Using the Human Gait for Authentication

Torkjel Søndrol

Master's Thesis
Master of Science in Information Security
30 ECTS
Department of Computer Science and Media Technology
Gjøvik University College, 2005

The MSc programme in Information Security
is run in cooperation with the Royal Institute
of Technology (KTH) in Stockholm.

Institutt for
informatikk og medieteknikk
Høgskolen i Gjøvik
Postboks 191
2802 Gjøvik

Department of Computer Science
and Media Technology
Gjøvik University College
Box 191
N-2802 Gjøvik
Norway

Abstract

This thesis presents a new method for verifying a person's identity using kinetic gait analysis. The gait data is collected using a device that can be attached to a person's leg, where it detects the leg's movement in horizontal, vertical and sideway direction as the person walks. These data are used in an attempt to authenticating the walking person.

This thesis presents how the gait collection device was designed and the software used for collecting the gait data. It also proposes several different methods for analysing the gait data in such a way that given two sets of data, it is to some extent possible to determine whether these datasets came from the same person. Three experiments has been conducted to determine the strength of these methods of analysis under different contexts. These experiments shows that though none of the methods described in this thesis can compete with methods like fingerprint recognition and face recognition, it is a potential in using the human movements for identity verification.

Sammendrag

Denne rapporten presenterer en ny metode for å verifisere en persons identitet ved hjelp av kinetisk ganglagsanalyse. Ganglagsdata samles inn ved å bruke en innretning festet til personens bein, hvor den registrerer beinets bevegelser i horisontal, vertikal og sidelengs retning mens personen går. Disse dataene benyttes så i et forsøk på å autentisere personen som går.

Denne rapporten forteller hvordan innretningen som registrerer ganglagsdata ble bygget og beskriver programvaren som benyttes for å samle inn ganglagsdata. Rapporten foreslår også flere ulike metoder for å analysere dataene på en måte slik at man gitt to sett av ganglagsdata med en viss grad av nøyaktighet kan si hvorvidt dataene kom fra samme person. Tre eksperimenter ble utført for å undersøke styrken til analysemetodene under ulike settinger. Eksperimentene viser at selv om metodene ikke kan konkurrere med teknologi som fingeravtrykk- og ansiktsgjenkjenning, er det et potensiale i bruken av personers bevegelser for å verifisere deres identitet.

Preface

"Highest Queen of State, Great Juno, comes; I know her by her gait."—Shakespeare.

The thesis you are about to read marks the end of my two years of study to become a master in information security. Throughout these two years, I have learnt an overwhelming amount of new and valuable information regarding information security and computers in general. One of the topics that caught my interest early on was the field of biometric authentication, and I realised that this would be the area where I should devote my attention during my master thesis research. I was therefore extremely happy when my supervisor presented me his idea for an authentication scheme based on the person's gait. Though a lot of new information is presented in this thesis, and it might be difficult to get a tight grip on at first, I hope you don't get lost reading it for the first time. If this happens, put it away for some time. Take a walk outside, get some fresh air and and smell the roses. Then try to read it again. I hope you find the field of gait based authentication as fascinating as I have, and that this will inspire you into thinking in new ways regarding solutions for authentication.

Several persons have been very helpful making this project possible. I would first of all like to thank my supervisor, professor Einar Snekkenes, for the idea for this project and for excellent help with what at first looked like an overwhelming and unsolvable problem. I would also like to thank the college's electrical engineering lab for lending the necessary equipment to create the Gait Collector. Of course I would also like to thank everyone participating in my experiments, even though it was "very embarrassing" walking around with a large box attached to their leg. During the process of analysing the results, I would also like to thank Frode Volden who performed a serie of statistical analysis that I had no knowledge of when this master project started. Several persons volunteered to give feedback on the thesis you are about to read as it was being written, which helped a lot getting it as understandable as possible to the public. I also owe these a thanks. Last, but not least, I would like to thank my great friends Ole Kasper Olsen, Anders Wiehe, Fredrik Skarderud and Ole Martin Dahl for their accompanying, eager discussions, and for morally supporting me during the whole master process.

Torkjel Søndrol, 30th June 2005

Contents

List of Figures

List of Tables

1 Introduction

1.1 Topic Covered by this Thesis

Confirming the identity of a person based on the way he walks is a fairly new area in computer science compared with more traditional biometric methods, like fingerprint and face recognition. Though most research in this area today concentrates on using video footage to detect different kinematic gait features[1], the field of wearable computing introduces the possibility of encapsulating motion capturing technology onto a person. This makes it possible to authenticate the person in a kinetic fashion[2] as he walks without requiring the attention from the person himself.

Keywords: biometric authentication, gait, data analysis, motion capturing, FAR/FRR calculations, ROC curves, Fast Fourier Transforms.

1.2 Problem Description and Motivation

There exists several different ways of identifying persons based on their biometric characteristics, such as their fingerprint, face and voice. Though this might be sufficient regarding the level of security in many situations, they require an actively participating user when it comes to the process of retrieving the biometric feature. Some persons might have hygienic issues related to placing their finger on a reader where thousands of other unknown have placed their fingers in advance, or it might be inconvenient in other ways. Persons who know they are being registered when performing a certain task, like giving away their signature or a voice print might perform different than if they did the same task in a more normal situation. Such problems might be avoided by continuously registering the person's movements (e.g. their walking feature), and automatically performing the verification process when it is necessary. A kinetic gait recognition technology that is attached to the person, like the one presented in this thesis, protects the privacy of the walking persons better than more kinematic methods to gait analysis, like gait analysis of surveillance video. It is therefore necessary to develop a method that automatically registers a person's movements as well as algorithms for analysing these movements to detect whether two given datasets belong to the same person. Such a method for authentication will create a more comfortable authentication process for the user.

1.3 Research Questions

This research concerns whether or not gait might be used as a method for authentication using motion capturing technology. The process of getting this answer involve answering

[1] The study of the geometry of a person's movement. Often done by analysing video footage [8].
[2] The study of the forces involved in a person's movement. Often by analysing acceleration data [8].

a few basic questions. These are:

- Is it possible to retrieve reliable and valid kinetic gait data?
- Given a set of gait data — is it possible to analyse these data and tell whether or not they belong to the same person?
- Is it possible to use the human gait to verify a person's identity?

1.4 Claimed Contributions

The contribution from this research is new knowledge about methods and algorithms for authenticating persons based on kinetic gait data as well as an assessment of their performance. The methods proposed in this report can verify a person's identity based on acceleration data. They have been tested through several experiments. An improved version of the prototype presented in this thesis might in the future be incorporated into the Windows XP login manager or similar authentication schemes to make it possible to log onto a computer using the way you walk. It might also be implemented as an access control mechanism to restricted areas.

1.5 Method

The project involved literature study, data collection, development of methods for automatic data analysis and an analysis of these methods. It was therefore necessary to select a mixed-method approach [14] as described below:

- Literature study – qualitative.
- Collecting data – quantitative.
- Algorithm development – qualitative.
- Analysing data – quantitative.
- Algorithm analysis – quantitative.

The literature study was necessary to get a better understanding of the previous work in the field of gait analysis and of motion detection technology. Through a qualitative assessment of which previously used methods where the most usable for this study, a set of new gait analysis methods were developed. These methods analyses the obtained gait data in a quantitative manner. A quantitative assessment was performed based on the results from these methods to determine which ones worked best in the settings of the experiments. This helped evaluate the strengths and weaknesses of the proposed algorithms.

1.6 Outline of the Report

This report begins with a summary of previously performed work in the area for gait analysis and motion capturing in Chapter 2. Chapter 3 is a short introduction to the theory behind authentication and the human gait itself. Chapter 4 describes the technology behind the gait collection device which was created, while Chapter 5 describes how this device was used for gathering gait data. Chapter 6 describes several potential methods for data analysis. Some of them were used, some were not. How the data analysis was

performed and which methods were used are described in Chapter 7. Then it is ready to get into the experiments that were performed. Chapter 8 is a description of the experiments that was performed and the rules they were following. Chapter 9 describes the experiments themselves while Chapter 10 summaries them. Potential further work is described in Chapter 11, while Chapter 12 concludes the project.

Four appendixes are included as well. Appendix A presents the software embedded into the gait collection device that was created, while Appendix B contain the Matlab code for one of the analysis methods described. Appendixes C and D shows the results from a series of statistical tests applied to the results from two of the performed experiments.

2 Related Work

This project have analysed the kinetic movements in the human gait in an attempt of verifying a person's identity. Since it was necessary to develop new algorithms for this task, a study of other projects performing analysis of human movements—and gait in particular—is in place. This chapter will describe some of the related work in the area of gait analysis and motion capturing technologies. The literature described in this chapter creates a baseline for the later described analysis methods.

2.1 Work Related to Gait Analysis

Lee and Grimson presented a study on how gait might be used to identify and classify persons [39, 40]. Their work was based on video capture, where they tried to fit several ellipses around various body parts as the person walked across the screen. They computed the mean and the standard deviation of how these ellipses changed over time, as well as the magnitudes and phases of each region. They could then apply the Fast Fourier Transformation (FFT) to detect the frequency of changes over time. They further used the Mahalanobis distance and the p-value calculated using the Analysis of Variance (ANOVA) method to rank features into various categories before further analysis was performed. To detect the preformance of their gait recongition, they used a cumulative matching score. They were able to show that different types of clothing affected the person's gait.

The use of accelerometers in gait analysis has not been performed to any large extent. Guta *et al* [23] connected three ADXL202 accelerometers from Analog Devices[1] to a DSP card from Texas Instruments[2]. This was used as a speedometer to detect the walking and running speed of the person who was wearing the device. They where able to determine the step length and whether the person was walking or running based merely on the step frequency derived from the acceleration data. After calibrating the device with certain data regarding the person's leg length, muscle strength and body mass they were also able to estimate the person's walking speed based on his step length. Ladetto *et al* [37, 38] were able to get a more accurate estimate of a person's step length using a combination of acceleration data and GPS[3] data.

Zhang *et al* at MiniSun[4] have created a device for measuring physical activity, such as posture, gait, limb movement and transitions [72]. The device consist of a micro computer which can be attached to a belt. This device is equipped with five small sensors that can be attached to the body using medical tape, and these sensors connects to the

[1] `http://www.analog.com`.
[2] `http://www.ti.com`.
[3] Global Positioning System.
[4] `http://www.minisun.com`.

micro computer through cables. The resulting data can be analysed in various ways to detect things like walking speed, distance and energy expenditure.

Morris, Benbasat and Paradiso at Massachusetts Institute of Technology developed a sensor system for clinical analysis that can fit into a shoe [10, 46, 57]. This is a system designed for continuous monitoring a person's gait which transmits data wireless in real time. Among other, their system consist of three gyroscopes and three accelerometers. Though they are performing clinical research, and therefore not working in the field of authentication, their architectural design is capable of detecting the distinctive motions a person make during locomotion. They are currently working on how to analyse the output from the sensors, as well as looking at different pattern recognition methods.

Neural network was not used in great extent for gait recognition until Su and Wu used the Genetic Algorithm Neural Network on ground reaction force (GRF) data[5] [66]. A genetic algorithm uses techniques such as inheritance, mutation, natural selection and recombination from natural systems to mimic survival-of-the-fittest organisms[6]. By using these techniques on a set of input data only allowing the "best fitted" datasets to survive, it was possible for Su and Wu to classify gait patterns with a success rate of up to 98.7%. Herrero-Jaraba *et al* [27] used another neural network known as a Self-Organising Map (SOM) [35, 36] to recognise persons in video footage. SOM was also used by Köhle and Merkl [34] to classify different gait patterns. They based their study on GRF data, as in [66], and calculated the FFT transform of this data as an input to the SOM algorithm. They had already shown in an earlier study [33] that it was possible to distinguish between healthy and pathological gait using supervised neural networks, and that such a map also gave a reasonable evaluation of the gait pattern itself. They chose to use the FFT, first of all since the amount of data received depended on the duration of each step, but also because it reduced the amount of data considerably, since they only needed to use the first 64 Fourier-coefficients from each dataset during the training process of the SOM.

Tanawongsuwan and Bobick has performed a research on how parameters such as stride length and cadence[7] varies at different gait speeds [67]. This is an interesting and important work due to the fact that most gait recognition methods requires constant and natural gait on level ground etc. to work properly, and often fail if these requirements are altered. There are also other works that look at how the gait changes during various walking speeds [32, 41]. The process of recognising persons in [67] was performed by a nearest neighbour algorithm with the Eucledian distance. This distance was also calculated by Mowbray and Nixon [48], who tried to model the periodic deformation of the gait. They used Fourier descriptors and the Eucledian distance to detect differences between different classes of gait.

Nadeau *et al* has performed a research on how climbing stairs differs from normal, level walking at healthy persons aged over 40 [49]. The result from this work, was the discovery that while the gait speed in average decreased from $1{,}16\frac{\mathrm{m}}{\mathrm{s}}$ to $0{,}46\frac{\mathrm{m}}{\mathrm{s}}$ and the cadence decreased from $105{,}4\frac{\mathrm{steps}}{\mathrm{min}}$ to $93{,}6\frac{\mathrm{steps}}{\mathrm{min}}$, the gait cycle period increased from 1145ms to 1304ms. A similar study was performed by Stacoff *et al* [64]. They looked at how stair walking changed with age.

[5] Pressure sensitive boards are mounted in the floor, and are capable of detecting the forces used when walking on them.

[6] Source: `http://www.wikipedia.org`.

[7] Harmony and proportion in motions, as of a well-managed horse [1913 Webster].

2.2 Work Related to Motion Capturing

Research in the field of context sensitive technology has grown rapidly. These technologies tries to determine in which context a particular item is being used (indoor, outdoor, while moving, etc). Such research often makes use of data accuired from accelerometers, and the chosen methods for analysis is therefore relevant for this research. A list of some of the work in the area of motion capturing is therefore presented.

Hinckley *et al* tested different methods for user interaction on a PDA [28]. Among other things, they fitted the PDA with a dual-axis accelerometer to detect how the user held the device. Calculating the magnitude of a Fast Fourier Transformation (FFT) in both directions at 10 Hz and a time window of 32 samples, they were able to detect walking.

Schmidt *et al* performed a similar experiment to Hinckley one year earlier, where they measured data from 8 different sensors to detect in which context they where used (laying on a table, held in hand, in a suitcase etc.) [60]. Among the sensors where two accelerometers. Due to the difficulties of analysing and mapping data from 8 different sensors, they started their analysis by clustered the data using the Self-Organizing Map [35, 36], which is suitable for noisy conditions. Their analysis process was performed off-line after the data was sent to a computer to detect the different contexts the device was used in. They also calculated the average, standard deviation, quartile distance, base frequency and first derivative of the data from the accelerometers.

The use of accelerometers and gyroscopes has been popular, both in context sensitive technology, but also in other fields of motion capturing. Welch and Foxlin have described different technologies for motion capturing, and describes the use of three accelerometers and three gyroscopes as [70]

> " [...] the closest thing to a silver bullet among all ammunition technologies [...]"

With this, they think that even though such a device does not exist today, combining three accelerometers and three gyroscopes on one single chip would be very close to an imaginative magical device with very strict regulations:

- It should be the size of a transistor.
- No other parts should be mounted in the environment or on the user.
- It should track all six degrees of freedom.
- It should have a resolution better than 1mm in position and 0.1 degree in resolution.
- It should run at 1,000Hz with a latency less than 1ms.
- It should not need a clear line of sight with its subject or anything else.
- It should not be affected by sound, light, heat or magnetic fields.
- It should track its target no matter how far or fast it goes.
- It should run without wires for three years on a coin sized battery.
- It should cost less than $1.

These regulations are created with positioning technology in mind, but most of them are relevant for other means of motion capturing. A device like the one described would indeed be a very suitable device for measuring a person's walking pattern, but until it is

available on the market, a similar device must be created manually. Verplaetse [68] recommend using accelerometers capable of measuring accelerations of $\pm 10g$ when measuring the foot's movements. He performed a test where a shoe fitted with an accelerometer was measuring the accelerations of normal gait, and achieved results ranging between 0.19 and 6.57g, where almost all recorded accelerations where located around the mean of 1.59g.

3 Theory

3.1 Biometric Authentication

Authentication from an information security point-of-view is the process of confirming the identity of a human being. The use of biometric features for authentication purposes is only one of three approaches. The other two are the use of hardware tokens like smart cards, keys and other items a user *has*, while the last is the use of passwords, pass-phrases and other secrets a user *knows*. The use of tokens and passwords has long been the most commonly used method for authentication, as the technology needed to measure the biometric features has been large and expensive. In recent years, though, this technology has become more compact and fallen dramatically in price. It has therefore been used more and more often as an alternative form of authentication, often in combination with passwords and tokens. Biometric authentication uses one of several different biometric features to determine a person's identity. Some of the most frequently used features today are fingerprints, face, voice and iris [55].

To authenticate a person, two actions are necessary. First, the user has to be *enrolled*. This requires the user to register his identity and the biometric feature he want to authenticate himself using. In a fingerprint system, this would be to register one or several of his fingerprints using a fingerprint reader. The system would then store a *template* of the fingerprint. This is a digital representation of the fingerprint, where its unique features have been identified and extracted.

When the enrollment has been performed, the user is known to the system. The next action would therefore be an authentication attempt. This could either been done through *identification* or *verification*. During a verification attempt the user will enter his identity along with his fingerprint. The system will then compare this fingerprint with the pre-stored template for this user; i.e. a one-to-one comparison. If an identification attempt is performed, the user will only enter his fingerprint, and the system will check this with all the fingerprints in the database; i.e. a one-to-many comparison. A graphical representation of the authentication process is shown in Figure 3.1.

The result from a comparison between two templates will yield in a *matching score*, which will be an indication of the similarity between the two compared templates. The system will have a preset *threshold* value determining how large this matching score must be for two templates to be recognised as identical. The accuracy of the system will depend on the threshold value. A small threshold value will tolerate a low similarity score, thus resulting in situations where two different persons might be recognized as the same person by the system; a *false acceptance*. A high threshold value will only tolerate high matching scores, which might result in enrolled persons not being recognized by the sys-

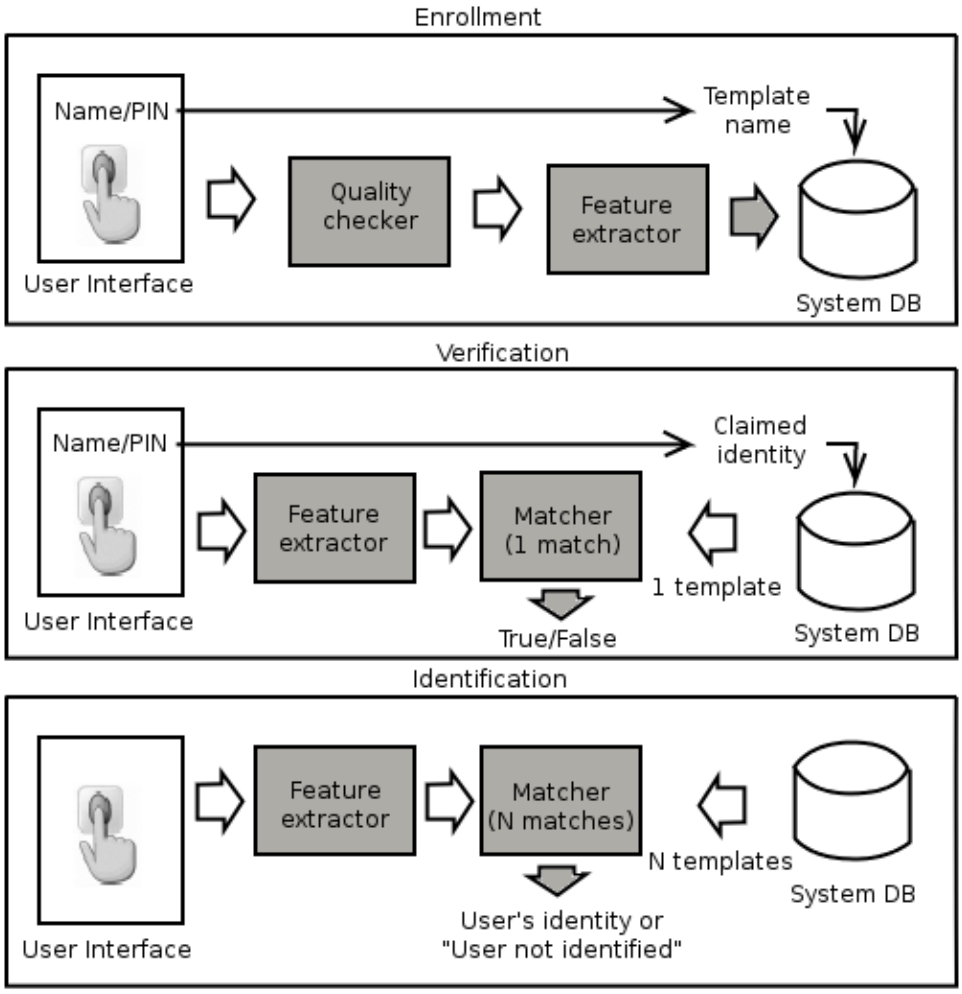


Figure 3.1: An illustration of the differences between the enrollment, verification and identification process. Source: [43].

tem; a *false rejection*. These are the two basic errors of any authentication system. The amount of false acceptances and rejections compared to the total number of authentication attempts on a system is known as the *false acceptance rate* (FAR) and *false rejection rate* (FRR). Using these rates makes it possible to compare the strengths of different authentication systems.

3.2 The Human Gait

One of the first studies of the human gait was made in the early 1900s when Marks [44] described how the process of walking could be divided into a serie of phases and looked at how various prosthetic designs of an amputee gait affected these phases.

Today, we divide the human gait into *gait cycles*, which are defined as the period from an initial contact of one foot to the following initial contact on the same foot [7, 13]. This period is possible to divide into three main tasks, which again is possible to divide into eight phases as illustrated in Figure 3.2. The first task is a *weight acceptance* period, which involves an *initial contact* phase and a *loading response* phase. During this task,

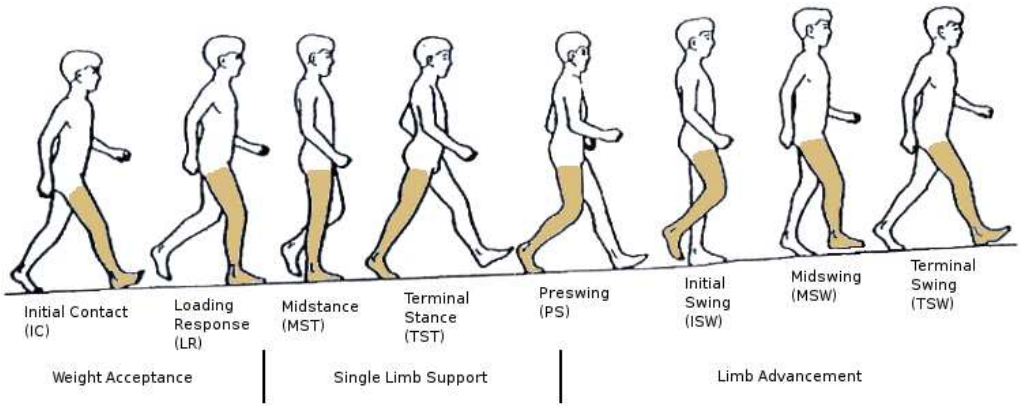


Figure 3.2: A complete gait cycle with its three tasks and eight phases displayed. Source: `http://www.orthoteers.co.uk/Nrujp~ij33lm/Orthgait.htm` (Last visited 09052005).

one foot is placed at the ground and the body weight is shifted to maintain stability and absorbing shock. The second task is a *single limb support* task consisting of a *midstance* phase, a *terminal stance* phase and a transition to the *preswing* phase. During this task, the contralateral foot is swung forward while the body weight is maintained on the stable foot. The last task is the *limb advancement* consisting of the *preswing* phase, the *initial swing* phase, the *midswing* phase and the *terminal swing* phase. During this task, the previously stable foot leaves the ground, shifting the body forward.

The study of identifying humans by their gait started in the field of psychology back in the 1970s. One early discover made by Cutting and Kozlowski [15] described how friends and family members could recognise each other by the way they walked, even if they only where observing light reflecting markers attached to several strategic parts of the walking person. A similar study was made by Johansson at nearly the same time with similar results [30]. Both studies also showed that determining the gender of the walking person was fairly easy even if the observer was not familiar with him.

The first attempt of automatically gait analysis was probably performed in 1994 by Niyogi and Adelson [53]. They described how changes in a 2 dimensional video footage of a walking person could be analysed.

