channels instead of 3. This would be the ideal solution but the input layer cannot be modified since we are using pretrained model. Therefore, we stack the five scalograms in the top of each other and then feed it to AlexNet CNN as shown in Figure 6.2. This will create discontinuities at the boundaries but the AlextNet CNN is deep enough to recognize these noise.

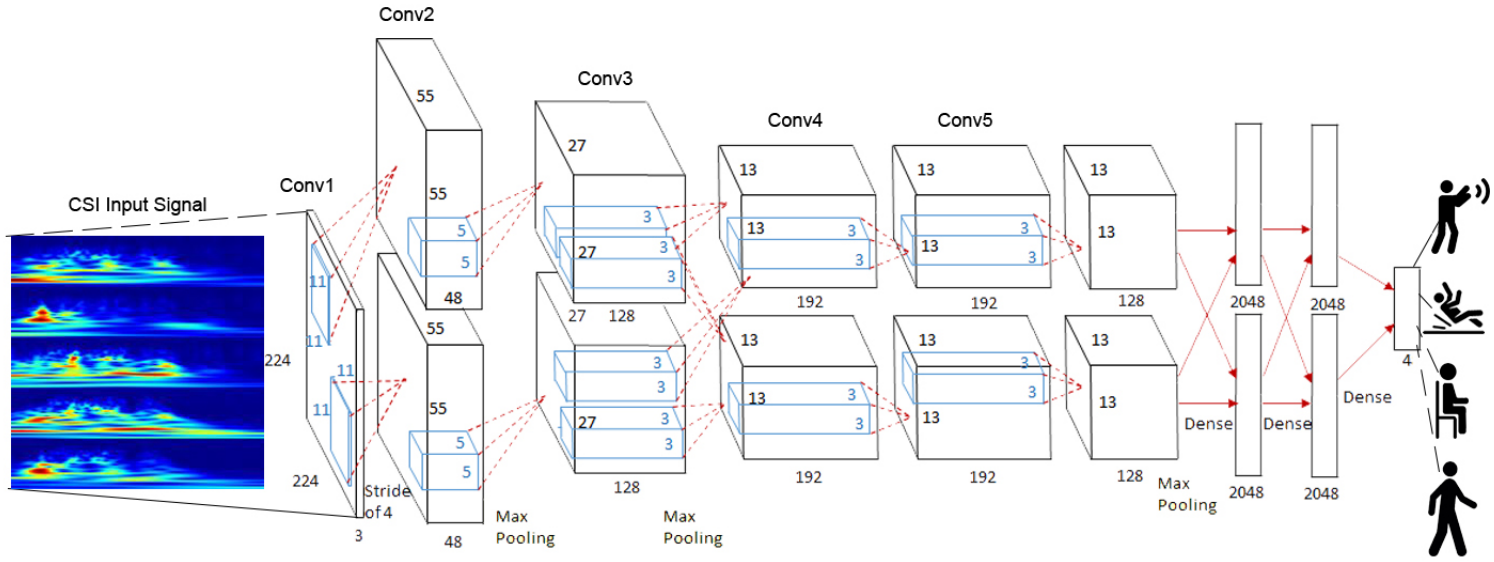


Figure 6.2: The AlexNet architecture

Chapter 7

Implementation And Evaluation

This chapter starts with the implementation details in Section 7.1. In Section 7.2, we present the data collection procedure and the evaluation setup. Then, Section 7.3 discusses the metrics used to evaluate the system. Section 7.4 presents and discusses the experimental results.

7.1 Implementation

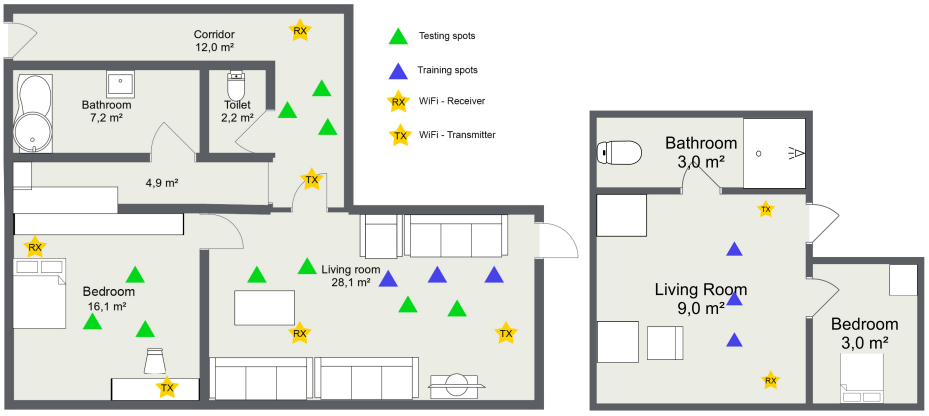
We implemented the system using two laptops: Lenovo Thinkpad X220 i5-2540M 4GB (transmitter), and Dell Latitude D630 Intel Core 2 Duo 2GB (receiver). Both of them are equipped with the Intel 5300 WiFi mini card and Ubuntu 12.04 with the Linux kernel version 3.2 installed. To collect the CSI data, we install the CSI tool released by Halperin et al [14]. The laptops are operated in monitor mode where it only monitors packets that have been sent by the transmitter. The packet transmission rate is set to 1000 packets/second. The transmission power is set to 15 dBm by default.

In the experiment, we use one antenna in the transmitter, and three antennas in the receiver. The transceivers are working in the frequency band of 5 GHz of 20 MHz channel bandwidth instead of 2.4 GHz because of the stability of the phase difference discussed in section 3.3 and also better movement speed resolution. The signal processing part such as PCA and CWT, and classification part are all done using MATLAB 2019 with the University student licences. The training process of AlexNet model is done on a single NVIDIA Tesla V100 GPU with 16 GB of RAM in Amazon Web Services (AWS) server.

