University of Southern Queensland

Faculty of Health, Engineering & Sciences

Cardiac Biometric Authentication Using Channel State Information

A dissertation submitted by

Matthew Armanasco

in fulfilment of the requirements of

ENG4112 Research Project

towards the degree of

Bachelor of Engineering (Honours) (Electrical & Electronic Engineering) / Bachelor of Science (Physical Sciences)

Submitted: October, 2022

Abstract

Owing to the pervasion of high-speed Wi-Fi signals in modern society, there has been a recent interest in utilising these signals for more than simply transferring information. By utilising a parameter known as Channel State Information (CSI), information about the physical environment in which a signal propagates can be obtained. This may be used to infer detail about the humans present in that environment. CSI measured between a standard, commodity hardware transmitter-receiver Wi-Fi communications link can provide highly accurate measurement of human characteristics. This project utilised CSI for human biometric authentication, measuring heart rhythm characteristics to identify a user.

Currently available human biometric authentication techniques include fingerprint scanning, retinal scanning, and facial recognition. The proposed system could be autonomous (requiring no user input or action), could be almost impossible to deceive, and requires no specialised hardware, only readily available Wi-Fi devices. Similar research has investigated the use of CSI to authenticate users based on respiration rhythm and gait, but these characteristics are easily replicable and they require action from the user. Research has shown that the human heart rhythm is highly unique to an individual, and thus this project fulfilled a research gap by investigating heart rhythm authentication. It was found that the system can accurately measure a human heart rhythm rapidly (in tests of 10 seconds or less), using a standard wireless home router Access Point (AP) and desktop PC. Users were required only to sit in between the signal path of the router and the PC for the duration of the test. Furthermore, it was found that authentication using a correlation algorithm is possible using these measurements, presenting a correct authentication rate of approximately 67%. It is believed that this accuracy can be improved with directional antennas, or by operation in ad-hoc mode (i.e. device-device, not device-AP).

University of Southern Queensland

Faculty of Health, Engineering & Sciences

ENG4111/2 *Research Project*

Limitations of Use

Certification of Dissertation

I certify that the ideas, designs and experimental work, results, analyses and conclusions set out in this dissertation are entirely my own effort, except where otherwise indicated and acknowledged.

I further certify that the work is original and has not been previously submitted for assessment in any other course or institution, except where specifically stated.

MATTHEW ARMANASCO

[redacted]

Acknowledgments

I would like to acknowledge and thank Dr John Leis for his excellent input and guidance in the development of this project. I would like to thank my family for their support, and I would like to thank my partner Letitia for abiding endless Wi-Fi outages, acting as test subject, and for her invaluable support.

MATTHEW ARMANASCO

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background Information

Channel State Information (CSI) is the information associated with a channel of a wireless communication link, describing how it propagates from transmitter to receiver. This technology has gained more attention in recent years due to advancements in processing speed and networking power. It has proved to be more effective than similar technologies, such as Software Defined Radio (SDR) and Received Signal Strength Indicators (RSSI), in applications regarding humans, such as human motion detection, activity recognition, and measurement (Al-qaness et al. 2019, p.3329). This is due to its stability, rich information capacity and device-free installation process. That is, many human observation technologies require setting up a cumbersome measurement device. CSI simply utilises the information already present within Wi-Fi signals which pervade almost every modern environment, without requiring the installation of additional sensors, such as cameras, acoustic sensors, accelerometers, wearable sensors, environment installed sensors, and smartphones (Al-qaness et al. 2019, p. 3329). Further, these types of sensors all require either a heavy, cumbersome installation, a device on the person, or direct, lit Line Of Sight (LOS), as is the case for cameras (Al-qaness et al. 2019, p. 3329). RSSI based censoring methods demonstrate many of the same benefits of CSI, however these methods are susceptible to interference from environmental factors and they are not as information-rich, meaning that they fail to accurately measure objects smaller than one metre, whereas CSI is highly accurate to a small scale of less than one metre (Yang, Zhou & Liu 2013, pp. 1-32). Halperin et al. (2010, pp. 159-170) presented a tool for extracting CSI from commercially available Intel 5300 series Network Interface Cards (NICs) on the 802.11n IEEE Wi-Fi protocol. This is an open-source tool built using proprietary firmware at Intel. This tool was a significant milestone in Wi-Fi based physical measurement research, as it allowed any user to extract and utilise information within the physical layer of the 5300 802.11n NIC and report it to higher layers. To understand the functionality of CSI and its subsequent use in human-based applications, an understanding of the 802.11n wireless Local Area Network (wLAN) system is required.

In 2009, the IEEE ratified the 802.11n wLAN standard, which was an advancement upon the 802.11a/b standard that introduced the capacity for use of multiple antennas at the physical layer, known as Multiple Input, Multiple Output (MIMO) (IEEE Standards Association 2009). This was based upon research done by Foschini, Gans, and Telatar in the mid 1990s which showed that by simply adding more antennas to a transmit/receive Wi-Fi system has the potential to linearly scale the throughput even whilst transmitting on the same Wi-Fi band at the same time (Halperin et al. 2010, pp. 159-170). This is a technique now used in all high-rate wireless systems today for higher signal to noise ratio (SNR), which is a measure if the received signal strength of a transmission relative to noise. As a benchmark, 802.11a normally transmits signals with 50mW of power, but the typical received signal strength is only 50pW, which is a 90dB (billion fold) loss, but is still higher than the signal floor for a 20MHz 802.11 channel of 0.1pW; the SNR is thus

$$SNR \qquad 10 \times \log_{10}\frac{50}{0.1} \approx 27dB \tag{1.1}$$

(Halperin et al. 2010, pp. 159-170).

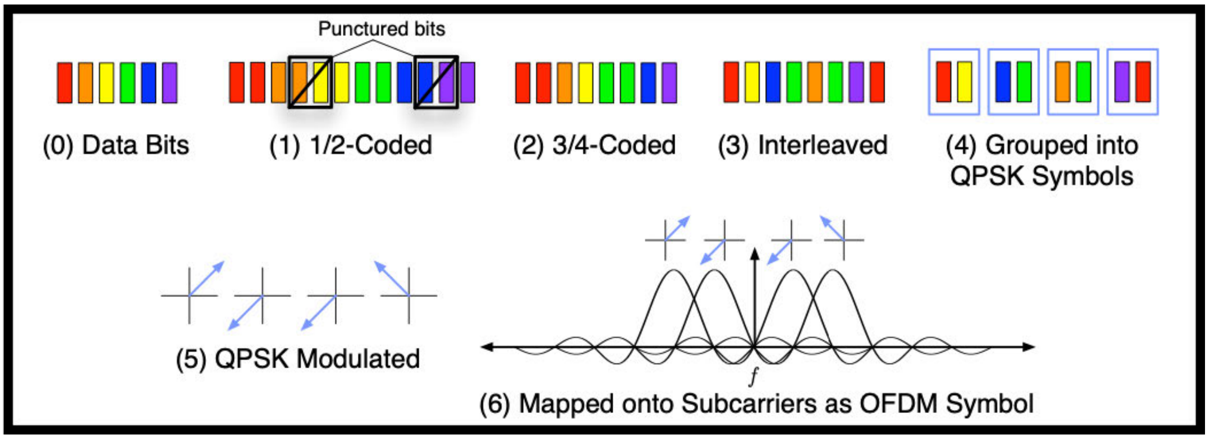


Figure 1.1: QPSK Modulation on OFDM (Halperin et al. 2010, pp. 159-170)

Naturally this 90dB loss indicates significant attenuation occurs. Signal attenuation for Wi-Fi is caused by two primary effects: path loss, and fading effects. Path loss is the way that a radiated signal spreads out over a wide area, causing the power to spread out also; path loss causes power to attenuate at least as quickly as the square of the distance travelled (Halperin et al. 2010, pp. 159-170). Fading effects include when obstacles such as water, metal, and glass reflect radio waves, otherwise known as shadowing (Halperin et al. 2010, pp. 159-170). A subsequent fading effect which is especially significant is multi-path fading, wherein the reflected copies of the signal are scattered and arrive at the receiver at different times after travelling along different paths (Halperin et al. 2010, pp. 159-170). These copies combine at the receiver, and they may add constructively or destructively depending on the relative phases; studies of this multi-path fading have shown variations can be as large as 15-20dB (Judd, Wang & Steenkiste 2008, pp. 118-131). Further, tiny fluctuations in path length can significantly affect signal strength as at Wi-Fi frequencies of 2.4GHz or 5GHz, full wavelengths (over which the phase completely cycles) are only 12cm and 6cm respectively. This means that multi-path effects can cause rapid signal changes, otherwise known as fast fading, as the environment changes around a stationary node. Multi-path effects depend on the phases of signals, and as such they are referred to as being highly frequency selective, as in some frequencies within a 20MHz communications channel can be wiped out entirely (known as deep fade) whereas some can be unaffected. The overall impact of multi-path effects is thus that received wireless signals exhibit great variability over time, frequency and space, which detracts from performance due to the likelihood of a single frequency experiencing deep fade bringing the entire transmission beneath the SNR required for a given communication scheme (Halperin et al. 2010, pp. 159-170). This variability does also engender possibilities of observing changes in these effects however.

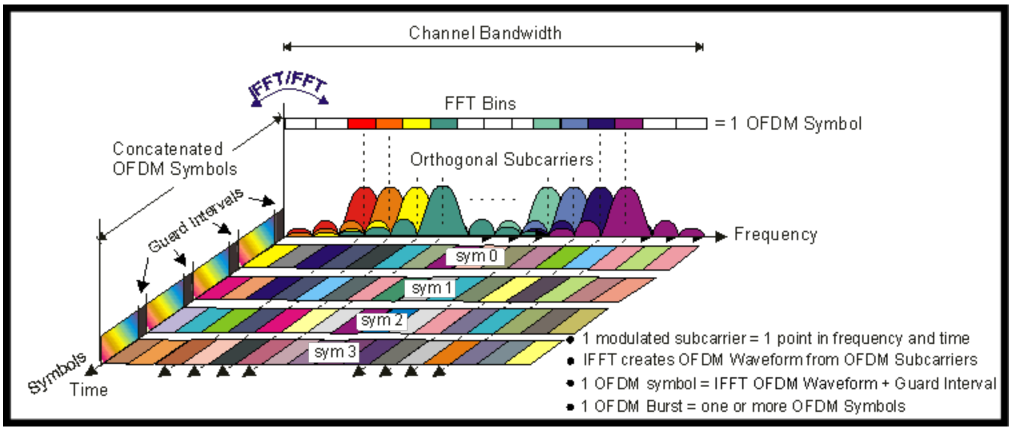


Figure 1.2: Orthogonal Frequency Division Multiplexing Frequency and Time Domain Diagram (Keysight Technologies, Inc. n.d.)

The main technique used to overcome these problems of variability in wireless communications is diversity. This is the process wherein information at the transmitter is split over multiple independently faded channels with redundancy, which reduces the likelihood that a single deep fade on a signal path will inhibit wireless communication entirely (Halperin et al. 2010, pp. 159-170). The challenge then is to find where these independently faded channels exist within the physical layer, by manipulating the time, frequency and spatial resources of the link. This is where Orthogonal Frequency Division Multiplexing (OFDM) is an essential technique. OFDM is the process within 802.11 wLAN of splitting a channel bandwidth into many subcarriers. Instead of transmitting a single subcarrier at a high-rate stream, many closely space orthogonal subcarriers, or symbols, are transmitted in parallel. All of these subcarriers use the same digital modulation scheme, such as Quadrature Phase Shift Keying (QPSK) at a low symbol rate. Each subcarrier channel is separated by a frequency guard band which minimises channel delay spread and intersymbol interference. An FFT is performed at the receiver to reconstruct the original data bits (Keysight Technologies, Inc. n.d.). Figure 1.1 shows OFDM diagrammatically. For example, an 802.11n physical layer based on OFDM might partition the 20MHz 802.11 channel into 64 subcarriers of 312.5kHz each, forming many individual narrowband channels, each using the same modulation scheme (QPSK), transmit power and coding. This improves computational and spectral efficiency and high aggregate data rates can be achieved, and hardware components can be shared across subcarriers. More importantly however, subdividing the channel using OFDM has created the independently faded channels that are required for diversity coding, as multi-path fading is frequency selective and thus the subcarriers will each experience different fades. The MIMO technique can therefore use OFDM to encode QPSK symbols onto the subcarriers of a signal (Halperin et al. 2010, pp. 159-170). Figure 1.2 shows subcarriers in the spectral domain using QPSK, with some interleaved redundancy using frequency diversity that can be used to correct for errors arising from fading reducing SNR on certain subcarriers. In 802.11n, 4 subcarriers at each edge of the channel are used as guard bands.

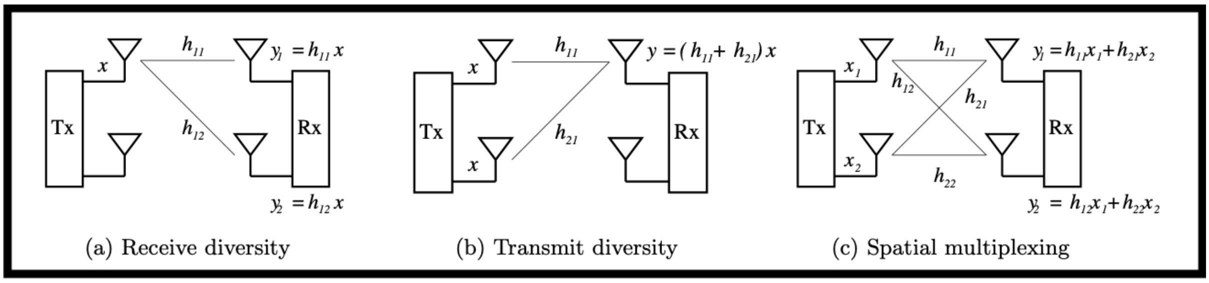


Figure 1.3: 2x2 Antenna Communication System Using Spatial Diversity (Halperin et al. 2010, pp. 159-170)

Spatial diversity is also used in the MIMO technique, which involves adding more antennas to an 802.11n transmitter or receiver to provide a new set of independently faded paths, even if the antennas are only centimetres apart. Doing so introduces spatial diversity into the system, which can be exploited to improve resilience to fading, and merely adding an extra antenna should double the signal power in an ideal scenario. This is known as spatial multiplexing, and a 2x2 system (2 transmitters and 2 receivers) is shown in Figure 1.3. Here, x_i and y_i represent transmitted and received signals respectively, and the channel gain h_{ij} is a complex number indicating a signal's attenuated amplitude and phase shift over the channel corresponding to the ith transmit antenna and the jth receive antenna. Note also that the received signals will have thermal RF noise coupled into them (Halperin et al. 2010, pp. 159-170). The result of OFDM coding for 802.11 wLAN is significantly improved performance consistency, mitigating impact of the significant variability that wireless signals suffer due to multi path fading.

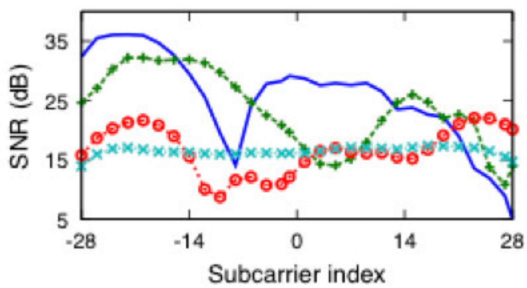


Figure 1.4: Example SISO CSI information for 52 Subcarriers for 4 Different Packets (Halperin et al. 2010, pp. 159-170)

A form of CSI therefore is describing the SNR of each subcarrier for each antenna, such as shown in Figure 1.4. In this way, RSSI is regarded as coarse grained channel information, whilst CSI is fine-grained channel information. RSSI measures the average SNR of the many received multi-path signals, whilst CSI describes the signal with detail demonstrating the quality of each OFDM subcarrier (Al-qaness et al. 2019, p. 3329).

This is the basis for the CSI authentication system. By leveraging the detailed information gathered from the many different paths of the OFDM subcarriers, information about the environment of the signals can be deduced. If a human forms part of this environment, then subsequently, information about the human's physicality can be deduced. Physicality detail about the human body that is unique to each person can therefore be deduced, and thus humans can be identified.

1.2 Research Objectives

The aim of this research is to develop an algorithm that can utilise channel state information derived from standard Wi-Fi hardware to rapidly and accurately authenticate a number of users, specifically by measuring human biometrics. The specific tasks undertaken comprise:

1. Extract CSI from commercial Wi-Fi hardware in a Linux environment using publicly available Intel 5300 NIC and software

2. Process CSI into data to be manipulated using MATLAB

3. Develop an algorithm in MATLAB that can detect the presence of a human in the environment of the Wi-Fi transmitter and receiver

4. Advance the algorithm to isolate and measure various physiological characteristics of a human

5. Advance the algorithm to have significant granularity with which one human can be authenticated as being different from another within a certain margin of confidence

6. Advance the algorithm to register biometric profiles of multiple users and authenticate a scanned human as matching one of the registered profiles

1.3 Overview of the Dissertation

The dissertation is organised as follows:

Chapter 2 evaluates related research done on CSI and on biometric authentication. A research gap is identified and methodology for this project is informed. The aim to authenticate users based on cardiac rhythm is determined based on previous research results and biometric authenticator findings.

Chapter 3 outlines the proposed methodology of the project based on the aims and related research. The ethical considerations of the project are evaluated. The method mainly involves configuring the Linux machine and the wireless network settings.

Chapter 4 detailed how the proposed methodology was implemented. Challenges associated with the network configuration and OS set-up are discussed. The physical environment set-up is outlined.

Chapter 5 discussed and analysed the results of the testing of the system. The development of the software used to carry out each function of the system is outlined as the project progressed. The results achieved with different algorithms are evaluated and a processing script is thus developed. An overall procedure for identifying a user based on their heart rhythm using CSI is proposed.

Chapter 6 concludes the project, evaluates what was accomplished based on the intended aims, and evaluates the effectiveness and feasibility of the system developed. Further work required to improve and validate the system and its findings thus far is suggested.