Lee and Grimson has also performed some research on how to identify persons based on video image capturing [39, 40]. They define gait as

> "[An] idiosyncratic[1] feature of a person that is determined by, among other things, an individual's weight, limb height, footwear, and posture combined with characteristic motion."

They also extends this definition to include

> "[The] appearance of the person, the aspect ratio of the torso, the clothing, the amount of arm swing, and the period and phase of a walking cycle."

Using these two definitions, it is clear that gait might be usable to identify persons. Although all humans move in the same basic pattern there are individual details in the relative timing and magnitudes of the motions. These variations have been studied much in clinical gait analysis, which in most cases tries to distinct pathological gait from normal gait, and not to identify humans [9].

[1] Of peculiar temper or disposition; belonging to one's peculiar and individual character (1913 Webster).

Recently, kinematic gait analysis has also been used in forensics. The murderer of Swedish Foreign Minister, Anna Lindh was identified, partly based on analysing gait data from surveillance cameras [20]. Earlier on, Lynnerup and Vedel [42] described a case where a similar procedure had been used to identify two bank robbers in Aalsgarde, Denmark. Also in the Norwegian NOKAS robbery, one of the suspects was identified, partly based on his physique and gait [25, 24]. Results from such analysis have not yet been refined in such a way that it can be used as evidence, but it might be a very strong circumstantial evidence.

The human gait has the advantage compared to other, more traditionally used biometric features, that it is not left behind, like for instance a fingerprint or a signature is. If the gait is measured using motion capturing, it is neither possible to capture using a digital camera, like the face. Therefore, the gait should be more difficult for an adversary to forge and hence, be a more secure method for authentication under certain situations. The human gait has also some disadvantages compared to the more traditional authentication methods. It's biggest weakness is that it is not as stable as many other biometrics. A change in footwear or clothing might disrupt the gait enough to hinder a person from being recognised by the system. Also, the gait differs whether the person walks normally, runs and walks up and down stairs. Analysing a person's gait based on video footage also has the potential of misuse, since it is possible to recognise persons without their knowledge or approval. This could be incorporated into todays video surveillance systems to recognise persons from a distance, which in worst case could resemble an Orwellian society [56]. However, this would very easy violate the Norwegian personal information legislation [2].

There are an overwhelming amount of different kinds of gait a person is able to perform. For simplification, this research focuses on what is defined as *normal walking*; walking at your normal walking speed along a flat surface. The average normal walking speed is at $1{,}32\frac{\mathrm{m}}{\mathrm{s}}$, which gives an average of 60 gait cycles each minute [13].

As a summary, one can recall Lee and Grimson's definition and remember that the human gait consists of many different elements which are characteristic for a person. The problem is to detect these features and analyse them in a way which will give a adequate result. It is also important to remember Tanawongsuwan and Bobick's work [67] and how the gait changes during different walking speeds and over time. These are the two main challenges which must be overcome before the human gait can be an applicable method for authentication. However, one must also recall the potential strengths such a method might have. It is not left behind like a latent fingerprint and does not require the person's attention during the acquiring process. This research will focus on measuring the legs acceleration during locomotion. There are however other walking features which can be captured using motion detection. One of these features is the angle of the foot, which might be possible to capture using a gyroscope.

4 Choice of Technology

4.1 The Gait Collector

Throughout this report, the device used for collecting a person's walking features is referred to as the Gait Collector. This is a device that was created for this research that is basically consisting of the following two components:

AVR Butterfly: AVR is a family of fairly cheap and feature-rich micro controllers from Atmel Corporation[1]. Their micro controllers are designed as Reduced Instruction Set Computers (RISC), which mean they are small, cheap computers designed to perform simple specific tasks efficiently. The size and price of these controllers makes them very suitable for this project, where the gait device should be small enough to be carried in your pocket or possible to attach to a leg without disturbing the person in any way. For this project, the AVR Butterfly [6] has been used, which is an evaluation board equiped with a ATmega169 micro controller [5] and a whole serie of other usefull features, such as a 100 segment LCD display, a 4Mbit dataflash memory, a Real Time Clock 32,768 Hz oscillator, a 4-way joystick with center pushdown button, a RS-232[2] level converter, a bootloader for programming and a built-in safety pin so it can be hanged on your shirt.

ADXL202: The ADXL202 is a low cost dual-axis accelerometer from Analog Devices[3] capable of detecting acceleration up to $\pm 2g$, which should be sufficient for detecting normal walking. The use of accelerometers for motion capturing is possible due to the recent advances in this technology, which makes it more and more popular in technologies measuring tilt, shock and vibration. The output from the ADXL202 is a digital signal whose duty cycles (another name for the signal's pulse width) are proportional with the acceleration. The output from the accelerometers are measured in microseconds using the Gait Collector, and are referred to as the "acceleration data" throughout this report. As a result of this, the acceleration terms used in this report does not imply the classical definition of $\frac{m}{s^2}$. In stead, it is a measure between 0 and circa 3000 that is proportional with the actual acceleration. These accelerometers where chosen primary due to their low cost. Consult Appendix A.2 for further description of the duty cycle outputs. Verplaetse's recommendation of using $10g$ accelerometers [68] was not read prior to the purchase of the accelerometers. These recommendations are however for attaching accelerometers to a shoe

[1] `http://www.atmel.com`
[2] The standard communication interface.
[3] `http://www.analog.com`

and not to the leg. Data obtained using the Gait Collector shows that during normal walking, most of the registered values are in the range of $\pm 2g$. A range of $\pm 10g$ would in this situation most likely introduce inaccuracy.

Two ADXL202 accelerometers has been attached to the AVR Butterfly's ATmega169 micro controller. These two accelerometers are mounted at a right angle of each other such that they are capable of collecting acceleration data along three orthogonal axes; the vertical acceleration, which is referred to as the X axis in this report, the forward movement, referred to as the Y axis and the sideway movement, referred to as the Z axis. Figures 4.1 and 4.2 shows the inside of the Gait Collector, displaying the positioning of the two ADXL202 accelerometers.

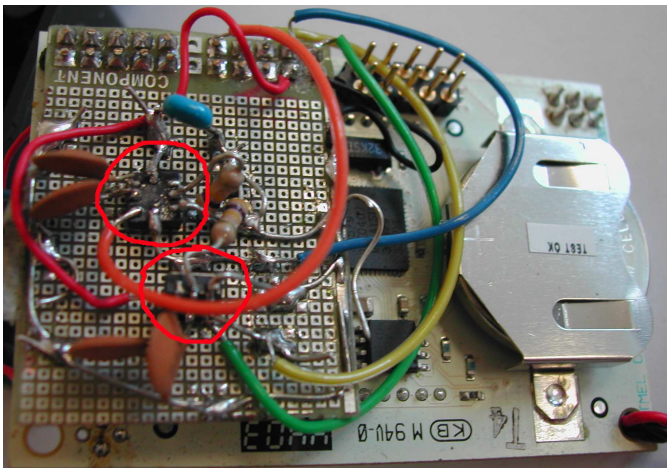


Figure 4.1: An overview of the inside of the Gait Collector illustrating how the two ADXL202 accelerometers (pointed out by the red circles) are soldered to the back of the AVR Butterfly.

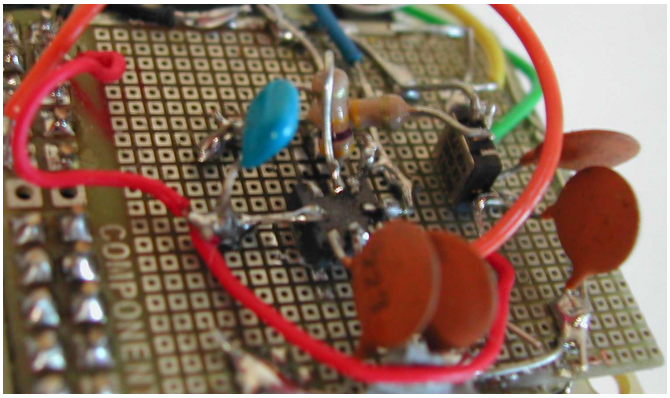


Figure 4.2: A close-up of the two accelerometers, illustrating their positioning related to each other.

A simplified electrical circuit diagram of the Gait Collector is shown in Figure 4.3. As shown in this diagram, there are also provided an In-System Programming (ISP) interface for programming the Gait Collector, an external power connector and a RS-232 as a computer interface. The software, which has been created for the Gait Collector is further described in Appendix A.

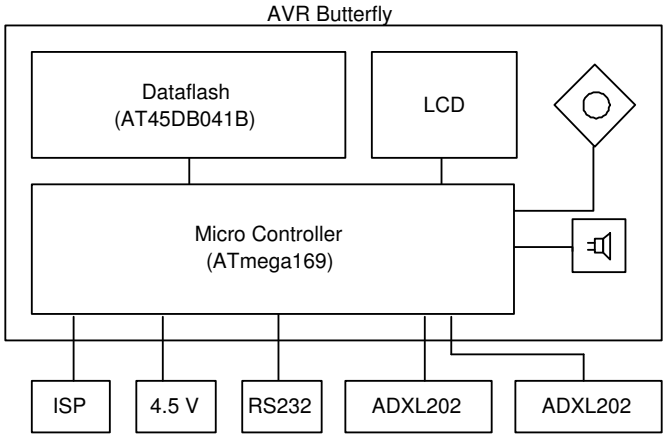


Figure 4.3: An electrical circuit diagram of the Gait Collector displaying how the AVR Butterfly and the two ADXL202 accelerometers are attached to each other.

The Gait Collector has been encapsulated in a plastic box measuring $5.4\text{cm} \times 8.2\text{cm} \times 3\text{cm}$. Some straps has also been provided to be able to attach the Gait Collector firmly to a person's leg. When attached, it appears as shown in Figure 4.4.

Figure 4.4: The Gait Collector attached to a person's leg in its ideal position just above the right foot's ankle. The picture shows both the Gait Collector itself, the necessary external battery supply and the strapping system used to attach it.

4.2 The Necessary Software

4.2.1 AVR Studio

Atmel has created a free and powerful development suite for their micro controllers, called AVR Studio. Using this tool, it is possible to create software using the AVR ASM language. It also provides the posibility of simulating the various micro controllers in software, making it possible to debug the code before transferring it to the controller. It also provides functionality for transferring the code to the micro controller using several

different interfaces.

4.2.2 WinAVR

Though AVR ASM is a powerful programming language, it gets complex and difficult to maintain if the code is large. In that case, the programming language C is a better choice. WinAVR (pronounced “whenever”) is a powerful, free tool for C programming on the AVR micro controllers. It is part of the free software movement, and available from `http://www.winavr.sourceforge.net`. Integrated into the WinAVR distribution is also the free text editor *Programmers Notepad*.

4.2.3 Serial interfaces

Using the RS-232 interface, it is possible to connect the Gait Collector to a computer. This makes it possible to communicate with the device, ordering it to perform tasks like to start collecting gait data or sending the obtained gait data to the computer. These tasks can be done through any serial interface software, like HyperTerminal[4], which is integrated into Microsoft Windows. Another good software is Terminal by Bray[5]. This program is also capable of storing the ouput from the serial port directly to file, and can therefore be recommended for retreiving the acceleration data from the Gait Collector.

4.2.4 Matlab

To ease the process of data analysis, there was a need for predefined functionality for performing as many of the necessary analysis operations as possible. The choice of software therefore fell on the Matlab package from MathWorks[6]. This package is suitable for developing both algorithms and applications for advanced mathematical operations fairly simple compared to programming languages, like C/C++ or Java.

[4]More info at `http://www.microsoft.com/technet/prodtechnol/windowsserver2003/library/ServerHelp/02c2459f-5b84-45fb-afab-610374d35994.mspx` (Last visited 29062005).

[5]Available from `http://bray.velenje.cx/avr/terminal/` (Last visited 08052005).

[6]`http://www.mathworks.com`.

5 Data Gathering

The data gathering process is performed by initiating the Gait Collector and then strapping it onto a participant's leg. During initiation, the Gait Collector is simply told the name or ID of the person participating in the experiment, whereas it starts to store the acceleration data onto its dataflash. The person can then participate in the experiment by walking with the Gait Collector strapped firmly to the leg. When enough walking data has been obtained, it can be transferred to a computer for further analysis.

5.1 The Raw Data Format

The raw data from the Gait Collector are in the form of duty cycles as shown in Figure 5.1. Here the length of $AccX$ and $AccY$ is a measure in microseconds proportional with the acceleration. When these values are to be analysed, they are transferred to a computer and stored in a file.

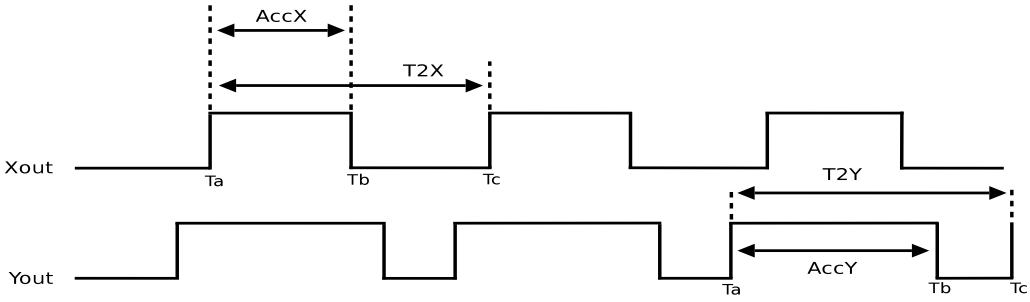


Figure 5.1: The ADXL202 Duty Cycle Output in X and Y axes.

After the data has been stored to file, the headers and other unnecessary information must be removed from the file to make it readable in Matlab. By doing this, all information in the file except the time and raw data from the accelerometers itself must be removed. To ease the analysis process, the acceleration data collected were the persons did not actually perform walking was manually deleted from the file.

The gait data collected from a person is stored in three files; one file with the plain raw data from the Gait Collector, one file with the headers removed to make it readable in Matlab and one file where data, which is not collected during walking, had been removed manually. Examples are shown in Listings 5.1–5.3. Listing 5.1 shows an example output from the file *alice1.log*, which contain raw data from Alice's first gait attempt. As shown here, it is possible to see when the gait attempt was performed, who performed it, and seven columns containing the output data from the Gait Collector. The *Sec* column

contains the time in which the data was stored to dataflash in seconds. It is meant to be a reference point during reading, and does not correspond exactly to the time the acceleration data was read from the accelerometers. The rest of the columns are values from the accelerometers' duty cycle outputs, as illustrated in Figure 5.1.

Listing 5.1: Example of content from the raw data file *alice1.log*.

```
Terminal log file Date: 18.03.2005 - 09:48:45
_____________________________________________

alice
Sec    AccX AccY AccZ T2X  T2Y  T2Z
  0    1412 1537 1158 3149 3286 3157
  0    1534 1521 1280 3195 3271 3211
  0    1421 1521 1167 3196 3258 3204
  ...
  117  1459 1521 1205 3196 3272 3212
  117  1458 1559 1204 3195 3296 3203
  117  1408 1520 1281 3195 3257 3203
```

The file without headers is shown in Listing 5.2. This file is readable in Matlab.

Listing 5.2: Example of content from the data file *alice1.dat*.

```
  0    1412 1537 1158 3149 3286 3157
  0    1534 1521 1280 3195 3271 3211
  0    1421 1521 1167 3196 3258 3204
  ...
  117  1459 1521 1205 3196 3272 3212
  117  1458 1559 1204 3195 3296 3203
  117  1408 1520 1281 3195 3257 3203
```

The file only containing acceleration data is shown in Listing 5.3. The acceleration data is organised in this way to ease the analysis process. By removing all data except the data containing walking features, it is not necessary to create methods for extracting these data automatically, which could increase the bias in the final data analysis. By organising the collected data into three files as illustrated here, it is possible to map the values from the different columns to the attributes they represent. It is also possible to discover exactly when the measurement was performed.

Listing 5.3: Example of content from the data file *alice1_.log*, which only contains gait data.

```
  36   1053 1217 1611 3196 3272 3196
  36   900  1369 1497 3234 3280 3196
  36   711  1369 1725 3196 3234 3196
  ...
  92   1001 1483 1648 3195 3271 3195
  92   1015 1331 1687 3196 3288 3196
  92   1053 1293 1649 3196 3272 3196
```

5.2 Noise and Biases During Data Gathering

There are basically three sources to bias during data gathering. These sources may cause reliability and validity problems in some extent. As defined in [12], reliability problems are the problems that will prevent the Gait Collector from yielding the same results after

repeatable measurements using the same person. Validity problems, on the other hand, is related to whether the Gait Collector actually measures what it was intended to measure.

5.2.1 Hardware and Software Related Problems

These problems are partly related to the hardware of the Gait Collector and its assembling. Such problems will give systematic errors, and therefore give the Gait Collector less reliability.

Accelerometer wrongfully oriented: Positioning the two accelerometers at *exactly* a right angle on each other is practically impossible with ordinary soldering equipment. They will therefore not represent a perfect system of coordinates as shown in Figure 5.3. However, if they are mounted with enough precision, it should be sufficient to detect the characteristics of a person's movement.

Data not being time correlated: The acceleration data is collected simultaneously in all three axis at the same time, as shown in Appendix A.2. The acceleration data is therefore not collected with equal time intervals, but such a method of gathering, makes it possible to collect data from all three axis at almost the same time.

5.2.2 Attachment Related Problems

These problems are related to the attachment of the Gait Collector. Such problems might result in the measurement of *where* the person has attached the Gait Collector rather than *how* the person walks. It therefore gives errors in the Gait Collector's validity.

Prototype design: The prototype has been designed with functionality in mind, which has been on the cost of comfort and size. As a result, it need 4.5 volts external power supply consisting of three AA batteries to collect gait data. This makes the Gait Collector a bit heavy and not very comfortable to wear, and makes it difficult to attach without moving to much along the leg while walking.

Gait Collector mounting: The Gait Collector is mounted to a person's leg using straps. Even though care is taken to mount the Gait Collector as firmly as possible, it is impossible to prevent it completely from moving during locomotion. This is due to the Gait Collector's size, straps and the fact that the skin itself is moving at a certain level.

5.2.3 Thermal Noise

These problems are related to how heat in the Gait Collector's micro electronics might cause wrongful measurements. This will mostly be a problem in validity, since the output from the accelerometers will vary both with the temperature and the acceleration.

Static noise: Even though the Gait Collector is placed stationary on the ground and not moved, it is a certain degree of difference from one measurement to another due to static disturbance. The level of static noise is however fairly easy to measure by placing the Gait Collector stationary on the ground. Figure 5.2 illustrates the Fast Fourier Transform of data collected when the gait Collector was placed stationary on its side. Even though this noise is affecting the results from the analysis, measures for reducing the level of static noise was not taking into account during this project. What is interesting here is the dominant spike at roughly 0.8Hz. Not much

work has been done trying to determine the cause of this spike, but it might be related to the sampling frequency of circa 16 samples each second.

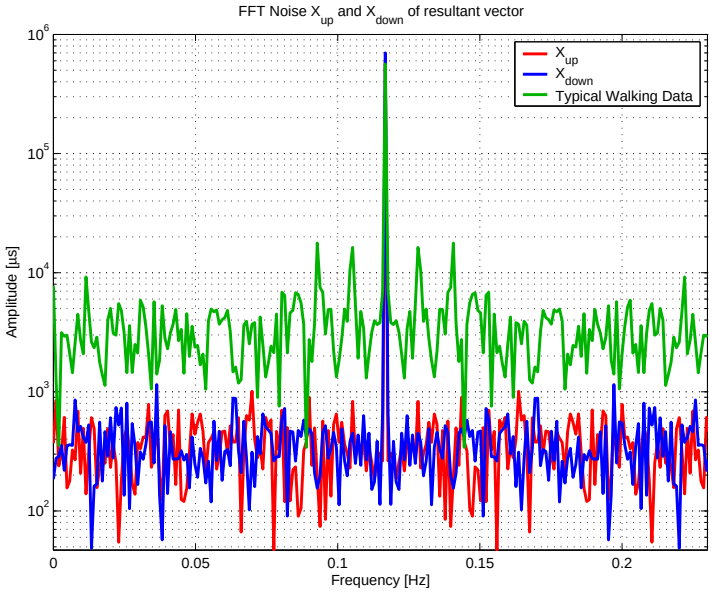


Figure 5.2: The graphs shows the degree of static noise affecting the data collected using the Gait Collector. It shows the FFT transforms of the resultant vector when the Gait Collector is placed stationary on the ground with the X-axis accelerometer facing upward (red) and downward (blue) The green curve shows the FFT transform of typical walking data. The noise in Y and Z axis shows similar tendencies.

5.3 An Overview of the Obtained Data

Data collected using the Gait Collector results in acceleration data in the three axes X, Y and Z as shown in Figure 5.3. To gain a better understanding of what to look for in these data, more details regarding the information provided by the raw data itself is necessary.

X-axis: The acceleration data in the X-axis will reflect the acceleration provided by the rising and lowering of the foot. It will show a downward spike as the foot leaves the ground, as well as an upward spike direction as the foot reaches its highest point as it is moved forward.

Y-axis: Assuming the person walks with constant speed, the acceleration in Y-axis will not be affected by the speed of the person. This acceleration will merely be a reflection of the foot moving forward during the single limb support phase, resulting in a small spike when the foot touches the ground. It will also to some extent reflect the acceleration of the foot moving forward.

Z-axis: The acceleration in Z-axis will reflect the sideways motion of the foot during movement. This will not reflect any large accelerations, but since this movement is not a strictly necessary movement in the same extent as the movements in X and Y axes, the Z axis movement might have a larger level of individual characteristics.

As data is collected using the Gait Collector, the various characteristics of a person's movement can be made visible by displaying them in graphs. Graphical visualisation is

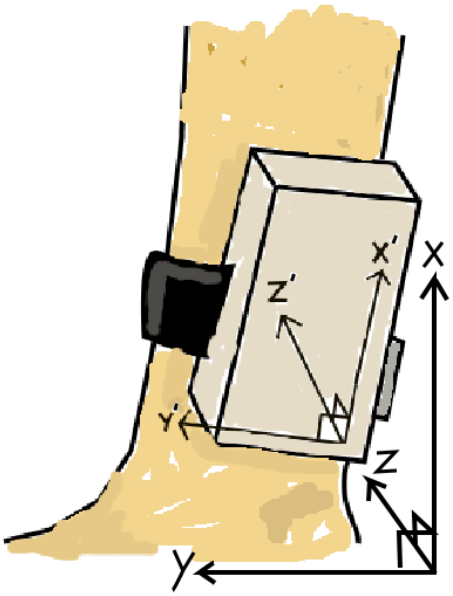


Figure 5.3: The Gait Collector attached to a leg with the three acceleration axes displayed. Also note how these axes are oriented compared with the physical axes.

also useful for illustrating the changes in time when applying some of the methods for analysis described later in Chapter 6, like the calculation of the resultant vectors and their the α and β angles (Section 6.3.1).

A typical dataset's resultant vectors and their orientations are shown in Figure 5.4. As shown, the data is fairly repeatable over several periods. BenAbdelader *et al* [9] claims the same thing in their report when persons are walking under the same conditions. By studying these periods closer, a more thorough understanding of how the foot is moving during a gait cycle is achieved. This is shown more closely in Figure 5.5, where acceleration data from three gait cycles are displayed. Remember that throughout this thesis, the term "acceleration" does not imply the classical notation of $\frac{m}{s^2}$. It is the duty cycle output from the accelerometers, which is measured in microseconds, and is proportional the acceleration. Consult Appendix A.2 for a more detailed description of this output.

A more detailed image of how the foot itself moves during one of the gait cycles is shown in Figure 5.5. The terminology used in this illustration is explained in Section 3.2. This illustration is an attempt of mapping the foot's movement with the various fluctuations in the X and Y axes acceleration data. As this mapping was not a major part of this project, it have not been validated. However, it gives an indication of which motions are associated with which parts of the obtained data.