7.2 Evaluation Setup

We collected a total of 330 training samples which is about 22 mins training time as shown in Table 7.1 of the four different activities performed by two people in two different living rooms ($7 \times 4\ m^2$, $3 \times 3\ m^2$), as illustrated in Figures 7.1a, 7.1b with the blue triangle spots. It is should be noted that the different activities are not equally distributed, and hence special attention must be paid when evaluating the performance of each individual activity.

To show the generality of the system, we conducted experiments in two untrained environments namely bedroom ($4 \times 4\ m^2$) and corridor ($2 \times 6\ m^2$) that is illustrated in Figures 7.1a with the green triangle spots. Both apartments are fully furnished with desks, sofas, and home appliances. Also, the system has been tested with 7 volunteers who are not included in the training set, and are varying in age, heights, weights and genders. The volunteers were invited to perform walking and clapping in the living room Figures 7.1a for a period of 36 seconds (9 samples). Due to safety purpose, the fall activity was done only for 12 seconds (3 samples) and so is the sit.



(a) Apartment 1 (b) Apartment 2

Figure 7.1: Floor plans

Table 7.1: Summary of training dataset

	Walk	Clap	Fall	Sit	Total
Living Room - 4×7 (m) (Env1)	54	54	36	36	180
Living Room - 3×3 (m) (Env2)	45	45	30	30	150
Total Samples*					330
Total Training Time (mins)					22

*Each sample is a 4 second period

7.3 Performance Measures

To evaluate the performance of the proposed method, we employ three standard metrics.

Accuracy: measures the proportion of correctly classified activities which can be expressed as follows:

$$\text{Accuracy} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \tag{7.1}$$

where T_p (true positive) and T_n (true negative) represent an outcome where the model correctly predicts positive and negative classes. F_p (false positive) and F_n (false negative) represent an outcome where the model incorrectly predicts positive and negative classes. However, accuracy is not a reliable metric since we have imbalanced dataset where the accuracy is biased towards the majority classes. This is commonly referred as the *accuracy paradox*.

Since the falling and sitting from the dataset Table 7.1 are less than walking and clapping, precision, recall and F-measure score are calculated.

Precision: measures the percentage of the results that are relevant. It is defined by the following:

$$\text{Precision} = \frac{T_p}{T_p + F_p} \tag{7.2}$$

Recall: or true positive rate measures the percentage of the *total* relevant results that are correctly classified. It is defined by the following:

$$\text{Recall} = \frac{T_p}{T_p + F_n} \tag{7.3}$$

Typically, improving in precision leads to reducing in recall and vice versa [48].

F1-measure: combines both of precision and recall with the following:

$$\text{F1-measure} = 2 \times \frac{precision \times recall}{precision + recall} \tag{7.4}$$

Since the dataset is the small, K-fold cross validation is utilized to evaluate the performance and stability of the classifier. Cross validation is used to check for overfitting and ability to predict unknown data. The three main steps of performing cross validation are:

1. The data is randomly shuffled and divided into k groups. One of the k subset is used as a test set.

2. The remaining k-1 groups are used for training. The model is fitted on the training dataset, and evaluated its performance on the test dataset.

3. Repeat the above steps until all the subsets are selected.

The general procedure of 5-fold cross validation is shown in Figure 7.2.

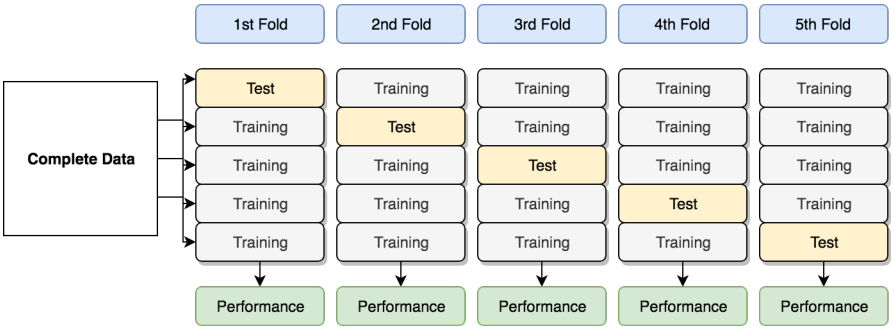


Figure 7.2: The scheme of 5-fold cross validation

7.4 Results & Discussion

7.4.1 Activity Recognition

The confusion matrix with the precision and recall of each activity is given in Table 7.2. The results show that the proposed system which is using both amplitude and phase differences achieves an average accuracy of 98.18% with the standard deviation of 1.27 using 5-fold cross validation across all activities. The AlexNet model showed best performance when classifying the walk and clap activity with the precision and recall of 100%. The fall and sit has a precision and recall of 97%, 94% and 94%, 97% respectively. There are two possible reasons that causing the model to misclassify the fall and sit activity:

- Even though falling results in a high peak in frequency for a short period of time as we have analyzed in section, it is likely that sitting down quickly is causing to have features similar to fall activity which results in a confusion.

- It is possible that after detection of the fall, the period which the person is transitioning to stop, that period is classified as a sit. Due to the static sliding window approach, similar features of different activities are generated. A solution would be to use a dynamic window approach which only extract features which the activity is happening but we will leave this for future work.

		Predicted				
		Walk	Clap	Fall	Sit	**Recall**
Actual	Walk	99	0	0	0	1
	Clap	0	99	0	0	1
	Fall	0	0	62	4	0.94
	Sit	0	0	2	64	0.97
	Precision	1	1	0.97	0.94	

Table 7.2: Confusion Matrix with AlexNet using 5-fold cross validation. The average classification accuracy is 98.18% ± 1.27

7.4.2 Impact of Window Size

Although it is intuitively clear that the larger window size, the better performance to recognize an activity since more information about the activity is captured. The results in Table 7.3 which shows the average accuracy and F-measure, and F-measure per activity under 5-fold cross validation confirm the intuition. With the window size of 4 seconds, the average accuracy is 98.18% ± 1.27, and the average F-measure is 97.73% across the four activities. However, the performance degrades when working on a window size of 3 seconds. The fall and sit F-measures have reduced by 7% while the walk and clap F-measures have slightly decreased by 1 to 2%. With the window size of 1 seconds, the performance has significantly degraded especially with the fall activity that has F-measure of 61.22%. Nevertheless, the fall and sit can be detected with the F-measure above 80% and average F-measure of 90% using 2 seconds window. By using a short window size, the system is able to save computation power while maintaining a reasonable performance.