Chapter 2

Literature Review

2.1 Chapter Overview

This chapter examines relevant research in the fields of CSI measurement and biometric authentication. By analysing the results of related work, the aims of this project are validated, and a research gap is identified – thereby justifying the undertaking of this project. The procedures undertaken in previous research is also used for informing structuring of the methodology of this project.

2.2 Wi-Fi Human Measurement

There is a significant amount of research using Wi-Fi signals to measure humans in some form. Some of this research that is especially pertinent to this study is reviewed and analysed. The findings of each paper are evaluated so as to determine their impact on this research project. Though these papers may only partly relate to this project, in that they are not using CSI for authentication but rather another human-involved process, their methodology may still prove to be quite relevant to formulating methodology for this project. The papers here effectively are performing the same or very similar measurements of humans, but using them for a different applications.

2.2.1 CSI-MIMO: An efficient Wi-Fi fingerprinting using Channel State Information with MIMO

Chapre et al. (2014, pp. 202-209) developed a novel location fingerprinting system using MIMO CSI that uses frequency and spatial diversity to locate people within an indoor environment. Their system uses either magnitude only or full complex CSI for dynamic and static environments respectively.

Dynamic environments were rooms with moving people. The researchers claim to have developed the first system for switching CSI measurement based on the mobility of a given environment.

The CSI data is aggregated over multiple antennas for each subcarrier to form a 'location signature' CSI-MIMO which consists of the difference between phase and magnitude between subcarriers. They create a radio-map, and they utilise multiple access points (APs) so as to reduce mean distance error and thus increase accuracy. This is done using a conventional Wi-Fi fingerprinting approach, wherein training is performed to determine the radio data at various locations within an environment. The collected data is then used to form location signatures which are stored in a map, forming a radio-map with various locations and their associated radio data. An unknown test point is used then, with radio data, and this is correlated with the radio-map to deduce the location of the test point. This approach may prove to be useful for this CSI authentication study. A radio-map could be analogous to what is required for forming a CSI based radio depiction of human body features, and this paper is a successful demonstration of how location data can be derived using CSI.

The article is also useful in demonstrating a mathematical procedure for manipulating CSI data. They show that a MIMO system on a network using OFDM in a narrow-band flat fading channel can be given by

$$y_i \qquad Hx_i + N_i \tag{2.2}$$

Where y_i and x_i are the received and transmitted signal vectors as explained in the Background Information section of this report, H denotes the channel matrix of CSI data, and N_i is the Gaussian noise vector. A training sequence that is known ($x_1, x_2, \ldots, x_n$) is transmitted and thus the channel matrix H can be estimated at the receiver side; the received signal y_i is given by

$$Y \qquad [y_1, y_2, \cdots, y_n \qquad HX + N \tag{2.3}$$

Hence, they estimate the channel matrix H with

$$\hat{H} \qquad \frac{Y}{X} \tag{2.4}$$

Which is the fine-grained physical layer complex information for N subcarriers, which can be interpreted as:

$$H \qquad [H_1, H_2, H_3, \cdots, H_N{}^{T} \tag{2.5}$$

Where

$$H_i \qquad |H_i|e^{j\sin(\angle H_i)} \tag{2.6}$$

And thus for MIMO systems, H_i has dimensions $p \times q$:

$$H_i \qquad \begin{bmatrix} h_{11} & h_{12} & h_{13} & \cdots & h_{1q} \\ h_{21} & h_{22} & h_{23} & \cdots & h_{2q} \\ h_{31} & h_{32} & h_{33} & \cdots & h_{3q} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ h_{p1} & h_{p2} & h_{p3} & \cdots & h_{pq} \end{bmatrix} \tag{2.7}$$

Therefore, the size of the channel matrix H is $p \times q \times N$ where h_{pq} is a complex number representing the amplitude and phase of each subcarrier. This matrix modelling of CSI subcarrier data can be used for wide array of CSI techniques, and as such this modelling will prove to be significant for usage in this report.

The system developed by Chapre et al. (2014, pp. 202-209) demonstrated an accuracy of 0.98m in a static environment and 0.31m in a dynamic environment.

2.2.2 Monitoring Breathing via Signal Strength in Wireless Networks

Patwari et al. (2014, pp. 1774-1786) investigate using RSSI as a method for measuring human respiration. Though this is not using CSI, the paper still presents useful information about how breathing can be measured using Wi-Fi signals. Though they use RSSI-based measurement, which is a coarse-grained information type derived from received Wi-Fi signals, their accuracy is quite strong, demonstrating the system exhibits reliable detection and frequency estimation with only 30 seconds of data, within 0.7 and 0.42 breaths per minute RMS error in several experiments. This indicates a promising potential reliability of CSI-based respiration measurement, which is proven to be more accurate and reliable than RSSI, being fine-grained information.

Examining the effectiveness of measuring human breathing using Wi-Fi signals is pertinent to this study as respiration is shown to be useful as a unique identifying quality between different people, as will be shown in the relevant literature. This means that it could be a parameter by which people can be authenticated using CSI.

The researchers implement an RSS measurement system wherein the user is position in the line of sight of radio links between many different wireless devices. Inhaling and exhaling is significant enough variation - given that the rest of the environment remains static - to engender a change in RSS. Then, the researchers estimate breathing rate based on the frequency at which the RSS is altered, due to a periodic change in its time delay. A comparison between actual breathing rate and power spectral density of the RSS link data shows that this estimation is very accurate, as shown in Figure 2.1. Naturally, owing to the utilisation of the coarse-grained metric RSSI, little information about a person's respiration can be deduced from this technique other than the breathing rate, and as such this system requires a far greater level of precision to be useful for authentication. The CSI authentication system can therefore build upon the concepts used to measure respiration within this paper and enhance the system using CSI's fine-grained information.

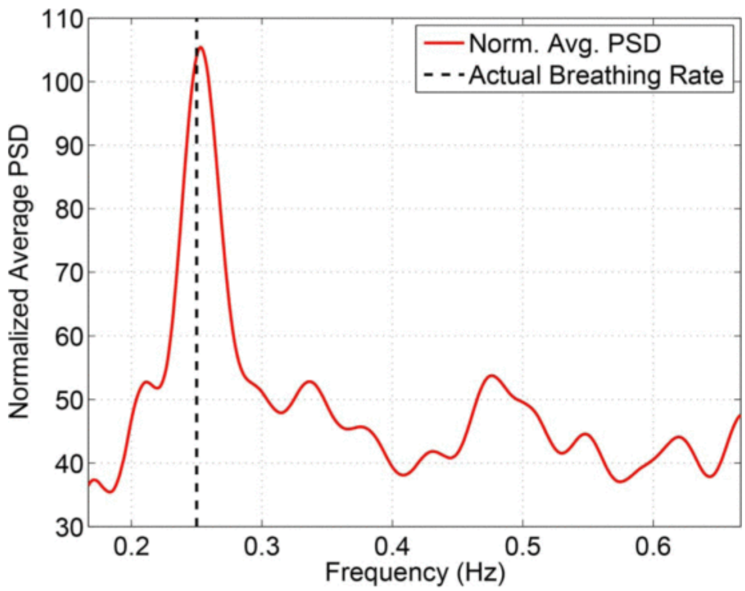


Figure 2.1: Respiration Measured Power Spectral Density vs Real (Patwari et al. 2014, pp. 1774-1786)

2.2.3 CSI-Based Authentication: Extracting Stable Features using Deep Neural Networks

Abyaneh, Pourahmadi and Foumani (2019) focussed their attention on the problem of rotation in their study. That is, when utilising CSI to authenticate a user, the signal paths due to their physical presence will be significantly different when they rotate their body, thus causing a correlation-based CSI method (i.e. comparing CSI to a training set) will not identify the user. They use both a correlation based approach and a deep neural network LocNet approach.

The method of authentication outlined in this research was simply to identify whether a user is standing in a specific 'secure' location, and features are extracted from the CSI data that can indicate the location of a user. This authentication technique is not robust and easily infiltrated by a malicious user. The primary focus of this research however was to identify how user rotation affects the accuracy of the location measurement, and from this investigation the researchers found that a deep neural network is more resilient to rotational changes than feature extraction using correlation. This is highly pertinent for this project, as these findings can be used to suggest future work in a deep neural network authenticator, if it is found that when using correlation for authentication, stability and resilience to positional adjustments is inadequate.

2.3 CSI Human Identification

There is a small amount of research currently available involving the use of CSI to identify/authenticate people. The findings are evaluated for the relevant papers to form a knowledge basis for this paper. The shortcomings of the research are evaluated also, to identify what aspects require improvement on/furthering within this paper. A broad view of the current literature can therefore be ascertained so that a specific gap in current knowledge can be identified, which this paper will aim to fulfil.

2.3.1 Device-Free Identification Using Intrinsic CSI Features

Wang et al. (2018, pp. 8571-8581) use CSI to identify users using either their gait (walking straight along a 6m line) or their respiration. The researchers argue in this paper that the most significant problem with using CSI for identification, and really any human measurement application by extension, is to extract high quality signals containing useful information from very noisy signals. They utilise a novel empirical mode decomposition (EMD) based identification system to decompose raw noisy CSI measurements into intrinsic mode functions (IMF) to extract multi-domain features.

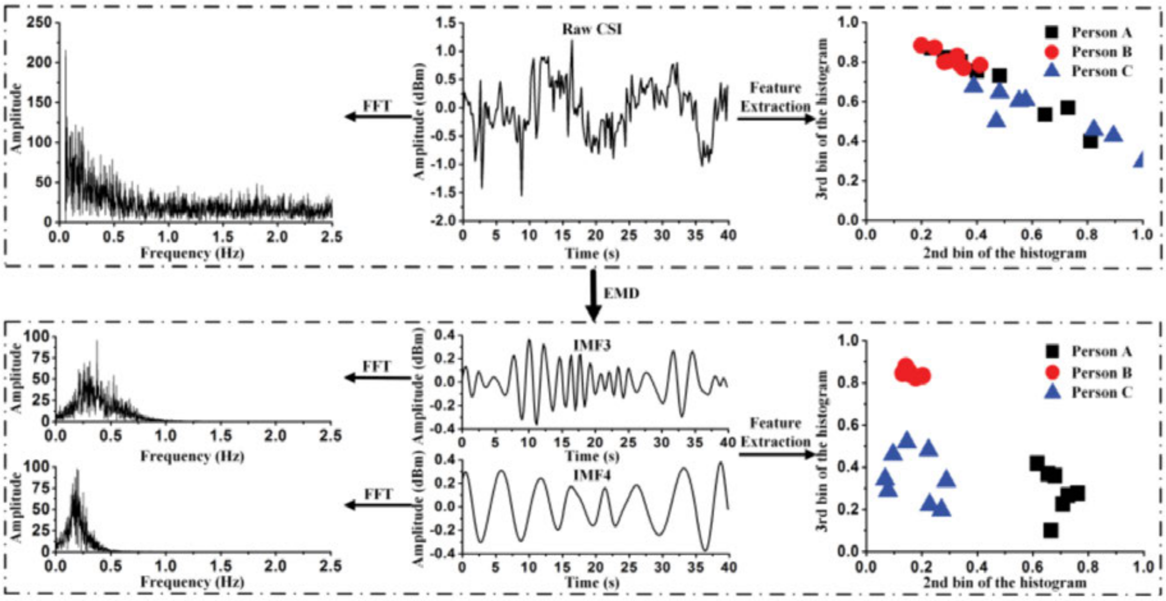


Figure 2.2: Respiration Feature Extraction from CSI with EMD-IMF (Wang et al. 2018, pp. 8571-8581)

The problem in feature extraction is that each persons characteristics will be quite different prior to identification and thus automatic adjustment and extraction is required. They use EMD to decompose a non-stationary and non-linear signal into a finite number of IMF components based on the local time scale characteristic of the signal, which form a complete and nearly orthogonal basis the original signal. These characterise the useful information in the raw signal. Figure 2.2 shows the performance of the system in separating breathing characteristics of three simultaneous users. Here, the time-domain and frequency-domain characteristics of the 3rd and 4th IMF components are shown, as well as the feature space spanned by only two normalised bins, the 2nd and 3rd bins of the time-domain histogram. The full framework of their system consists of a CSI collection and pre-processing module, an EMD based IMF components extraction module, a feature extraction module, and an identification module. This is shown in Figure 2.3.

Wang et al. (2018, pp. 8571-8581) use principal component analysis (PCA) to convert a set of correlated CSI channel measurements into a set of a linearly uncorrelated signals. Their system can be interpreted as a machine-learning system.

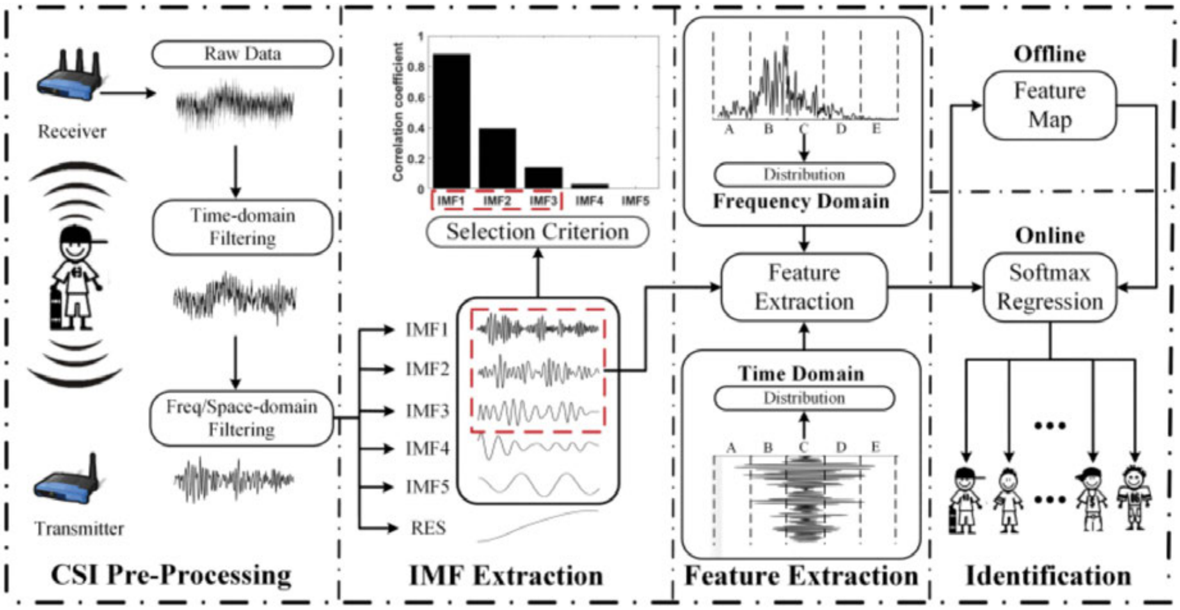


Figure 2.3: Softmax Regression Based PCA scheme for CSI identification (Wang et al. 2018, pp. 8571-8581)

For the respiration identification testing, they used a frequency of 5.32GHz on the transmitter continuously emitting signals. The receiver collected CSI measurements at a sample frequency of 100Hz. There were 3 antenna pairs and 30 channels for each antenna. They used a sample size of 10 people, with 8 training samples and 8 testing samples for each. Each sample is a time series with length 3000. This means their overall testing time would have been 24 seconds, not including processing time. They use a time domain mean filter of length 10. For each component, they select the IMF with the largest correlation coefficient to calculate time-domain features, select the 3 IMFs corresponding to the 3 largest correlation coefficients, and add them to calculate frequency-domain features. The receiver and transmitter are both equipped with an Intel 5300 NIC running ubuntu 11.04. For the gait test, they use 5.32GHz, 30 channels at a sampling rate of 2500Hz for 3 antenna pairs, and a time domain mean filter length of 5, with 11 principal components from the joint frequency spatial domain CSI measurements, compared to 10 for the respiration.

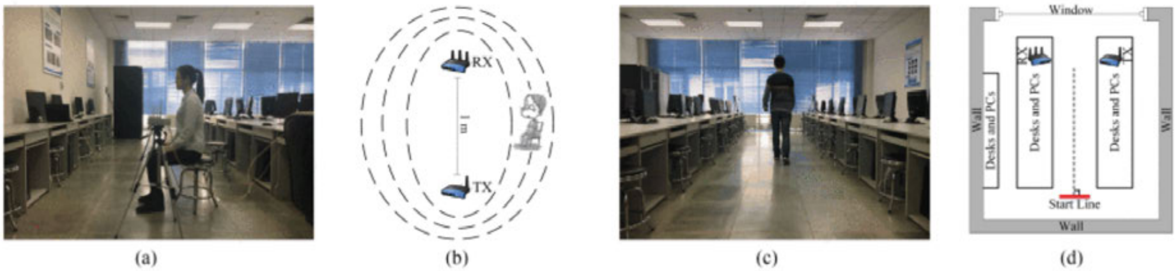


Figure 2.4: Testing Environment for CSI Identification (Wang et al. 2018, pp. 8571-8581)

They found that with the IMF framework, using both time domain and frequency domain components, they could achieve 97.5% accuracy for respiration identification and 90.4% accuracy with gait. That is, they use histograms to extract features from the reconstructed time series by the IMFs in both the time domain and frequency domain. They tested the robustness of this framework by changing the environment, including putting boxes around transceiver and changing clothes of the users. The system maintained accuracy through these changes owing to the fact that the system measures dynamic features in the user, which are not altered. When testing an untrained person (intruder), they see the probability vector is nearly random so by setting a threshold they easily identified an intruder.

This study demonstrates that CSI can be used to achieve a high degree of accuracy for device free identification. The significant learning points of the study are that an EMD based decomposition module significantly improves CSI identification compared to using raw signals. This study forms a strong foundation for further research in the CSI device-free identification field. The methodology of this report can be advanced by utilising a more robust identification feature and by improving the sensing speed of the system.

Utilising gait as an identifying unique feature is feasible to a certain degree, however an accuracy of 90% is not sufficient for highly sensitive applications, such as secure entry to a private location. Further, requiring a testing space of 6m walking space is not practical for most use cases. As such, gait can be considered an inappropriate measurement technique.