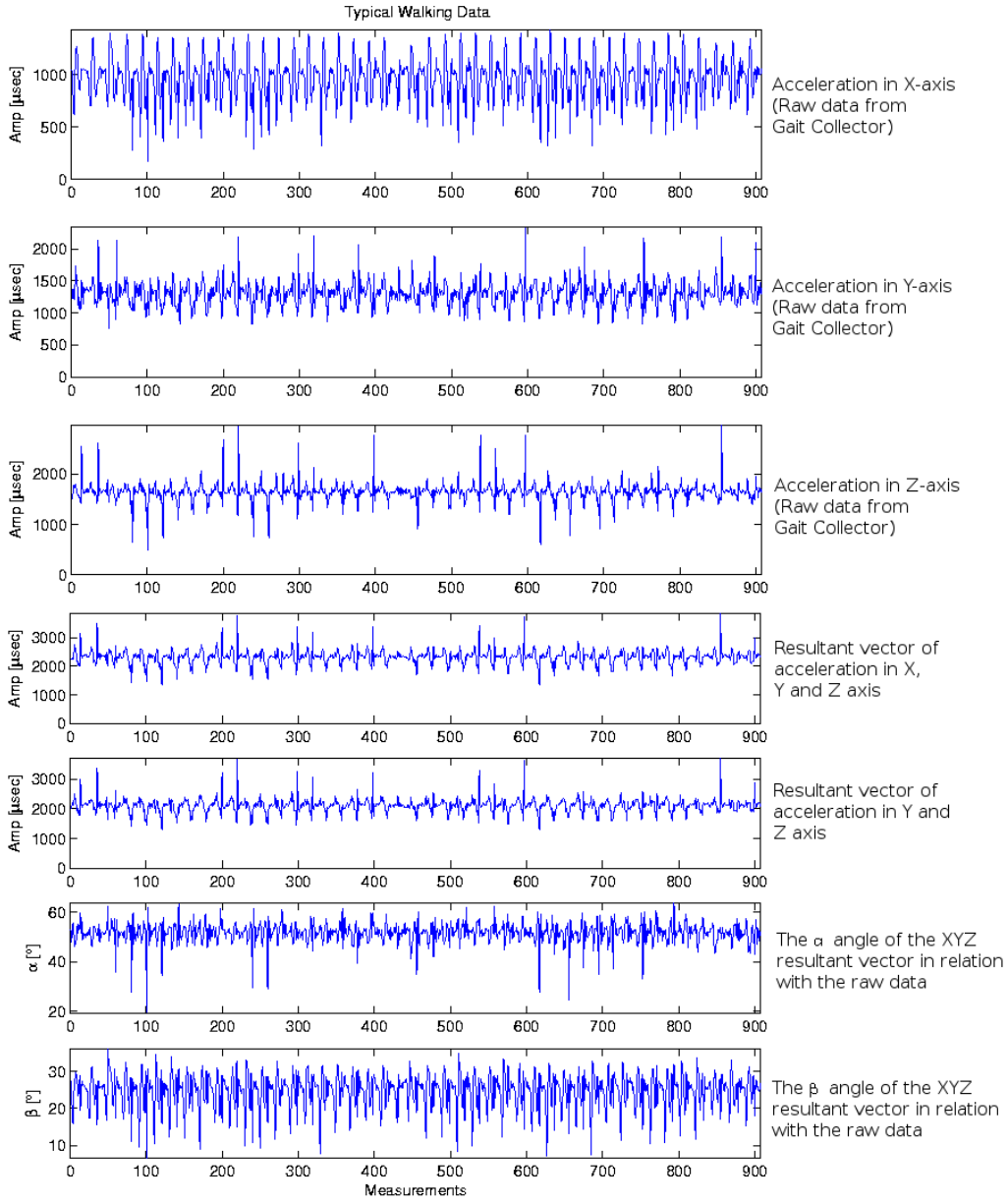


Figure 5.4: Shows the data gathered from the Gait Collector compared with the calculated resultant vectors and their orientation with the Gait Collector's data over circa 900 measurements, or roughly a time period of one minute. As shown, the foot moves in a fairly stable manner over time.

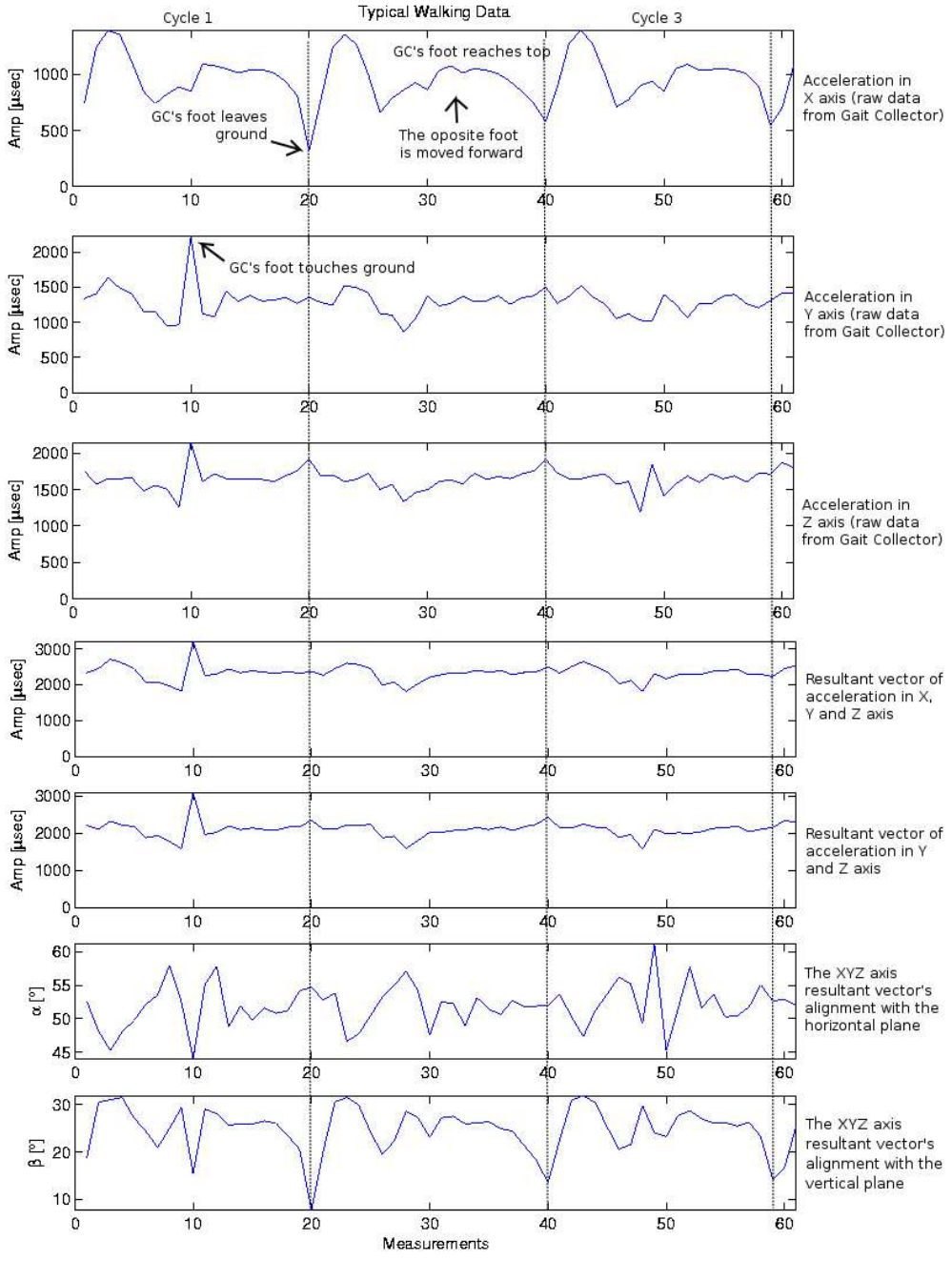


Figure 5.5: Shows the data gathered from the Gait Collector over three gait cycles compared with the calculated resultant vectors and their orientation with the Gait Collector's data. How the resultant vectors and the α and β angles were calculated are shown in Section 6.3.1.

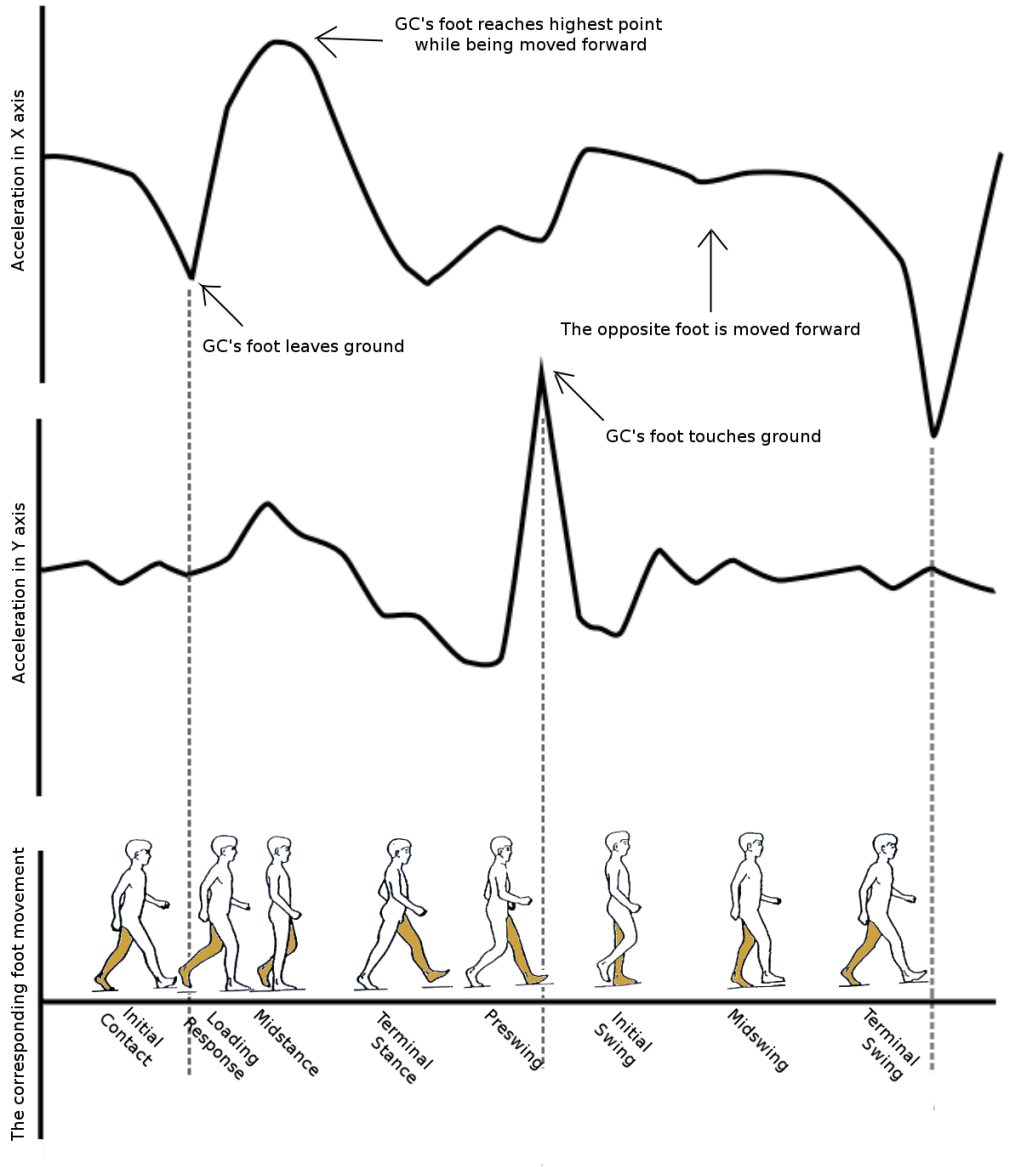


Figure 5.6: Shows how the foot itself moves during a gait cycle and how this corresponds with the X and Y axis acceleration data. The foot with the Gait Collector attached is coloured brown. The various phases of the gait cycle is described closer in Section 3.2. The walking person is borrowed from `http://www.orthoteers.co.uk/Nrujp~ij33lm/Orthgait.htm` (Last visited 09052005).

6 Data Analysis Theory

This chapter proposes several methods that might be usable to prepare and analyse gait data in a manner that will help determine whether two datasets belong to the same person. Most of the methods described in this chapter were successfully used by others performing gait analysis, though not on the same kind of data as has been obtained during this project. An overwhelming amount of available methods for signal processing and data analysis which might be applied to this research. However, it is possible to organise them into the following four categories:

Data pre-processing methods: When examining the output from the Gait Collector, it is not particularly structured. The pre-processing methods all have the purpose of simplifying the data by reducing noise and other irregularities, clustering data or similar. This is a necessary step to ease the later process of comparing datasets. It will help overcome some of the limitations in the raw data, and give new information regarding how the Gait Collector and the leg it is attached to move during locomotion.

Data analysis methods: After applying one or several pre-processing methods, the process of analysing data to detect distinction and behaviour can begin. This will distinguish one person from another.

Comparison methods: After the data analysis has been performed on two different datasets, the result will be two arrays of values representing the gait characteristics of the persons which are analysed. The task is now to compare these two datasets to find differences and equalities among them.

Similarity detection methods: The previously described methods for data comparison gives a vector representation of the similarity between two datasets. The big question is then how to get one single number out of this vector, which will tell something about the similarities in the two datasets. This value can be compared with a pre-defined threshold value to determine whether or not they the two datasets are from the same person.

6.1 The Framework for Data Analysis

To make the process of analysing the collected acceleration data more structured, a framework is needed. This prevents a blindly us of the methods that look most suitable without really being able to compare the results from the different methods afterward. Such a framework will also make it possible to obtain valid and reliable FAR and FRR

data for the various analysis methods. The whole framework consists of a series of operations as illustrated in Figure 6.1 and described below:

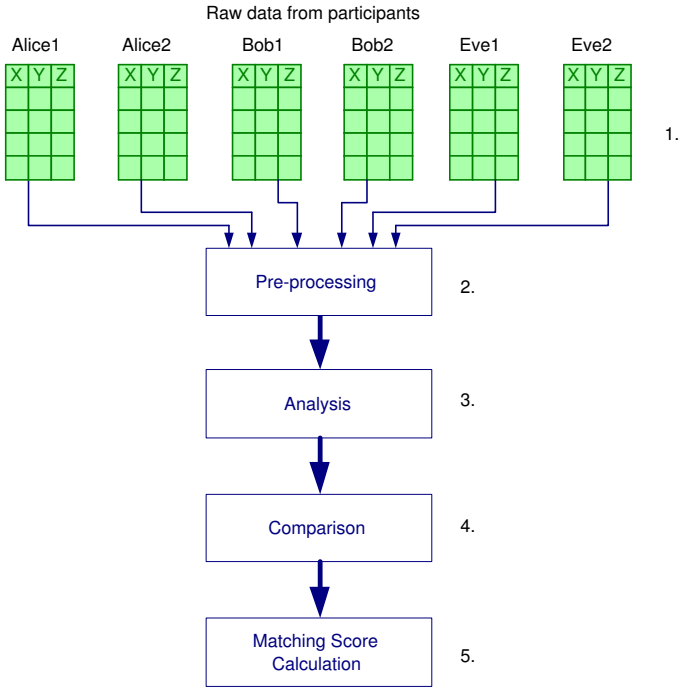


Figure 6.1: The framework for analysing data collected using the Gait Collector to determine whether two datasets belong to the same person. All proposed methods for analysis described in the following chapters follows this framework.

1. Obtain two sets of gait data from a number of different persons.
2. Select one or more methods for pre-processing the data, and apply these to the datasets.
3. Select a preferred method for data analysis, and apply it to each of the pre-processed datasets.
4. Select a method for data comparison and apply it to each possible permutation of two datasets from the analysis process. This will give a serie of comparisons. Some of them are comparisons of the same person, other are comparisons of two different persons.
5. Select a method for detecting the result from the comparisons, and use this method to obtain a similarity score from each comparison. This score will tell whether a certain combination of data simplification, data analysis and data comparison is capable of detecting whether or not two datasets belong to the same person.

6.2 False Acceptance Rate and False Rejection Rate Calculation

The previous section has proposed a framework that was used to create a structured way of analysing the gait data. To determine which of several different data analysis methods is the most suitable the FAR and FRR were calculated for each of the methods. Though the experiments described later in this report are not of a large enough scale

for a thorough quantitative estimate of the FAR and FRR rates, they are used to give a preliminary impression of the extent of acceptances and rejects achieved in a given situation. Throughout this thesis, the FAR and FRR rates are calculated based on O'Gorman's definition [54]. He define FAR in a fingerprint authentication system as

> "[The] ratio of the number of instances of pairs of different fingerprints found to (erroneously) match to the total number of match attempts."

while FRR is being defined as

> "[The] ratio of the number of instances of pairs of the same fingerprint found not to match to the total of match attempts."

Based on these definitions, the FAR is being defined by the formula

$$\mathrm{FAR} = \frac{\mathrm{Number\ of\ false\ acceptances}}{\mathrm{Total\ number\ of\ comparisons}} \tag{6.1}$$

while FRR is being defined as

$$\mathrm{FRR} = \frac{\mathrm{Number\ of\ false\ rejections}}{\mathrm{Total\ number\ of\ comparisons}} \tag{6.2}$$

If the FAR and FRR rates for a series of threshold values are obtained, it is possible to plot these rates as a *Receiver Operating Curve* (ROC curve), where one axis displays FAR rates and the other displays FRR rates. A typical ROC curve will form an "elbow" toward the origin. Based on such a curve, it is possible to make tradeoff decisions regarding what FAR/FRR values that is desirable for a system. Figure 6.2 illustrates a typical ROC curve. It shows the results from Bobick and Johnson's gait research [11] where they performed gait recognition based on video images. Such ROC curves will be presented for each of the performed experiments as an indication of the relations between their obtained FAR and FRR values.

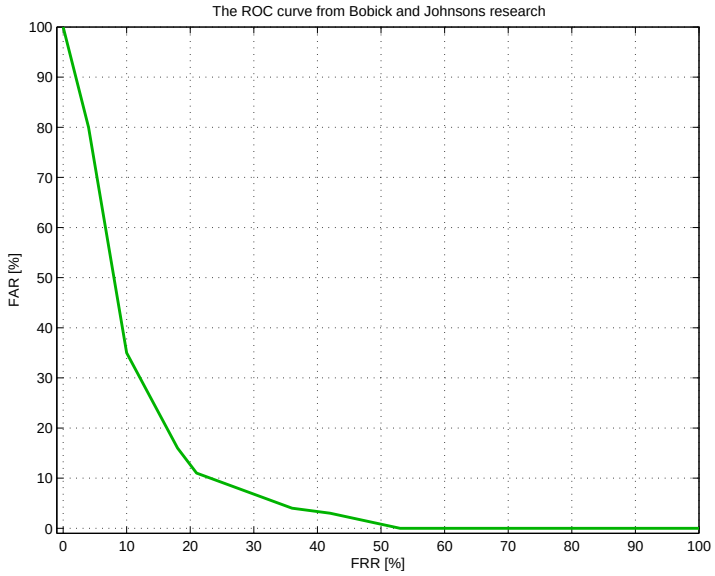


Figure 6.2: A ROC curve illustrating the results from Bobick and Johnson's gait research [11].

6.3 Possible Data Analysis Methods

This section will briefly list several different methods for performing data analysis on the gait data. Some of the described methods were combined into new gait analysis methods and tested in three experiments. Other methods where postponed due to the limited time available, or they did not suit the scale of the experiments. How the various analysis methods were combined are explained in Chapter 7, while the results from the experiments are described in Chapter 9.

6.3.1 Pre-processing Methods

These methods are meant to simplify the raw output from the Gait Collector and remove some of the noise and irregularities that exist in this output.

Resultant Vector Calculation

The raw data from the Gait Collector can be analysed without any further manipulation, but it has some weaknesses. For instance, the output from the various accelerometers will vary depending on how the Gait Collector is oriented on the leg. To overcome this problem, it is possible to calculate the resultant vector of the Gait Collector's output. Since the accelerometers are mounted at a straight angle to each other, it is possible to treat the acceleration data as the vectors x, y and z. Then it is possible to re-compose them into one vector, $|\overrightarrow{V_1}|$ using Pythagoras rule twice [19]:

$$|\overrightarrow{V_1}| = \sqrt{x^2 + y^2 + z^2} \tag{6.3}$$

Later in the thesis, this vector is referred to as the XYZ resultant vector. This method for determining the length of a vector applies also to the resultant vector for only two of the acceleration vectors as well, like y and z:

$$|\overrightarrow{V_2}| = \sqrt{y^2 + z^2} \tag{6.4}$$

Later in the report, this vector has been referred to as the YZ resultant vector. In addition to simplifying the data analysis part to only treating one set of data, it also has the advantage that $|\overrightarrow{V_2}|$ should not be affected by the orientation of the Gait Collector. This should only affect its orientation with the horizontal and vertical plane of the Gait Collector; the angles α and β as shown in Figure 6.3.

Calculating the Resultant Vector's Orientation

Figure 6.3 shows how the resultant vector $|\overrightarrow{V_1}|$ is oriented with the Gait Collector's axis. During locomotion, this vector will change its orientation as the leg with the Gait Collector attached accelerates upwards and downward, forward and backward, and sideways. These changes in orientation are reflected by two angles; the α angle, which represent $|\overrightarrow{V_1}|$'s orientation with the horizontal plane of the Gait Collector and the β angle, which represent $|\overrightarrow{V_1}|$'s orientation with the vertical plane of the Gait Collector. Once the XYZ resultant vector has been calculated, as shown earlier in this section, the α angle can be found using the formula

$$\alpha = \sin^{-1}\left[\frac{z}{\overrightarrow{V_2}}\right] \tag{6.5}$$

Similarly, β is found using the formula

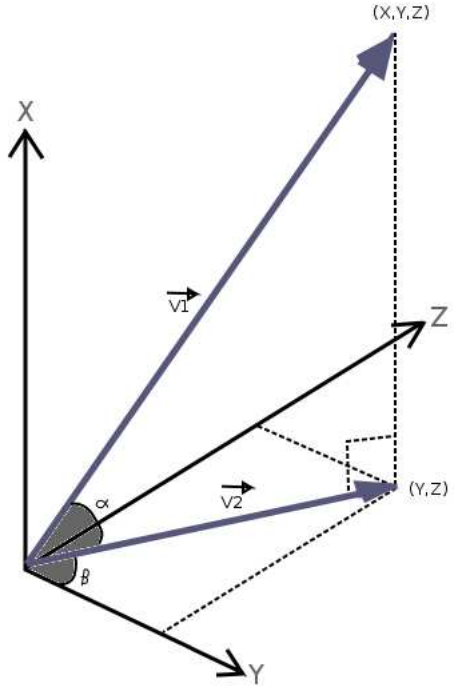


Figure 6.3: How the acceleration in X, Y and Z axis can be re-composed into two resultant vectors. It shows how these vectors are oriented relative to the Gait Collector.

$$\beta = \sin^{-1}\left[\frac{x}{\overrightarrow{V_1}}\right] \tag{6.6}$$

In these equations x and z are raw acceleration data from the Gait Collector in the X and Z axes.

Autocorrelation

Autocorrelation is a function which detects non-randomness in data, and is capable of finding appropriate time series if the data is non-random [51]. This is a method used by BenAbdelkader *et al* [9], when they looked at how stride[1] and cadence could be used to identify persons. Although their study was based on kinematic video data and this thesis covers kinetic acceleration data, the use of autocorrelation might be relevant. The autocorrelation r_k for a signal $Y_1, Y_2, ..., Y_n$ at the time $X_1, X_2, ..., X_n$ at the time k is defined as [51]:

$$r_k = \frac{\sum_{i=1}^{N-k}(Y_i - \overline{Y})(Y_{i+k} - \overline{Y})}{\sum_{i=1}^{N}(Y_i - \overline{Y})^2} \tag{6.7}$$

Neural Networks and Kohonen's Self-Organising Map (SOM)

Neural networks are inspired by how biological nervous systems, like the brain processes information. They are trained for tasks, like pattern recognition or data classification through a learning process much like people learn to perform tasks. Through this process, the neural network is learning how to match a serie of input patterns to a known output

[1] The distance covered by a step (WordNet).

pattern. It uses this information later to match more complex and new input patterns to unknown outputs. Neural networks were used by [66] to classify gait patterns. For more information on the subject of neural networks, consult [31, 65, 71].

Kohonen's SOM [35, 36], often referred only to as the Self-Organising Map, is a variant of the neural network that performs unsupervised learning. It tries to detect an underlying structure in the data, where nodes which are close to each other in the dataset interacts differently than nodes which are considered further apart. It takes the data from different sensors and places them in a 2-dimensional grid to simplify the further analysis, resulting in a representation of the data as a topological map. This method has been applied by [27, 33, 34, 60, 66] to detect differences both in data gathered by motion detection devices like accelerometers, and in video data. However, most of the research today uses the SOM to classify various kinds of gait patterns. It has not been used much to authenticate persons. Figure 6.4 shows how Köhle and Merkl classified various gait characteristics in a SOM [34].