Window	Accuracy	F-measure per activity				F-measure
(Sec)	(%) ± std	Walk	Clap	Fall	Sit	
4	**98.18 ± 1.27**	**100**	**100**	**95.38**	**95.52**	**97.73**
3	94.45 ± 2.34	99.4	98.48	88.80	88.80	93.87
2	91.49 ± 2.88	97.65	98.49	82.32	84.86	90.83
1	80.77 ± 1.77	87.64	94.5	61.22	75.79	79.79

Table 7.3: Effect of window size on average accuracy, F-measure and F-measure per activity under 5-fold cross validation

7.4.3 Impact of Transmission Rate

A range of transmission rates have been evaluated to analyze the performance of the system. Different rates starting from 1000 Hz to 50 Hz were achieved by down-sampling the transmission rate of 1000 Hz which is used through the experiments. The results in Table 7.4 show that the performance remains relatively steady from 1000 Hz to 100 Hz with the average accuracy and F-measure of 97% , and then starts deteriorating at 50 Hz with average accuracy of 39%. This is because the Nyquist frequency of 50 Hz is 25 Hz, and the activities go up to 60 Hz as discussed in section 5.4. As a result, high frequency components are not captured with the 50 Hz, and hence the model cannot reliably classify the performed activities. Nevertheless, the 1000 Hz is used through the experiments to increase the measurement resolution and reduce the noises associated with analog-to-digital converter (ADC) [25].

Fs	Accuracy	F-measure per activity				F-measure
(Hz)	(%) ± std	Walk	Clap	Fall	Sit	
1000	**98.18 ± 1.27**	**100**	**100**	**95.38**	**95.52**	**97.73**
500	97.89 ± 1.73	99.50	98.99	95.38	95.50	97.34
200	97.27 ± 1.97	98.99	98.48	95.38	94.81	96.92
100	96.66 ± 2.91	97.96	98.51	93.85	94.74	96.26
50	39.69 ± 5.18	52.25	44.10	26.10	24.82	36.81

Table 7.4: Performance comparison for different transmission rates under 5-fold cross validation

7.4.4 Impact of Amplitude & Phase Differences

The performance of having only amplitudes, phase differences only, amplitude and phase differences is evaluated under the 5-fold cross validation. The results in Table 7.5 show that the overall average accuracy and F-measure across all the activities is about 95% with or without the phase differences. Using both amplitude and phase differences have improved the average accuracy by 2% and the average F-measure by 1% higher than using only amplitude. The average F-measure and accuracy of using phase differences only is 2% less than using amplitude and phase differences. The increase is relatively small because the three antennas of the received signal is capturing most of the activity signal. Hence, the effect of distance and people on the performance is studied in sections 7.4.5 and 7.4.6.

	Accuracy	F-measure per activity				F-measure
	(%) ± std	Walk	Clap	Fall	Sit	
Amp & Ph Diff	**98.18 ± 1.27**	**100**	**100**	**95.38**	**95.52**	**97.73**
Amp	96.66 ± 1.97	99.50	99.00	93.02	92.65	96.04
Ph Diff	96.36 ± 3.14	99.50	98.98	91.60	92.54	95.65

Table 7.5: Performance with and without phase differences using 5-fold cross validation in terms of average accuracy, F-measure per activity and average F-measure

7.4.5 Impact of Distance

To evaluate the effect of distance which is the distance between the performed activity and the transceivers on the system performance, true positive rate (TPR) is used to detect the presence of the activity to the total number of times the activity is performed. The results in Figure 7.3 show that the activity range detection across the four different activity.

The walk activity in Figure 7.3a shows that the proposed system detects walking up to 5 meters using amplitude and phase differences with TPR of 80% while the TPR of using only phase differences has dropped significantly to 11% at 5 meters. The system could not detect the walk activity at 5 meters when using only amplitude. For the clap activity, the results in Figure 7.3b show that the system could fully detect clapping up to 5 meters with the TPR of 100% with or without the phase differences. The fall activity in Figure 7.3c is detected up to 3 meters and then the TPR has dropped to 83% when using amplitude and phase differences while the TPR of using only amplitude or phase differences is 33% and 66% respectively. The sit activity in Figure 7.3d is detected up to 4 meters and then the TPR has dropped to 66% using amplitude and phase differences, 50% using phase difference only and 33% using amplitude only.

While each activity has a difference in detection range, the results show that using both amplitude and phase differences have overall improved the detection range. This is because the phase difference utilizes antenna diversity and the sum of sensory sensitivity is the sum of two antenna's sensitivity [60]. It is observed that the clapping activity is detected with the high TPR even with the long distance compared to large movements such as walking and falling. This is because when the user is far away from the transceivers, the reflected signal decreases and the model tends to classify to clapping when the signal reflection is less.