Respiration demonstrates a greater degree of reliability at 97.5%, and requires only that the user sits in the correct location. However, requiring that the user sits for 8 testing samples totalling 24 seconds is impractical for most applications, and prohibits CSI identification from being a worthwhile alternative to existing device-based identification systems. Furthermore, a respiration based authentication system leaves room for system attack/infiltration purely based on a person's breathing not being as unique as other metrics; respiration can be emulated quite closely by purely observing a user and breathing at the same rate, in the same way (same thoracic/abdominal breathing pattern). Wang et al. (2018, pp. 8571-

8581) did not evaluate this concern, and though it may seem unrealistic, it is as serious a concern as severing a finger to breach fingerprint authentication, or removing an eyeball to breach retinal scanning authentication. Further, one's breathing patterns vary greatly as mood, activity level and environment varies. Wang et al. (2018, pp. 8571-8581) did not address this variation.

Thus this project will aim to advance upon Wang et al. (2018, pp. 8571-8581) research by improving the authentication time, aiming for a sub-10 second authentication. Further, a more robust authentication metric will be determined and utilised to circumvent the vulnerabilities associated with respiration-based authentication.

2.4 Cardiac Based Authentication

Recent advancements in medical science have shown that a person's cardiac rhythm is a highly unique identifier, akin to a fingerprint. There has been recent research done into utilising an electrocardiogram (ECG) as a biometric authenticator, and this research shows that it is nearly impossible to counterfeit, unobtrusive, cost-effective, and it is present in every living individual; fingerprint, retina scanning, respiration, and gait are all exclusive to people with disabilities prohibiting that biometric. Making it applicable to CSI authentication is the fact that it is a dynamic characteristic, in the sense that its uniqueness stems from how it varies over time, rather than being a static characteristic. As such, it is promising as an alternative to respiration or gait based CSI authentication. There is, at the time of writing, no existing research into measuring ECG using CSI for authentication purposes. Therefore, the effectiveness of this biometric for device free identification will be investigated.

2.4.1 Individual Identification via Electrocardiogram Analysis

Fratini et al. (2015, p. 78) performed a survey into methodology for indentification via ECG using 100 different research papers. This study is highly pertinent to the proposed CSI ECG authentication scheme as, though the method of feature extraction is to be through wireless signals, the exact features that will be used to determine a user's identity must be determined. Fratini et al. (2015, p. 78) outlines which features of the complex ECG are most effective for identification and how to determine these features once they are measured through a given method.

This paper outlines two distinct types of ECG biometric identification: fiducial and non-fiducial approaches, depending on the need to identify precise points in the heartbeat. The authors propose that there is a significant problem associated with fiducial approaches:

> "temporal, amplitude and morphological features require accurate detection of fiducially and the achieved results are clearly dependent on the recognition procedure."

As such non-fiducial approaches have been proposed in recent years that overcome this problem by not requiring fiducials recognition. Non-fiducial approaches therefore reduce error rate and computational effort, they do not require identification of ECG waves and they have the potential to take fine features into account which could be lost using fiducials. All forms of non-fiducial ECG identification are based on the assumption that ECG is a quasi-periodic signal, in that it is very repetitive. The authors divide the approaches into three main categories: autocorrelation based, phase space based and frequency based.

The autocorrelation based approaches simply took a period of time of an ECG waveform and performed a normalised autocorrelation over a window of lags. This extracts information about an ECG's characteristics, in that it is shift invariant and highlights non-random patterns. The QRS complex especially maintains a strong invariance in shape and time width.

The frequency based features approach include a paper by Loong et al. (2010, pp. 340-345) which utilised a linear predictive coder (LPC) to model the frequency content of the ECG recordings. The survey also included Wang et al. (2018, pp. 8571-8581) EMD-IMF based approach, as well as the PCA approach that Wang et al. (2018, pp. 8571-8581) used for feature extraction. They outline that PCA reduces the feature space dimensionality by performing Eigen-analysis on the covariance matrix of the original features.

Overall, the survey argues that the most significant findings are that most studies ignore the variability of ECG in the long term (e.g. variability induced by age, iterate exercise) and due to pathological effects, that most studies are conducted on a population of just a few dozen, that most studies do not and should include data on subjects in different states (e.g. relaxed, stressed, before during and after exercise) and along a period of several years, and that it is not yet clear which is the best feature set and classification scheme for ECG biometrics.

2.5 Chapter Summary

After performing an evaluation of the research relevant to CSI based identification, an understanding of appropriate methodology has been gained. After evaluating Wang et al. (2018, pp. 8571-8581), which is the most similar research performed to date, a literature gap has been identified in the speed of authentication as well as the reliability of the method. As such, ECG is proposed as a unique identifier with which people can be identified using CSI. ECG is considered an appropriate identifier as it is dynamic, meaning it is measurable with CSI and invariant to environmental changes, it maintains the privacy of the user (compared to facial scanners), it is within all people and is not exclusive to disabilities, and it is extremely difficult to counterfeit. The ECG cannot be observed and replicated as gait and respiration can to a certain degree. A problem associated with the ECG is that it varies with mood and activity level, and age, fitness and condition in the long term, which can be investigated with extensive testing.

Chapter 3

Research Design and Methodology

3.1 Chapter Overview

This chapter outlines the ethical considerations of undertaking the project before detailing the resources and method that will be used for conducting testing. The ethical considerations discussed relate to privacy concerns of a sensitive physiological measurement and equality of effectiveness for users of different body compositions and sex. The primary aspects of the method outlined are the installation of the hardware, configuration of the Linux kernel, and setting up the physical environment.

3.2 Ethical Considerations

The primary ethical consideration of applying CSI to ECG for authentication is the necessity to measure and store ECG information about users. ECG, especially in its raw form, can give away highly private information, such as the fitness level, approximate age, heart health, etc. If the data base of trained sets of ECG information were to be leaked/hacked, the consequences would be even more severe than if standard passwords/fingerprints were leaked, as the ECG is both an authentication method and a telling physical condition indicator. This ethical consideration is mitigated by the fact that raw ECG signals are not the ideal form for CSI measurements, and thus the processed ECG signals without much of the telling information should be all that is necessary to store for authentication.

Other ethical considerations are associated with the necessity for the Linux 802.11n CSI tool to be run on an unencrypted network (security protocols turned off). This presents a safety risk in that people with malicious intent could access the network if they are nearby enough and observe CSI changes due to a user's heartbeat. If they developed a fiducial system to measure fine details about the ECG signals, they could garner sensitive personal health information about users. This tool requires no security on the network as it was developed using proprietary Intel firmware that the authors of the tool do not have access to anymore as they no longer work at Intel, and the tool was not originally designed with headroom for managing security protocols.

The development of this technique for revealing information about a physical environment from readily available Wi-Fi signals also engenders considerable ethical concerns regarding malicious use and privacy. Furthering this technology furthers the risk that a person with malicious intent can garner even more information of value to them by infiltrating a wireless network. Similarly, furthering this technology makes unknown Wi-Fi networks dangerous, as any Wi-Fi network that a person is in the physical vicinity of could have the means to gather information about them. There is also growing concern about the amount of information that users are required to divulge to technology and to companies, especially when considering recent controversies regarding data leaks and companies secretly selling user data. There is certainly a privacy risk involved in measuring and storing information related to a user's heart, and many people may not be comfortable divulging information about their heart to a device; encryption and safe data storage are paramount for implementing a technology like the CSI heart authenticator. These are problems already faced by existing technologies such as facial scanners, retinal scanners and fingerprint scanners, and certainly strategies for mitigating these ethical problems can be learned from approaches to these technologies.

Lastly, there is an ethical dilemma within utilising the stable rhythm of the heart for authentication. People with a pacemaker could find that the proposed system is not suitable for them, as the pacemaker would reflect/interfere with the microwave signals and might cause them to be unusable for identification. Further, there is a possibility that people with irregular heart rhythms due to health conditions could find the system unsuitable for them. The system relies on a heart rhythm recorded some time in the past to match a heart rhythm recorded at the time of authentication, and thus people whose heart rhythm is changing significantly might not be able to match a previously recorded sample.

3.3 Resources

- Intel 5300 802.11n Network Interface Card
- Computer with free PCI-E x1 lane or PCI-E x16 lane with adapter, running Ubuntu Linux kernel version between 3.2 (e.g. Ubuntu 12.04) and 4.2 (e.g. Ubuntu 14.04.4)
- 802.11n capable AP with security protocols turned off (e.g. no WPA2)
- Testing environment with minimal variability

A second hand 802.11n NIC was acquired as these cards are not used in new devices anymore. The device was only inexpensive at approximately $20 AUD. The NIC was installed in a standard form factor desktop environment in the PCI-E x1 lane of an ITX motherboard. A NIC to PCI-E x1 adapter card was required which was also inexpensive. Two antennas were installed on the A and B channels of the network card.

3.4 Method

The first step in the methodology is to install the 5300 NIC, either into a network card slot if it is available, or into a PCI-E adapter and into a PCI-E slot. The next step is then to configure the 802.11n Linux CSI tool made by Halperin et al. (2010, pp. 159-170). This requires configuration of the network drivers in Ubuntu Linux. It also requires that the security protocols on the Wi-Fi network upon which the CSI authentication will be performed are disabled. This means that the authentication system should be placed on an independent network that is not connected to the internet and is not connected to any personal devices.

The system should be able to communicate with the AP through the Linux tool's proprietary drivers, and record channel state information from wireless communication. A simple ping will be used to form the packets for communication. The CSI can then be logged to a file on the Ubuntu system, which will be transferred to a different computer for pre-processing in MATLAB. Then, the raw CSI data can be processed with one of the various schemes outlined in the literature review. Testing is required to determine which of these schemes will be best suited to this project, considering the dynamic spatial path frequency coupled measurements taken with ECG signals and also considering time and resources. The most likely schemes are either an autocorrelation, linear predictive coder or EMD-IMF approach on a non-fiducial scheme. Autocorrelation is the simplest approach and EMD-IMF may be the most complicated, which must be evaluated alongside performance. Non-fiducial details especially pertaining to the QRS complex will be investigated, as Fratini et al. (2015, p. 78) found that these are the most reliable and most detailed ECG characteristics. The QRS complex pattern is shown in Figure 3.1.

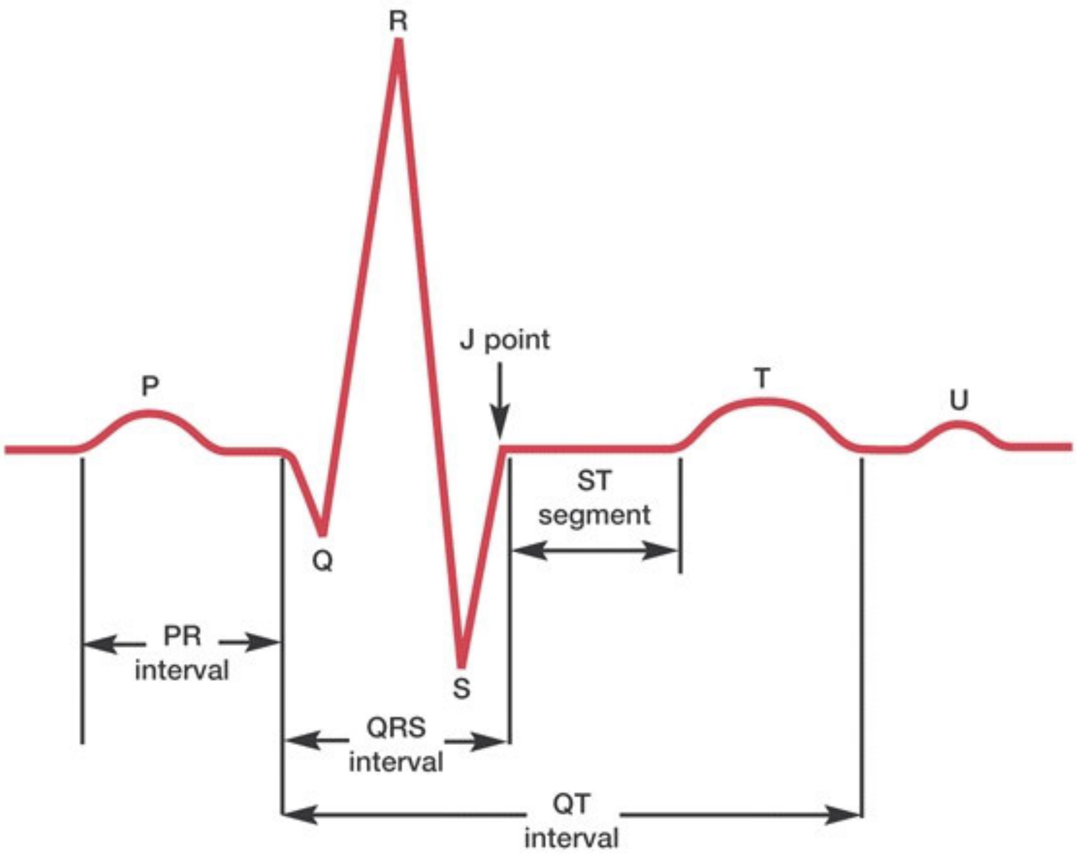


Figure 3.1: Basic Pattern of Electrical Activity Across the Heart (QRS Complex) (Ashley & Niebauer 2004, p. 19)

The testing environment will be comprised of a Netgear Nighthawk RAX80 802.11ax MIMO AP with the Intel 5300 802.11n NIC computer placed 2-5m away at the same elevation. Users will stand in between the two devices within the field path of the signals as demonstrated in Figure 2.4. Testers will be relaxed, not having exercised recently.

One method of overcoming the variability associated with ECG in the short term is to store not just one set of ECG characteristics per user, but to store many sets that correspond to a user's different moods, activity levels, etc. The problem with this is that it reduces the uniqueness of each set and thus can reduce the maximum number of users the system can handle whilst maintaining accuracy. Another strategy to overcome variability in the long term is to iteratively adjust the ECG measurements for every successful authentication. That is, set a threshold for identification (e.g. if the user matches a stored data set within x% confidence) and if the user's ECG characteristics fall within this threshold, adjust the stored data set to match the new characteristics. The problem introduced with this strategy is that for this adjustment to be significant enough to be worthwhile, the matching confidence threshold must be somewhat less than 100% which may introduce more inaccuracy into the system.

3.5 Chapter Summary

This chapter detailed the ethical considerations of the project, the resources required and the method was to be used. The privacy of the data is of utmost importance. The method primarily involves configuring the Linux environment and the analysis software.

Chapter 4

Application of Methodology

4.1 Chapter Overview

This chapter contains how the designed method was applied in practice, as well as how the system should be tested. The wireless network configuration is discussed as well as the Linux environment and the physical environment set up.

4.2 Configuration of the Linux Machine

The first step in the making of this project was acquiring and installing an Intel 5300 802.11n NIC into a computer running a specific version of Ubuntu Linux. This card is no longer installed in new computers, and thus a second hand product was the easiest route to obtaining a working card.

Figure 4.1: Intel 5300 802.11n NIC installed into PCI-E x16 NIC Adapter

This card required installation into a PCI-E x1 to PCI-E x16 network card adapter so that it could be installed into an available computer which only had a PCI-E x16 slot available on its motherboard. Two Wi-Fi antennas were installed into slots A and B; the third slot, C, did not have an antenna space on this network adapter, and thus was left unconnected. This simply means that any signals picked up on antenna C can be ignored as noise. This card was then installed into a test bench.

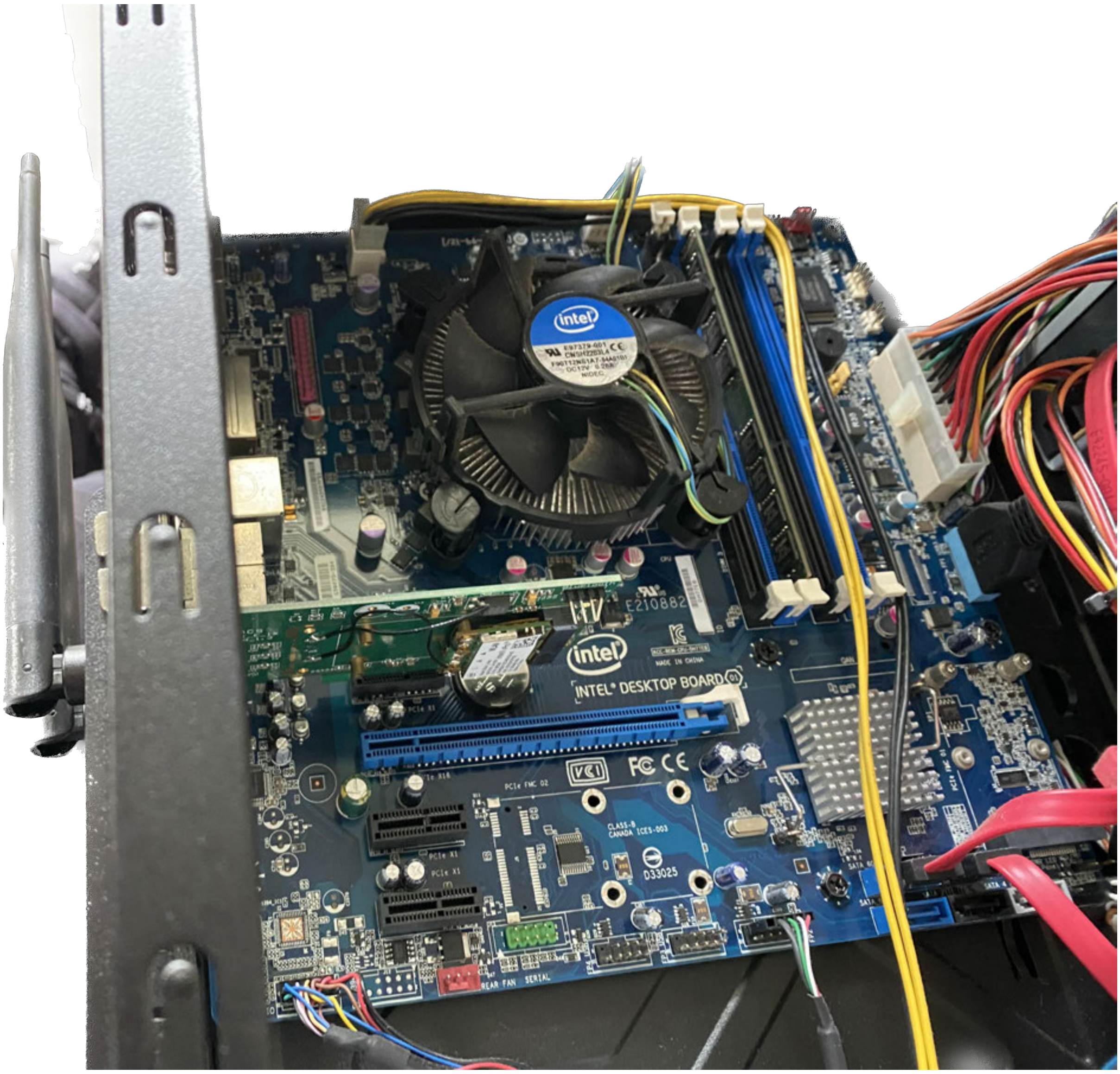


Figure 4.2: Intel 5300 NIC on Adapter Board on Motherboard

4.3 Configuration of Ubuntu Linux, Software, and Physical Environment

Now the operating system can be installed on the 5300 machine. The developers of the Linux 802.11n CSI tool, Halperin et al. (2010, pp. 159-170), explain that the tool is expected to work on Linux Operating systems based on an upstream Linux kernel version between 3.2 (e.g. Ubuntu 12.04) and 4.2 (e.g. Ubuntu 14.04.4). Ubuntu 14.04 LTS OS was used running Kernel Linux 3.13.0-24-generic. Next the build tools, Linux development headers, and Git client were installed. The initialisation of the CSI tool is detailed in Appendix C. To maintain greater control over network settings on the OS, the Network Manger Graphical User Interface (GUI) was disabled and prevented from booting, and command-line network utility iw was installed instead. Ubuntu's wireless drivers iwldvm and iwlwifi were prevented from booting automatically at system start up so that the CSI tool drivers can be loaded instead.