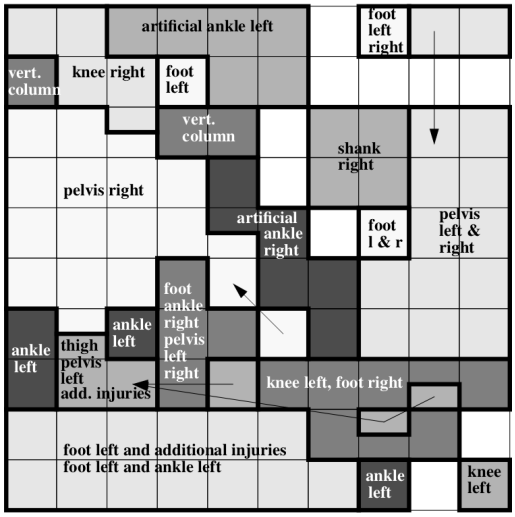


Figure 6.4: An example showing how Köhle and Merkl [34] were able to classify various gait patterns using a SOM.

Principal Components Analysis (PCA)

PCA is a technique for simplifying a data set by transforming it to a new coordinate system. This is done such that the greatest variances in any of the datasets lies on the first axis, the second greatest variance on the second axis and so on[2]. A similar method was applied by Chellappa *et al* [21, 73]; they used eigenvector analysis to recognise human facial images. This method was derived even further when Huang *et al* used eigenvector[3] analysis along with canonical space transforms to perform gait analysis based on video capture [29] as shown in Figure 6.5. A similar analysis was also made by Nixon *et al* [52].

[2]Source: `http://www.wikipedia.org`

[3]A vector of a linear operator that gives a scalar multiplied with itself when it is being operated by the operator (`http://www.wikipedia.org`).

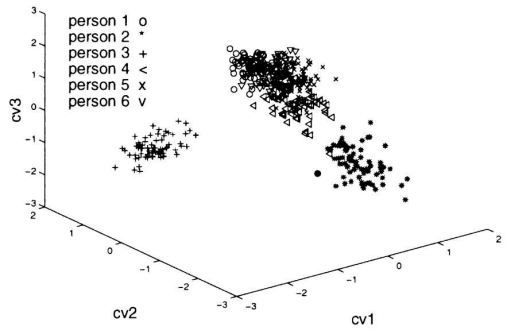


Figure 6.5: An example showing how Huang *et al* were able to use PCA and canonical space transforms to recognise the walking pattern of various people [29].

Hidden Markov Model (HMM) and Dynamic Time Warping (DTW)

HMM [59] is a statistical model which treats the system as a Markov process, trying to determine a set of unknown, hidden parameters using observed parameters. Such a model is suitable as an input for further pattern recognition analysis. HMM is also a much used method during speech recognition, often in combination with the Dynamic Time Warping (DTW) method, which adjusts speech signals of different speeds such that they can be compared more easily [18, 74]. The problem with HMM for this project, though, is that the model requires a predefined Markov chain of various states as shown in Figure 6.6. Each state has one or more transitions to other states with a specific weight that is representing the probability of transiting between these states. Creating such a graph would require more resources than what was available for this project. A DTW analysis of gait data was applied in Sethuraman and Prem's patent of an automatic gait detection system [61].

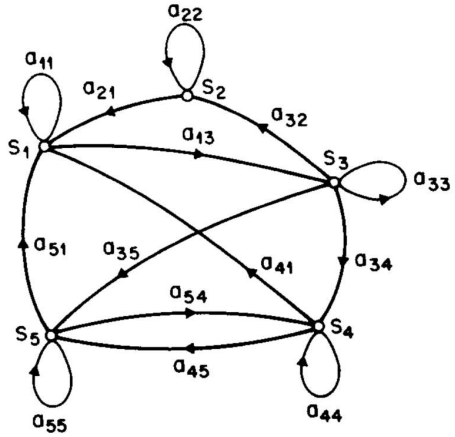


Figure 6.6: An example of a Markov chain with five states and several transitions between these states [59].

6.3.2 Data Analysing Methods

These methods are meant to find characteristic behaviour in the collected gait data and thence detecting features which to some extent will be distinct from one person to another.

The Fast Fourier Transformation (FFT)

The FFT is based on a discovery made by the French mathematician Jean-Baptiste Fourier. He proclaimed a controversial theory in 1807 that any periodic signal can be represented as a sum of enough sinusoidal waves with proper phase and amplitude [63]. The FFT is based on this theory. It is a faster version of the Discrete Fourier Transformation (DFT), which takes a discrete signal represented in the time-domain and transforms it into a discrete signal represented in the frequency-domain. The Fourier transform is therefore a very suitable method for discovering the frequency of a signal buried in noise, as shown in Figure 6.7.

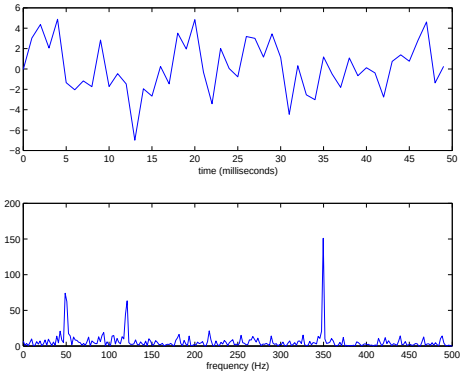


Figure 6.7: The upper graph shows the signal $\sin(2\pi 50t) + \sin(2\pi 120t) + \sin(2\pi 350t)$, which has been altered with random noise. The bottom graph shows the FFT applied to this signal, detecting its amplitude as three distinctive peaks at 50, 120 and 350Hz.

The DFT is defined by the following following formula, where $x_0, ..., x_{n-1}$ are complex numbers:

$$f_j = \sum_{k=0}^{n-1} x_k e^{-(2\pi i/n)jk} \qquad j = 0, ..., n-1. \tag{6.8}$$

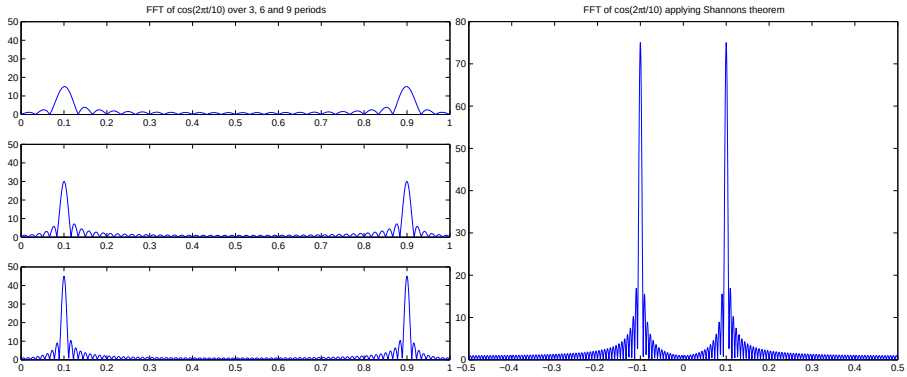
Applying this formula to an input signal will give its representation as harmonics of different frequencies. The FFT is therefore a suitable tool for breaking one signal into more simple, consistent signals. Note that the results from an FFT can vary dramatically depending on the number of measurements it covers and the number of periods in the signal which is represented, as shown in Figure 6.8. One must also make sure the signal's spectrum is entirely below $\frac{f_s}{2}$ (also known as the Nyquist frequency), since the sampling frequency must be at least twice the highest frequency component, where f_s is the sampling frequency. This is known as Shannon's Sampling Theorem [62].

As shown in Figure 6.8, the results from a FFT transform is a symmetrical curve. For signal processing, it is often common to manipulate this curve to only examine half of it. However, for simplification, this has not been done throughout this project.

The FFT of gait data has been used for analysis by [33, 34, 39, 60]. Note that FFT is a suitable transformation both for simplifying the data and to detect differences. It might therefore be an efficient method for data analysis.

Stride Frequency

The stride frequency is the basic information provided by the accelerometers. This can be found by detecting two local maximum values in the X axis acceleration data. It



(a) (b)

Figure 6.8: FFT calculations of $\cos(2\pi t/10)$ — (a) How the transformation changes when calculated over 3, 6 and 9 periods. (b) The transformation when applying Shannon's theorem [1].

will indicate two points where the foot with the Gait Collector attached is at its highest point while being moved forward, and the time between two of such points will be equivalent with the gait cycle time. One could possibly also calculate the stride frequency based on the acceleration data in Y axis by looking at the peaks as the foot touches the ground. However, by looking back at Figure 5.4, the Y axis data does not to appear be as consistent as the X axis data. When two points have been obtained, the stride frequency will be the number of measurements between these two points. If the number of measurements captured each second is known as well, it is also possible to get the base frequency in seconds.

The calculation of the stride frequency was part of Davis' process of finding differences in children and adult walking styles [16].

Gait Cycle Time

The calculation of gait cycle time was one of several biometric features measured by Davis and Taylor [17] to recognise human walking, and is one of the most fundamental descriptors in the human locomotion. To detect the period of a gait cycle, we will first have to calculate the stride frequency as shown earlier in this subsection. The gait cycle time T_l can then be calculated in seconds by the formula

$$T_l = \frac{1}{\text{stride frequency}} \tag{6.9}$$

Stride Length

Though Guta *et al* [23] do not describe the formula for detecting the stride length or the experiment where this discover was made in any close detail, they claim there is a strong correlation between the stride frequency and the stride length. They also say that the stride length increases with the stride frequency. In adition, Davis [16] claims it is possible to distinct between children and adult walkers by looking at the stride frequency and stride length together.

Average

When the stride frequency has been calculated as described earlier in this section, one can calculate its average. Schmidt *et al* calculated the average of the accelerometer output in their research to detect whether the device that was used was laying still on a table [60].

Therefore averaging the stride frequency over a smaller time period might tell whether the person is standing still, running or climbing stairs. It might also to be integrated into a process of verifying a person's identity.

6.3.3 Methods for Data Comparison

These methods are meant to take two sets of data and compare them. These methods will therefore give an indication of how similar the two compared datasets are.

Spectrum Subtraction

By first calculating the FFT transform of two spectrums A and B and then sorting them by their amplitude, it is possible to subtract the amplitude values from these two spectrums as a measure of comparison using the formula

$$Diff_i = |A_{amp_i} - B_{amp_i}| \tag{6.10}$$

Local Minima Detection

When comparing data from the same person, there will in most cases be several points in the datasets with similar characteristics, while they will differ in smaller or greater extent in other parts of the datasets. The parts with large similarities will therefore show up as distinctive minimas in the result of a spectrum subtraction, as described earlier in this section. By using the *polygamma function* to transform the dataset into its logarithmic domain, these minimas will be even more distinctive. The polygamma function ψ is defined as the derivative of the gamma function $\Gamma(x)$ by the formula[4]

$$\psi(x) = \frac{d(\log(\Gamma(x)))}{dx} \tag{6.11}$$

The gamma function $\Gamma(x)$ is defined by the integral[5]

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt \tag{6.12}$$

Figure 6.9 shows the results from two analysis processes; one where two datasets from the same person has been analysed, and one where gait data from two different persons has been analyses. As shown, there are a massive amount of local minimas in the comparison of gait data from the same person.

As shown in Figure 6.9, there will be a lot of tiny peaks and minima in the data, which will disturb the local minima detection algorithm. It is therefore necessary to apply a smooth filter on the data to remove these peaks. The "Moving Average smooth filter"[6] in Matlab is therefore used on the results from the polygamma function.

After smoothing the data, the local minima detection itself can be performed. For this, the local minima function created by Serge Koptenkos[7] was applied with a filter value of 35. The Matlab code for performing the local minima detection is listed in appendix B. The results from this minima detection is a new dataset with 1's in the areas with minimas in the compared graph and 0's elsewhere.

[4]Source: `http://www.mathworks.com/access/helpdesk/help/techdoc/ref/psi.html` (Last visited 29062005).

[5]Source: `http://www.wikipedia.org`

[6]Source `http://www.mathworks.com/access/helpdesk/help/toolbox/curvefit/ch_data6.html` (Last visited 29062005).

[7]Available from `http://www.mathWorks.com/matlabcentral/fileexchange/loadFile.do?objectId=3170&objectType=file` (Last visited 06052005)

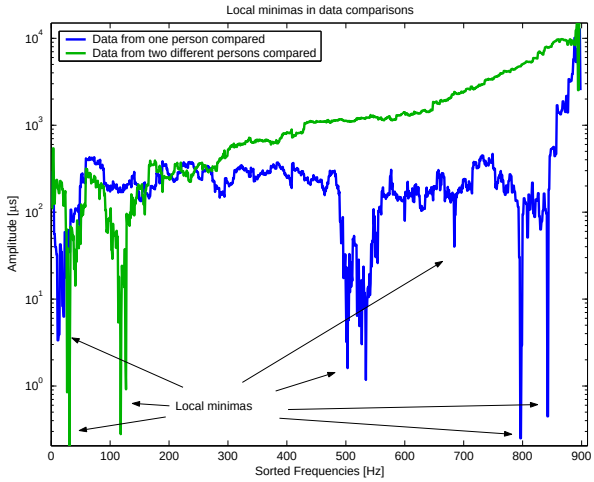


Figure 6.9: The amount of local minimas which occurs after a spectrum subtraction process is a possible method for analysis.

Euclidean Distance

The Euclidean distance is the straight line between two points in space. Its formula is defined as:

$$ED = \sqrt{\sum_{n=0}^{N} (x_n - y_n)^2} \tag{6.13}$$

This method of analysis was performed by [48] to classify images and by [67] to classify gait patterns. This method would both compare the two datasets and give a similarity score. After such a method for comparing, any further similarity detection is therefore unnecessary.

Analysis of Variance (ANOVA)

ANOVA is a collection of statistical methods which splits the observed variance into different parts. This method for analysis was used by Lee and Gimson [39] to get the p-value. By selecting all features with a p-value below a given threshold, they were able to rank different gait features after a FFT had been applied to the data.

Mahalanobis Distance

The Mahalanobis distance is a statistical method based on the correlation between variables. It tries to find the similarities between a known sample set and an unknown one. It is somewhat similar to the Euclidean distance, but also uses the correlations of the datasets. Given a set of values $\mu = (\mu_1, \mu_2, \mu_3, ...\mu_p)$ and the covariance matrix Σ for the vector $x = (x_1, x_2, x_3, ...x_p)$, the Mahalanobis distance is defined as[8]:

$$D_M(x) = \sqrt{(x - \mu)' \Sigma^{-1} (x - \mu)} \tag{6.14}$$

The Mahalanobis distance was used by Lee and Grimson [39] to measure the similarities between two data sequences and to remove bias of parameters with large variance. It

[8] Source: `http://www.wikipedia.com.`

was also used by Davis and Taylor [17] to determine which class a given walking feature would fit into.

Convolution

Convolution [63] is a mathematical way of combining two signals into a third signal. The resulting signal will imply the amount of overlap between the two signals. Given the two functions f and g, the convolution between the two functions is written $f * g$, and defined as[9]:

$$(f * g)(t) = \int f(\tau)g(t-\tau)d\tau \tag{6.15}$$

6.3.4 Methods for Similarity Determination

To be able to use the proposed methods in an authentication system, it is necessary to get a similarity score when two sets of gait data are compared. The methods listed here will take the results after two datasets have been compared and return a score telling how equal the two datasets are.

Area Calculation

The calculation of the integral of a curve will give the area between the curve and the X axis given that all values in the curve are larger than zero. This statement is true if the curve was calculated using the spectrum subtraction method described in Subsection 6.3.3. Then the results will be an indication of how closely related these two datasets are. The result will be 0 if the two compared datasets are completely identical, and grow toward infinity as they have lesser and lesser in common.

Standard Deviation

By calculating the standard deviation for a set of data from one of the three axes, it is possible to tell whether there are movement in that direction. The size of the standard deviation might also be distinctive from one person to another. This calculation was used by Schmidt *et al* [60] to detect movement in a device fitted with accelerometers.

Given a dataset $x_1, ...x_N$, it is possible to calculate its standard deviation σ using the formula[10]:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2} \tag{6.16}$$

Interquartile Range

Sorting a dataset will make it possible to calculate the distance from the values in the first quartile to the values in the third quartile, whereas each quartile contains 25% of the data. This might be distinctive from one person to another, but will also tell whether the person is moving or standing still. Schmidt *et al* [60] used this method to determine whether a device fitted with accelerometers was lying still or being moved.

[9] Source: `http://www.wikipedia.org`.
[10] Source: `http://www.wikipedia.org`.

7 The Selected Data Analysis Methods

The previous chapter listed several different methods for pre-processing several sets of gait data, analysing them, comparing the analysed data sets, and finally detecting a similarity score. Though many of these methods has been used successfully in various research, not all are as relevant for this research. There are several reasons for this. One is that almost all previous research has been analysing kinematic movement from various sources of kinematic video data. There has been very limited work on identifying persons based on kinetic acceleration data prior to this project. Therefore a vast part of this research have involved making an assessment of which methods are the most suitable to treat the gathered kinetic data.

7.1 The Methods used for Data Analysis

This section describe how several of the previously described analysis methods were combined using the framework from Section 6.1 to create methods for gait data analysis. An overview of the analysis methods developed is shown in Figure 7.1.

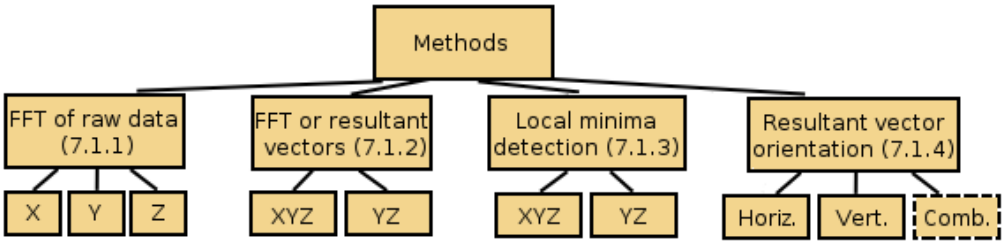


Figure 7.1: A three illustrating the various methods developed for analysing gait data. The stippled method was only performed in the preliminary experiment.

7.1.1 Comparing FFT of the Raw Data

These are the three simplest analysis methods, and therefore the most obvious ones to start with. Using the terminology described earlier, the methods are as following:

Data pre-processing and analysis: None of the pre-processing methods are being used. The FFT transform from Section 7.2 is respectively being applied to the acceleration data in each of the axes X, Y and Z. Each axis is analysed separately.

Comparison: The results from the FFT calculations are sorted by their amplitudes, and the spectrum subtraction method from Section 6.3.3 is applied to get a vector representation of their similarity.

Similarity score calculation: The area beneath the results from the comparison is calculated using the area calculation method described in Section 6.3.4.

These methods consist of comparing the FFT for each of the Gait Collector's X, Y and Z axis directly. Note that it is the acceleration in the *Gait Collector's* XYZ plane that is being analysed. This plane is not oriented in the same way as the physical XYZ plane (recall Figure 5.3 back on page 21). These methods are therefore vulnerable for validity flaws due to the fact that how the Gait Collector is aligned while attached to a person's leg will affect the measurement. Because of this essential weakness, it is not very applicable method in real life unless the Gait Collector is encapsulated in such a way that it can not be attached to the leg in any other manner than the correct one.

7.1.2 Comparing FFT of the Resultant Vectors

These two methods are improvements of the direct comparison described in the previous section (Section 7.2) that tries to overcome the problem with the Gait Collector's orientation. The methods are as following:

Data pre-processing and analysis: The pre-processing part of these methods involve calculating the XYZ and YZ resultant vectors as described in Section 6.3.1. This results in two methods for analysis. The resultant vectors are analysed separately using the FFT transform from Subsection 7.2.

Comparison: The resulting datasets from the FFT calculations are sorted by their amplitudes and the spectrum subtraction method from Section 6.3.3 is applied.

Similarity score calculation: The area beneath the resulting graphs from the comparisons are calculated using the area calculation method from Section 6.3.4.

Someone might ask why looking at the YZ resultant vector and not the XY and XZ resultant vectors? This is merely a question of time and resources. This would require more resources than what were available for this research, since all analysis methods which are based on the resultant vector calculations would have to been performed on the new resultant vectors as well. Therefore, the analysis of the XY and YZ vectors are left open to future research. Who knows—maybe they will give even better results than the analysis of the YZ resultant vector?

7.1.3 Detecting Local Minima after Calculating FFT from Resultant Vector

These two methods are somewhat similar to the analysis of the resultant vector described in the previous section. By studying the various graphs after the XYZ and YZ resultant vectors were calculated one can observe a vast amount of spikes, or local minima in the graphs where data from the same person are compared (as shown previously in Figure 6.9). These minima occur in areas of the two sorted FFT transforms where parts of these transforms have similarities in both amplitude and frequencies. When the transforms are subtracted, these equalities show up as a serie of local minimas. Though this observation is very interesting, there were not much time available to determine exactly *why* these minimas occur. This is therefore left as an idea for further research. The methods detecting local minima are as following:

Data pre-processing and analysis: The XYZ and YZ resultant vectors are calculated as described in Section 6.3.1. The two resultant vectors are analysed separately.

After the resultant vector has been calculated, the FFT transform from Section 7.2 is applied to each of the vectors.

Comparison: The resulting datasets from the FFT calculation are sorted by their amplitudes, and the spectrum subtraction method described in Section 6.3.3 is applied. Then the local minima detection itself can be applied. By using the local minima detection from Section 6.3.3 the amount of spikes in the graphs are detected. This gives new vectors with 1's in the corresponding areas of the FFT transform where there are minimas and 0's elsewhere.

Similarity score calculation: The areas beneath the resulting vectors from the local minima detection are calculated using the area calculation method from Section 6.3.4. However, for the other described analysis methods, a small similarity score would equal great similarity. For this method, a large score will equal great similarity. To make the results more consistent with the rest of the analysis methods, the inverse of the similarity score is calculated using the equation

$$New\ similarity\ score = \frac{1}{\text{Old similarity score}} \times 10,000 \tag{7.1}$$

The result was multiplied with 10,000 to avoid too small scores.

7.1.4 Comparing FFT of the Resultant Vectors' Orientation

The last three methods examines how the XYZ axis resultant vector changes its horizontal and vertical orientation with the Gait Collector over time. The analysis is performed as following:

Data pre-processing and analysis: The XYZ resultant vectors is first calculated as described in Section 6.3.1. Then the method for determining this resultant vector's orientation is applied (which also is described in Section 6.3.1). These horizontal and vertical orientation are further analysed separately. In addition of comparing the orientation of the resultant vector. It was also attempted to combine the information regarding the horizontal and vertical orientation with the information regarding the length of the vector in a third analysis method. Here, the results from the horizontal and vertical orientation are multiplied with the XYZ resultant vector itself to check whether this give a better results than any of the three methods separately. The FFT from Section 7.2 is calculated for each of the three methods to detect the frequencies of changes of the data over time.

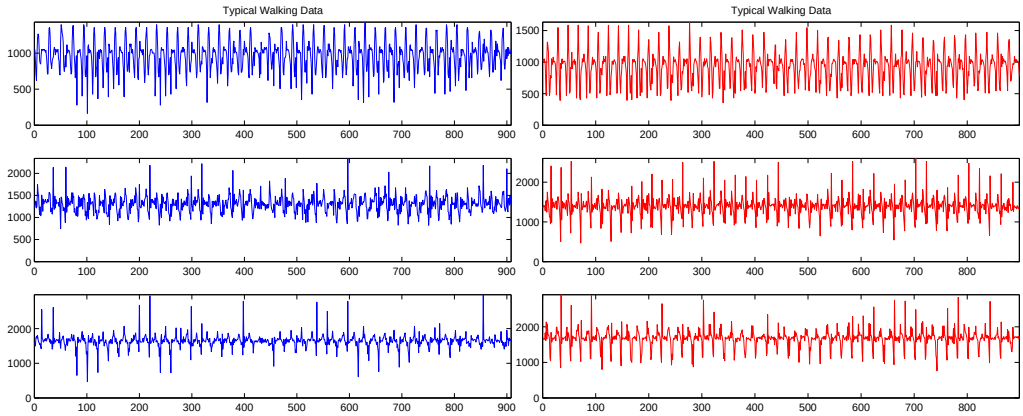
Comparison: The resulting datasets from the FFT calculations are sorted by their amplitudes, and the spectrum subtraction method from Section 6.3.3 are applied.

Similarity score calculation: The area beneath the resulting graphs from the comparisons a calculated using the area calculation from Section 6.3.4.