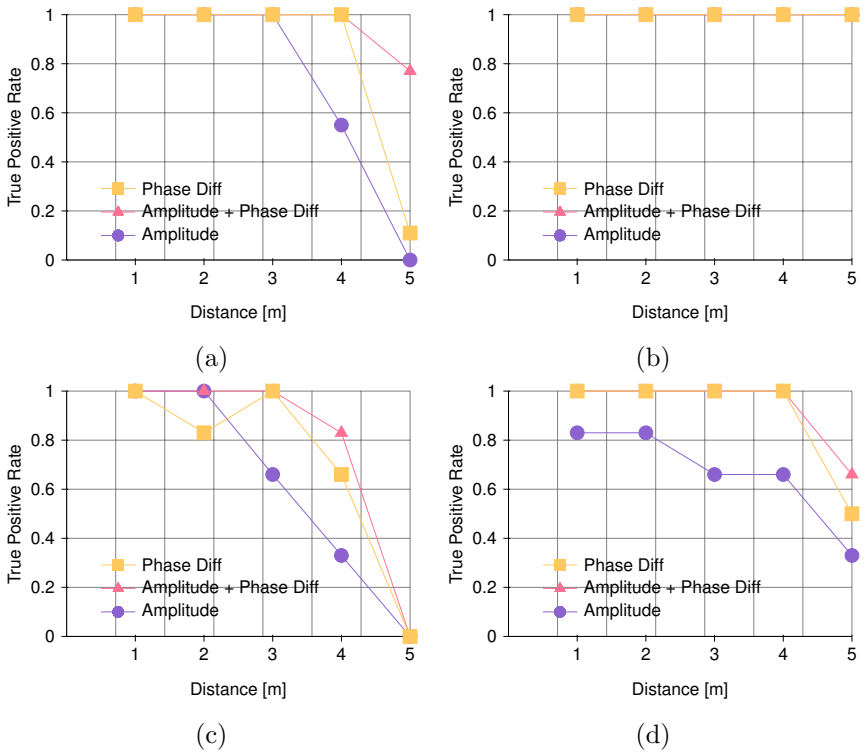


Figure 7.3: Effect of distance on true positive rate using amplitude, phase differences, amplitude and phase differences across the four different activities: a) Walk, b) Clap, c) Fall, d) Sit

7.4.6 Impact of Different People

Another goal of this study is evaluate the generalizability of the proposed method across different people. We evaluated with 7 different people who are not included in the training set, and varies in gender, age, height and weight. The results in Figure 7.4 show that the performance of the system of 7 different people across different activities using with and without phase differences. While using amplitude only yields to excellent results for most people, using amplitude and phase differences have improved the performance for some people. For instance, the clapping for subject index 7 has F-measure of 80% when using only amplitude but the F-measure has increased to 100% when using both amplitude and phase differences. Also, the sit activity for subject index 7 has improved by 20% when using the amplitude and phase differences. In some cases, the activity could not be detected using the amplitude. For example, the fall activity for subject 2 is not detected using amplitude but using phase differences have improved the F-measure to 80%. One possible reason is that different people have different ways of performing the activity, and using only amplitude can lead to

uncertainty between activities such as fall and sit. Because phase difference is more sensitive than amplitude [60], more reliable results can be provided.

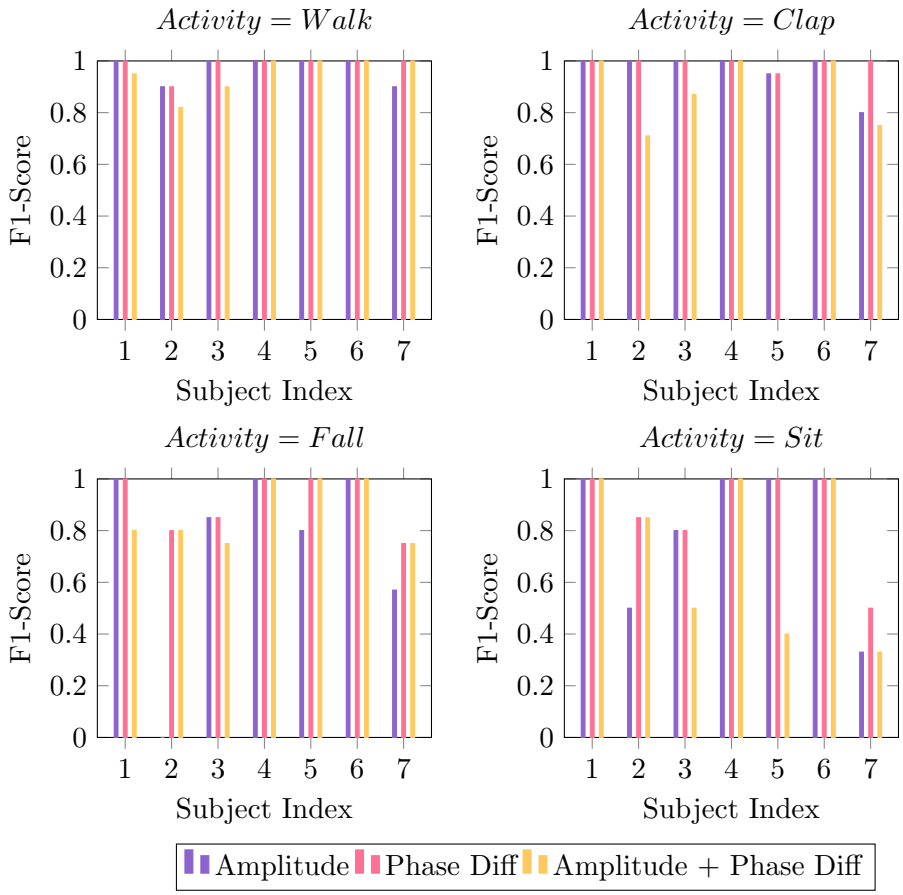


Figure 7.4: Impact of different people on the performance across the different activities using amplitude, phase difference, and amplitude & phase difference.

7.4.7 Impact of Untrained Environments

We have evaluated the system in two untrained environments - bedroom of size 16 m^2 and corridor of size 12 m^2 as shown in Figure 7.1a, and compared the performance when the phase difference is used. The results in Table 7.6 and Table 7.7 show the accuracy, F-measure per activity and average F-measure of both environments bedroom and corridor respectively. The accuracy has improved when using the phase difference in both environments. It is observed that using amplitude and phase difference in both environments have improved the accuracy and average F-measure of the different activities. For instance, the F-measure of fall in bedroom has increased by

about 20% and the F-measure of sit has increased by 7% when using amplitude and phase difference compared to using only amplitude. Using only phase difference has slightly improved the accuracy 3% compared to using amplitude but the combination of both amplitude and phase difference have improved the accuracy by 6% compared to amplitude only.

Since the corridor is narrower than the bedroom, the effect of multipath is much stronger, and hence it has a strong impact in feature extraction. As a result, the accuracy of using amplitude in corridor is much less than the accuracy in bedroom. However, using amplitude and phase difference has significantly improved the performance to an accuracy of 93% and average F-measure of 91.42% compared to accuracy of 71% and average F-measure of 67.81% when using amplitude only.