After loading the modified iwlwifi wireless driver back in with CSI logging enabled data can be collected. Note that the CSI tool does not work for encrypted access points, as the firmware for the Intel 5300 NIC did not have enough code room for the beamforming software paths required for measuring CSI as well as the encryption software paths required for connecting to encrypted access points (running WPA2 for example). As such, the AP's encryption must be disabled. Further, the AP used to perform this experiment (Netgear RAX80) was capable of Wi-Fi protocol 802.11ax, beyond 802.11n, and thus the mode needed to be limited. This was done by limiting the channel mode to 433Mbps, allowing for up to 802.11n speeds. Then when booting the CSI machine, the custom wireless driver must be loaded, and the wireless devices must be brought online using the iw command line utility. The AP can then be connected to, and logging can begin. Using a second terminal window, the process ID (PID) of the logging process is found and then the priority of that process is set to -19. Then the AP is pinged at a frequency high enough to allow for high fidelity measurement of the heart (i.e. 0.02s delay), and these ping packets act as the carrier which has its CSI manipulated by the physical environment of its propagation. This changing CSI is what is measured. The full list of commands used to initialise the logging process are listed in Appendix C.3. The CSI machine is placed upon a desk so that the 5300's antennas are raised to the same elevation as the AP's antennas, which are placed exactly 2.0m away.

It must be noted that though the pinging interval is defined, this is only the interval at which a packet is transmitted to the AP. This packet must then travel to the AP, and then the AP must transmit a reply to the PC. The time of propagation is always going to be variable due to the sensitivity of Wi-Fi signals, and thus the desired sampling frequency is never going to be representative of the reality of how often the heart is sampled. This is certainly a factor that contributes to lessened accuracy of measurement and is simply engendered by using Wi-Fi signals for measurement. The only methods for mitigating this variability are ensuring as little interference as possible in the test environment and by using higher gain communications equipment (e.g. directional antennas), though errors in these Wi-Fi signals are inevitable and their round trip time will always vary. This is why the sampling frequency of the pinging is not considered in-depth in this system, and this has implications for frequency domain analysis (i.e. using a Fast Fourier Transform).

One interesting obstacle faced in the configuration of the CSI tool each time it was set up was that the modem speed settings first needed to be set to 54Mbps to get the Linux CSI machine to connect to the Wi-Fi network. By repeatedly running the command 'sudo iw dev wlan0 connect {my SSID}' and checking whether an IPv6 address is stored with the command 'sudo ls /sys/kernel/debug/ieee80211/phy0/netdev:wlan0/stations/', a connection is attempted and the success of that connection is checked (as the Network Manager GUI was disabled). It could then be observed that the Linux machine very easily connected to the network on 54Mbps but when switching to 433Mbps speeds to allow for 802.11n High Speed protocol communications for CSI data collection, the Linux machine was very reluctant to connect to the network. Repeatedly switching back to 54Mbps speeds and retrying the connection was required before the system would connect to the network, and sometimes a modem reboot was required before the machine would connect to the network. The cause of this temperamental connectivity is unclear, but this was a barrier to easily testing the system.

Figure 4.3: Physical setup

Note that the AP and the test antennas are positioned perpendicular to one another, and they are positioned with a physical interface blocking propagation in one direction. This was done intentionally to provide maximum coupling between the orthogonally radiated signals out of the cylindrical Wi-Fi antennas and the AP. As outlined in the Literature Review, the person under test should position themselves in-between the curved signal path of the antennas and the AP in order to have their body physically interfere with the Wi-Fi signals. As such, the people under test were positioned as shown in Figure 4.4.

Figure 4.4: Position of person under test

To standardise testing, testers were asked to remove all but the thinnest most layer of clothing on their upper body. Each tester was asked to sit in the same chair with the height level adjusted so that the antennas are level with their heart.

4.4 Chapter Summary

This chapter discussed the configuration of the wireless network for CSI measurement as well as the set-up of the software on the Linux machine and the physical testing environment set-up. Initialising the communications channel link proved to be difficult, and the communication packets were discussed in terms of reliability for sampling information and variability.

Chapter 5

Testing Results and Discussion

5.1 Chapter Overview

This chapter outlines the testing and results from each aspect of the testing phase of the project. This includes the preliminary testing data collection phase, the biometric detection phase, and lastly the authentication optimisation phase. MATLAB code is developed as the project's phase advances.

5.2 Preliminary Testing

The first step in processing the gathered CSI data is to parse the data to a computer running MATLAB, which was done with a USB drive. A MATLAB script is thus required. The Linux Tool provides proprietary software, namely a MEX-file compiled from provided software read_bfee.c to unpack the binary CSI format. This file is first recompiled using MATLAB.

The first test performed was a simple ping to the AP IP address from the CSI machine. After loading this csi.dat file into the MATLAB script, it is unpacked with read_bf_file which delivers pertinent information regarding the CSI, namely Nrx and Ntx, which shows that the AP is transmitting the return pings on one antenna and the CSI machine is using three antennas to receive them (remembering that antenna C is disconnected). It also shows a received signal strength indicator (RSSI) for each of the antennas, indicating on average how strongly their netlink to the AP was, measured during packet preamble. These RSSIs are relative to an internal reference. It also shows rate, which is a hexadecimal value that is read using a format outlined in the source code. The rate is 0x8455, indicating that the modulation scheme is indeed OFDM, but the bit rate is only 12Mbps. Lastly, the csi field of the struct outlines the CSI matrix normalised to an internal reference, where the dimensions are Ntx x Nrx x 30, where the dimension of 30 is across the 30 subcarriers in the OFDM channel.

Now, these parameters can be combined to compute CSI in absolute dBm rather than relative to Intel's internal reference. This is done with get_scaled_csi.m, which combines the RSSI and Automatic Gain Control (AGC) values together to get RSS in dBm, and measured noise is used to compute SNR. The result is a matrix of dimension 1x3x30 representing the SIMO channel condition for this link, in normalised linear voltage space. The full source code is listed in Appendix D.1. Finally, the MATLAB toolbox command db() is used to convert from linear (voltage) scale into logarithmic (power) scale. Now, a plot of the SNR of each of the subcarriers can be formed, as shown in Figure 5.1.

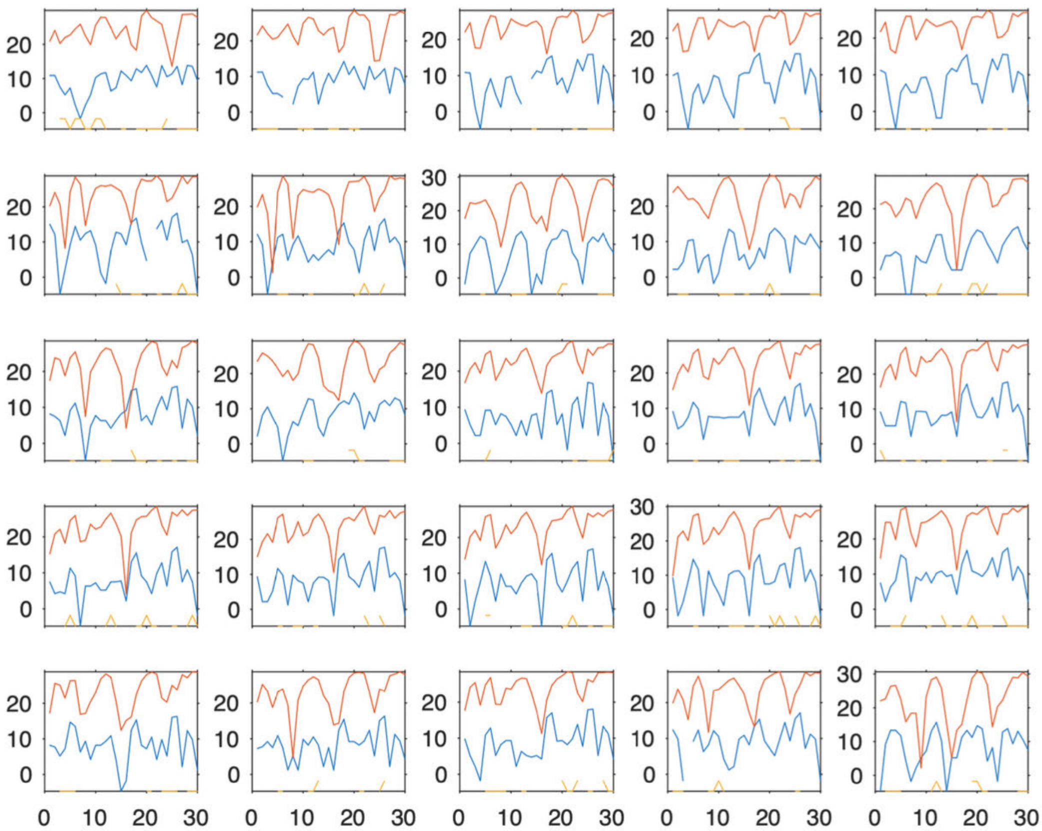


Figure 5.1: Preliminary CSI Testing Results from 25 Packets for 30 OFDM Subcarriers (SNR vs Subcarrier Index 1x3 SIMO)

This plot shows the effective SNR produced by the four 802.11 modulation schemes (BPSK, QPSK, 16QAM and 64QAM), and we can observe that some packets experience significant frequency selective fading. That is, taking the packet from row 3, column 1 for example, some subcarriers exhibit a large drop in SNR on antenna A and B, but some subcarriers exhibit a large drop in SNR on A but a rise on

B, and vice versa. This result from the preliminary testing was promising, as it indicated that the channel link is not so strong that is impervious to interference from the physical environment, but it showed that the strength was strong enough to withstand manipulation and still provide good dB SNR.

5.3 Biometric Detection

The next phase of testing and development was to, as planned, detect a cardiac rhythm using the measured strength of the OFDM subcarriers. To capture the rhythm of the heart, effectively the system is observing the fast fading of the Wi-Fi signals caused by the motion of the heart beat. This required inverting the algorithms dimensions so as to demonstrate how each subcarrier varies over the transmission of different packets, rather than showing how all 30 subcarriers behaved for each packet. It is also required a significant number more packets (i.e. less delay between local host pings) to increase sampling frequency. This is done using a second MATLAB script which is listed in Appendix D.2.

This phase of development showed that measuring cardiac rhythm with CSI is highly sensitive not only to the user's position, but their posture and body angle as well. This is because the movement of the heart is only a small physical environment change, and thus it is difficult to observe signal path impact due its movement. The finding that body angle affects CSI human measurement coincides with the findings made by Abyaneh, Pourahmadi and Foumani (2019) in their extensive CSI rotation investigation. After extensive testing, it was found that the optimal position for impacting the CSI using one's cardiac movement is approximately 10cms from the tips of the antennas, at an angle of approximately 45˚. Further, movement in the environment needed to be kept to an absolute minimum. There were no wireless devices switched on nearby, and the user was required to stay as still as humanly possible for the duration of the test. A quality measurement of the heart rhythm of the 20-year-old male, who will be referred to as User A, is shown in Figure 5.2, which shows the heart beat detected using CSI on at least one antenna of all 30 subcarriers. Note that some subcarriers are strongly affected by the movement of the heart and some are only weakly affected, as some subcarriers are taking paths directly through the heart where its movement can directly affect the signals. Some signals are taking paths that might only briefly travel through the heart, or may miss the heart entirely. Note also that the sample index spacing is not representative of the true time delay of each sample as prescribed by the sampling frequency, as the sample period is dependent on the round-trip time of the wireless signal (which itself is dependent on the antennas, the physical space, and the processing delay of the router and the PC, amongst various other factors).

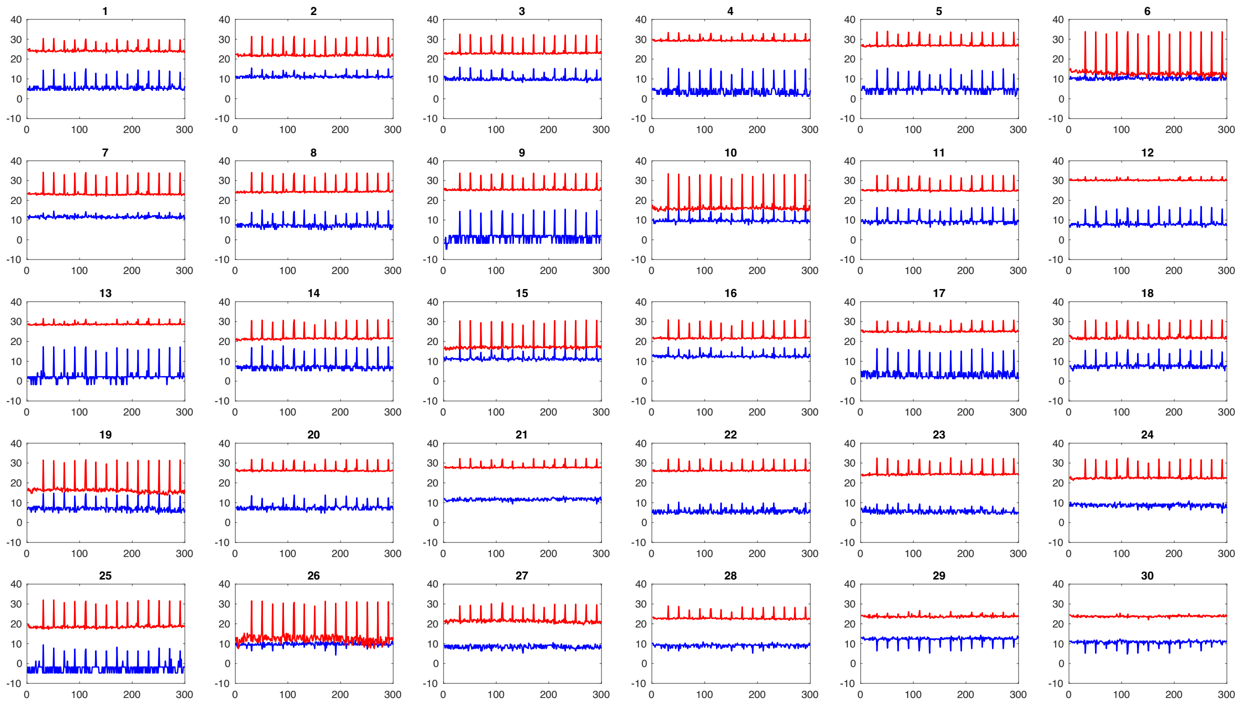


Figure 5.2: Quality Heart Rhythm Measurement Using CSI for 30 Subcarriers (SNR vs Sample Index) (User A - Male, age 20)

Figure 5.3 shows one of the subcarriers from Figure 5.2. Further tests were performed on another two subjects. More users would present a more accurate assessment of the performance of the system, but only two other testers were available at the time of testing. These testers were a 20-year-old female (User B) and a 49 year-old-male (User C). The users are tabulated in table 5.1, alongside a subjective assessment of the difficulty of acquiring quality measurements for each.