7.2 Graphical Example of Data Analysis

To ease the understanding of how the methods proposed in this chapter were used during data analysis, a step-by-step example of one of the analysis processes is given. This example shows a graphical representation of how the raw data from the acceleration in X axis from Section can be used to verify a person's identity through data analysis.

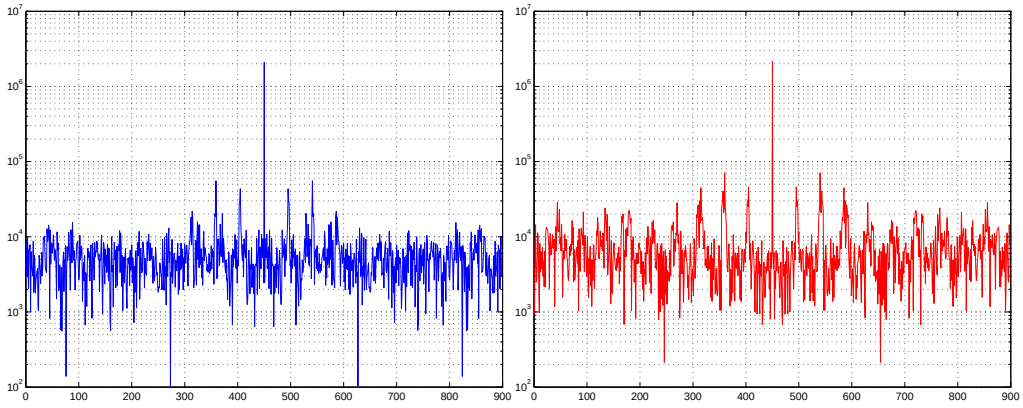
The first step in the process is to get a set of raw data from two different persons—let's say from the two volunteers Alice and Eve. The datasets shown in this example were obtained as part of the preliminary experiment described later in Section 9.1. The gait data collected from Alice and Eve represents their walking features in X, Y and Z axes and are shown in Figure 7.2.



Alice's raw gait data Eve's raw gait data

Figure 7.2: Raw gait data collected from Alice and Eve. The three graphs represents the leg's acceleration in X axis at the top, Y axis in the middle and Z axis at the bottom as the two persons walks for roughly one minute.

The FFT transform from Section is applied to the parts of the gait data representing the acceleration in X axis and the results are shown in Figure 7.3. This transform shows how the frequencies of the gait data's amplitude changes rather than how the amplitude changes over time.

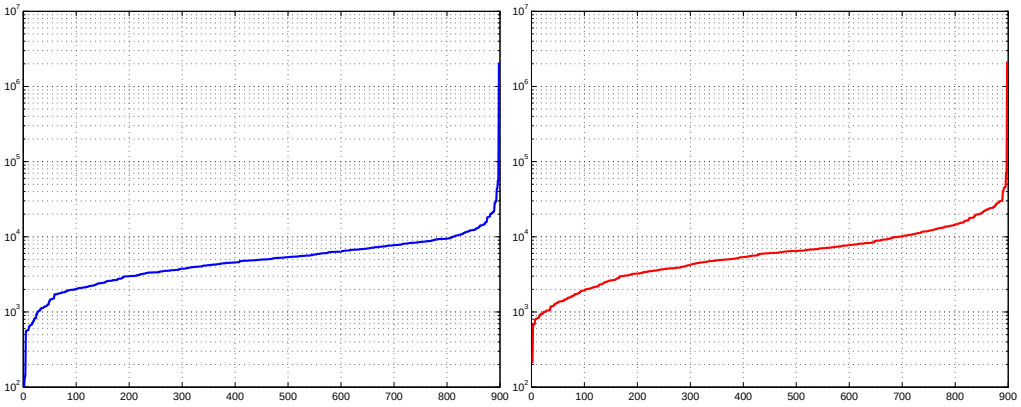


FFT of Alice's X axis gait data FFT of Eve's X axis gait data

Figure 7.3: The FFT transform has been applied to Alice and Eve's X axis acceleration data.

After the FFT transform has been calculated for Alice and Eve, the resulting datasets are sorted by their amplitude. By doing this the noise in the FFT transform is somewhat reduced, and the smooth vectors shown in Figure 7.4 are achieved.

After the FFT transforms has been sorted, they are ready to be compared. This is done simply by subtracting the two curves from each other, and the result is shown in Figure



The sorted FFT of Alice's X axis gait data The sorted FFT of Eve's X axis gait data

Figure 7.4: The FFT transform for Alice and Eve's acceleration data has been sorted by their amplitude.

7.5. This curve is a representation of the similarities in Alice and Eve's gait. By calculating the area between this curve and the X-axis, it is possible to achieve a score representing how similar their walking pattern is. If their walking features were completely identical, they would get a similarity score of 0. The comparison of Alice and Eve's walking features, however are not completely identical. Calculating the area of their comparison gives a similarity score of 690.03. Section 9.1 will later in this thesis show that when using this method for comparing gait data from the same person, one will obtain values closer to 300, which indicates that a score close to 700 is a good score.

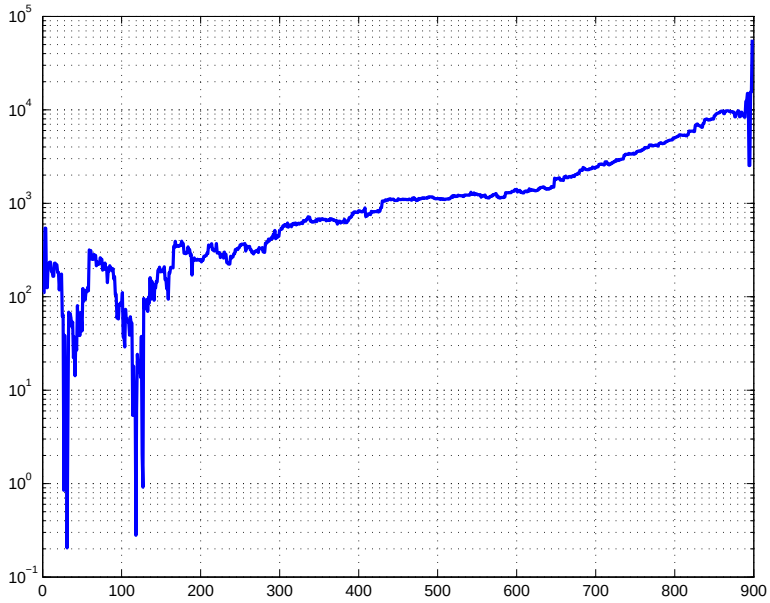


Figure 7.5: The result after subtracting Alice and Eve's sorted FFT datasets from each other.

This was a short introduction to the principle behind the data analysis described earlier in this chapter. Though there are some degree of variation in the various analysis methods, they all follow these basic steps.

8 Experiment Design

The process of determining which analysis methods that are the most suitable was done by performing three experiments as described in Table 8.1.

Preliminary Experiment	
Motivation	Chapter 6 describes a large amount of potential ways to analyse data collected with the Gait Collector. This experiment was performed to ease the process of designing the analysis methods described in Chapter 7.
Number of participants	Three persons participated.
Experiment outline	Two sets of gait data were collected from each person. They walked roughly 35 meters at a straight line, turned and walked back. Then they waited roughly one hour for the next gait attempt. This gave a total of 6 datasets to compare.
Large Scale Experiment	
Motivation	A more large scale experiment was performed to observe how the methods designed during the preliminary experiment worked on a larger set of gait data.
Number of participants	20 persons participated; 9 female and 11 male.
Experiment outline	Each person made one gait attempt walking circa 35 meters, turning and walking back. By looking for the delay in the acceleration data where they turned, it was possible to manually split each dataset in two parts. Each of them were compared, resulting in 40 datasets. Since the Gait Collector was not removed from the persons' legs between each of their obtained datasets, it was a possible validity problem. The analysis methods might detect *where* on the leg the Gait Collector was attached rather than *how* the persons walked.
Long Term Experiment	
Motivation	This experiment tested how the developed methods worked on data collected over a longer period of time. This also made it possible to examine the potential validity problem from the large scale experiment.
Number of participants	Four persons participated.
Experiment outline	Each person made five gait attempts over a period of one week. This gave 20 datasets to compare.

Table 8.1: An overview of the three experiments that were performed.

8.1 Rules of the Game

In order to get valid and reliable data for the experiments, a set of rules were defined. Not only did these rules create more reliable and valid data—they also simplified the analysis process, since they made it possible to control some of the factors which might have interfered with the results. The rules defined for the experiments are as follows:

Information Registered: Some personal information were registered on the participants of the large scale experiment. This made it possible to detect whether persons with similar physical characteristics have similar gait features. To protect the persons' privacy, their names were not registered—they only received a unique ID number, so they could be distinguished from each other. The registered information was the persons' gender, height, weight, foot size and leg length.

Gait Collector Mounting: The Gait Collector had to be mounted in a similar way on all participants. This implied mounting it on the right leg as shown in Figure 8.1. It was mounted as close to the ankle as possible, and tighten firmly enough to reduce movement relative to the leg.

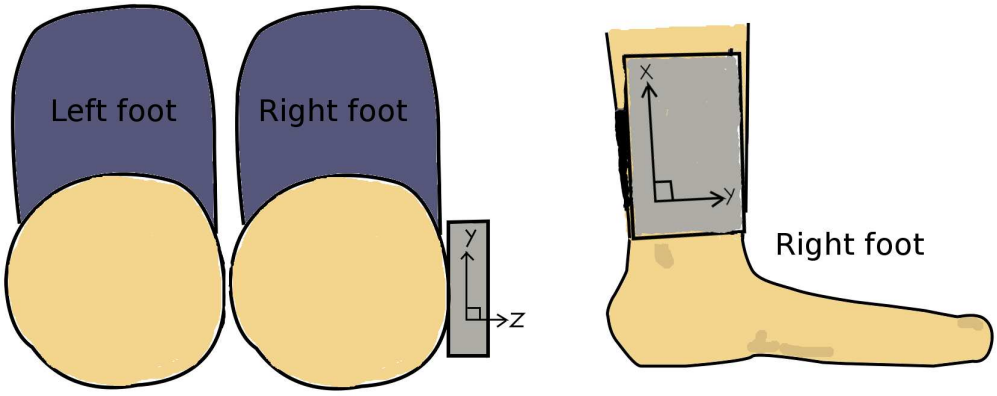


Figure 8.1: An ideal horizontal and vertical alignment of the Gait Collector to the foot. The Gait Collector is attached at the outside of the right leg just above the ankle.

Walking Distance: The participants walked a total distance of roughly 70 meters. They performed normal walking at a plane and tiled surface, walking half the distance in a straight line, turned around and walked back.

Population: The experiments were performed using volunteers from the college's campus. The majority of the subjects were in other words between 20 and 40 years old.

9 The Performed Experiments and their Results

As described in Chapter 8, three experiments were performed during this project. The results from each experiment are described in this chapter.

When the experiments are presented, they show how the similarity scores obtained when comparing gait data from the same person relates to comparing gait data from two different persons. There will almost always be occasions where a comparison of different persons will get better similarity scores than when comparing data from the same person. This gives overlaps in the two groups as shown in Figure 9.1, which results in false acceptances and false rejects. By comparing the amount of such overlaps and the FAR and FRR rates of a method, it is possible to determine which method is more suitable than the other. When the FAR and FRR rates are presented in this chapter, the FRR rates illustrates the rate of wrongful rejections given a threshold with no wrongful acceptances. The presented FAR rates shows what rate of wrongful acceptances the method would yield if no-one were wrongful rejected. Figures 9.1 and 9.2 illustrates this concept more clearly. The formulas for calculating the FAR and FRR are thoroughly explained in Section 6.2.

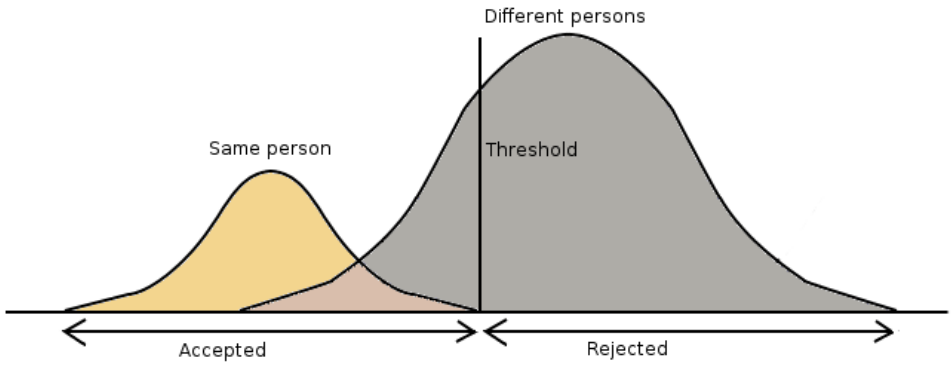


Figure 9.1: A threshold value where no-one are wrongfully rejected, but many are wrongfully accepted.

9.1 The Preliminary Experiment

Chapter 7 describe the various data analysis methods that were created during this research in detail. The purpose of the preliminary experiment was to create these methods.

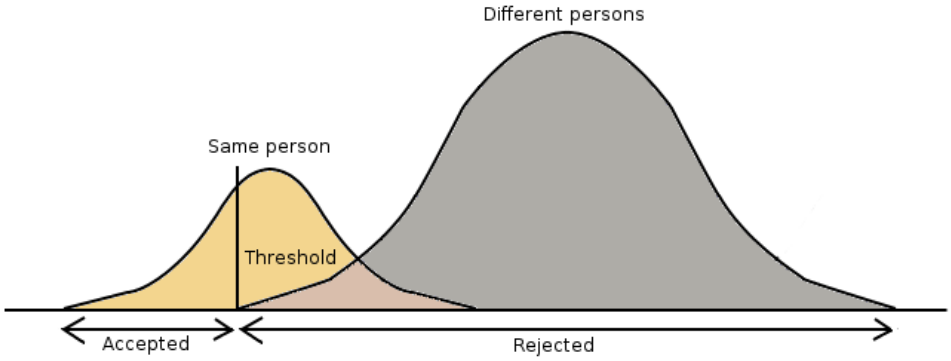


Figure 9.2: A threshold where no-one are wrongfully accepted, but many are wrongfully rejected.

Because of this, the physical characteristics of the participants that participated in the experiment were not registered. The experiment was based on data collected from three different persons, each performing two attempts of normal walking. Throughout this section, these persons have been named *Alice*, *Bob* and *Eve*. The results from each analysis method have been presented as follows:

A similarity graph where the vectors obtained after the data comparisons are plotted. This shows how comparisons of the same person relates to comparisons of different persons. It is here desirable to have comparisons of different persons clustered in other areas of the graph than comparisons of the same persons.

A similarity table showing the similarity scores of each possible comparison of Alice, Bob and Eve's gait data. This is shown in Table 9.1. Comparisons of the same person are grouped in the upper rows of the table, while comparisons of different persons are grouped in the bottom rows. If there are no overlapping in these groups for a given method, it indicate a suitable method for analysis.

A FAR and FRR table where FAR and FRR values for the various methods are compared. How these rates were obtained are explained in the beginning of this chapter. This is shown in Table 9.2.

A ROC curve illustrating the relations between the FAR and FRR values for the various methods for all possible threshold values. It therefore illustrate what tradeoff that has to be made when choosing the appropriate threshold for any given method. This is shown in Figure 9.13.

	Raw Data			Resultant Vector		Local Minima		Orientation		
	X	**Y**	**Z**	**XYZ**	**YZ**	**XYZ**	**YZ**	**Horizontal**	**Vertical**	**Combined**
Data from the same person	195.44	272.26	205.66	293.36	265.69	33.90	59.52	5.92	6.04	72.69
	266.8	278.73	458.4	348.8	311.14	40.98	64.94	7.62	6.68	120.65
	620.96	385.82	542.39	599.39	503.48	47.85	76.34	13.16	11.66	228.88
Data from different persons	**388.89**	**301.34**	**212.24**	605.24	537.22	51.81	**42.55**	**3.82**	**7.16**	**28.87**
	397.45	395.04	**268.26**	777.47	696.06	57.80	**52.91**	**6.22**	**10.46**	**44.21**
	436.92	480.24	**298.73**	781.0	788.22	58.48	**62.50**	**7.69**	11.86	**92.47**
	511.25	500.74	**421.08**	919.49	929.04	59.88	**66.23**	**7.73**	12.73	**108.79**
	529.78	850.25	827.04	1054.5	974.51	68.49	**74.63**	**8.80**	12.87	**152.9**
	610.18	1078.1	893.68	1060.5	1132.7	68.49	86.21	**10.52**	13.86	**164.71**
	628.03	1193.6	1005.2	1193.0	1320.3	69.44	96.15	**11.06**	19.43	**180.84**
	690.03	1230.2	1260.4	1466.0	1465.2	90.91	98.04	**11.80**	19.67	272.75
	823.64	1292.5	1286.6	1475.2	1590.3	96.15	105.26	**12.51**	20.11	337.26
	851.79	1418.6	1294.6	1777.4	1860.1	105.26	138.89	15.44	23.56	381.45
	1086	1563.6	1415.9	1896.8	1922.8	156.25	138.89	17.72	24.26	409.08
	1186.5	1642.6	1671.9	2185.5	2192.6	200.00	217.39	20.98	24.40	501.33

Table 9.1: The **similarity table** showing similarity scores obtained by various analysis methods tested in the preliminary experiment. Small similarity scores is the same as great equality for any given comparison. Results after comparing gait data from the same person is displayed in the upper rows, while results after comparing gait data from two different persons are displayed in the bottom rows. When a comparison of two different persons have given a better score than when comparing data from the same person, it has been **emphasised**. An ideal analysis method will obviously not have any of such overlapping.

9.1.1 Comparing FFT of the Raw Data

As described in Section 7.2, these methods analyses the FFT of each of the axes X, Y and Z directly without any other methods for data pre-processing. This analysis is therefore vulnerable to validity biases, since the Gait Collector's orientation will affect the results from the analysis.

Direct Comparison of FFT Data from X Axis Acceleration

The FFT of the movements in X axis shows the frequency of how the foot is moved up and down during locomotion. As shown in the first column of Table 9.1, there is a comparison of data from the same person that gets a fairly large similarity score. It is impossible to tell whether this difference is due to Alice, Bob and Eve walking in a similar manner, or because the Gait Collector was misaligned during some of the measurements.

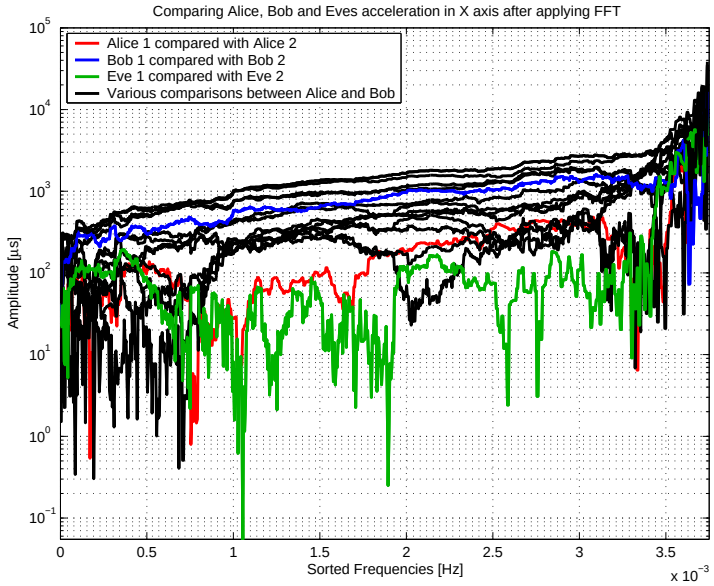


Figure 9.3: The result of all possible comparisons of the FFT of Alice, Bob and Eve's acceleration in the X axis gait data. The coloured graphs are Alice, Bob and Eve compared with themselves, while the black graphs are different comparisons among them.

A result of this experiment is that a threshold value larger than 621 would give a FRR of 0%, but it would also give a FAR rate of at least 40%.[1]. On the other hand, a threshold value less than 388 would give a FAR value of 0%, but also a FRR value of at least 6.7%[2].

Direct Comparison of FFT Data from Y Axis Acceleration

The FFT of the movements in Y axis shows how the foot is moved forward and backward during locomotion.

What can be observed in the second column of Table 9.1 is that there are very small differences in the comparisons of data from the same person and from different persons. As a result, setting a threshold value larger than 386 would give a FRR of 0% and a FAR of at least 6.7%[3]. A threshold value smaller than 300 on the other hand, would give a

[1] FAR: $\frac{6 \text{ false acceptances}}{15 \text{ comparisons}} = 40\%$ FRR: $\frac{0 \text{ false rejections}}{15 \text{ comparisons}} = 0\%$
[2] FAR: $\frac{0 \text{ false acceptances}}{15 \text{ comparisons}} = 0\%$ FRR: $\frac{1 \text{ false rejections}}{15 \text{ comparisons}} = 6.7\%$
[3] FAR: $\frac{1 \text{ false acceptances}}{15 \text{ comparisons}} = 6.7\%$ FRR: $\frac{0 \text{ false rejections}}{15 \text{ comparisons}} = 0\%$

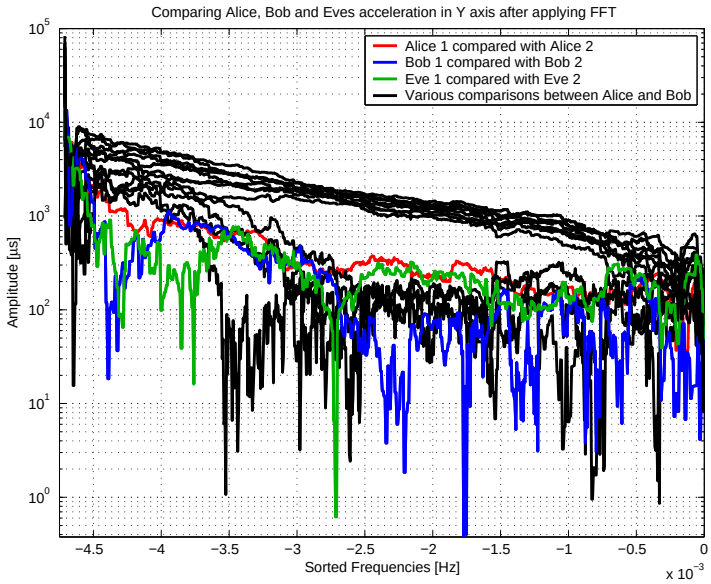


Figure 9.4: The result of all possible comparisons of the FFT of Alice, Bob and Eve's acceleration in Y axis gait data. The coloured graphs are Alice, Bob and Eve compared with themselves, while the black graphs are different comparisons among them.

FAR of 0% and a FRR of at least 6.7%[4].

Direct Comparison of FFT Data from Z Axis Acceleration

The FFT of the movements in Z axis shows the frequency of how the foot is moved sideways during locomotion.

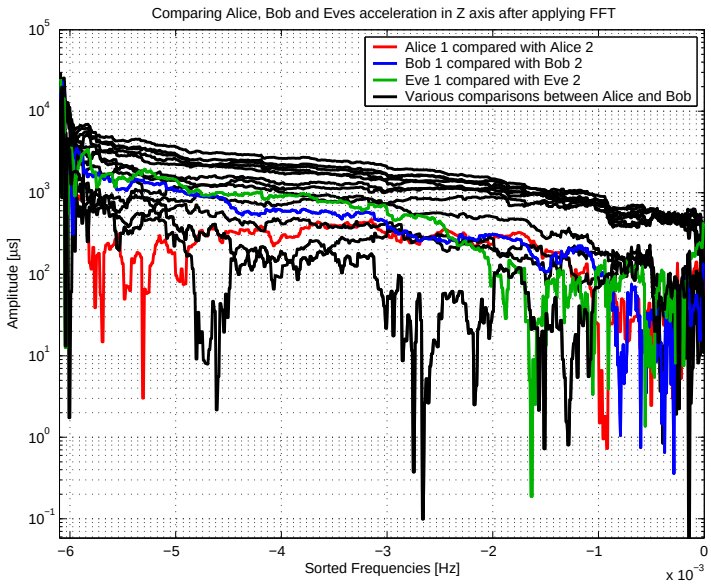


Figure 9.5: The result of all possible comparisons of the FFT of Alice, Bob and Eve's acceleration in Z axis gait data. The coloured graphs are Alice, Bob and Eve compared with themselves, while the black graphs are different comparisons among them.