Table 7.6: Environment1 (Bedroom): performance with and without phase differences in terms of average accuracy, F-measure per activity and average F-measure

	Accuracy	F-measure per activity				F-measure
		Walk	Clap	Fall	Sit	
Amp & Ph Diff	**96**	**100**	**100**	**92.31**	**90.91**	**95.80**
Amp	90	94.74	**100**	72.73	83.33	87.70
Ph Diff	93	**100**	**100**	85.71	80	91.43

Table 7.7: Environment2 (Corridor): performance with and without phase differences using 5-fold cross validation in terms of average accuracy, F-measure per activity and average F-measure

	Accuracy	F-measure per activity				F-measure
		Walk	Clap	Fall	Sit	
Amp & Ph Diff	**93**	**100**	**100**	**85.71**	**80**	**91.42**
Amp	71	71.43	78.26	61.54	60	67.81
Ph Diff	76	80	**100**	40	70.59	72.65

7.4.8 Computation Time

Another goal of this research is to develop an online classification of the human activity recognition. Since the window size used in the experiment is 4 seconds with the sampling rate of 1000 Hz, the sequences of the activity signal are processed using four main methods: CSI extraction and interpolation, PCA, CWT, and classification. Table 7.8 shows the average computation time of our proposed method using 2.5 GHz Intel Core i7 machine. The results show that the average CPU time to classify a four second activity signal is about 1.2 seconds which is suitable for real-time application. 79.17% of the CPU time is consumed by the CSI extraction and Interpolation method. This is because the extraction and interpolation method involves extracting the CSI data and interpolating for each subcarrier which there are 90 subcarriers in total from three received antennas. The PCA method only takes 1.58% of the CPU time. Then, 12.5% of the CPU time is spent on CWT method which involves generating 5 scalograms at different scales. Finally, classification of the scalograms takes 6.75% of the CPU which corresponds to only 0.081 seconds.

Table 7.8: Average computation time of classifying activities

Methods	Average CPU Time	
	(sec)	(%)
CSI extraction & Interpolation	0.95	79.17
PCA	0.019	1.58
CWT	0.15	12.5
Classification	0.081	6.75
Total Time	**1.2**	

Chapter 8

Conclusions and Future Work

8.1 Conclusions

In this thesis, we design and implement an online low-cost yet accurate human activity recognition system that uses commodity WiFi devices to recognize activities. The key contribution of our work is the use of amplitude and phase difference to recognize performed by different people and can be applied in untrained/complex environments.

In chapter 3, we begin by studying the correlation amplitudes of the subcarriers, and exploit the feasibility of the phase and phase difference between antennas in activity recognition. We show that the subsequent subcarriers are more correlated than the further subcarriers which motivated applying PCA method introduced in chapter 4. Furthermore, we discuss how the phase difference between subsequent antennas is more stable than the phase information due to the shared clock between antennas.

In chapter 4, we discuss the preprocessing techniques used to remove the internal hardware noises, and improve the extracted features of amplitude and phase difference. We apply a 1D linear interpolation to have an even sampled CSI data which is important in performing time-frequency analysis and obtain the scalograms. The CSI signal is segmented into fixed windows with the certain overlap to enable an online classification of the activity. We also apply PCA and a moving average to reduce the dimensions and noises.

In chapter 5, we discuss why time-frequency analysis is applied and in particular CWT, and then describe the process of selecting the scale and the wavelet family. We then discuss how to apply the CWT to the principle components of amplitude and phase difference.

In Chapter 6, we discuss why CNN is chosen, and how AlexNet CNN architecture is used to classify the scalograms obtained from the previous chapter.

We analyze the classification results of the proposed system in chapter 7. The results show that the system achieves an average accuracy of 98.18% with standard deviation of 1.27 using 5-fold cross validation across the four activities (walk, clap, fall and sit). The system shows its capability to obtain high performance under transmissions rates as low as 100 Hz, and under window size as low as 2 seconds. In long distance, the system is able to detect walking with the TPR of 80% at the distance of 5 meters when using amplitude and phase difference compared to TPR of 55% at the distance of 4 meters when using only amplitude. In untrained environments, using amplitude and phase difference has significantly improved the performance to an accuracy of 93%, 96% and average F-measure of 91.42%, 95.80% compared to accuracy of 71%, 90% and average F-measure of 67.81%, 87.70% respectively when using amplitude only. The system shows its generalizability across 7 different people with the F-measure above 80%. It also shows its capability to be run in real-time application with an average computation time of 1.2 seconds.

8.2 Future Work

In order to reach the full potential of WiFi activity recognition, several challenges and topics need to be investigated before applying to the real-world.

- Even though the proposed system is able to recognize activities performed by different people, only a single user can perform the activity in one environment. In real-world, there may be more than one user in the same environment which each one of them is performing different activities. The main challenge is how to determine the number of people and recognize their activities separately?

- Sometimes high recognition accuracy is limited to the size and complexity of the environment. The proposed system achieves high accuracy up to 4 meters and then the performance degrades. One of the possible ways to increase accuracy is by using multiple APs and combining the information to recognize an activity. However, which AP needs to be selected and how can one fusion multiple AP needs requires further research.

- The purposed system works when one person is performing one activity at a time. However, sometimes people perform more than one activity at the same time. For instance, talking on the phone while walking or eating while watching TV. Distinguishing the signals due to activities of different parts of the body is another challenge.

Bibliography

[1] H. Abdelnasser, M. Youssef, and K. A. Harras. Wigest: A ubiquitous wifi-based gesture recognition system. In *2015 IEEE Conference on Computer Communications (INFOCOM)*, pages 1472–1480, April 2015.

[2] S. Arshad, C. Feng, Y. Liu, Y. Hu, R. Yu, S. Zhou, and H. Li. Wi-chase: A wifi based human activity recognition system for sensorless environments. In *2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, pages 1–6, June 2017.

[3] D. Banerjee. *Pll Performance, Simulation and Design*. Dog Ear Publishing, 2006.