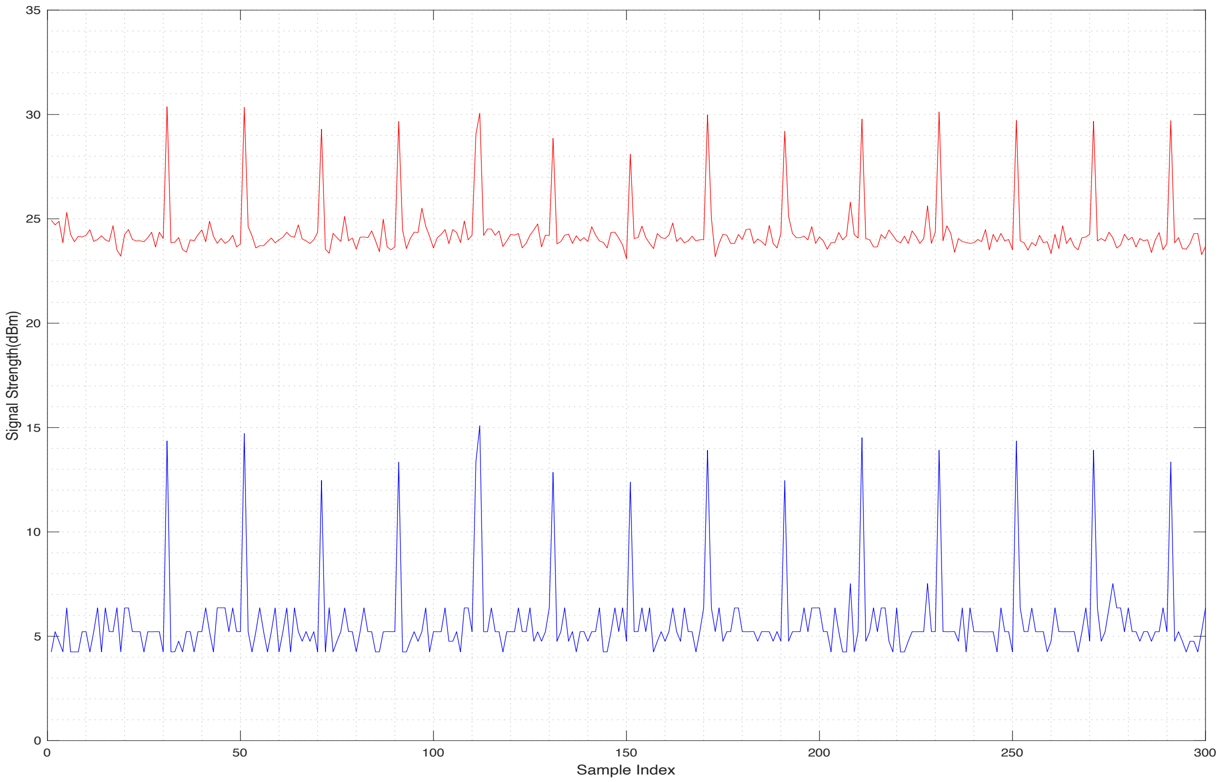


Figure 5.3: Signal Strength Variation Over Time for One Subcarrier on Two Antennas for Quality Heart Rhythm Measurement (SNR vs Sample Index) (User A – Male, age 20)

For comparative purposes, a real ECG is shown in Figure 5.4.

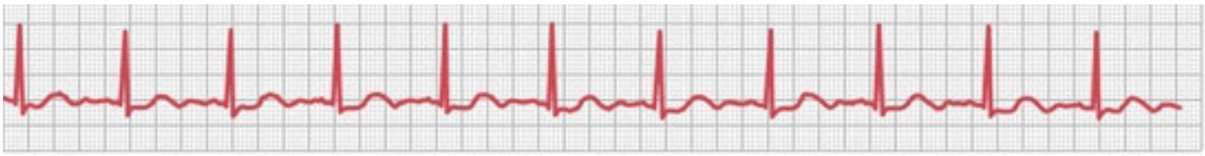


Figure 5.4: Real ECG recorded with ECG machine (Ashley & Niebauer 2004, p. 29)

Based on a superficial visual comparison of the real ECG and the system's CSI heart rhythm measurement, the performance of the system in measuring heart rhythm seems very strong (although fluctuation due to noise is noticeable). This shows promise that the measurement performance is strong enough to allow for extraction of fine non-fiducial details for authentication.

Table 5.1: User's Physical Description and Measurement Difficulty

User	Physical Description				Measurement Difficulty
	Sex	Age	Height(cm)	Weight(kg)	
User A	Male	20	177	91	Easiest; tests often, i.e. 1 in 2, result in good measurement of heart
User B	Female	20	164	62	Difficult; many tests required before good data achieved
User C	Male	48	184	85	Moderate; a few tests before good measurement

After repeated testing, it was found that despite eliminating as much variability in the environment as possible, the sensitivity and temperamental nature of the microwave frequency signals resulted in high quality heart rhythm measurements being difficult to obtain consistently. The final 200 samples of each sample space are clipped as it was found that for this configuration and sample interval, data quality begins to degrade in the final 2/5 of the test. This is likely because the environment conditions have changed too significantly to continue accurately tracking the heart rhythm after the first 3/5 of a test, and the user has likely accidentally moved slightly. This means that data in these portions of the tests were not always consistent when the heart was being measured, meaning erroneous correlation values would be measured.

The tests were repeated for each participant until at least three quality reference sample space sets were achieved. This was a compromise reached between correlation test validity and time taken for repeated tests. On average, every one in three tests performed measured the heart rhythm accurately for some portion of the test period on some subcarriers. Approximately one in 8 tests performed accurately measured the heart rhythm for all or nearly all of the duration of the test on all or nearly all of the 30 subcarriers. This low rate of successful testing, when also considering the difficulty in initialising the system for 433Mbps speeds, meant that repeatedly testing the system was difficult and time consuming, and as such for each participant only three high quality sample space sets were measured. The data for each of these quality tests for 30 subcarriers are listed in Appendix D.3.

An unexpected obstacle that was encountered during the testing of the system is that metal objects (e.g. jewellery) near the heart of a user impacted the signals significantly. Users wearing necklaces were required to remove them before the heart rhythm could be measured. This is because the metal causes multipath reflections that interfere with the subcarriers travelling close to the heart, causing destructive interference due to the phase difference at the receiver.

Based on the subjective assessment of the difficulty of achieving good measurement for each user shown in Table 5.1, User B was the most difficult to achieve good measurement with, followed by User C being somewhat difficult, followed by User A being relatively easy. The ease and consistency of measuring the heart rhythm is certainly going to be due to how isolated the heart rhythm is compared to an ideal system. This ideal system would be the receiver and transmitter inside an anechoic chamber with only the heart of the user inside the chamber (which is of course impossible). The more physical media that the signals must travel through which is not moving due to the beating of the heart is going to weaken the signal (due to path loss and reflections). Women by nature on average have more breast tissue than men, which could be the reason for which User B presented good quality data less often than the male testers. This is an issue that should be investigated in further work and could present a real ethical issue for the effectiveness equality of the system for all users. The quality of data derived from testing User C indicates that age does have an impact upon the effectiveness of the system. Age being the reason for the difficulty experienced with User C is further supported by the fact that User C was taller and weighed less, indicating less tissue surrounding the heart. This could also engender an ethical issue regarding the performance of the system for all people. Again, though, this is simply a subjective assessment of data quality based on difficulty of measurement. The data will be investigated more thoroughly after implementing an authentication algorithm to numerically analyse it. Also, there may simply be biological differences in the testers and their hearts which may affect the system's performance. This variability can only be eliminated with many more trials for a varied user base.

5.3 Authentication Algorithm

The first step in developing an algorithm to authenticate the users was to develop a new MATLAB script that takes an input file containing the data for 30 subcarriers over the duration of the test. It was decided that an autocorrelation algorithm would be the algorithm tested as it was the fastest and easiest to implement within the time constraints of the project. Other algorithms such as the LPC and EMD-IMF discussed in the literature review are left as future work for the system.

Initially, the intention for the system was to cross correlate each subcarrier of the test data against all of the OFDM subcarriers of the reference data, to eliminate spatial domain variations (as there was not a structured methodology for positioning a user, tests were performed with the user simply placing their chest close to the antennas). After testing this algorithm, the performance of the system was very poor, and much of the correlation yielded erroneous data. For example, many of the subcarrier cross correlations resulted in NaN values for the entire correlation lag period. To remedy this, a feature was implemented to isolate portions of the subcarrier data that demonstrated very high quality heart rhythm measurements, in order to avoid correlating wildly different data yielding NaN values or very poor correlation maximum values. This feature was successful in removing parts of the subcarrier data over the test duration that did not capture the heart rhythm prominently, which was based on the amplitude of the peaks of the heart rhythm compared to the noise (effectively the heart signal's modulation's SNR). However, this did not improve the performance of the algorithm when comparing each subcarrier to the rest, and in some ways worsened the validity of the data. By removing portions of the test duration on some subcarriers and not others, some correlations were comparing a small sample space of the periodic heart signal against a much larger sample space without accounting for a larger lag period, resulting in erroneous results. Further, some subcarriers did not capture the heart rhythm within the strength margin of the elimination feature at all, and as such were removed entirely. This therefore lowers the number of total averages used to calculate the percentage match, reducing the accuracy of the algorithm. This algorithm's reliability and ability to measure very minute differences in very complicated signals depends on averaging many subcarriers over many samples over two antennas over multiple tests in a data base. Removing some of these averages reduces the ability for the algorithm to evaluate this fine grain data.

After making these findings, it was decided that the autocorrelation algorithm would simply correlate each subcarrier with its reference counterpart – for example, subcarrier 1 test vs subcarrier 1 reference, subcarrier 2 test vs subcarrier 2 reference, et cetera.

The autocorrelation MATLAB script performs the following processes:

1. Load the data from the test, as well as from the reference test to be tested against, each comprised of samples for the duration of the test over 30 subcarriers over two antennas.

2. Separate multi-dimensional data into respective antennas and subcarriers.

3. Cross correlate data 1 against data 2, across all subcarriers and both antennas, with a maximum lag period of 20 samples. This is to accommodate for slight phase differences of the heart signal, without unnecessarily comparing against the full delay of the test duration, as the heart is periodic.

4. Find the maximum peak value of the correlation of each subcarrier of the test data correlated against the reference data, iterated over all subcarriers, for each antenna. Take the mean of these maximums to be the percentage match. Iterate this process over stored tests of all users stored

in the data base (A, B and C). Store the percentage matches of the test data against all stored data sets in separate arrays for each user.

5. Eliminate any matches that are 100% (this is simply to eliminate the possibility of a false positive match by erroneously including the test data itself in the data base when correlating test data). Take the mean of the array of means for the match against each user to be the average percentage match against that user. The highest percentage match is thus the identity of the user, and the percentage matches against each user are printed.

After running the correlation script for each test sample, the percentage match of each of the tests against each of the users (omitting correlating each sample against itself) has been determined. The results are shown in table 5.2, where green highlighted results show a correct identification and red highlighted results show an incorrect identification. Note that here we are correlating each of the three tests for a given user against the other two tests, as well as against all three of the tests for the other two users, and comparing the average correlation results. It is this way that the fine grained differences in heart signals can be analysed – comparing, for example, the modulation imposed on just one signal by a user's heart would simply not have the fidelity in this system to find minute differences across different users. However, by examining how 30 different signals are affected across two different antennas, minute differences can be examined over such a large sample space. This is why the 802.11n protocol is imperative not just for capturing CSI but for the success of this project overall, as OFDM allows us to access a far greater sample space compared to just a single wireless communications channel link.

Table 5.2: Correlation Testing Results

		User A Percentage Match (%)	User B Percentage Match (%)	User C Percentage Match (%)	Correct Matches
User A	Test 1	92.56	91.04	91.82	2/3
	Test 2	96.48	94.63	96.70	
	Test 3	97.33	95.30	96.86	
User B	Test 1	86.61	86.92	85.96	1/3
	Test 2	95.99	93.98	96.29	
	Test 3	94.56	92.88	94.22	
User C	Test 1	95.48	93.81	96.59	3/3
	Test 2	96.61	94.69	97.34	
	Test 3	96.99	95.01	98.03	
Overall Match Rate					**6/9 (67%)**

Thus, the overall correct match rate of the system using the autocorrelation algorithm (comparing only respective subcarriers) was 2/3 or 67%. This was based on taking the highest percentage match of each test to be the user that the system has identified as the tester. A correct identification rate of 67% indicates that the system is identifying the users more often than not, which indicates that the algorithm is indeed comparing the heart rhythm measurements and finding differences in the signals. Evidently, User C yielded the strongest correlation results with a correct identification in all three test cases.

It must be noted that there are flaws in identifying users with this method. The first of which is that the highest percentage match is taken to be the identified user, without considering whether the match is strong enough. Examining User B Test 1, the highest percentage match was 86.92%, which would not be considered strong enough to consider a correct identification, as was the case in the system designed by Wang et al. (2018). An improvement to the security of this system would be to include a threshold beneath which authentication is denied. It is also noticeable that User B exhibited lower percentage match values overall compared to A and C, whilst User C exhibited higher percentage match values. This coincides with the subjective assessment of the difficulty of attaining quality measurements, except for User C having the greatest percentage matches. User B's low percentage matches could be attributed to User B being female, increasing the minimum the distance between the hart and the antennas due to breast tissue, or it may have simply been a biological heart difference, or a difference in the physical environment (frequency domain or spatial) at the time of testing. These variables can be eliminated with more extensive testing with many more users.

It must also be noted that it is possible that User C exhibited the best correct identification purely because their tests exhibited the strongest percentage match overall, rather than just with the other User C tests. This can be seen in the higher values seen in any correlation using a test of User C compared to other correlations (i.e. between A and B or vice versa). This could mean that User C's tests contained less random noise, causing overall higher average correlation against any other data set. This is a problem related to correlation itself, and is another aspect that could be analysed by performing a greater number of tests over many different users. This problem may also be eliminated by incorporating a frequency domain analysis (i.e. Fast Fourier Transform) feature into the system.

5.4 Final Processing Configuration

After capturing data, processing it and finally analysing the data using a correlation algorithm, the cardiac biometric authentication system has been developed. As an overall description of the operations performed by the system at a broad perspective, the process of authenticating a user with the system uses the following steps:

1. Clear the physical environment of obstacles.
2. Disable network security and switch to the 802.11n IEEE protocol on the 2.4GHz band. Disable any wireless devices/communication aside from the Linux CSI PC.
3. Turn on the CSI PC, enable the modified wireless drivers and connect to the Wi-Fi network. Check the connection was successful by checking for the IPv6 address stored in the stations folder of the PHY information in the kernel (i.e. in /sys/kernel/debug/ieee80211/phy0/netdev:wlan0/stations/). If the link is not being

established, revert to lower speeds (i.e. earlier protocol such as 802.11g), and try again. Use dhclient wlan0 to renew IPv4 address.

4. Begin logging CSI and prioritize the process ID of the logging process.

5. Position user under test with the left side of their chest next to the PC's antennas, with their chest at approximately at 45° angle, approximately 5-10cm away. Begin pinging the router from the PC from another terminal process at a frequency high enough for good fidelity (i.e. 0.02s delay), for a set test duration. After pinging completes, end the logging process. The captured CSI information will be stored in a .dat file.

6. Use a USB drive to transfer the CSI data to a separate computer for processing with MATLAB. Run the processing script to store the subcarrier strength over the test duration. Repeat steps 4-6 for all desired user tests.

7. After forming a reference user test data set, run the analysis script for each test. Record the results and observe if the users were correctly identified.

5.5 Chapter Summary

This chapter discussed the results of the testing of the system. The final data collection and analysis procedure is outlined as determined by navigating challenges faced in the testing the system. When correlating each subcarrier over the test duration against all other subcarriers, as initially planned, the performance of the system was very poor. This was resolved by simply correlating each subcarrier against itself from the reference data set. The heart rhythm of the users was able to be accurately measured, and users were able to be identified correctly 67% of the time.

Chapter 6

Conclusion

6.1 Discussion of System Feasibility and Effectiveness

The performance of the system as well as its various flaws have been discussed as the progression of the system's development was detailed. As an overall evaluation, the effectiveness and feasibility of the system is discussed in this section.

The system can accurately measure heart rhythms with CSI and identify users 67% of the time. However, there are flaws in the validity of this result. Firstly, with only three users and only three tests each, there simply isn't enough data to rule out the many variabilities within the system that could contribute to this result. Many more tests are required before the true performance of the system can be ascertained, with many more users. More tests were not able to be performed however, due to the difficulty of initialising the system and difficulty of recording quality measurement of the heart. These difficulties certainly impede the feasibility of the project for more extensive testing and real-world implementation. The suspected reduced performance of the system for female testers is certainly a problem for future implementation that should be investigated further (this hypothesised problem due to more flesh in the area surrounding the heart would also apply for users with high body fat levels, another aspect that requires investigation).

The assessment of the system's effectiveness is further impacted by requiring the user to be very still with their very close to the antenna at a very specific angle. Orientating users in this manner reduces the effectiveness of the system as some users may simply find it difficult to position themselves in this specific way and stay very still, especially when considering many repeated tests might be required before the system captures the heart rhythm. However, is expected that with improvements to the sensitivity of the system and improved isolation of the heart rhythm the system should be able to perform without such close proximity to the heart.

By far the largest problem associated with the feasibility of this system is that the only known method for capturing CSI data – which is by using the 802.11n Intel 5300 NIC Linux CSI software developed by Halperin et al. (2010, pp. 159-170) – relies on hardware that is no longer produced and no longer installed in new devices, as well as on a networking protocol that has long since been succeeded. As time goes on, this NIC will become increasingly more scarce, despite being a widely available and inexpensive device currently. Further, the software being capable of operating only on the 802.11n protocol also means that it is possible that future commodity Wi-Fi hardware will not be able to accommodate the networking requirements. As Intel is the sole owner of the production of the NIC, the future viability of this system depends entirely on whether or not the company would produce them.

Despite its significant drawbacks, the feasibility of the system is bolstered by the fact that it uses very affordable, widely available hardware, and by the fact that it uses only Wi-Fi signals which are ubiquitous. Installing the CSI logging in a device-device manner, rather than device-AP as done in this system, would allow for the two devices to operate on their own private network within standard home or business Wi-Fi, eliminating security concerns due to not having security protocols enabled. Naturally, the data of the users must still be encrypted and protected in some way. The argument for the effectiveness of the system is strengthened by the speed of testing – correlation for strong results requires only a short time period of testing. Further, it is expected that shorter test durations would yield stronger results if compared to longer test durations, as there would be less time for environmental fluctuations and the user's positioning/body to modulate the signal.

6.2 Conclusion

This project has achieved its two primary aims – measuring a human heart rhythm with CSI and identifying a person based on their measured heart rhythm. All of the six Research Objectives have been achieved.