[4]FAR: $\frac{0\ false\ acceptances}{15\ comparisons} = 0\%$ FRR: $\frac{1\ false\ rejections}{15\ comparisons} = 6.7\%$

The third column of Table 9.1 show three comparisons of data from different persons receiving similarity scores lower than 300. This indicate that two different persons with similar walking features might have been compared in various ways. Derived from the table, we can see that a threshold value larger than 543 would give a FRR of 0% and a FAR of at least 26.7%[5] while a threshold value smaller than 212 would give a FAR of 0% and a FRR of at least 13.3%[6].

9.1.2 Comparing FFT of the Resultant Vectors

As described in Section 7.1.2, these methods analyses the FFT of the XYZ and YZ resultant vectors. They therefore have less validity biases than the previously described direct comparison of X, Y and Z axes, since the orientation of the Gait Collector will not affect the results in the same extent.

Comparing FFT of XYZ Resultant Vectors

This analysis looks at the acceleration data in all three axes X,Y and Z at once. The resultant vector will be a representation of the foot's movement in the horizontal and vertical plane, creating a summary of the acceleration forces created during the movement of the foot.

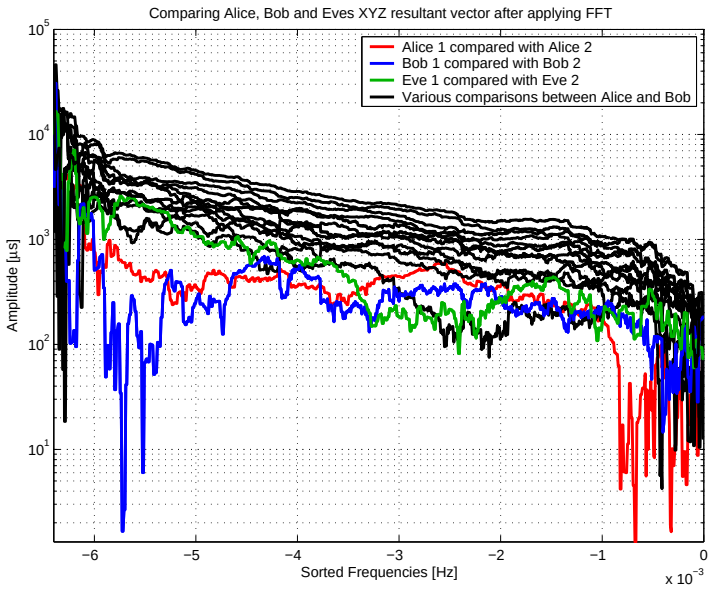


Figure 9.6: The result of all possible comparisons of the FFT of Alice, Bob and Eve's XYZ resultant vector. The colored graphs are Alice, Bob and Eve compared with themselves, while the black graphs are different comparisons among them.

As shown in the fourth column of Table 9.1, no comparison between two different persons yield in a similarity score smaller than when two datasets from the same person is compared. A result of this is that a threshold between 600 and 605 would give FAR and FRR of 0%[7]. However, though there are no overlapping, the difference in the two closest similarity scores from the two groups is not very large.

[5] FAR: $\frac{4 \text{ false acceptances}}{15 \text{ comparisons}} = 26.7\%$ FRR: $\frac{0 \text{ false rejections}}{15 \text{ comparisons}} = 0\%$
[6] FAR: $\frac{0 \text{ false acceptances}}{15 \text{ comparisons}} = 0\%$ FRR: $\frac{2 \text{ false rejections}}{15 \text{ comparisons}} = 13.3\%$
[7] FAR: $\frac{0 \text{ false acceptances}}{15 \text{ comparisons}} = 0\%$ FRR: $\frac{0 \text{ false rejections}}{15 \text{ comparisons}} = 0\%$

Comparing FFT of YZ Resultant Vectors

This analysis looks at the acceleration data in the Y and Z axes as one. It is therefore a reflection of how the foot moves in the horizontal plane, and will not be affected by the horizontal orientation of the Gait Collector.

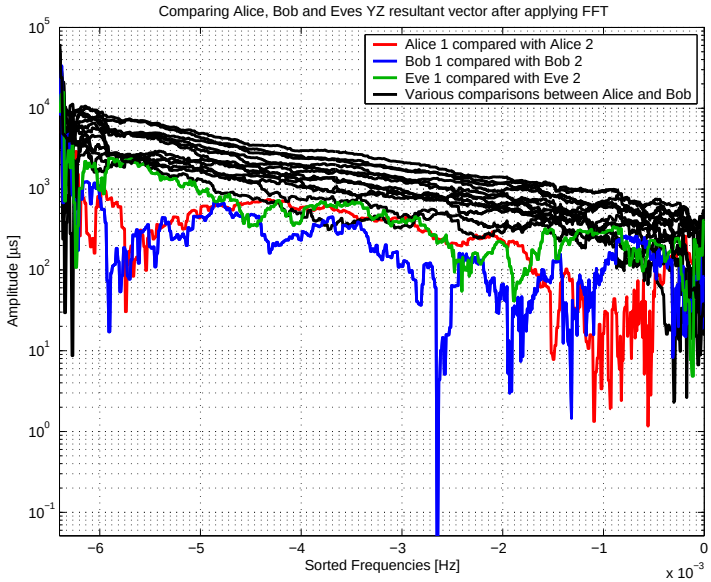


Figure 9.7: The result of all possible comparisons of the FFT of Alice, Bob and Eve's YZ resultant vector. The colored graphs are Alice, Bob and Eve compared with themselves, while the black graphs are different comparisons among them.

As shown in the fifth column of Table 9.1, just as with the calculation of the XYZ resultant vector, no comparison between two different persons yield a similarity score smaller than when two datasets from one person are compared. A result of this is that a threshold value between 504 and 537 would give FAR and FRR of 0%[8].

9.1.3 Detecting Local Minima after Calculating FFT from Resultant Vector

As described in Section 7.1.3, several spikes will appear in the vector that is a result of two resultant vectors from the same person are being subtracted. These methods therefore searches for local minimas after two resultant vectors has been compared.

Local Minima Detection on the FFT of XYZ Resultant Vector

The input to the local minima detection algorithm are shown in Figure 9.6. After the local minima detection have been applied, no comparisons between the same person gives similarity scores larger than comparisons between different persons. These results are shown in the sixth column of Table 9.1. A threshold value of between 52 and 48, will achieve a FAR rate and a FRR rate of 0%[9]. The two comparisons in Table 9.1 that received matching scores of 47.85 and 51.81 are shown in Figure 9.8. These are the two comparisons in each of the two groups that got matching scores closest to each other. What might be observed in this figure is that though this method of analysis gives FAR and FRR of 0% with the appropriate threshold value, the method is detecting a vast

[8] FAR: $\frac{0 \text{ false acceptances}}{15 \text{ comparisons}} = 0\%$ FRR: $\frac{0 \text{ false rejections}}{15 \text{ comparisons}} = 0\%$
[9] FAR: $\frac{0 \text{ false acceptances}}{15 \text{ comparisons}} = 0\%$ FRR: $\frac{0 \text{ false rejections}}{15 \text{ comparisons}} = 0\%$

amount of local minima in the comparison between two different persons which might be avoided with a more improved method of analysis. This would most likely give an even better method of analysis.

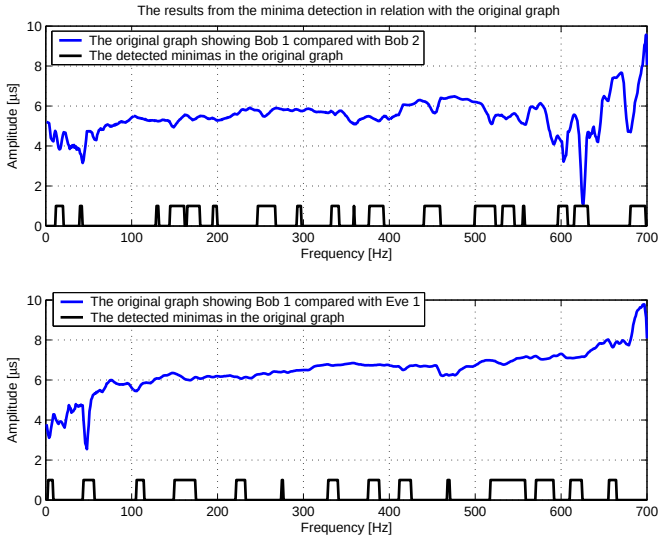


Figure 9.8: The results from the local minima detection on the comparison between Bob's two walking attempts in the upper graph and between Bob and Eve's first walking attempt in the lower graph. The blue curves are the result after subtracting the two compared spectrums after applying the FFT transform on the XYZ resultant vector, while the black areas at the bottom indicates the positions on the blue curves where local minimas were detected.

Local Minima Detection on FFT of the YZ Resultant Vector

The input to the local minima detection algorithm is shown in Figure 9.7. As shown in Table 9.1, there are occasions where comparisons between different persons give better matches than comparisons between the same person. Therefore, a threshold larger than than 76 would give a FRR of 0% and a FAR rate of at least 33.3%[10]. In the same manner, a threshold smaller than 43 would give a FRR of 20% and a FAR of 0%[11].

A graph visualising the comparison between the comparisons giving similarity scores of 131 and 235 from Table 9.1 is shown in Figure 9.9. This graph most likely illustrates why this method of analysis fails. Here, a large amount of small and therefore unnecessary spikes are detected in the comparison of different persons, which in a large extent reduces their similarity score. This method must therefore be improved before it is applicable in a more realistic environment.

9.1.4 Comparing FFT of the Resultant Vectors' Orientation

As described more closely in Section 7.1.4, this methods looks at how the XYZ resultant vector changes its horizontal and vertical orientation with the Gait Collector.

Comparing the FFT of the Horizontal Orientation of the XYZ Resultant Vector

The horizontal orientation of the XYZ axes' resultant vector in relation with the Gait Collector is reflected by the α angle shown in Figure 6.3 back on page 29.

As shown in the eight column of Table 9.1, there are several comparisons of datasets

[10] FAR: $\frac{5 \text{ false acceptances}}{15 \text{ comparisons}} = 33.3\%$ FRR: $\frac{0 \text{ false rejections}}{15 \text{ comparisons}} = 0\%$

[11] FAR: $\frac{0 \text{ false acceptances}}{15 \text{ comparisons}} = 0\%$ FRR: $\frac{3 \text{ false rejections}}{15 \text{ comparisons}} = 20\%$

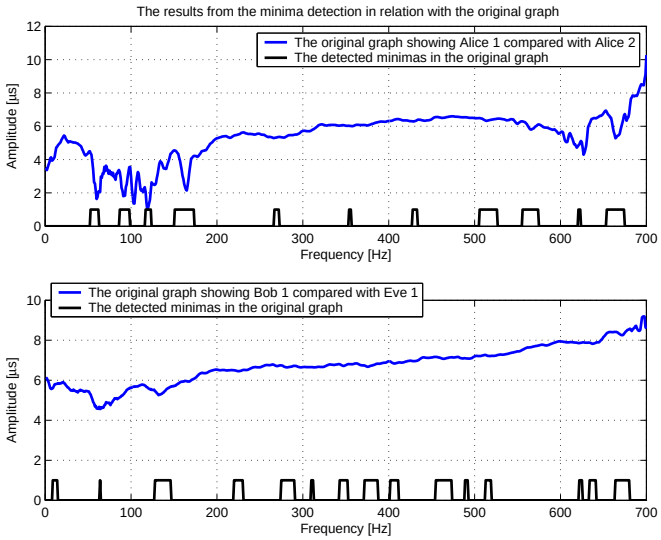


Figure 9.9: The results from the local minima detection on the comparison between Bob's two walking attempts in the upper graph and between Bob and Eve's first walking attempts in the lower graph. The blue curves are the result after subtracting the two compared spectrums after applying the FFT transform on the YZ resultant vector, while the black areas at the bottom indicates the positions on the blue curves where local minimas were detected.

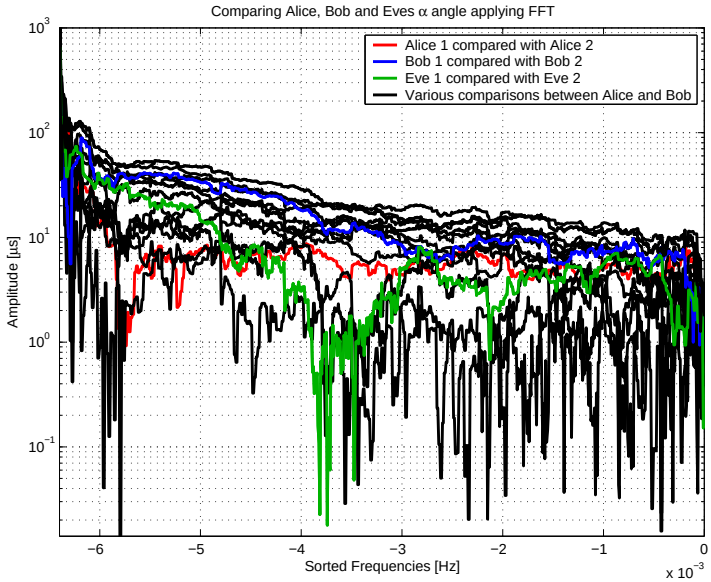


Figure 9.10: The result of all possible comparisons of how Alice, Bob and Eve's XYZ resultant vector is aligned with the Gait Collector's horizontal axis. The coloured graphs are Alice, Bob and Eve compared with themselves, while the black graphs are different comparisons among them.

coming from two different persons which achieves better similarity scores than when comparing datasets from the same person. This might indicate that there are not a very great extent of individuality in how the α angle changes related to the horizontal plane of the Gait Collector. A threshold less than 3 would give a FAR of 0%, but it would also

give a FRR of at least 20%[12]. At the same time, a threshold larger than 13 would give a FRR of 0%, but it would also give a FAR of at least 60%[13].

Figure 9.10 shows, just as in the table, that there are not a great extent of characteristics in the comparisons among the same person compared with comparisons among different persons, and that such a method for verifying a person's identity therefore might not be as usable as the other methods.

Comparing the FFT of the Vertical Orientation of the XYZ Resultant Vector

The vertical orientation of the XYZ resultant vector in relation with the Gait Collector is reflected by the β angle. The results from this comparison are shown in Figure 9.11.

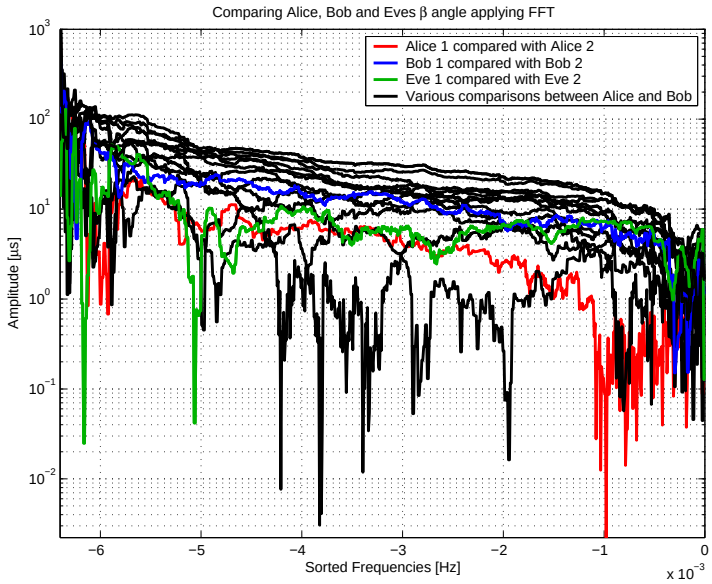


Figure 9.11: The result of all possible comparisons of how Alice, Bob and Eve's XYZ resultant vector is aligned with the Gait Collector's vertical axis. The coloured graphs are Alice, Bob and Eve compared with themselves, while the black graphs are different comparisons among them.

Figure 9.11 shows, just as in the ninth column of Table 9.1, that there are some overlapping in the comparison of the same person and of different persons. Using the resultant vector's alignment to verify a person's identity might therefore not be very usable compared with other methods. Though there are only a couple of situations in which comparisons among datasets from two different persons gives better similarity scores than when comparing datasets from the same person, the various scores are very close to each other. As a result, a threshold of at least 11.7 would give a FRR of 0%, but also a FAR of at least 13.3%[14]. At the same time, a threshold less than 7 would give a FAR of 0% and a FRR of at least 6.7%[15]. Using this information compared with the information in the graph 9.11, it appears to be a greater extent of individuality in how the resultant vector is aligned with the Gait Collector's vertical axis than with the horizontal axis. Though this might be a more promising result, it is still worse than only

[12] FAR: $\frac{0 \text{ false acceptances}}{15 \text{ comparisons}} = 0\%$ FRR: $\frac{3 \text{ false rejections}}{15 \text{ comparisons}} = 20\%$
[13] FAR: $\frac{9 \text{ false acceptances}}{15 \text{ comparisons}} = 60\%$ FRR: $\frac{0 \text{ false rejections}}{15 \text{ comparisons}} = 0\%$
[14] FAR: $\frac{2 \text{ false acceptances}}{15 \text{ comparisons}} = 13.3\%$ FRR: $\frac{0 \text{ false rejections}}{15 \text{ comparisons}} = 0\%$
[15] FAR: $\frac{0 \text{ false acceptances}}{15 \text{ comparisons}} = 0\%$ FRR: $\frac{1 \text{ false rejections}}{15 \text{ comparisons}} = 6.7\%$

looking at the FFT of the resultant vector, as shown in Subsection 9.1.2.

Comparing the FFT of the Horizontal and Vertical Orientation and the XYZ Resultant Vector at Once

This experiment was an attempt of combining how the resultant vector was aligned with the resultant vector itself to look at whether this gives a better or worse result than examining either of the components alone.

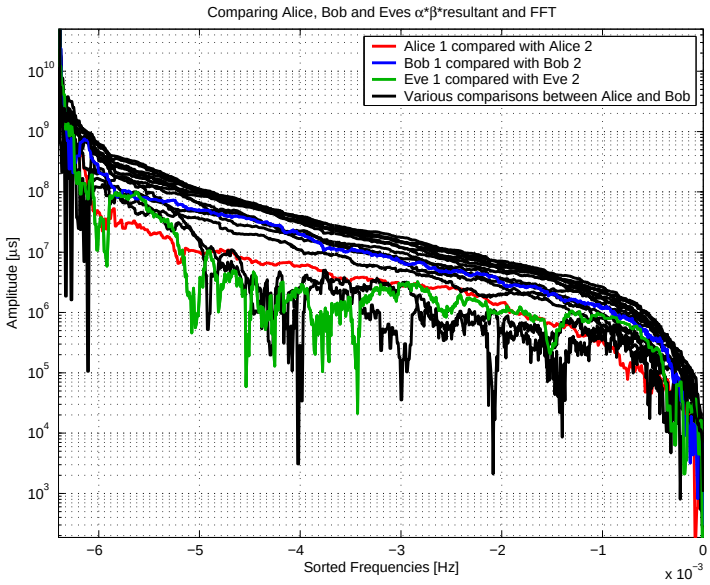


Figure 9.12: The result of all possible comparisons of how Alice, Bob and Eve's XYZ resultant vector is aligned with the Gait Collector as well as its length. The colored graphs are Alice, Bob and Eve compared with themselves, while the black graphs are different comparisons among them.

As shown in the tenth column of Table 9.1, as one would expect, the in-distinctiveness of the α and β angles only works as noise on the resultant vector data, and the result is an overall worse set of similarity scores compared with analysing the resultant vector alone. If one chooses a threshold value smaller than 28 in this result, we achieves a FAR of 0%, but also a FRR of 20%[16]. At the same time, a threshold value of at least 229 gives a FRR of 0% and a FAR of 46.7% [17]. It therefore appears to be more useful to analyse the resultant vector alone rather than analysing it along with its orientation.

9.1.5 Experiment Summary

Table 9.2 displays a summary of the results from the preliminary experiment. The methods with low FAR and FRR values and little overlap indicates good methods for analysis. As a conclusion, it seems like comparing the FFT of the XYZ and YZ resultant vectors are the most suitable methods for analysis. There are also some indications that detecting the local minima after the resultant vectors has been subtracted, has the potential of becoming a good analysis method. However, as shown in the results from these experiments, a lot of redundant minimas are detected which in a great extent disrupts the results from these analysis methods. Therefore, better methods for detecting the local minimas must be developed before the local minima detection can be considered an applicable method

[16] FAR: $\frac{0 \text{ false acceptances}}{15 \text{ comparisons}} = 0\%$ FRR: $\frac{3 \text{ false rejections}}{15 \text{ comparisons}} = 20\%$

[17] FAR: $\frac{7 \text{ false acceptances}}{15 \text{ comparisons}} = 46.7\%$ FRR: $\frac{0 \text{ false rejections}}{15 \text{ comparisons}} = 0\%$

Method	Percentage of Overlaps	FAR Rate	FRR Rate
Comparing FFT of X axis data	50%	40%	6.7%
Comparing FFT of Y axis data	8.3%	6.7%	6.7%
Comparing FFT of Z axis data	33.3%	26.7%	13.3%
Comparing FFT of the XYZ Resultant Vector	0%	0%	0%
Comparing FFT of YZ Resultant Vector	0%	0%	0%
Local Minima Detection on the FFT of XYZ Resultant Vector	0%	0%	0%
Local Minima Detection on the FFT of YZ Resultant Vector	41.7%	33.3%	20%
Comparing the FFT of the Horizontal Orientation of the XYZ Resultant Vector	75%	60%	20%
Comparing the FFT of the Vertical Orientation of the XYZ Resultant Vector	16.7%	13.3%	6.7%
Comparing the FFT of the Horizontal and Vertical Orientation and the XYZ Resultant Vector at Once	58.3%	46.7%	20%

Table 9.2: The results from the preliminary experiments. The **percentage of overlaps** displays how many comparisons of data from two different persons that received better matching scores than comparisons of data from the same person. The **FAR** and **FRR** summarises these calculations throughout this section.

for analysis. The other methods for analysis shows rather large FAR and FRR values, which indicates they are not very suitable for analysis. It is also noticeable that trying to combine the horizontal and vertical orientation with the resultant vector by multiplication introduces more noise into the results than by analysing either of them separately. This method of analysis has therefore not been included in the further experiments.

Figure 9.13 shows the ROC curves for the preliminary experiment. It is not easy to read these curves, since only 15 comparisons were made for each method. The methods with curves close to origin are however the most promising methods from this experiment. These are the same methods that received 0% FAR and FRR rates in Table 9.2.

9.2 The Large Scale Experiment

A large scale experiment was conducted to detect whether the results from the preliminary experiments are reflected when the same analysis methods are being used on gait data collected from a larger crowd. During the large scale experiment, 20 volunteers (11 male and 9 female) where participating. They each followed the rules from Section 8.1, and the data from each walking attempt was manually split in two halves, giving two datasets from each participant—a total of 40 datasets to compare. In addition to registering the participants' gender, their height, weight, leg length, foot length and shoe type was registered. The question of how the person's physical characteristics affected their gait was not analysed in this chapter, but some statistical tests has been applied to them. This is described closer in Appendix C.

The results from all comparisons during the preliminary experiment were possible to list in a set of tables and graphs since only six datasets were compared. This gave a total of 15 comparisons. In the large scale experiment however, 40 datasets were compared, which gave a total of 780 comparisons. These tables are therefore not listed in this re-

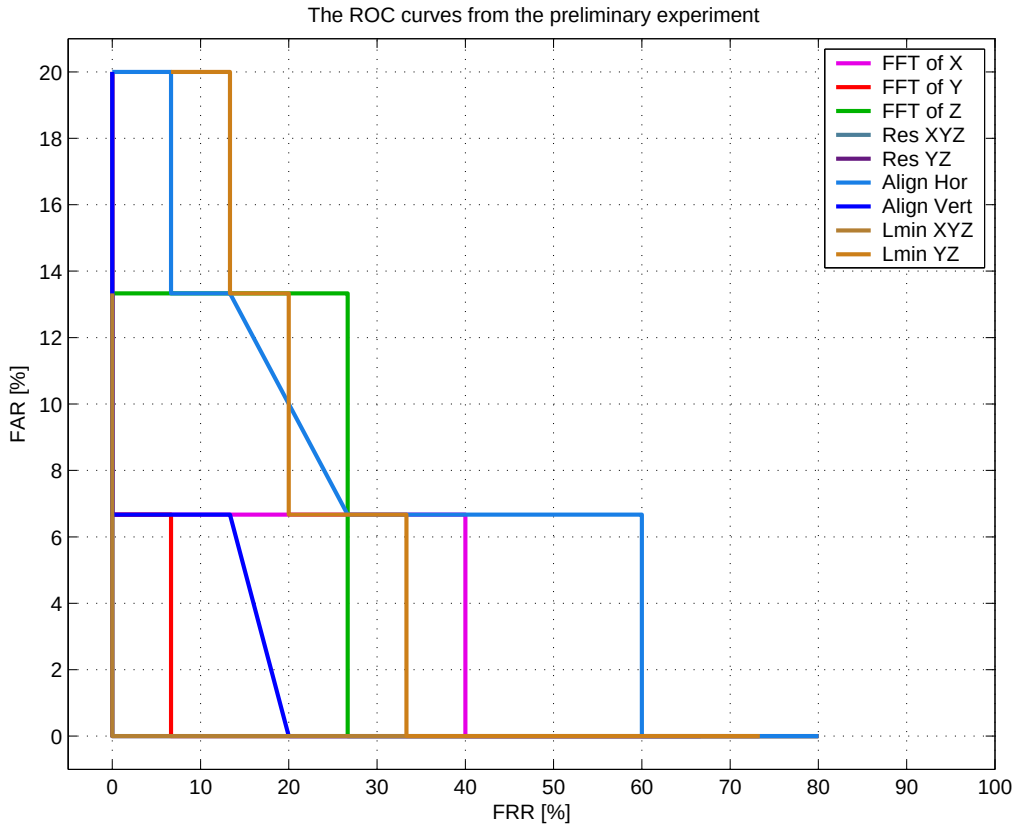


Figure 9.13: The ROC curves for the analysis methods tested during the preliminary experiment. Since there were not performed any more than 15 comparisons for each method, the curves are very difficult to read.

port because of their size, but the results are available at `http://www.torkjel.com/results/results_analysis.xls`.