[4] Paul Bourke. Interpolation methods. `http://paulbourke.net/miscellaneous/interpolation/`, 12 1999. Accessed: 2019-07-30.

[5] Christina Brogrdh, Ulla-Britt Flansbjer, Christina Espelund, and Jan Lexell. The 6-minute walk test indoors is strongly related to walking ability outdoors in persons with late effects of polio. *European Journal of Physiotherapy*, 15(4):181–184, 2013.

[6] Andreas Bulling, Ulf Blanke, and Bernt Schiele. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)*, 46(3):33, 2014.

[7] Chandralika Chakraborty and PH Talukdar. Issues and limitations of hmm in speech processing: a survey. *International Journal of Computer Applications*, 141(7):975–8887, 2016.

[8] Jen-Yin Chang, Kuan-Ying Lee, Yu-Lin Wei, Kate Ching-Ju Lin, and Winston Hsu. We Can "See" You via Wi-Fi - WiFi Action Recognition via Vision-based Methods. *arXiv e-prints*, page arXiv:1608.05461, Aug. 2016.

[9] 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*, pages 1–1, 2018.

[10] Tahmid Z. Chowdhury. *Using Wi-Fi channel state information (CSI) for human activity recognition and fall detection*. PhD thesis, University of British Columbia, 2018.

[11] T. Z. Chowdhury, C. Leung, and C. Y. Miao. Wihacs: Leveraging wifi for human activity classification using ofdm subcarriers' correlation. In *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 338–342, Nov 2017.

[12] G Arun Francis, P Karthigaikumar, and S Dhivya. Modulation and demodulation process of orthogonal frequency division multiplexing. *Middle-East Journal of Scientific Research*, 23(2):362–366, 2015.

[13] H. P. Gupta, H. S. Chudgar, S. Mukherjee, T. Dutta, and K. Sharma. A continuous hand gestures recognition technique for human-machine interaction using accelerometer and gyroscope sensors. *IEEE Sensors Journal*, 16(16):6425–6432, Aug 2016.

[14] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. Tool release: Gathering 802.11n traces with channel state information. *SIGCOMM Comput. Commun. Rev.*, 41(1):53–53, Jan. 2011.

[15] Jiawei Han, Jian Pei, and Micheline Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.

[16] Z. He and L. Jin. Activity recognition from acceleration data based on discrete consine transform and svm. In *2009 IEEE International Conference on Systems, Man and Cybernetics*, pages 5041–5044, Oct 2009.

[17] Samer Hijazi, Rishi Kumar, and Chris Rowen. Using convolutional neural networks for image recognition. *Cadence Design Systems Inc.: San Jose, CA, USA*, 2015.

[18] Y. Jia. Diatetic and exercise therapy against diabetes mellitus. In *2009 Second International Conference on Intelligent Networks and Intelligent Systems*, pages 693–696, Nov 2009.

[19] V. P. G. Jimenez, M. J. F. . Garcia, F. J. G. Serrano, and A. G. Armada. Design and implementation of synchronization and agc for ofdm-based wlan receivers. *IEEE Transactions on Consumer Electronics*, 50(4):1016–1025, Nov 2004.

[20] Ian Jolliffe. *Principal component analysis*. Springer, 2011.

[21] Nopphon Keerativoranan, Azril Haniz, Kentaro Saito, and Jun ichi Takada. Mitigation of CSI temporal phase rotation with b2b calibration method for fine-grained motion detection analysis on commodity wi-fi devices. *Sensors*, 18(11):3795, nov 2018.

[22] A. M. Khan, Y. Lee, S. Y. Lee, and T. Kim. A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer. *IEEE Transactions on Information Technology in Biomedicine*, 14(5):1166–1172, Sep. 2010.

[23] A. M. Khan, Y. K. Lee, and S. Y. Lee. Accelerometer's position free human activity recognition using a hierarchical recognition model. In *The 12th IEEE International Conference on e-Health Networking, Applications and Services*, pages 296–301, July 2010.

[24] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[25] Silcon Labs. Improving adc resolution by oversampling and averaging. `https://www.cypress.com/file/236481/download`, 7 2013. Accessed: 2019-07-30.

[26] O. D. Lara and M. A. Labrador. A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys Tutorials*, 15(3):1192–1209, Third 2013.

[27] Rajalakshmi Nandakumar, Bryce Kellogg, and Shyamnath Gollakota. Wi-Fi Gesture Recognition on Existing Devices. *arXiv e-prints*, page arXiv:1411.5394, Nov. 2014.

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

[29] Sameera Palipana, David Rojas, Piyush Agrawal, and Dirk Pesch. Falldefi: Ubiquitous fall detection using commodity wi-fi devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(4):155:1–155:25, Jan. 2018.

[30] A. J. PAULRAJ, D. A. GORE, R. U. NABAR, and H. BOLCSKEI. An overview of mimo communications - a key to gigabit wireless. *Proceedings of the IEEE*, 92(2):198–218, Feb 2004.

[31] E. Perahia and R. Stacey. *Next Generation Wireless LANs: 802.11n and 802.11ac*. Cambridge University Press, 2013.

[32] Ronald Poppe. A survey on vision-based human action recognition. *Image and Vision Computing*, 28(6):976 – 990, 2010.

[33] Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel. Whole-home gesture recognition using wireless signals. In *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking*, MobiCom '13, pages 27–38, New York, NY, USA, 2013. ACM.

[34] Matthias Schulz. *Teaching Your Wireless Card New Tricks: Smartphone Performance and Security Enhancements Through Wi-Fi Firmware Modifications*. PhD thesis, Technische Universität, Darmstadt, 2018.

[35] Matthias Schulz, Jakob Link, Francesco Gringoli, and Matthias Hollick. Shadow wi-fi: Teaching smartphones to transmit raw signals and to extract channel state information to implement practical covert channels over wi-fi. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '18, pages 256–268, New York, NY, USA, 2018. ACM.

[36] Matthias Schulz, Daniel Wegemer, and Matthias Hollick. Nexmon: The c-based firmware patching framework. https://nexmon.org, 2017.