The first aim has been achieved as the heart rhythm of three different users were able to be measured with enough granularity to clearly observe the ECG-like rhythm in plots. This heart rhythm can be measured rapidly with an automated processing script. The system requires that the user remains still and positions their chest very close to the antennas, with no metal objects impeding the signal and as little environmental interference as possible. Achieving a test where the heart rhythm is accurately measured for the full duration of the test often takes many attempts. The difficulty of obtaining a high quality measurement limits the effectiveness and feasibility of the system, but the quality of the good measurements when they were able to be captured exceeded expectations. Examples of a good quality measurement are shown in Figures 5.2 and 5.3

The second aim of this project was achieved – the identity of users was able to be ascertained from the measured heart rhythms. This is outlined in Table 5.2. User A was correctly identified in two of three cases, User B in one of three cases, and User C in three of three cases. These results indicate the system is more effective than randomly guessing the identity of the users. Further, it should be noted that the system is not simply always correctly identifying a certain user and failing to ever identify another. If User C was correctly identified in all cases but User B was unable to be identified, for example (if User A or User C were always a greater match than User B tested against User B), the validity of the results would certainly be less convincing. This would indicate that effects other than the heart rhythm are modulating the signal to a greater degree when performing the tests, and that environmental interference is more significant to the correlation algorithm than the heart's modulation. It is hypothesised that User C was able to be identified correctly more often than the other users as this user's tests showed greater match percentages derived from correlation in all of the test cases they were a part of, not just when testing against themselves. This could be because these tests contained the least amount of noise, improving the correlation in all cases. Certainly, many more users and tests are required to eliminate variability and truly ascertain the effectiveness of the system.

Though this rate of correct identification is not strong enough for considering the system a truly secure authentication method, and notwithstanding the various drawbacks to the system's effectiveness and feasibility, the results of this project show there is a promising possibility for CSI-based cardiac authentication to supplant existing biometric authenticators and enhance cybersecurity. The flexibility of Wi-Fi signals as a measurement device allows for the system to possibly be far more versatile than existing authentication technique. This system also has a significant advantage over other CSI-based authenticators as well as existing biometric authenticators in that it would require no specific action from the user if the sensitivity could be improved to a point where the user would not be required to sit directly next to the antennas, as long as they are in their vicinity. This engenders the possibility for online logins to be performed at home or in the office autonomously based on the present Wi-Fi network measuring the heart rhythm of a user. Another possible future scenario if the technology could be improved is by having a device-device link established between the ceiling and floor of a door for example, for rapid and autonomous unlocking based on a user's heart rhythm measured as they approach the door. There are significant improvements required before the system can approach this level of security and measurement fidelity, and there are barriers to the feasibility of its wide-scale implementation, but the results of this project show that CSI-based cardiac authentication is possible and achievable.

6.3 Further Work

The future work of this project lies mainly in improving the fidelity of its measurement, performing further testing, and improving the feasibility of its use.

There are two main avenues available for undertaking the improvement of the measurement of the heart rhythm. These are improving the RF equipment and improving the networking configuration. By changing the style of antenna used on the Linux PC to attain more directionality, it is believed that the heart measurement would improve. This would be because the signal would have a greater SNR, thus being more impacted by the significant movement of the heart rather than by random environmental noise. This should isolate the individuality of the fine-grain movements of the heart more effectively. Further, as discovered with networking diagnostics, the system was only transmitting packets at speeds of 12Mbps. Transmitting more data more often may result in an increased ability to capture fine-grained movements of the heart. Lastly, and perhaps most importantly, it is believed that by switching the network configuration to run in device-device operation mode rather than device-AP far greater measurement fidelity can be achieved. This configuration would allow for far greater control of the packet transmission, processing, ping response, and other networking parameters that are controlled by the AP in the current state of the system. These parameters could thereby be optimised for the greatest fidelity of measurement – perhaps a shorter round-trip time could be achieved, to the effect of enabling analysis of the sampling frequency of the system by eliminating variability in the sample period.

There is perhaps the greatest amount of future work to be done in the testing aspect of the system. Most importantly, the results derived from this project should be verified by comparison with a real ECG machine. Secondly, the results will be greatly improved (at least in their validity) by performing a far greater number of tests with a far greater number of users. Users selected should come from a diverse array of backgrounds, with an ideal group of testers being very diverse in weight, age, race, sex, fitness, body composition, height, and health. The system should be tested on people with pacemaker devices and people with heart conditions, to examine the effectiveness of the system in these cases and garner an understanding of its equitability for real-world implementation. Tests should also be performed on users a number of times on different days, at different times, with the users being in different moods and levels of alertness. Tests should be done directly following exercise, sleep, caffeine consumption, alcohol consumption, and various other conditions that could affect their heart rhythm. Furthermore, other comparative algorithms should be tested, such as the LPC or EMD-IMF algorithms discussed as well as correlation over different subcarriers. Importantly, the FFT and other frequency domain investigative analysis techniques should be tested as to whether or not they can identify unique aspects of a person's heart rhythm that can be compared for authentication, whilst also hopefully eliminating noise that causes invalid correct correlation data.

For improving the feasibility of the system for implementation, the most important improvements will be to streamline the data processing and to decrease the temperamentality of the signals regarding the positioning of the user. The system should automatically parse a recorded test to the MATLAB processing script and thereby directly into the database. Users should be prompted and directed through performing a series of reference test to initialise their identity into the system. Then when performing an authentication test, the CSI should be automatically transferred to the processing algorithm and then automatically into the authentication algorithm before showing the result to the user. This would improve this system to the effect of fully and thoroughly achieving Research Objective 6. Furthermore, the sensitivity of the system should be improved following improvements made to its measurement and authentication aspects, and as such the difficulty of achieving good quality measurements should lessen. This would improve the ease of use and thus feasibility of the system. Lastly, the future longevity of the equipment required for this system to operate should be investigated, given that the 5300 NIC is no longer manufactured and the requisite software kernel (Ubuntu 14.04 LTS) is approaching end of life. If continued testing on this hardware and software is determined unfeasible, further investigation should be done into the possibility of extending CSI measurement capabilities to other network cards and operating systems.

References

Abyaneh, AY, Pourahmadi, V & Foumani, AHG 2019, ‘CSI-based authentication: Extracting stable features using deep neural networks’, *Transactions on Emerging Telecommunications Technologies*, vol. 31, no. 10, viewed 17 May 2022, <https://onlinelibrary.wiley.com/doi/10.1002/ett.3795>.

Al-qaness, MAA, Elaziz, MA, Kim, S, Ewees, AA, Abbasi, AA, Alhaj, YA & Hawbani, A 2019, ‘Channel State Information from Pure Communication to Sense and Track Human Motion: A Survey’, *Sensors*, vol. 19, no. 15, p. 3329, viewed 10 May 2022, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6696212/>.

Ashley, EA & Niebauer, J 2004, *Cardiology Explained*, Remedica, London, viewed 4 October 2022, <https://www.ncbi.nlm.nih.gov/books/NBK2204/>.

Chapre, Y, Ignjatovic, A, Seneviratne A & Jha, S 2014, ‘CSI-MIMO: Indoor Wi-Fi fingerprinting system’, *39th Annual IEEE Conference on Local Computer Networks*, pp. 202-209, viewed 15 May 2022, <https://ieeexplore.ieee.org/document/6925773>.

Fratini, A, Sansone, M, Bifulco, P & Cesarelli, M 2015, ‘Individual identification via electrocardiogram analysis’, *BioMedical Engineering OnLine*, vol. 14, p. 78, viewed 26 May 2022, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4535678/>.

Halperin, D, Hu, W, Sheth, A & Wetherall, D 2010, ‘Predictable 802.11 packet delivery from wireless channel measurements’, *ACM SIGCOMM Computer Communication Review*, vol. 40, no. 4, pp. 159-170, viewed 10 May 2022, <https://dl.acm.org/doi/abs/10.1145/1851275.1851203>.

IEEE Standards Association 2009, *IEEE Standard for Information technology-- Local and metropolitan area networks-- Specific requirements-- Part 11: Wireless LAN Medium Access Control*

(MAC)and Physical Layer (PHY) Specifications Amendment 5: Enhancements for Higher Throughput, 802.11n-2009, IEEE, Manhattan, New York, viewed 12 May 2022, <https://standards.ieee.org/ieee/802.11n/3952/>.

Judd, G, Wang, X & Steenkiste, P 2008, ‘Efficient channel-aware rate adaptation in dynamic environments’, *ACM International Conference on Mobile Systems, Applications, and Services*, pp. 118-131, viewed 12 May 2022, <https://dl.acm.org/doi/10.1145/1378600.1378615>.

Keysight Technologies, Inc. n.d., *Concepts of Orthogonal Frequency Division Multiplexing (OFDM) and 802.11 WLAN*, Keysight Technologies, Inc., Santa Rosa, California, viewed 12 May 2022, <https://rfmw.em.keysight.com/wireless/helpfiles/89600b/webhelp/subsystems/wlan-ofdm/content/ofdm_basicprinciplesoverview.htm>.

Loong, JLC, Subari, KS, Besar, R & Abdullah, MK 2010, ‘A New Approach to ECG Biometric Systems: A Comparitive Study between LPC and WPD Systems’, *World Academy of Science, Engineering and Technology, International Journal of Biomedical and Biological Engineering*, vol. 4, no. 8, pp. 340-345, viewed 26 May 2022, <https://publications.waset.org/9556/a-new-approach-to-ecg-biometric-systems-a-comparitive-study-between-lpc-and-wpd-systems>.

Patwari, N, Wilson, J, Ananthanarayanan, S, Kasera, SK & Westenskow, DR 2014, ‘Monitoring Breathing via Signal Strength in Wireless Networks’, *IEEE Transactions on Mobile Computing*, vol. 13, no. 8, pp. 1774-1786, viewed 15 May 2022, <https://ieeexplore.ieee.org/document/6587032>.

Wang, J, Zhao, Y, Fan, X, Gao, Q, Ma, X & Wang, H 2018, ‘Device-Free Identification Using Intrinsic CSI Features’, *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8571-8581, viewed 23 May 2022, <https://ieeexplore.ieee.org/abstract/document/8404072>.

Yang, Z, Zhou, Z & Liu Y 2013, ‘From RSSI to CSI: Indoor localization via channel response’, *ACM Computing Surveys*, vol. 46, no. 2, pp. 1-32, viewed 10 May 2022, <https://dl.acm.org/doi/abs/10.1145/2543581.2543592>.

Appendix A

Project Specification

ENG4111/4112 Research Project

Project Specification

For: Matthew Armanasco

Title: RF Biometric Authentication using Channel State Information

Major: Electrical and Electronic

Supervisors: Dr. John Leis

Enrollment: ENG4111 – ONC S1, 2022

ENG4112 – ONC S2, 2022

Project Aim: To develop an algorithm that can utilize channel state information derived from standard Wi-Fi hardware to rapidly and accurately authenticate a number of users, specifically by measuring human biometrics.

Programme: Version 1, 14th March 2022

1. Conduct initial background research on RF authentication, biometric measurement, and channel state information algorithms. Identify suitable human biometrics for measurement in authentication.
2. Review existing research into human measurement using channel state information.
3. Review available and suitable firmware that meets the requirements of the project.
4. Conceptualize and design a biometric measurement system, including suitable selection of hardware, software pseudo-code planning, and positioning of sensors. Construct the system with commercial off-the-shelf hardware.
5. Configure the system for extraction of channel state information from Wi-Fi subcarriers. Make preliminary observations of the way that human bodies alter the channel state information.
6. Develop an algorithm to target the selected human biometrics using a filter based approach to the extracted channel state information.
7. Develop an algorithm to identify the correlation of measured signal data and stored data pertaining to various different users. Use this correlation to make a judgement as to the identity of the person; i.e., authenticate the user.
8. Process and evaluate experimental data, especially as compared to an ideal, simulated permutation of the channel state information authentication.

If time and resource permit:

9. Refine the authentication algorithm, especially aiming for greater speeds in authentication as well as capacity for a greater number of stored authenticated users.

Appendix B

Risk Assessment

UNIVERSITY OF SOUTHERN QUEENSLAND

University of Southern Queensland **Offline Version**

USQ Safety Risk Management System

Note: This is the offline version of the Safety Risk Management System (SRMS) Risk Management Plan (RMP) and is only to be used for planning and drafting sessions, and when working in remote areas or on field activities. It must be transferred to the online SRMS at the first opportunity.

Safety Risk Management Plan – Offline Version			
Assessment Title:	ENG4111/2 Thesis Project	Assessment Date:	1/06/2022
Workplace (Division/Faculty/Section):		Review Date:(5 Years Max)	Click here to enter a date.
Context			
Description:			
What is the task/event/purchase/project/procedure?	Cardiac Biometric Authenticaiton using Channel State Information		
Why is it being conducted?	To fulfil honours project requirement of Bachelor of Engineering (Honours)		
Where is it being conducted?	Student place of residence		
Course code (if applicable)	ENG4111	Chemical name (if applicable)	
What other nominal conditions?			
Personnel involved	Student and volunteer beta testers		
Equipment	Wireless router, computer		
Environment	Home office		
Other			
Briefly explain the procedure/process	Stand in between router and computer and have ECG measured with wifi signals		
Assessment Team - who is conducting the assessment?			
Assessor(s)	Matthew Armanasco		
Others consulted:	Dr John Leis		

Eg 1. Enter Consequence

Eg 2. Enter Probability

Eg 3. Find Action

	Consequence				
Probability	Insignificant No Injury 0-$5K	Minor First Aid $5K-$50K	Moderate Med Treatment $50K-$100K	Major Serious Injuries $100K-$250K	Catastrophic Death More than $250K
Almost Certain 1 in 2	M	H	E	E	E
Likely 1 in 100	M	H	H	E	E
Possible 1 in 1000	L	M	H	H	H
Unlikely 1 in 10 000	L	L	M	M	M
Rare 1 in 1 000 000	L	L	L	L	L
Recommended Action Guide					
E=Extreme Risk – Task *MUST NOT* proceed					
H=High Risk – Special Procedures Required (See USQSafe)					
M=Moderate Risk – Risk Management Plan/Work Method Statement Required					
L=Low Risk – Use Routine Procedures					

Step 1 (cont)	Step 2	Step 2a	Step 2b	Step 3			Step 4				
Hazards: From step 1 or more if identified	*The Risk:* What can happen if exposed to the hazard without existing controls in place?	*Consequence:* What is the harm that can be caused by the hazard without existing controls in place?	*Existing Controls:* What are the existing controls that are already in place?	*Risk Assessment:* Consequence x Probab lity = Risk Level			*Additional controls:* Enter additional controls if required to reduce the risk level	*Risk assessment with additional controls:*			
				Probability	Risk Level	ALARP? Yes/no		Consequence	Probability	Risk Level	ALARP? Yes/no
Example											
Working in temperatures over 35º C	Heat stress/heat stroke/exhaustion leading to serious personal injury/death	catastrophic	Regular breaks chilled water ava lable loose clothing fatigue management policy.	poss ble	high	No	temporary shade shelters essential tasks only close supervision buddy system	catastrophic	unlikely	mod	Yes
Working near computers /devices that may catch fire	Destory all progress thus far, cause serious burns, injury	Major	Ventilated and cool room, ensure computers are free from dust	Rare	Low	Yes		Select a consequence	Select a probability	Select a Risk Level	Yes or No
No network security	Sensitive ECG/health data could be accessed by malicious person. Network is vulnerable to online attacks.	Moderate	Disconnect network from internet. Only have security off when required for testing. Store only non-fiducial ECG characteristics.	Unlikely	Moderate	No	Encrypt ECG data.	Major	Rare	Low	Yes
ECG informatio n cannot be extracted with CSI	No significant results in time for project completion.	Insignificant	Aim for increasingly superficial measurements if no results as time goes on (e.g. aim to just observe periodicity of ECG at least)	Poss ble	High	Yes		Select a consequence	Select a probability	Select a Risk Level	Yes or No
		Select a consequence		Select a probabi ity	Select a Risk Level	Yes or No		Select a consequence	Select a probability	Select a Risk Level	Yes or No
		Select a consequence		Select a probabi ity	Select a Risk Level	Yes or No		Select a consequence	Select a probability	Select a Risk Level	Yes or No
		Select a consequence		Select a probabi ity	Select a Risk Level	Yes or No		Select a consequence	Select a probability	Select a Risk Level	Yes or No
		Select a consequence		Select a probabi ity	Select a Risk Level	Yes or No		Select a consequence	Select a probability	Select a Risk Level	Yes or No
		Select a consequence		Select a probabi ity	Select a Risk Level	Yes or No		Select a consequence	Select a probability	Select a Risk Level	Yes or No
		Select a consequence		Select a probabi ity	Select a Risk Level	Yes or No		Select a consequence	Select a probability	Select a Risk Level	Yes or No

Step 1 (cont)	Step 2	Step 2a	Step 2b	Step 3			Step 4				
Hazards: From step 1 or more if identified	*The Risk:* What can happen if exposed to the hazard without existing controls in place?	*Consequence:* What is the harm that can be caused by the hazard without existing controls in place?	*Existing Controls:* What are the existing controls that are already in place?	*Risk Assessment:* Consequence x Probab lity = Risk Level			*Additional controls:* Enter additional controls if required to reduce the risk level	*Risk assessment with additional controls:*			
				Probability	Risk Level	ALARP? Yes/no		Consequence	Probability	Risk Level	ALARP? Yes/no
Example											
Working in temperatures over 35º C	Heat stress/heat stroke/exhaustion leading to serious personal injury/death	catastrophic	Regular breaks chilled water ava lable loose clothing fatigue management policy.	poss ble	high	No	temporary shade shelters essential tasks only close supervision buddy system	catastrophic	unlikely	mod	Yes
		Select a consequence		Select a probabi ity	Select a Risk Level	Yes or No		Select a consequence	Select a probability	Select a Risk Level	Yes or No
		Select a consequence		Select a probabi ity	Select a Risk Level	Yes or No		Select a consequence	Select a probability	Select a Risk Level	Yes or No
		Select a consequence		Select a probabi ity	Select a Risk Level	Yes or No		Select a consequence	Select a probability	Select a Risk Level	Yes or No
		Select a consequence		Select a probabi ity	Select a Risk Level	Yes or No		Select a consequence	Select a probability	Select a Risk Level	Yes or No
				Select a probabi ity	Select a Risk Level	Yes or No		Select a consequence	Select a probability	Select a Risk Level	Yes or No
				Select a probabi ity	Select a Risk Level	Yes or No		Select a consequence	Select a probability	Select a Risk Level	Yes or No
				Select a probabi ity	Select a Risk Level	Yes or No		Select a consequence	Select a probability	Select a Risk Level	Yes or No
				Select a probabi ity	Select a Risk Level	Yes or No		Select a consequence	Select a probability	Select a Risk Level	Yes or No
				Select a probabi ity	Select a Risk Level	Yes or No		Select a consequence	Select a probability	Select a Risk Level	Yes or No

Step 5 - Action Plan (for controls not already in place)			
Additional controls:	*Resources:*	*Persons responsible:*	*Proposed implementation date:*
			Click here to enter a date.
			Click here to enter a date.
			Click here to enter a date.
			Click here to enter a date.
			Click here to enter a date.
			Click here to enter a date.
			Click here to enter a date.
			Click here to enter a date.
			Click here to enter a date.
			Click here to enter a date.
			Click here to enter a date.
			Click here to enter a date.