9.2.1 FAR and FRR Results from the Analysis

Table 9.3 contains the FAR and FRR values obtained from the large scale experiment. The FAR and FRR values were calculated as described in the beginning of this chapter. Two threshold rates has been listed for each method. The first is the threshold value that gave a FRR rate of 0%, while the second is the threshold value that gave a FAR rate of 0%. As seen in the table, all analysis methods gives rather large FAR rates. This is a result of a rather large amount of overlaps in the comparisons of the same persons and the comparisons of different persons. The large rates can probably also be contributed to the rather large skewed distribution of these groups. There are 20 comparisons in the first group and 760 in the second. This would also explain the rather small FRR rates, since even though all comparisons of gait data from the same persons are wrongfully rejected, one does not get a higher FRR rate than 2.6%[18]. Due to the rather large extremes in the FAR/FRR rates, it is fairly difficult to say anything regarding the strength of the different methods of analysis looking at these rates alone.

[18] $\frac{20 \text{ false rejects}}{780 \text{ rejects}} = 2.6\%$

Method	Threshold	FAR	FRR
Comparing FFT of X axis data	> 244	52.7%	0.0%
	< 45	0.0%	2.6%
Comparing FFT of Y axis data	> 966	83.2%	0.0%
	< 94	0.0%	2.3%
Comparing FFT of Z axis data	> 395	81.8%	0.0%
	< 30	0.0%	2.4%
Comparing FFT of the XYZ Resultant Vector	> 958	86.8%	0.0%
	< 50	0.0%	2.6%
Comparing FFT of the YZ Resultant Vector	> 968	86.5%	0.0%
	< 38	0.0%	2.6%
Local Minima Detection on the FFT of XYZ Resultant Vector	> 1250	94.6%	0.0%
	< 60	0.0%	2.6%
Local Minima Detection on the FFT of YZ Resultant Vector	> 208	74.1%	0.0%
	< 56	0.0%	2.6%
Comparing FFT of the Horizontal Orientation of the XYZ Resultant Vector	> 12.1	64.0%	0.0%
	< 1.8	0.0%	2.4%
Comparing FFT of the Vertical Orientation of the XYZ Resultant Vector	> 5.6	48.6%	0.0%
	< 1.3	0.0%	2.4%

Table 9.3: The FAR and FRR results from the large scale experiments at a given set of threshold values.

9.2.2 The Distribution of Data within the Groups

When analysing the results from the experiment, the results from the various comparisons were sorted into two groups; the first group consisted of comparisons of the same person, while the second group consisted of comparing two different persons. The overall purpose of the experiment was to detect whether there are any significant differences in the matching scores within these two groups. The mean values and standard deviations for each group were calculated to get an indication of how the various matching scores obtained from the different experiments were distributed within their respective groups. The results are shown in Table 9.4. It is desirable to have a certain distance between the mean values of the two groups. This would indicate that the matching scores obtained when comparing two different persons would cluster around another range of values than comparisons of the same person. Also it is desirable with a small standard deviation, since this indicates how scattered the distribution of data are within one group. Large standard deviations would in other words imply large overlaps among the two groups.

As Table 9.4 shows, the obtained mean values have a certain degree of difference in the mean values for the two groups, which indicates there are some extent of differences in the two groups. However, the calculated standard deviations are rather large—specially for the methods detecting the local minima. It is therefore reasonable to conclude that there are a vast amount of overlap within the two groups for these methods.

9.2.3 Experiment Summary

As the results from this experiments show along with the data in Table 9.5, there are a vast amount of overlaps within the two groups. The two methods with the smallest degree of overlaps are when comparing the FFT of the X axis data and the comparison of the vertical orientation of the XYZ resultant vector. These are also the two methods

Method	Mean		Std. Deviation	
	Same	Different	Same	Different
Comparing FFT of X axis data	127.06	262.73	51.69	139.41
Comparing FFT of Y axis data	283.40	553.03	241.36	452.70
Comparing FFT of Z axis data	149.46	235.66	99.32	147.88
Comparing FFT of the XYZ Resultant Vector	272.95	488.56	258.79	402.93
Comparing FFT of YZ Resultant Vector	266.73	490.44	255.50	421.86
Local Minima Detection on the FFT of XYZ Resultant Vector	166.80	329.49	257.38	1098.41
Local Minima Detection on the FFT of YZ Resultant Vector	123.70	397.41	38.71	1321.78
Comparing the FFT of the Horizontal Orientation of the XYZ Resultant Vector	5.33	10.42	2.65	5.78
Comparing the FFT of the Vertical Orientation of the XYZ Resultant Vector	3.11	6.60	1.26	3.70

Table 9.4: Shows the mean values of standard deviations obtained within the two groups. **Same** are the results from comparing gait data from the same persons, while **Different** contains the results from comparing gait data from two different persons.

Method	Overlapping
Comparing FFT of X axis data	54.1%
Comparing FFT of Y axis data	85.4%
Comparing FFT of Z axis data	83.9%
Comparing FFT of the XYZ Resultant Vector	89.1%
Comparing FFT of YZ Resultant Vector	88.8%
Local Minima Detection on the FFT of the XYZ Resultant Vector	97.2%
Local Minima Detection on the FFT of the YZ Resultant Vector	76.1%
Comparing the FFT of the Horizontal Orientation of the XYZ Resultant Vector	65.5%
Comparing the FFT of the Vertical Orientation of the XYZ Resultant Vector	49.3%

Table 9.5: The table shows the amount of overlap within comparisons of two different persons and comparisons of the same persons.

with the “smallest” FAR values, if one can refer to values larger than 45% as small values. However, as stated earlier in this chapter, the rather enormous FAR values obtained might partly be contributed to a very skewed distribution of the two groups.

When comparing the extent of overlapping in the large scale experiment with the similar results from the preliminary experiments, it is clear that all analysis methods have a larger difficulty of verifying a person’s identity than what first appeared (Section 9.1.5). In the preliminary experiment several of the methods did not have any overlapping at all among the two groups, like the comparison of the resultant vectors. However, in the large scale experiment, these are among the methods that had the greatest extent of overlapping. They are therefore not as suitable analysis methods as it first appeared. It can also be observed that analysing the resultant vector’s alignment had some of the worst results in the preliminary experiment. However they were among the best methods in this experiment.

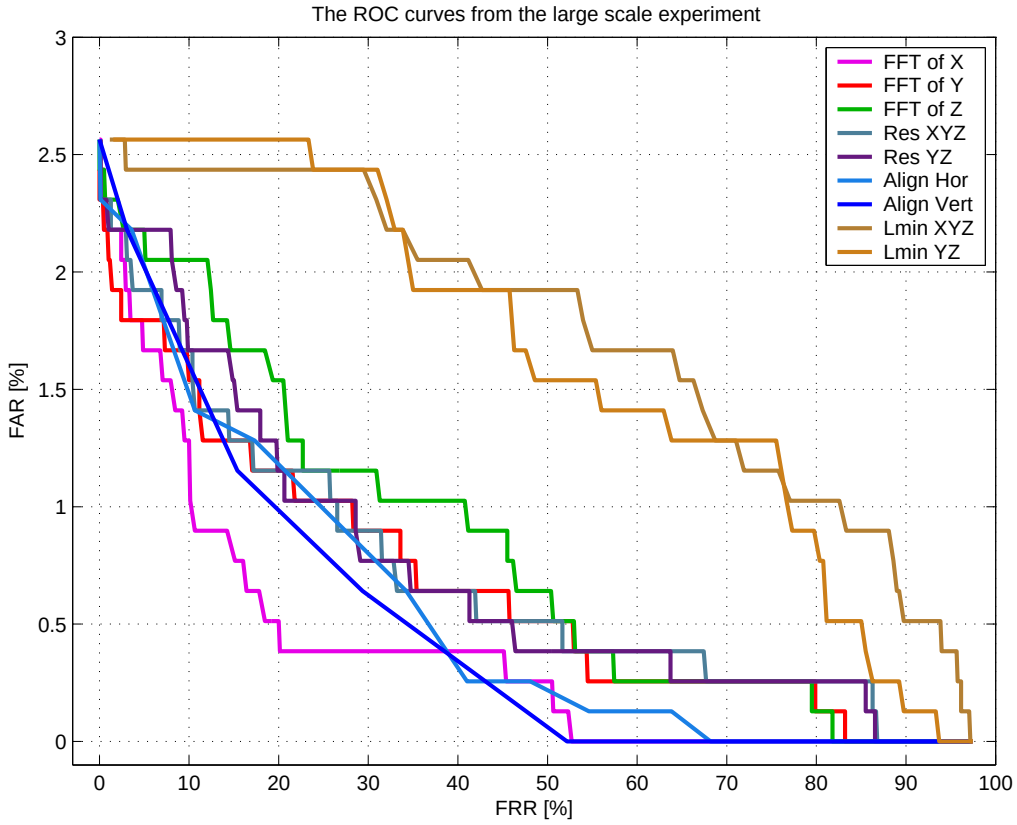


Figure 9.14: The ROC curves for the analysis methods tested during the large scale experiment. The reason for not getting FAR values larger than 3% is the skewed distribution of comparisons; 20 comparisons of the same person and 760 comparisons of different persons.

Figure 9.14 shows the ROC curves for the large scale experiment. What can be observed here is that the most promising analysis methods will be the ones with values closest to origin. This confirms the observations from Table 9.5—that the FFT transform of X axis data and the orientation of the resultant vector might be the most useful methods. However, it is also clear from this figure that the methods performing local minima detection does not manage to verify a person’s identity in any large extent. This might

be due to the fact that these methods are detecting far to many unnecessary minima.

9.3 The Long Term Experiment

A long term experiment was performed to detect how the data analysis methods work on gait data obtained over a longer period of time. Its main purpose was however to determine whether the fact that the participants only performed one walking attempt during the large scale experiment introduced validity problems. This was done by having four participants making five walking attempts over a period of one week. As with the large scale experiment, the results from this experiments are fare to large to be fitted into a set of tables or graphs. They can therefore be found at `http://www.torkjel.com/results/results_analysis.xls`.

As with the preliminary and large scale experiments, the FAR and FRR are calculated using the method described in the beginning of this chapter. While there were a skewed distribution of comparison of the same persons and of different persons, it is slightly more even than in the large scale experiment. Here, we have 40 comparisons of the same person and 150 comparisons of different persons.

9.3.1 FAR and FRR Results from the Analysis

Method	Threshold	FAR	FRR
Comparing FFT of X axis data	> 308	44.7%	0.0%
	< 57	0.0%	21.1%
Comparing FFT of Y axis data	> 1527	61.6%	0.0%
	< 123	0.0%	19.5%
Comparing FFT of Z axis data	> 887	77.4%	0.0%
	< 51	%	21.1%
Comparing FFT of the XYZ Resultant Vector	> 2000	68.4%	0.0%
	< 99	0.0%	19.5%
Comparing FFT of the YZ Resultant Vector	> 1878	65.8%	0.0%
	< 60	0.0%	21.1%
Local Minima Detection on the FFT of XYZ Resultant Vector	> 357	70.5%	0.0%
	< 62	0.0%	20.0%
Local Minima Detection on the FFT of YZ Resultant Vector	> 236	61.1%	0.0%
	< 55	0.0%	21.1%
Comparing FFT of the Horizontal Orientation of the XYZ Resultant Vector	> 12.8	40.0%	0.0%
	< 2	0.0%	21.1%
Comparing FFT of the Vertical Orientation of the XYZ Resultant Vector	> 12.6	61.1%	0.0%
	< 1.7	0.0%	21.1%

Table 9.6: The FAR and FRR results from the large scale experiments at a given set of threshold values.

Table 9.6 contains the FAR and FRR values obtained from the results of the long term experiment. As with the large scale experiment, two threshold rates has been listed for each method. How these threshold values were calculated is described in the beginning of this chapter. The first is the threshold value that gave a FRR rate of 0%, while the second is the threshold value that gave a FAR rate of 0%. This table shows that the obtained FAR and FRR rates are rather moderate compared with the corresponding results from the large scale experiment (Section 9.2.1). This is probably partly since this experiment have a slightly more even distribution of the data in the two groups. It is therefore possible

to achieve a FRR rate of 21.1%[19] when setting a threshold where all comparisons of gait data from the same person are being rejected. This compared to a FRR rate of only 2.6%[20] in the large scale experiment.

9.3.2 The Distribution of Data within the Groups

Method	Mean		Std. Deviation	
	Same	Different	Same	Different
Comparing FFT of X axis data	162.93	296.76	60.79	156.97
Comparing FFT of Y axis data	360.54	904.81	349.59	625.22
Comparing FFT of Z axis data	205.48	267.98	194.54	199.51
Comparing FFT of the XYZ Resultant Vector	392.04	858.19	458.71	787.88
Comparing FFT of YZ Resultant Vector	389.26	891.06	446.45	790.36
Local Minima Detection on the FFT of XYZ Resultant Vector	127.93	271.92	59.80	849.90
Local Minima Detection on the FFT of YZ Resultant Vector	118.16	293.00	46.43	856.22
Comparing the FFT of the Horizontal Orientation of the XYZ Resultant Vector	6.33	13.67	2.53	6.40
Comparing the FFT of the Vertical Orientation of the XYZ Resultant Vector	5.12	8.78	2.74	4.47

Table 9.7: Shows the mean values of standard deviations obtained within the two groups. **Same** are the results from comparing gait data from the same persons, while **Different** contains the results from comparing gait data from two different persons.

Table 9.7 contains the results after calculating the mean values and standard deviation of the similarity scores within the two groups. The first thing noticeable is that the mean values and standard deviation values has increased significantly compared with the corresponding results from the large scale experiment. This might be due to the changes in the participants' gait patterns the week. An interesting observation is however that the standard deviation for the methods analysing the local minima has a much smaller standard deviation in this experiment compared with the large scale experiment. The reason for this is unknown, but it gives much better results than in the previous experiment. Other than the alignment of the resultant vector, the results from the other analysis methods are somewhat conclusive with the corresponding results from the large scale experiment regarding which methods are better than the other. A problem with the large scale experiment was that the volunteers did not remove the Gait Collector from their leg between the gathering of each data sets. Although the volunteers attached the Gait Collector to their leg according to the specifications in Section 8.1, small differences in their individual attachments occurred, that could bias the results from the data gathering. This could result in validity problems if it was not the way the person walked that was being measured, but rather how the Gait Collector was attached to the person's leg. Since the persons in the long term experiment actually did remove the Gait Collector

[19] FRR: $\frac{40 \text{ false rejects}}{190 \text{ rejects}} = 21.1\%$

[20] FRR: $\frac{20 \text{ false rejects}}{780 \text{ rejects}} = 2.6\%$

from their leg between each measurement, re-attaching it in a slightly other manner for the next gait attempt, the results from this experiment actually is an indication that the Gait Collector are measuring the way the persons walks and not the way it is attached. The fact that the results from this experiment and the large scale experiment shows similar tendencies is an indication that the large scale experiment really measured how the persons walked.

9.3.3 Experiment Summary

Method	Overlapping
Comparing FFT of X axis data	56.7%
Comparing FFT of Y axis data	78.0%
Comparing FFT of Z axis data	98.0%
Comparing FFT of the XYZ Resultant Vector	86.7%
Comparing FFT of YZ Resultant Vector	83.3%
Local Minima Detection on the FFT of the XYZ Resultant Vector	89.3%
Local Minima Detection on the FFT of the YZ Resultant Vector	77.3%
Comparing the FFT of the Horizontal Orientation of the XYZ Resultant Vector	50.7%
Comparing the FFT of the Vertical Orientation of the XYZ Resultant Vector	77.3%

Table 9.8: The table shows the amount of overlap within the comparisons of two different persons and the comparisons of data from the same persons.

Table 9.8 shows that just as in the large scale experiment, there are a large amount of overlapping within the groups of persons compared with themselves and persons compared with each other. Still, the amount of overlapping is smallest when comparing the FFT of the X axis data and when comparing the horizontal orientation of the XYZ resultant vector. These results are somewhat conclusive with the results from the large scale experiment. Though the obtained FAR/FRR rates are smaller than the ones from the large scale experiment, this is probably due to a more even distribution of comparisons within the two groups. It is therefore difficult to say anything about the strength of the analysis methods based on the FAR/FRR results alone other than that, they all obviously have large difficulties in determining the difference among certain comparisons.

Figure 9.15 shows the ROC curves for the long term experiment. This experiment shows a slightly less scattered distribution of ROC curves than the large scale experiment did. Even the local minima calculation starts to get more promising results. This might be due to the fact that more gait data were collected and analysed for each person giving more reliable data analysis. From the ROC curves it is also clear that detecting the orientation of the resultant vector still are two of the analysis methods with the best results.

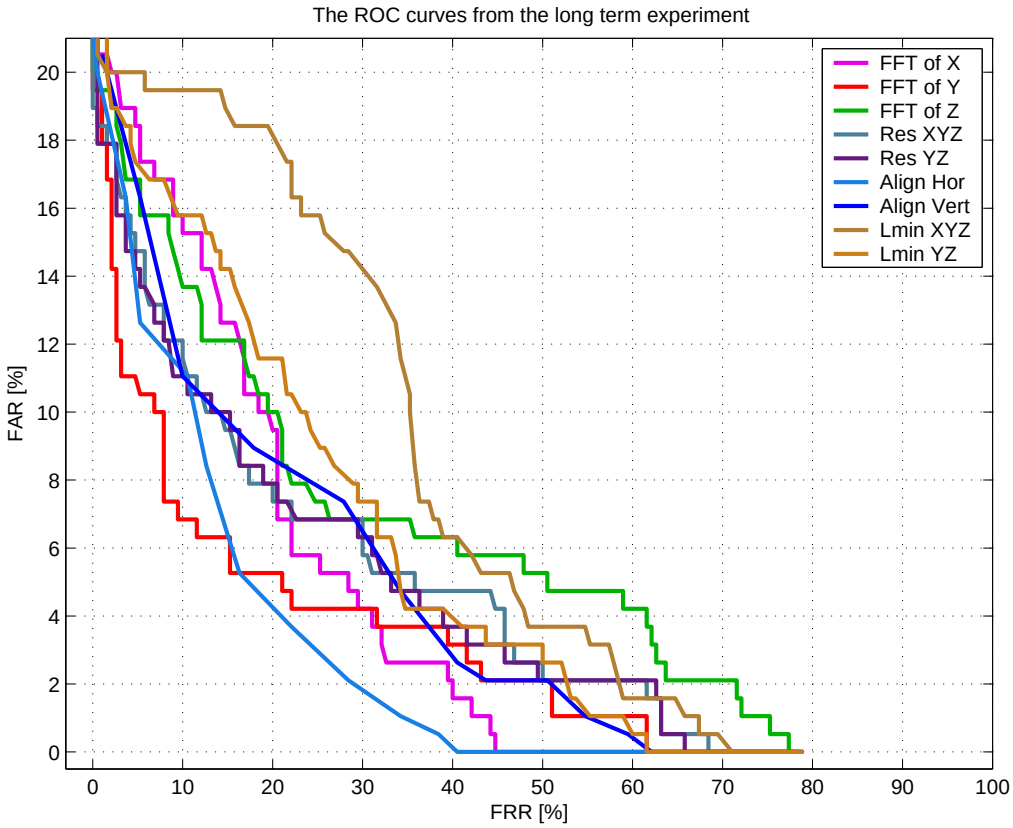


Figure 9.15: The ROC curves for the analysis methods tested during the large scale experiment. The reason for not getting FAR values larger than 20% is the skewed distribution of comparisons; 40 comparisons of the same person and 150 comparisons of different persons.

10 Discussion of the Experiments

Three experiments has been performed during this project. The first was a *preliminary experiment*, which was performed to ease the process of selecting the most appropriate analysis methods. Then a *large scale* experiment was performed to detect how the developed analysis methods worked in a larger setting and to improve the confidence of the results from the preliminary experiment. Lastly, a *long term* experiment was performed to identify whether the fact that the Gait Collector was not removed from the participants' legs between each obtained dataset during the large scale experiment introduced validity errors. It also gave an impression of how the analysis methods worked on gait data collected over a longer period of time.

Although the scale of the experiments performed were not large enough to get an thorough quantitative view of the FAR and FRR values for the various methods, it is possible to get some indications from these calculations. One obvious observation is that the results from the preliminary experiment is not very conclusive with the results from the large scale and long term experiments. The preliminary experiment indicated several strong methods, like the calculation of the XYZ and YZ resultant vector and the local minima detection. These methods however are not among the best ones in the two last experiments. Here, the alignment of the XYZ resultant vector and the FFT of the X and Y axes are the best methods. It is however important to remember that when analysing the FFT of the raw data from the Gait Collector, it is critical that the Gait Collector is oriented the same way each time it is attached to the person's leg. A new method for analysis which in some clever way combine several of the best analysis methods could probably give a stronger method of analysis. This will be left as an exercise for further work.

Depending on the level of certainty that is necessary, none of the methods appear to be very accurate at verifying a person's identity after the large scale and long term experiments has been performed. However, the best methods might still be sufficient for low-security authentication applications.

Although it is difficult to compare the FAR and FRR values from the experiments performed with similar research, the ROC curves from the long term experiment has been compared with the results from Bobick and Johnson's gait analysis [11] in Figure 10.1. It might seems like most of the methods from this experiment are better than Bobick and Johnson's ROC curve, but this might be due to the limited comparisons available for this experiment, which prevented FAR values larger than 23%.

As shown in the preliminary experiment, the local minima detection has a tendency of detection several unnecessary tiny minimas that should not have been included in the further analysis process. If these methods were refined to only detect the minima which

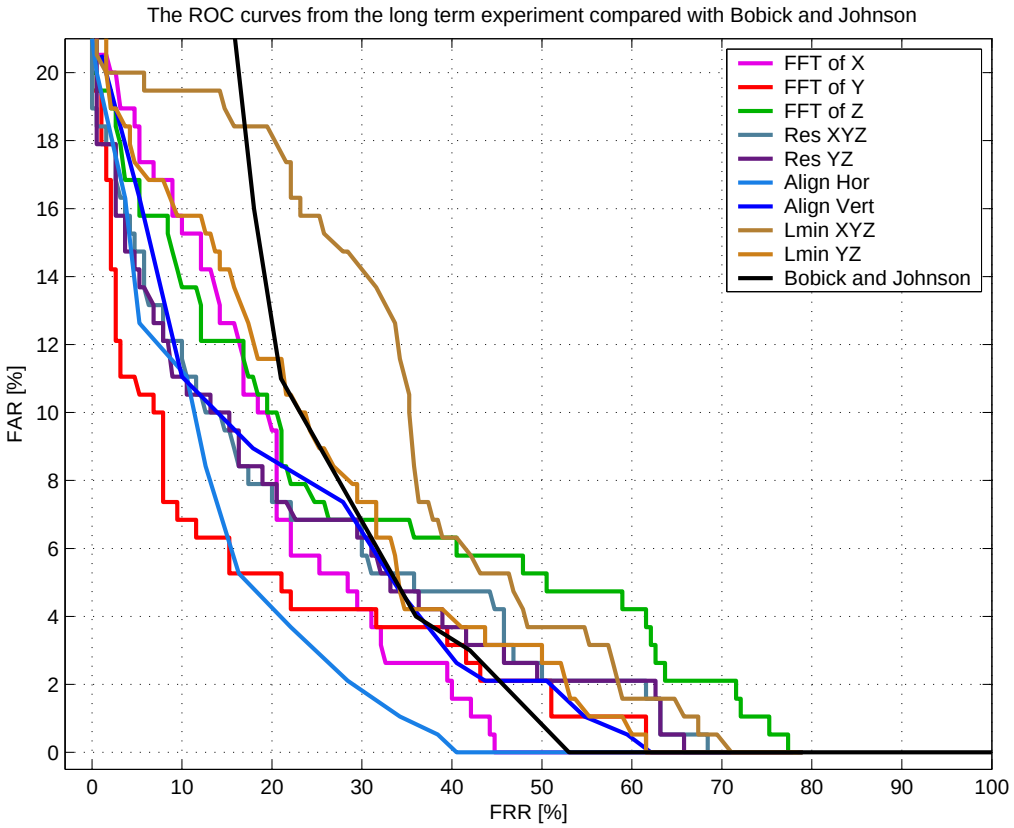


Figure 10.1: A comparison of the ROC curves from the long term experiment with Bobick and Johnson's ROC curve [11].

had characteristics, they might have some potentials.