[37] Martin Schüssel. Angle of arrival estimation using wifi and smartphones. In *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, volume 4, page 7, 2016.

[38] Ervin Sejdi, Igor Djurovi, and Jin Jiang. Timefrequency feature representation using energy concentration: An overview of recent advances. *Digital Signal Processing*, 19(1):153 – 183, 2009.

[39] Souvik Sen, Božidar Radunovic, Romit Roy Choudhury, and Tom Minka. You are facing the mona lisa: Spot localization using phy layer information. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*, MobiSys '12, pages 183–196, New York, NY, USA, 2012. ACM.

[40] Muhammad Shoaib, Stephan Bosch, Ozlem Incel, Hans Scholten, and Paul Havinga. A survey of online activity recognition using mobile phones. *Sensors*, 15(1):2059–2085, jan 2015.

[41] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In *CVPR 2011*, pages 1297–1304, June 2011.

[42] S. Sigg, U. Blanke, and G. Trster. The telepathic phone: Frictionless activity recognition from wifi-rssi. In *2014 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 148–155, March 2014.

[43] Rubén Solera-Ureña, Jaume Padrell-Sendra, Darío Martín-Iglesias, Ascensión Gallardo-Antolín, Carmen Peláez-Moreno, and Fernando Díaz-de María. Svms for automatic speech recognition: a survey. In *Progress in nonlinear speech processing*, pages 190–216. Springer, 2007.

[44] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[45] B. Y. Su, K. C. Ho, M. J. Rantz, and M. Skubic. Doppler radar fall activity detection using the wavelet transform. *IEEE Transactions on Biomedical Engineering*, 62(3):865–875, March 2015.

[46] Jit Ken Tan. *An adaptive orthogonal frequency division multiplexing baseband modem for wideband wireless channels*. PhD thesis, Masters thesis, MIT, 2006.

[47] Zengshan Tian, Ze Li, Mu Zhou, Yue Jin, and Zipeng Wu. Pila: Sub-meter localization using csi from commodity wi-fi devices. *Sensors*, 16(10), 2016.

[48] Kai Ming Ting. *Precision and Recall*, pages 781–781. Springer US, Boston, MA, 2010.

[49] Christopher Torrence and Gilbert P Compo. A practical guide to wavelet analysis. *Bulletin of the American Meteorological society*, 79(1):61–78, 1998.

[50] Asaf Tzur, Ofer Amrani, and Avishai Wool. Direction finding of rogue wi-fi access points using an off-the-shelf mimoofdm receiver. *Physical Communication*, 17:149 – 164, 2015.

[51] J.-W. van Dijk, K. Tummers, C. D. A. Stehouwer, F. Hartgens, and L. J. C. van Loon. Exercise therapy in type 2 diabetes: Is daily exercise required to optimize glycemic control? *Diabetes Care*, 35(5):948–954, mar 2012.

[52] Deepak Vasisht, Swarun Kumar, and Dina Katabi. Decimeter-level localization with a single wifi access point. In *Proceedings of the 13th Usenix Conference on Networked Systems Design and Implementation*, NSDI'16, pages 165–178, Berkeley, CA, USA, 2016. USENIX Association.

[53] Michalis Vrigkas, Christophoros Nikou, and Ioannis A. Kakadiaris. A review of human activity recognition methods. *Frontiers in Robotics and AI*, 2, nov 2015.

[54] H. Wang, D. Zhang, Y. Wang, J. Ma, Y. Wang, and S. Li. Rt-fall: A real-time and contactless fall detection system with commodity wifi devices. *IEEE Transactions on Mobile Computing*, 16(2):511–526, Feb 2017.

[55] Wei Wang, Alex X. Liu, and Muhammad Shahzad. Gait recognition using wifi signals. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '16, pages 363–373, New York, NY, USA, 2016. ACM.

[56] Wei Wang, Alex X. Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. Understanding and modeling of wifi signal based human activity recognition. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, MobiCom '15, pages 65–76, New York, NY, USA, 2015. ACM.

[57] 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)*, pages 1230–1239, June 2017.

[58] Yi Wang, Xinli Jiang, Rongyu Cao, and Xiyang Wang. Robust indoor human activity recognition using wireless signals. *Sensors*, 15(7):17195–17208, jul 2015.

[59] Y. Wang, K. Wu, and L. M. Ni. Wifall: Device-free fall detection by wireless networks. *IEEE Transactions on Mobile Computing*, 16(2):581–594, Feb 2017.

[60] C. Wu, Z. Yang, Z. Zhou, K. Qian, Y. Liu, and M. Liu. Phaseu: Real-time los identification with wifi. In *2015 IEEE Conference on Computer Communications (INFOCOM)*, pages 2038–2046, April 2015.

[61] Yaxiong Xie, Zhenjiang Li, and Mo Li. Precise power delay profiling with commodity wifi. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, MobiCom '15, pages 53–64, New York, NY, USA, 2015. ACM.

[62] Tong Xin, Bin Guo, Zhu Wang, Pei Wang, and Zhiwen Yu. Freesense: human-behavior understanding using wi-fi signals. *Journal of Ambient Intelligence and Humanized Computing*, 9(5):1611–1622, Oct 2018.

[63] Jian Bo Yang, Minh Nhut Nguyen, Phyo Phyo San, Xiao Li Li, and Shonali Krishnaswamy. Deep convolutional neural networks on multichannel time series for human activity recognition. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, pages 3995–4001. AAAI Press, 2015.

[64] Q. Yang, Z.H. Zhou, Z. Gong, M.L. Zhang, and S.J. Huang. *Advances in Knowledge Discovery and Data Mining: 23rd Pacific-Asia Conference, PAKDD 2019, Macau, China, April 14-17, 2019, Proceedings. Part II.* Number pt. 2 in LNCS sublibrary: Artificial intelligence. Springer International Publishing, 2019.

[65] Zheng Yang, Zimu Zhou, and Yunhao Liu. From rssi to csi: Indoor localization via channel response. *ACM Comput. Surv.*, 46(2):25:1–25:32, Dec. 2013.