Step 6 - Approval					
Drafter s name:	Matthew Armanasco			Draft date:	29/05/2022
Drafter s comments:					
Approver s name:	Dr John Leis	Approver s title/position:	Professor/supervisor/examiner		
Approver s comments:					
I am satisfied that the risks are as low as reasonably practicable and that the resources required will be provided.					
Approver s signature:	Dr John Leis			Approval date:	7/6/2022

Appendix C

Linux Environment Configuration

C.1 Installation of Tool

The following commands are used to install the CSI Linux tool.

```
sudo apt-get install gcc make linux-headers-$(uname -r) git-
core

sudo apt-get install iw
echo iface wlan0 inet manual | sudo tee -a
/etc/network/interfaces
sudo restart network-manager

echo blacklist iwldvm | sudo tee -a
/etc/modprobe.d/csitool.conf
echo blacklist iwlwifi | sudo tee -a
/etc/modprobe.d/csitool.conf
```

Next the CSI Tool Linux source tree is obtained, which includes modifications to the system's wireless driver. Then the appropriate tag in the repository for the specific upstream kernel version (3.13.0-24-generic) is checked out.

```
CSITOOL_KERNEL_TAG=csitool-$(uname -r | cut -d . -f 1-2)
git clone https://github.com/dhalperi/linux-80211n-csitool.git
cd linux-80211n-csitool
git checkout ${CSITOOL_KERNEL_TAG}
```

C.2 Configuration of Linux

The following commands are used to configure the Linux kernel to work properly with the CSI tool.

```
make -C linux-80211n-csitool-supplementary/netlink
sudo modprobe -r iwlwifi mac80211
sudo modprobe iwlwifi connector_log=0x1
sudo ifconfig wlan0 up
sudo ifconfig eth0 up
sudo iwconfig wlan0 essid TP-LINK_DD9C //link your router name
sudo ls
/sys/kernel/debug/ieee80211/phy0/netdev:wlan0/stations/
sudo cat
/sys/kernel/debug/ieee80211/phy0/netdev:wlan0/stations/bc:46:9
9:d6:dd:9c/rate_scale_table
sta_id 0
echo 0x1c90e | sudo tee
/sys/kernel/debug/ieee80211/phy0/netdev:wlan0/stations/bc:46:9
```

```
9:d6:dd:9c/rate_scale_table
sudo iwconfig wlan0 essid TP-LINK_DD9C
```

C.3 Initialisation after boot

The following commands are ran each time the system is booted before beginning CSI collection.

```
sudo modprobe wlan0 up
sudo modprobe eth0 up
sudo iw dev wlan0 connect {MY SSID}
sudo ls
/sys/kernel/debug/ieee80211/phy0/netdev:wlan0/stations/
sudo dhclient wlan0
sudo linux-80211n-csitool-supplementary/netlink/log_to_file
csi.dat

// open a separate terminal window

pgrep log_to_file
sudo renice -19 -p pid(pid of log_to_file process)

ping -i 0.02 -l 500 192.168.1.1
```

Appendix D

MATLAB Scripts and Measurements

D.1 Processing Script

The following code is used to process the CSI data after it is captured in the Linux machine. The code has been written from scratch apart from proprietary functions "read_bf_file" and "get_scaled_csi" provided by Halperin et al. (2010, pp. 159-170).

```
clear all
close all

% read in desired test file
filename = '500_MB_02_3.dat'

% unpack binary CSI format with proprietary function read_bf_file made by Halperin
et al. (2010, pp. 159-170)
csi_trace = read_bf_file(filename);
length = length(csi_trace);
m = ceil(sqrt(length));
n = m;
num_subc = 30;

csi_array = zeros(2,3,30,29);
for i = 1:1:length
    % iterate over each entry of the CSI data (i.e. each packet)
    csi_entry = csi_trace{i};
    % scale CSI data to absolute dBm rather than reference with proprietary
function read_bf_file made by Halperin et al. (2010, pp. 159-170)
    csi = get_scaled_csi(csi_entry);
    for j=1:1:size(csi,1)
        csi_dim(j,:,:) = csi(j,:,:);
    end

    % squeeze to eliminate single dimension
    d = db(abs(squeeze(csi_dim(1,:,:)).'));

    if size(csi,1) == 2
        csi_array(:,:,:,i) = csi;
        d = db(abs(squeeze(csi_dim(2,:,:)).'));
    elseif size(csi,1) == 1
        csi_array(1,:,:,i) = csi; % this skips the 2nd TX when only 1x3x30
    end

    test_1(:,i) = d(:,1);
    test_2(:,i) = d(:,2);
    test_3(:,i) = d(:,3);
end


seq_1 = test_1;
seq_2 = test_2;


% save the processed CSI data for analysis in the authentication script

filename_mat = filename
filename_mat(end-2) = 'm'
save(filename_mat, 'seq_1', 'seq_2')
```

D.2 Authentication Script

The following code is used to analyse the processed CSI data and attempt to authenticate users. This is done with correlation before averaging over all reference tests, both antennas and all subcarriers. The overall percentage match of the test data to each user is displayed.

```
clear all
close all
clc
test_file = '500_MA_02_1.mat'
num_tests = 3;
num_subc = 30;
x = 1:300; % eliminate last 200 samples of test

% iterate over all the registered reference tests for each user
for k = 1:1:num_tests
    matfilename = sprintf('500_MA_02_%d.mat',k)

    % load CSI data for each antenna for each of the stored reference data
    % sets
    Var1_1 = 'seq_1';
    Var2_1 = 'seq_2';
    data_1_file1 = load(matfilename, Var1_1).(Var1_1);
    data_2_file1 = load(matfilename, Var2_1).(Var2_1);

    % load CSI data for each antenna for the test data set
    Var1_2 = 'seq_1';
    Var2_2 = 'seq_2';
    data_1_file2 = load(test_file, Var1_2).(Var1_2);
    data_2_file2 = load(test_file, Var2_2).(Var2_2);

    data_1 = data_1_file1;
    data_2 = data_2_file1;

    data_1 = data_1(:,x);
    data_2 = data_2(:,x);

    data_1_file1 = data_1_file1(:,x);
    data_1_file2 = data_1_file2(:,x);

    data_2_file1 = data_2_file1(:,x);
    data_2_file2 = data_2_file2(:,x);

    for i = 1:1:num_subc
        % correlate each subcarrier with each subcarrier. lags of 20 to account
        % for the heart beat pulses not matching in time exactly.
        [try_disA(i,:), try_lagsA(i,:)] = xcorr(data_2_file1(i,:),
data_2_file2(i,:),20,'normalized');
        [try_disB(i,:), try_lagsB(i,:)] = xcorr(data_1_file1(i,:),
data_1_file2(i,:),20,'normalized');
        % find the peak correlation value and store in an array iterated
        % over all tests.
        [peakcorrA(i), peakcorr_locA(i)] = max(try_disA(i,:));
        [peakcorrB(i), peakcorr_locB(i)] = max(try_disB(i,:));
    end

    % take the percentage match for each antenna to be the average of the
    % peak correlation values
    percent_match_A = mean(peakcorrA,'omitnan')
    percent_match_B = mean(peakcorrB,'omitnan')

    % take the overall match of the test data against this user to be the
    % average of the two antennas
```

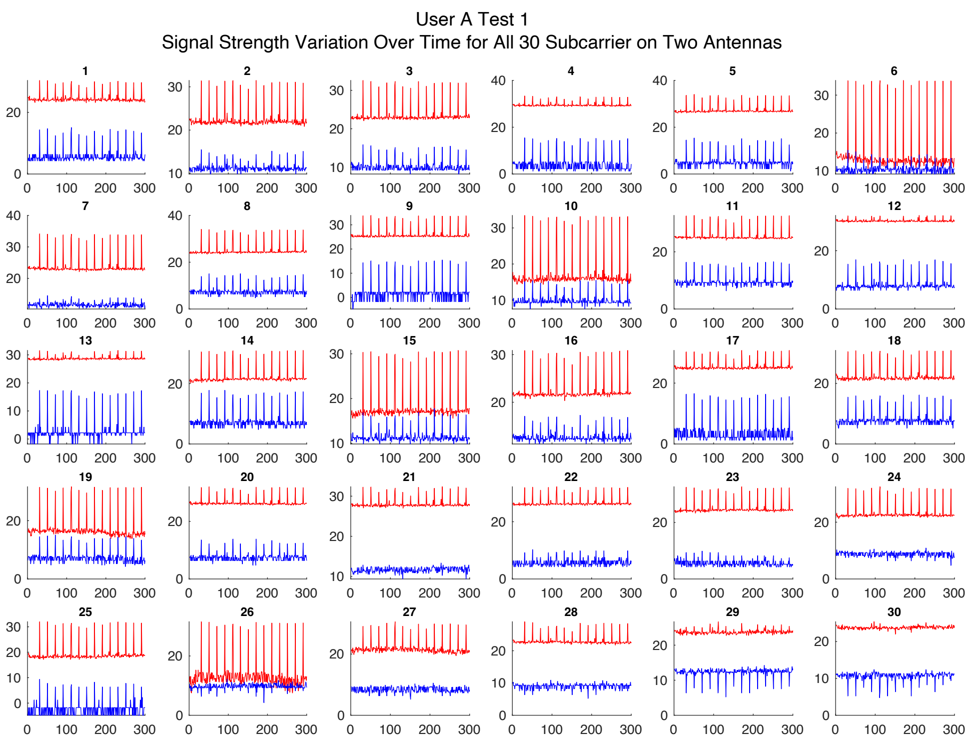
```
    match_arrayMA(k) = mean([percent_match_A, percent_match_B])*100;
end

% iterate over all the registered reference tests
for k = 1:1:num_tests
    matfilename = sprintf('500_LB_02_%d.mat',k)

    % load CSI data for each antenna for each of the stored reference data
    % sets
    Var1_1 = 'seq_1';
    Var2_1 = 'seq_2';
    data_1_file1 = load(matfilename, Var1_1).(Var1_1);
    data_2_file1 = load(matfilename, Var2_1).(Var2_1);

    % load CSI data for each antenna for the test data set
    Var1_2 = 'seq_1';
    Var2_2 = 'seq_2';
    data_1_file2 = load(test_file, Var1_2).(Var1_2);
    data_2_file2 = load(test_file, Var2_2).(Var2_2);

    data_1 = data_1_file1;
    data_2 = data_2_file1;

    data_1 = data_1(:,x);
    data_2 = data_2(:,x);

    data_1_file1 = data_1_file1(:,x);
    data_1_file2 = data_1_file2(:,x);

    data_2_file1 = data_2_file1(:,x);
    data_2_file2 = data_2_file2(:,x);

    for i = 1:1:num_subc
        % correlate each subcarrier with each subcarrier. lags of 20 to account
        % for the heart beat pulses not matching in time exactly.
        [try_disA(i,:), try_lagsA(i,:)] = xcorr(data_2_file1(i,:),
data_2_file2(i,:),20,'normalized');
        [try_disB(i,:), try_lagsB(i,:)] = xcorr(data_1_file1(i,:),
data_1_file2(i,:),20,'normalized');
        % find the peak correlation value and store in an array iterated
        % over all tests.
        [peakcorrA(i), peakcorr_locA(i)] = max(try_disA(i,:));
        [peakcorrB(i), peakcorr_locB(i)] = max(try_disB(i,:));
    end
    % take the percentage match for each antenna to be the average of the
    % peak correlation values

    percent_match_A = mean(peakcorrA,'omitnan')
    percent_match_B = mean(peakcorrB,'omitnan')

    % take the overall match of the test data against this user to be the
    % average of the two antennas
    match_arrayLB(k) = mean([percent_match_A, percent_match_B])*100;
end

% iterate over all the registered reference tests
for k = 1:1:num_tests
    matfilename = sprintf('500_MB_02_%d.mat',k)

    % load CSI data for each antenna for each of the stored reference data
    % sets
    Var1_1 = 'seq_1';
    Var2_1 = 'seq_2';
    data_1_file1 = load(matfilename, Var1_1).(Var1_1);
    data_2_file1 = load(matfilename, Var2_1).(Var2_1);

    % load CSI data for each antenna for the test data set
    Var1_2 = 'seq_1';
```

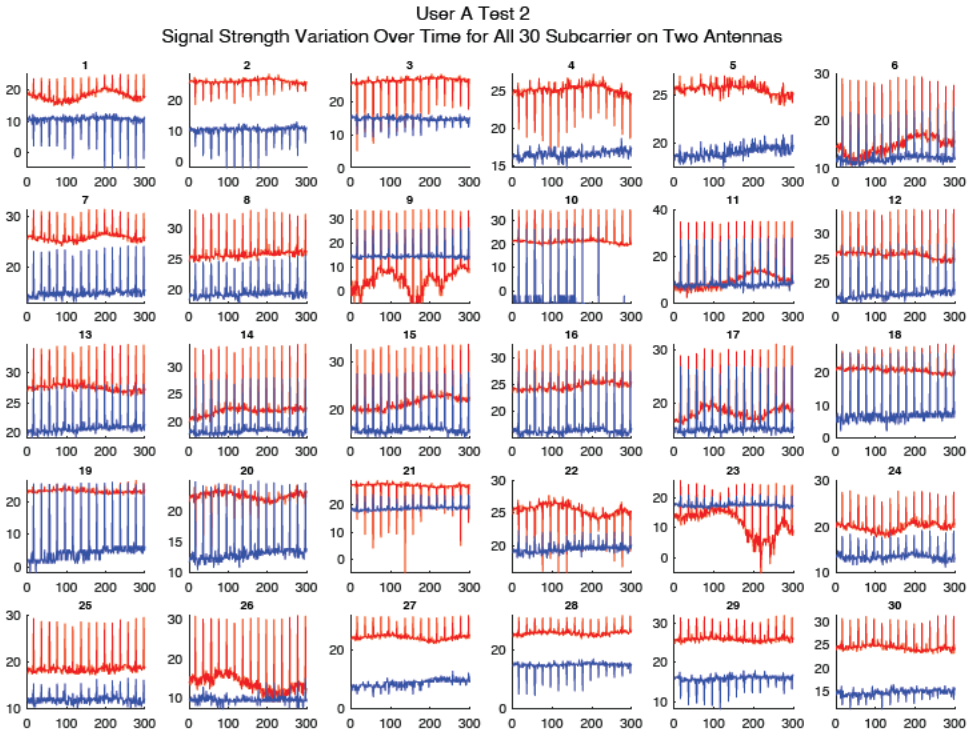
```
    Var2_2 = 'seq_2';
    data_1_file2 = load(test_file, Var1_2).(Var1_2);
    data_2_file2 = load(test_file, Var2_2).(Var2_2);

    data_1 = data_1_file1;
    data_2 = data_2_file1;

    data_1 = data_1(:,x);
    data_2 = data_2(:,x);

    data_1_file1 = data_1_file1(:,x);
    data_1_file2 = data_1_file2(:,x);

    data_2_file1 = data_2_file1(:,x);
    data_2_file2 = data_2_file2(:,x);

    for i = 1:1:num_subc
        % correlate each subcarrier with each subcarrier. lags of 20 to account
        % for the heart beat pulses not matching in time exactly.
        [try_disA(i,:), try_lagsA(i,:)] = xcorr(data_2_file1(i,:),
data_2_file2(i,:),20,'normalized');
        [try_disB(i,:), try_lagsB(i,:)] = xcorr(data_1_file1(i,:),
data_1_file2(i,:),20,'normalized');
        % find the peak correlation value and store in an array iterated
        % over all tests.
        [peakcorrA(i), peakcorr_locA(i)] = max(try_disA(i,:));
        [peakcorrB(i), peakcorr_locB(i)] = max(try_disB(i,:));
    end

    % take the percentage match for each antenna to be the average of the
    % peak correlation values
    percent_match_A = mean(peakcorrA,'omitnan')
    percent_match_B = mean(peakcorrB,'omitnan')

    % take the overall match of the test data against this user to be the
    % average of the two antennas
    match_arrayMB(k) = mean([percent_match_A, percent_match_B])*100
end

% note that in future work it would be possible to shorten this script and
% include the iteration code only once, whilst iterating over all the users
% in the database. this approach was taken in this project for ease of
% implementatin regarding the file iteration.

figure
for i = 1:1:num_subc
    subplot(5,6,i)
    hold on
    plot(data_2_file2(i,:),'Color','red')
    plot(data_1_file2(i,:),'Color','blue')
    title(i,fontsize=9)
end

sgtitle('Figure 2: Signal Strength Variation Over Time for All 30 Subcarrier on Two
Antennas',FontSize=13)


% eliminate perfect matches, and find the overall match of the test sample
% against all the users.
match_arrayLB(find(match_arrayLB==100)) = NaN;
match_arrayMA(find(match_arrayMA==100)) = NaN;
match_arrayMB(find(match_arrayMB==100)) = NaN;
percent_match_MA = mean(match_arrayMA,'omitnan')
percent_match_LB = mean(match_arrayLB,'omitnan')
percent_match_MB = mean(match_arrayMB,'omitnan')

% display the overall matches.
```