[66] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee. A survey on behavior recognition using wifi channel state information. *IEEE Communications Magazine*, 55(10):98–104, Oct 2017.

[67] Yunze Zeng, Parth H. Pathak, and Prasant Mohapatra. Analyzing shopper's behavior through wifi signals. In *Proceedings of the 2Nd Workshop on Workshop on Physical Analytics*, WPA '15, pages 13–18, New York, NY, USA, 2015. ACM.

[68] Daqing Zhang, Hao Wang, Yasha Wang, and Junyi Ma. Anti-fall: A non-intrusive and real-time fall detector leveraging csi from commodity wifi devices. In Antoine Geissbühler, Jacques Demongeot, Mounir Mokhtari, Bessam Abdulrazak, and Hamdi Aloulou, editors, *Inclusive Smart Cities and e-Health*, pages 181–193, Cham, 2015. Springer International Publishing.

[69] Shugang Zhang, Zhiqiang Wei, Jie Nie, Lei Huang, Shuang Wang, and Zhen Li. A review on human activity recognition using vision-based method. *Journal of Healthcare Engineering*, 2017:1–31, 2017.

[70] X. Zheng, J. Wang, L. Shangguan, Z. Zhou, and Y. Liu. Smokey: Ubiquitous smoking detection with commercial wifi infrastructures. In *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, pages 1–9, April 2016.

[71] Z. Zhou, Z. Yang, C. Wu, W. Sun, and Y. Liu. Lifi: Line-of-sight identification with wifi. In *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, pages 2688–2696, April 2014.

[72] Y. Zhuo, H. Zhu, and H. Xue. Identifying a new non-linear csi phase measurement error with commodity wifi devices. In *2016 IEEE 22nd International Conference on Parallel and Distributed Systems (ICPADS)*, pages 72–79, Dec 2016.

[73] Y. Zhuo, H. Zhu, H. Xue, and S. Chang. Perceiving accurate csi phases with commodity wifi devices. In *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, pages 1–9, May 2017.

Appendix A

Confusion matrices

	Predicted			
	Walk	Clap	Fall	Sit
Walk	99	0	0	0
Clap	0	99	0	0
Fall	0	0	62	4
Sit	0	0	2	64

Table A.1: Confusion Matrix of 4 seconds window size - Dataset: 7.1

	Predicted			
	Walk	Clap	Fall	Sit
Walk	165	0	0	0
Clap	0	162	1	2
Fall	2	1	119	10
Sit	0	1	16	115

Table A.2: Confusion Matrix of 3 seconds window size - Dataset: 7.1

	Predicted			
	Walk	Clap	Fall	Sit
Walk	229	0	2	0
Clap	0	229	0	2
Fall	9	1	156	32
Sit	0	4	23	171

Table A.3: Confusion Matrix of 2 seconds window size - Dataset: 7.1

	Predicted			
	Walk	Clap	Fall	Sit
Walk	443	8	33	11
Clap	1	481	6	7
Fall	62	21	251	128
Sit	10	13	68	371

Table A.4: Confusion Matrix of 1 seconds window size - Dataset: 7.1

	Predicted			
	Walk	Clap	Fall	Sit
Walk	99	0	0	0
Clap	0	99	0	0
Fall	0	0	62	4
Sit	0	0	2	64

Table A.5: Confusion Matrix of 1000 Hz sampling rate - Dataset: 7.1

	Predicted			
	Walk	Clap	Fall	Sit
Walk	98	1	0	0
Clap	0	98	0	1
Fall	0	0	62	4
Sit	0	0	1	65

Table A.6: Confusion Matrix of 500 Hz sampling rate - Dataset: 7.1

	Predicted			
	Walk	Clap	Fall	Sit
Walk	98	1	0	0
Clap	1	97	0	1
Fall	0	0	62	4
Sit	0	0	2	64

Table A.7: Confusion Matrix of 200 Hz sampling rate - Dataset: 7.1

	Predicted			
	Walk	Clap	Fall	Sit
Walk	96	3	0	0
Clap	0	99	0	0
Fall	1	0	61	4
Sit	0	0	3	63

Table A.8: Confusion Matrix of 100 Hz sampling rate - Dataset: 7.1

	Predicted			
	Walk	Clap	Fall	Sit
Walk	58	17	6	18
Clap	20	41	16	22
Fall	26	11	15	14
Sit	19	18	12	17

Table A.9: Confusion Matrix of 50 Hz sampling rate - Dataset: 7.1

	Predicted			
	Walk	Clap	Fall	Sit
Walk	5	4	0	0
Clap	0	9	0	0
Fall	0	1	4	1
Sit	0	0	3	3

Table A.10: Confusion Matrix of untrained environment (Corridor) - Amplitude only

	Predicted			
	Walk	Clap	Fall	Sit
Walk	6	0	2	1
Clap	0	9	0	0
Fall	0	0	2	4
Sit	0	0	0	6

Table A.11: Confusion Matrix of untrained environment (Corridor) - Phase difference only

	Predicted			
	Walk	Clap	Fall	Sit
Walk	9	0	0	0
Clap	0	9	0	0
Fall	0	0	6	0
Sit	0	0	2	4

Table A.12: Confusion Matrix of untrained environment (Corridor) - Amplitude & Phase difference

	Predicted			
	Walk	Clap	Fall	Sit
Walk	9	0	0	0
Clap	0	9	0	0
Fall	1	0	4	1
Sit	0	0	1	5

Table A.13: Confusion Matrix of untrained environment (Bedroom) - Amplitude only

	Predicted			
	Walk	Clap	Fall	Sit
Walk	9	0	0	0
Clap	0	9	0	0
Fall	0	0	6	0
Sit	0	0	2	4

Table A.14: Confusion Matrix of untrained environment (Bedroom) - Phase difference only

	Predicted			
	Walk	Clap	Fall	Sit
Walk	9	0	0	0
Clap	0	9	0	0
Fall	0	0	6	0
Sit	0	0	1	5

Table A.15: Confusion Matrix of untrained environment (Bedroom) - Amplitude & Phase difference