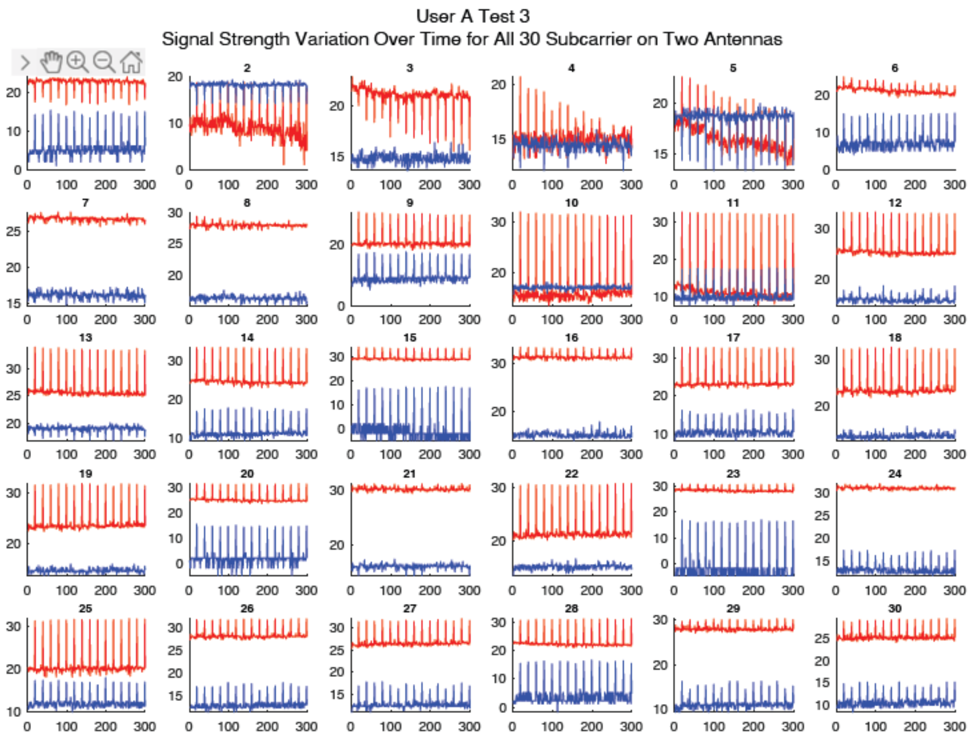
```
overall_match_str = sprintf('MA = %.2f  LB = %.2f  MB = %.2f',percent_match_MA,
percent_match_LB, percent_match_MB)

figure
plot(data_1_file2(:,1:end/10))
hold on
plot(data_2_file2(:,1:end/10))
xlabel('OFDM Subcarrier Number',FontSize=12)
ylabel('Signal Strength (dBm)',FontSize=12)
title('Figure 1: Signal Strength Variation over Different Subcarriers for 30
Signals on Two Antennas',FontSize=13)
grid minor

figure
plot(data_1_file2(1,:),'Color','blue')
hold on
plot(data_2_file2(1,:),'Color','red')
xlabel('Sample Index',FontSize=12)
ylabel('Signal Strength(dBm)',FontSize=12)
title('Figure 3: Signal Strength Variation Over Time for One Subcarrier on Two
Antennas',FontSize=13)
grid minor
```

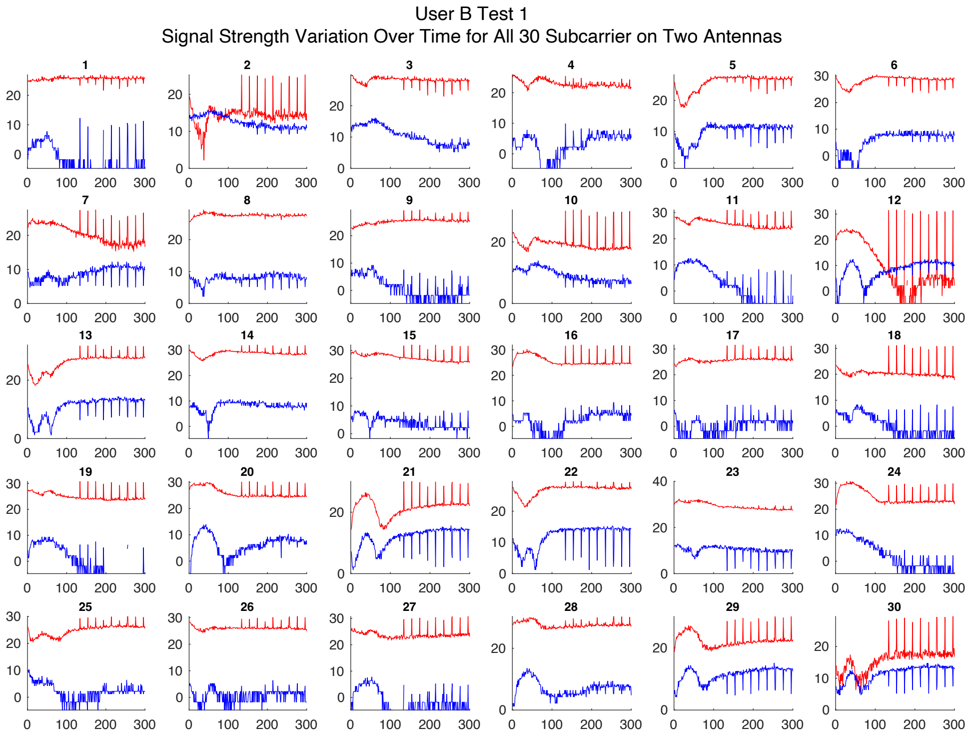
D.3 Output Heart Rhythm Measurements

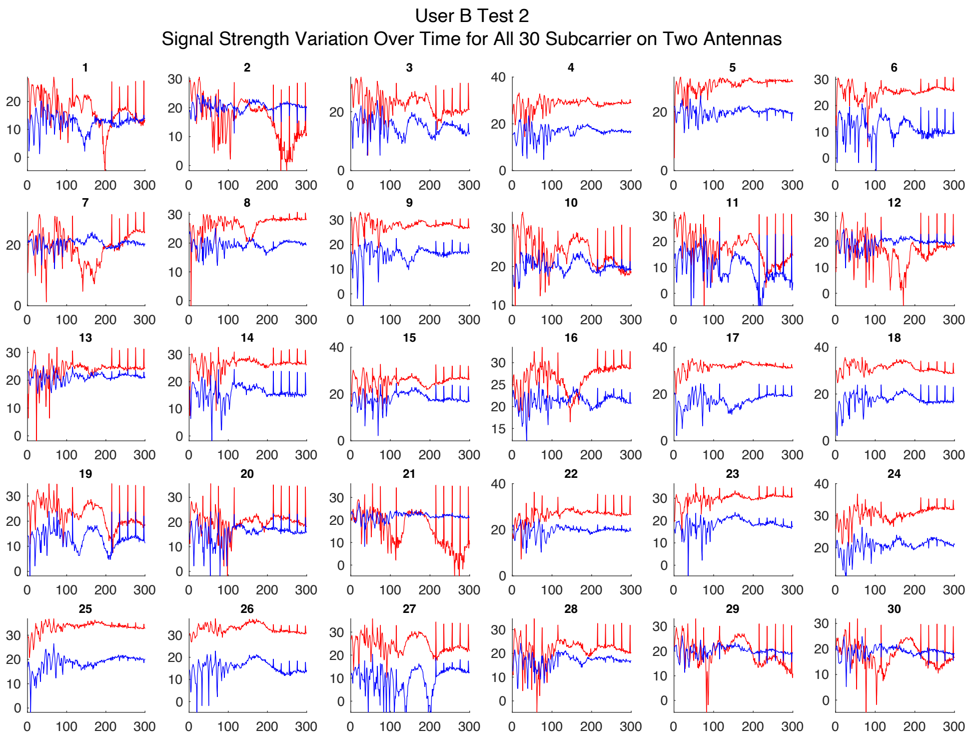
This appendix shows all of the measured heart rhythms from each of the tests done on each of the users.

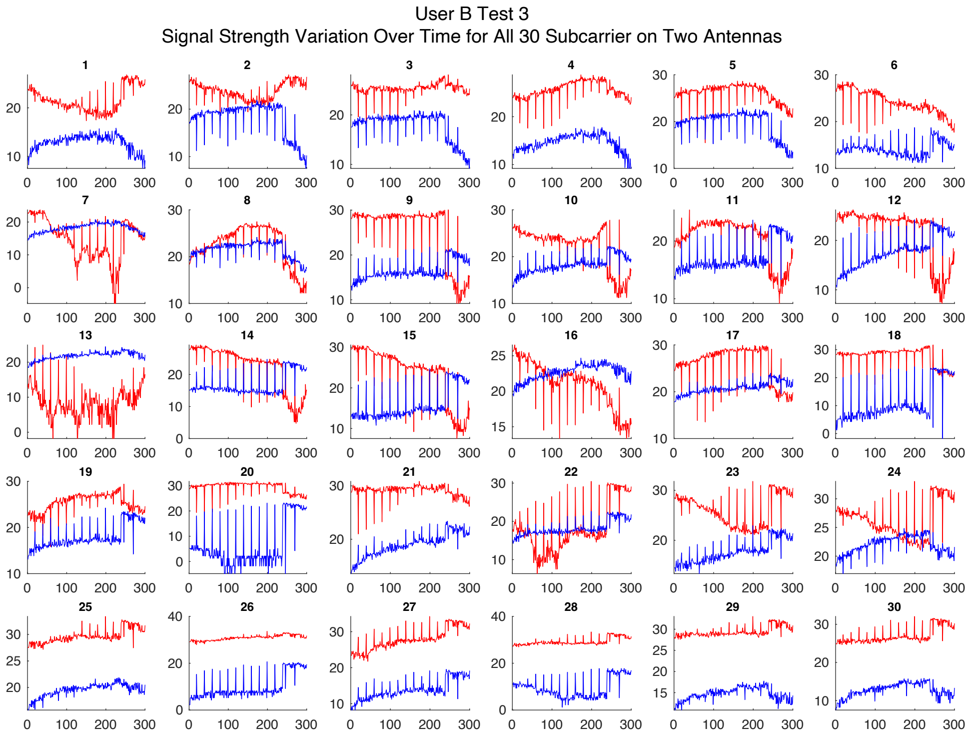


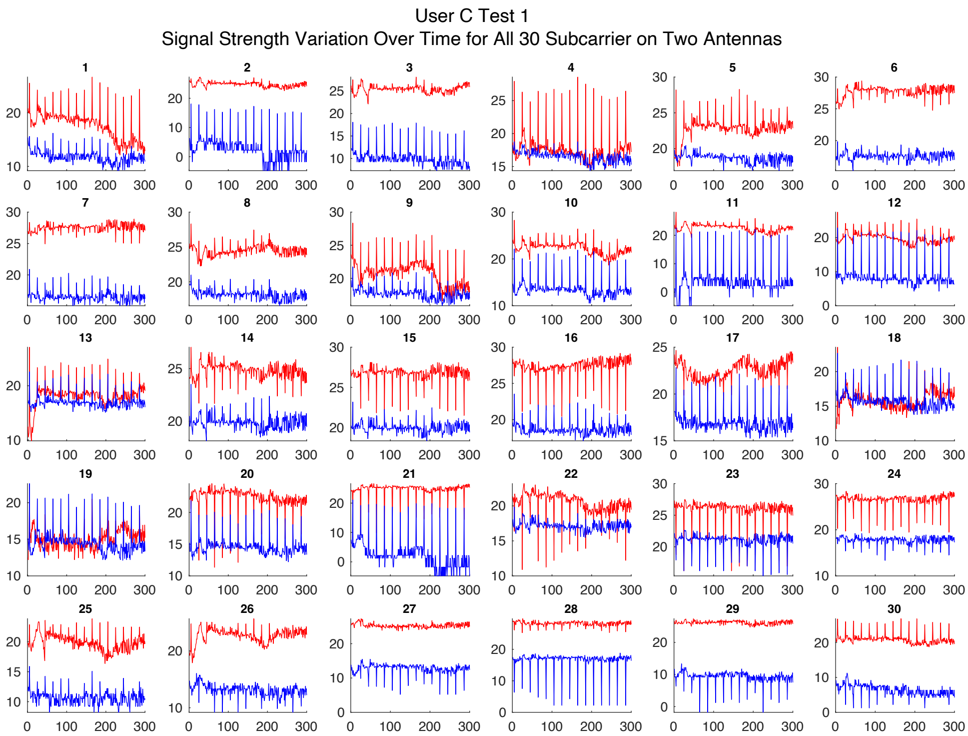

User A Test 2
Signal Strength Variation Over Time for All 30 Subcarrier on Two Antennas

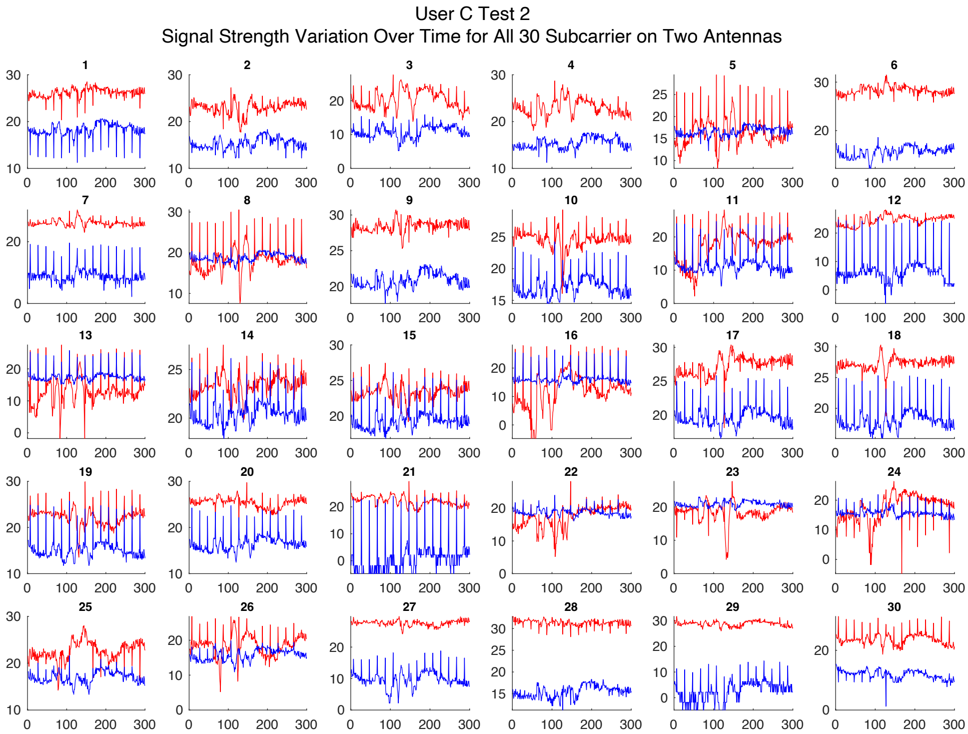

User A Test 3
Signal Strength Variation Over Time for All 30 Subcarrier on Two Antennas

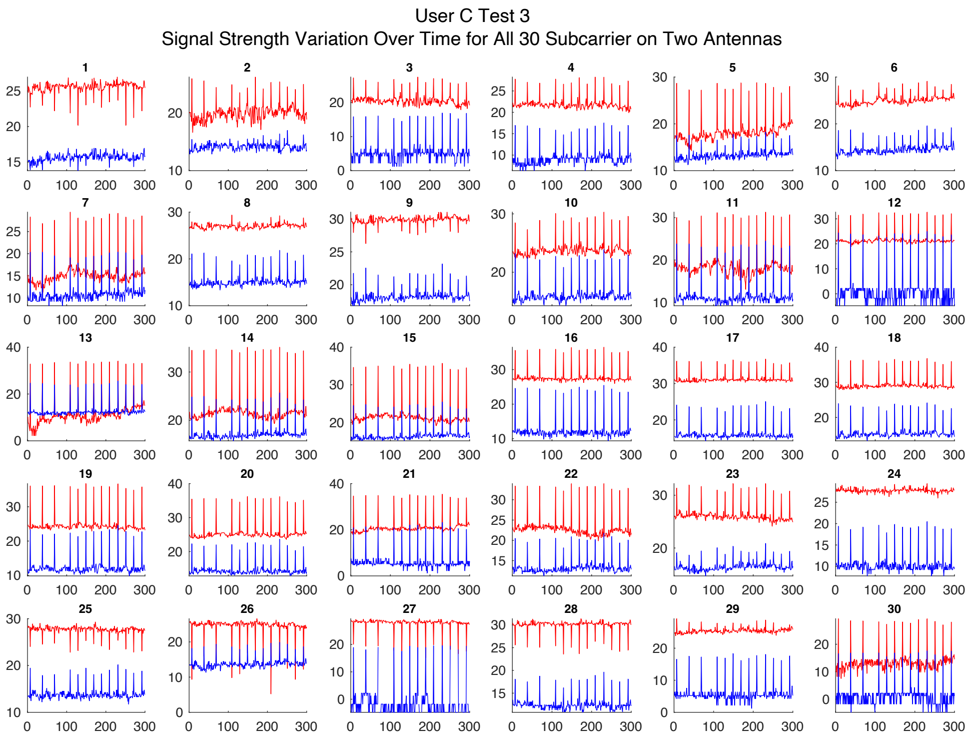

User B Test 1
Signal Strength Variation Over Time for All 30 Subcarrier on Two Antennas


User B Test 2
Signal Strength Variation Over Time for All 30 Subcarrier on Two Antennas


User B Test 3
Signal Strength Variation Over Time for All 30 Subcarrier on Two Antennas
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
0 100 200 300


User C Test 1
Signal Strength Variation Over Time for All 30 Subcarrier on Two Antennas
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
0 100 200 300


User C Test 2
Signal Strength Variation Over Time for All 30 Subcarrier on Two Antennas


User C Test 3
Signal Strength Variation Over Time for All 30 Subcarrier on Two Antennas