Before the kinetic gait analysis can be a usable method for authentication, better methods for analysis must be found. [54] uses FAR rates lower than 0.001% and FRR rates ranging between 5% and 20% as examples of what would be preferred in a military installation using fingerprint recognition, while consumer applications should have FRR rates around 0.5%.

11 Further Work

This project have opened a new door in the field of kinetic gait authentication. As the task of developing methods for gait analysis was performed, new problems and possibilities for further work were discovered. This chapter lists the next steps of research which should follow this thesis.

Calibration: The accelerometers connected to the Gait Collector is not being calibrated prior to the data collection phase. This is due to the potential difficulties this would be regarding the usability of the device; a calibration according to Analog Devices' specification [69] would require each of the two accelerometers to be placed horizontal during the calibration procedure to get a force of 0g affection during calibration. Finding an alternative way of automatically calibrating the accelerometers without the user's attention would be preferable. If automatically calibration was incorporated into the Gait Collector's routines, there might be possible to create new and maybe better analysis methods. This will probably also removes validity issues in the gait data.

Location Detection: It might not only be of interest *who* the person wearing the Gait Collector is, but also *where* he is. In this case it might be interesting to detect where he has walked, in a manner like *"The person walked 10 paces at a straight line, turned 40 degree right, walked another 10 paces before walking up a stair with 12 steps"*. This could be integrated with pre-stored data of the building where the user was to detect where on this location he was. It could also be used to detect whether he was indoor or outdoor and perhaps be incorporated into some sort of physical access control.

Compensate for the Physical Plane: When the Gait Collector is attached onto a user's leg, it will not only be misaligned with the physical coordinate system, as shown in figure 5.3—it will also follow the leg's rotation as it is moved during the gait cycle. Finding a method for overcoming this problem would be of interest, since the detection of the leg's acceleration in the physical coordinate system is not affected by how the Gait Collector is attached on the leg. Detecting this would require information regarding the orientation of the Gait Collector, which would require the incorporation of gyroscopes as well as accelerometers. If such a problem was solved, it might be possible to create new and maybe better analysis methods.

Incorporate the Authentication Process onto the Gait Collector: At the present time, the Gait Collector only contains a minimum of functionality—only necessary for

collecting acceleration data, storing it and transferring it to a PC on request. It might however be of interest to integrate the whole analysis and authentication process onto the Gait Collector. This would enable it to continuously checking whether the correct person was using the Gait Collector, and for instance sending out a loud alarm signal if it detected the wrong person using it. The Gait Collector could then be used to protect equipment which a person was wearing, like his mobile phone, PDA or similar. The current prototype of the Gait Collector has however a very limited programmable storage space of only 16K bytes. Fitting advanced algorithms for analysis, like the FFT and similar might be difficult.

Gait Collector Code Optimising and Power Management: Not too much attention has been taken into optimizing the code on the Gait Collector regarding the speed in which it reads from the accelerometers, and the limitation of data written to the dataflash and its battery consumption. At the present time, the dataflash is filled in less than 10 minutes, which is sufficient for prototyping purposes, but must be optimized for a final product. This might be possible by not storing unnecessary data. Another source which might be optimized during data storing is the storing procedure itself, which at the present time is performed with a printf() command. This command only accepts ASCII data, where a character is represented by 1 byte of data—sufficient for representing 65 536 different characters. Each reading of acceleration data is written as a block of 5 bytes, whereas the acceleration data itself might be as much as 4 byte and the last byte is to separate two readings. This might be optimized, by defining our own character set consisting of only 4 bits, which is necessary to represent all digits between 0 and 9 as well as 5 other special characters, like a separation character, a read error character, a newline character and an end-of-file character. This optimizing would make each reading consisting of 20 bits in stead of 80 as it is at the present stage.

It would also be necessary to optimise the code regarding power management such that it did not need three AA batteries to run efficiently. This would have to be done by creating clever routines that prevented the Gait Collector from reading acceleration data when the user was not walking or moving in other ways. It would also be of interest to make the Gait Collector rechargeable e.g. through USB interface.

Reduction of Biases: Section 5.2 contain several sources to error in the collected data. Though this project has not focused on liming the amount of noise in the measurements, solutions for handling these problems must be taken into account in a final version of the Gait Collector. This applies both in the physical design of the Gait Collector to prevent a wrongful attachment to the leg, during the gathering of acceleration data an during analysis of these data.

It would be relevant to determine the source of the dominant spike in Figure 5.2. This might be done by changing the accelerometers' sampling frequency to check whether the frequency of the spike changes as well.

FAR and FRR Calculation: The experiments described in this thesis is nowhere large scaled enough to get an accurate quantitative measure of the FAR and FRR rates for the Gait Collector. It would therefore be of interest to perform an experiment of a larger scale, which focused on the detection of such data. This would imply

collecting acceleration data from a minimum of 100 persons and plotting the Detection Error Tradeoff [45] and Receiver Operating Characteristic [43] curves for different threshold values. We would then get a more accurate impression of the FAR and FRR rates which the Gait Collector produces.

Secure Gait Collector's Data: This is necessary in a real-life authentication system. Unless a secured channel is established between the computer which receives gait data and the Gait Collector which transfers them, the data analysis software has no guarantee that the received data has not been tampered with in transmit. It is therefore necessary to establish an encrypted channel between the analysis software and the Gait Collector following cryptographic standard, like the ones specified by the Public Key Infrastructure [58] or similar.

Real-Time Transmission: At the present time, the data transmission from the Gait Collector to a PC requires connecting them together with a serial cable. By fitting the Gait Collector with equipment for wireless data transmission, it might be possible to transfer the data in real time analysing the data as the person walks. It could then be possible to synchronise the gait data analysis with captured video data of the person as well to get new an more accurate understandings of the leg's movement during gait. It would also be preferable to use a USB interface rather than a serial interface with the computer, since almost no portable computers are produced with a serial interface anymore.

Using Gyroscopes: As described previously in this chapter, the use of accelerometers makes it difficult to detect the Gait Collector's orientation in relation with the physical coordinate system. It is somewhat possible to detect when the person is standing still, since the resultant vector calculated from the Gait Collector's XYZ axis then will be pointing straight down along the physical X axis. However, as the foot moves in an arc during its gait cycle, the Gait Collector is also rotating and changing its orientation with the physical plane constantly. It would therefore be necessary to detect how it is angled at all time. Solving this should probably be done using gyroscopes. Gyroscopes are also capable of detecting the angular acceleration in which the leg moves, which gives new and perhaps distinctive gait data not detected by the accelerometers at the present stage.

New Mounting Possibilities: We have in this project measured gait data by attaching the Gait Collector to the person's leg close to the ankle. But how would a person's walking features look like if it was mounted on the tip of the person's shoe, on his knee, on his hip or in his pocket? Would other mountings give more distinctive gait data? A more thorough research on this would be of interest—after all, it would be more convenient to keep the Gait Collector in your pocket rather than strapped to your leg.

Other Methods for Analysis: There are several potential methods for analysis described in chapter 6 which have not been considered feasible to achieve within the time limits of this project. Methods which have been used with success in similar data analysis projects are different Artificial Neural Networks, like the Self-Organizing Maps and Genetic Algorithm Neural Networks. Hidden Markov Models are popular in voice recognition, and might be applicable in the analysis of gait data as well.

Investigate Privacy: Gait analysis implemented in video surveillance introduces the possibility of identifying persons without their knowledge. This has the potential of creating privacy issues. The possibilities and challenges by violating a person's privacy through kinetic gait analysis would also be interesting to investigate. The reduced possibility of violating a person's privacy when using kinetic gait analysis compared with gait analysis through kinematic video data might possibly be an extra motivation for using this as an authentication scheme.

Statistical Analysis: Appendixes C and D shows the results after a serie of statistical analysis methods have been applied to the results from the large scale and long term experiments. However, they are not discussed in any great detail. It would therefore be of interest to perform a more thorough research on the results from these analyses. Also, it would be interesting to apply other statistical analysis methods to the results of the experiments. This might make it possible to discover unknown strength and weaknesses in the methods developed though this research.

Biomechanical Gait Model: A thorough biomechanical model of the human gait would also be very interesting to create. Such a model could help explaining whether the human gait is a suitable authentication method. A biomechanical[1] gait model could perhaps also help answering some of the phenomenas discovered in this project where no suitable answers yet has been provided.

One of these phenomenas is the model of the leg's acceleration in Figure 5.6. A suggestion for a mapping has been provided, but it this has not been validated. This however, might be possible through for instance video observations or similar.

Another phenomena is that the local minimas detection shows how a serie of spikes occur in the comparison of the same person and not in comparison of different persons. Exactly why does this happen? This would be interesting to discover, since it might help improving this analysis method.

Fool the Gait Collector: This thesis has not focused on how it is possible for an adversary to be recognised as an enrolled user in a gait authentication system. It would therefore be very interesting to determine how easy it is to fool the Gait Collector. Is it enough to walk next to the person you want to impersonate to get a gait pattern similar to his? Is it possible to learn to walk like a person in any other manners? For the Gait Collector to be a workable authentication technology, some research must be performed in this area.

[1] The mechanics of a part or function of a living body, such as of the heart or of locomotion (`http://www.dictionary.com`).

12 Conclusion

This project started by asking one simple question: *"Is it possible to authenticate a person by the way he walk?"* As the project moved forward, it was clear that this was a much more difficult question than first thought, and a wide area of new questions were identified. *"Is it possible to collect reliable and valid gait data from a person by means of motion capture?", "What kind of work has already been performed in this area?", "Is it possible to create methods that manages to verify a person's identity with an adequate level of certainty?"* The list goes on, as shown by the proposed further work in Chapter 11. This research project have shown that although the use of motion capturing to verify a person's identity by his gait is a novel area of information security, it shows some promising tendencies. This would however require that even more of the proposed methods of analysis are tested and that other motion capturing techniques are incorporated into the Gait Collector. The use of gyroscopes, for instance, would probably help determining how the Gait Collector was aligned on the foot, resulting in even more raw data to incorporate into the analysis methods.

The experiments performed in this research shows that it is possible to collect reliable and valid movement data from a person using accelerometers. They also show that it is possible to detect relations between a person and his gait using kinetic acceleration data. Through fairly simple analysis methods is it possible to distinguish him from other persons. However, for gait authentication to achieve acceptable security performance, more work is needed and better methods for analysis must be developed and assessed.

Bibliography

[1] FFT Tutorial. University of Rhode Island Department of Electrical and Computational Engineering ELE 436: Communication Systems. http://www.ele.uri.edu/~hansenj/projects/ele436/fft.pdf (Last visited 03062005).

[2] Lov om Behanling av Personopplysninger (Personopplysningsloven/Privacy Act), 2001. http://www.lovdata.no/cgi-wift/wiftldles?doc=/usr/www/lovdata/all/nl-20000414-031.html&dep=alle&kort+,+titt=personopplysningsloven& (Last visited 22062005).

[3] Douglas G. Altman. *Practical Statitics for Medical Research*. Chapman & Hall/CRC, 1999.

[4] Analog Devices. *AT45DB041B Datasheet*. http://www.atmel.com/dyn/resources/prod_documents/doc3443.pdf (Last visited 05052005).

[5] Analog Devices. *ATmega169 Datasheet*. http://www.atmel.com/dyn/resources/prod_documents/doc2514.pdf (Last visited 05052005).

[6] Atmel Coorperation. *AVR Butterfly Quick Start User Guide*, 2003. http://www.atmel.com/dyn/resources/prod_documents/doc4249.pdf (Last visited 05052005).

[7] Ed Ayyappa. Normal Human Locomotion, Part 1: Basic Concepts and Terminology. *Journal of Porsthetics and Orthetics*, 9:10–17, 1997.

[8] Willem Back and Hilary M. Clayton, editors. *Equine Locomotion*, chapter 3, pages 55–76. W.B. Saunders Company, first edition, December 2000.

[9] Chiraz BenAbdelkader, Ross Cutler, and Larry Davis. Stride and Cadence as a Biometric in Automatic Person Identification and Verification. In *IEEE International Conference on Automatic Face and Gesture Recognition*. Microsoft Research, 2002.

[10] Ari Y. Benbasat, Stacy J. Morris, and Joseph A. Paradiso. A Wireless Modular Sensor Architecture and its Application in On-Shoe Gait Analysis. In *IEEE Sensors*, 2003.

[11] Aaron F. Bobick and Amos Y. Johnson. Gait recognition using static activity-specific parameters. In *Proceedings of Computer Vision and Pattern Recognition Conference (CVPR 2001)*, Kauai, Hawaii, December 2001.

[12] Edward G. Carmines and Richard A. Zeller. *Relability and Validity Assessment*. Number 17 in Quantitative Application in the Social Science. Sage Publications Inc., 1979.

[13] Carl W. Chan and Andrew Rudins. Foot Biomechanics During Walking and Running. *Mayo Clinic Proceedings*, 69:448–61, 1994.

[14] John W. Creswell. *Research Design - Quantitative, Qualitative and Mixed Methods Approaches*. SAGE Publications, Lincoln, second edition, 2003.

[15] James E. Cutting and Lynn T. Kozlowski. Recognizing Friends by their Walk: Gait Perception Without Familiarity Cues. *Bulletin of the Psychonomic Society*, 1977.

[16] James W. Davis. Visual Categorization of Children and Adult Walking Styles. In *International Conference on Audio- and Video-based Biometric Person Authentication*, pages 295–300, June 2001.

[17] James W. Davis and Stephanie R. Taylor. Analysis and Recognition of Walking Movements. In *International Conference on Pattern Recognition*, pages 315–318. Quebec City, Canada, August 2002.

[18] John R. Deller, John H.L. Hansen, and John G. Proakis, editors. *Discrete-Time Processing of Speech Signals (IEEE Press Classic Reissue)*, chapter 11. Wiley-IEEE Press, reprint edition, September 1999.

[19] Charles H. Edwards and David E. Penny. *Calculus With Analytic Geometry*. Prentice Hall, fifth edition, 1999.

[20] Cecilie Enger. Slik Går de Rett i Baret. *Dagens Næringsliv*, April 24th 2005.

[21] Kamran Etemad and Rama Chellappa. Discriminant Analysis for Recognition of Human Face Images. *Journal of the Optical Society of America A*, 14(8):1724–1733, August 1997.

[22] Andy Field. *Discovering Statistics Using SPSS for Windows*. SAGE Publications, 2000.

[23] Gábor Guta, Attila Érsek, Norbert Gosztonyi, Sándor Melo, and Dr. István Szabó. Implementing a Speedometer for Walking and Running with the TMS320F243 DSP Controller. Application Report, University of Debrecen, April 2000.

[24] Frode Hansen, Nicolai Heyerdahl, and Sonja Holtermann. Er Dette Samme Mann? *Dagbladet*, May 8th 2004.

[25] Frode Hansen, Nicolai Heyerdahl, Sonja Holtermann, and Ulf André Andersen. Hun Risikerer Livet. *Dagbladet*, April 20th 2004.

[26] Anna Hart. Mann-Whitney Test is not Just a Test of Medians: Differences in Spread can be Important. *British Medical Journal*, 323:391–393, August 2001.

[27] José Elías Herrero-Jaraba, Carlos Orrite-Uruñuela, David Buldain, and Armando Roy-Yarza. Human Recognition by Gait Analysis using Neural Networks. In *Proceedings of the International Conference on Artificial Neural Networks*, volume 2415, pages 346–369, 2002.

[28] Ken Hinckley, Jeff Pierce, Mike Sinclair, and Eric Horvitz. Sensing Techniques for Mobile Interactionxs. In *Proceedings of the 13th Annual ACM Synposium on User Interface and Technology*, pages 91–100, 2000.

[29] Ping S. Huang, Chris J. Harris, and Mark S. Nixon. Recognising Humans by Gait via Parametric Canonical Space. *Artificial Intelligence in Engineering*, 13, 359–336 1999.

[30] Gunnar Johansson. Visual Motion Perception. *Scinetific American*, 232:76–80, 85–88, June 1975.

[31] Amit Khajanchi. Artificial Neural Networks: The Next Intelligence. `www.usc.edu/org/techalliance/Anthology2003/Final_Khajanch.pdf` Last visited 27062005.

[32] Chris Kirtley, Michael W. Whittle, and R.J. Jefferson. Influence of Walking Speed on Gait Parameters. *Journal of Biomechanical Engineering*, pages 282–288, 1985.

[33] Monika. Köhle and Stefan Holzreiter. Assessment of Gait Patterns Using Neural Networks. In *J. Biomech*, volume 26, pages 645–651, 1993.

[34] Monika Köhle and Dieter Merkl. Things we Observed when Watching People Walk: Classification of Gait Patterns with Self-Organizing Maps. In *The 7th Annual Conference on Neural Networks*, April 1996.

[35] Teuvo Kohonen. The Self-Organizing Map. In *Proceedings of the IEEE*, volume 78, pages 1464–1480, September 1990.

[36] Teuvo Kohonen, Samuel Kaski, Panu Somervuo, Krista Lagus, Merja Oja, and Vesa Paatero. *Biennial Report 2002-2003*, chapter 8, pages 113–122. CIS, February 2004.

[37] Quentin Ladetto. On Foot Navigation: Continuous Step Calibration using both Complementary Recursive Prediction and Adaptive Kalman Filtering. Technical report, Geodetic Laboratory, Institure of Geometic, Swiss Federal Institute of Technology, 2000.

[38] Quentin Ladetto, Vincent Gabglio, Bertrand Merminod, Phlilippe Terrier, and Yves Schutz. Human Walking Analysis Assisted by DGPS. Technical report, Geodetic Laboratory, Institure of Geometic, Swiss Federal Institute of Technology, 2000.

[39] Lily Lee and Eric Grimson. Gait analysis for recognition and classification. In *IEEE Conference on Face and Gesture Recognition*, pages 155–161, 2002.

[40] Lily Lee and Eric Grisom. Gait Appearance for Recognition. *Lecture Notes in Computer Science*, 2359:143–154, 2002.

[41] Jennifer L. Lelas, Gregory J. Merroman, Patrick O. Riley, and D. Casey Kerrigan. Predicting Peak Kinematic and Kinetic Parameters from Gait Speed. *Gait and Posture 17*, pages 106–112, June 2003.

[42] Niels Lynnerup and Jens Vedel. Person Identification by Gait Analysis and Photogrammetry. *Journal of Forensic Science*, 50(1), January 2005. Technical Note.

[43] Davide Maltoni, Dario Maio, Anil K. Jain, and Salil Prabhakar. *Handbook of Fingerprint Recognition*. Springer Verlag, second edition, 2003.

[44] A.A. Marks. *Manual of Artificial Limbs*. New York, 1905.

[45] Alvin F. Martin, George Doddington, Terri Kamm, Mark Ordowski, and Mark Przybocki. The DET Curve in Assessment of Detecting Task Perfomance. In *Eurospeech*, volume 4, pages 1895–1898, 1997.

[46] Stacy J. Morris and Joseph A. Paradiso. Shoe-Integrated Sensor System for Wireless Gait Analysis and Real-Time Feedback. In *Second Joint IEEE/EMBS and BMES Joint Meeting*, 2002.

[47] Harvey Motulsky. *Analyzing Data with GraphPad Prism*. GraphPad Software Inc., 1999.

[48] Stuart D. Mowbray and Mark S. Nixon. Automatic Gait Recognition via Fourier Descriptors of Deformable Objects. In *Proceedings of Audio Visual Biometric Person Authentication*, pages 566–573. Guildford, 2003.

[49] Sylvie Nadeau, Bradford J. McFadyen, and Francine Malouin. Frontal and Sagittal Plane Analyses of the Stair Climbing Task in Healthy Adults Aged Over 40 Years: what are the Challenges Compared to Level Walking? *Clinical Biomechanics*, 18:950–959, December 2003.

[50] New York University. *SPSS for Microsoft Windows for new Users of SPSS V.9.0 on Microsoft Windows at NYU*, fourth edition, September 1999.

[51] NIST/SEMATECH. *e-Handbook of Statistical Methods*. NIST/SEMATECH, 2005. `http://www.itl.nist.gov/div898/handbook/` (Last visited 04032005).

[52] Marks S. Nixon, John N. Carter, John. M. Nash, Ping S. Huang, Dave Cunado, and Sarah V. Stevenage. Automatic Gait Recognition. In *IEEE Colloquium on Motion Analysis and Tracking*, pages 3/1–3/6, London, UK, 1999.

[53] Sourabh A. Niyougi and Edward H. Adelson. Analyzing and Recongizing Walking Features in XYT. In *IEEE Computer Society International Conference on Computer Vision and Pattern Recongition*, pages 469–474, June 1994.

[54] Lawrence O'Gorman. *Biometrics: Personal Identification in Networked Society*, chapter 2: Fingerprint Verification. Springer Verlag, first edition, 1999.

[55] Tricia Olsson. Strengthening Authentication with Biometric Technology. Technical report, SANS Institute, August 2003.

[56] George Orwell. *1984*. Plume, Centennial Edition edition, May 2003.

[57] Joseph A Paradiso and Stacy J. Morris. A Compact Wearable Sensor Package for Clinical Gait Monitoring. *Motorola Offspring Journal*, 2002.

[58] Radia Perlman. An Overview of PKI Trust Models. *IEEE Network*, 13(6):38–43, November/Desember 1999.

[59] Lawrence R. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. In *Proceedings of the IEEE*, volume 77, pages 257–285, February 1989.

[60] Albrecht Schmidt, Kofi Asante Aidoo, Antti Takaluoma, Urpo Tuomela, Kristof Van Laerhoven, and Walter Van de Velde. Advanced Interaction in Context. In *Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing*, volume 1707 of *Lecture Notes In Computer Science*, pages 89–101. Springer Verlag, 1999.

[61] Panchanathan Sethuraman and Kuchi Prem. Pat. WO2004099942 - Gait Recognition System, November 2004. http://v3.espacenet.com/textdoc?DB=EPODOC&IDX=WO2004099942&F=0&QPN=WO2004099942 (Last visited 05052005).

[62] Claude E. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27, 1948.

[63] Ph.D. Steven W. Smith. *The Scientist and Engineer's Guide to Digital Signal Processing*. California Technical Pub., 1997.

[64] Alex Stacoff, Christian Diezi, Gerhard Luder, Edgar Stüssi, and Inés A. Kramers-de Quervain. Ground Reaction Force on Stairs: Effects of Stair Inclination and Age. *Gait and Posture 21*, pages 24–38, November 2005.

[65] Christos Stergiou and Dimitrios Siganos. Neural Networks. Webpage. http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html/ (Last visited 27062005).

[66] Fong-Shin Su and Wen-Lan Wu. Design and Testing of Genetic Algorithm Neural Network in the Assessment of Gait Patterns. In *Medical Engineering & Physics*, volume 22, pages 67–74, February 2000.

[67] Rawesak Tanawongsuwan and Aaron Bobick. Performance Analysis of Time-Distance Gait Parameters under Different Speeds. In *4th International Conference on Audio- and Video Based Biometric Person Authentication*, Guildford, UK, June 2002.

[68] Chris Verplaetse. Internal Proprioceptive Devices: Self-Motion Sensing Toys and Tools. *IBM Systems Journal*, 35(3&4):639–650, May 1996.

[69] Harvey Weinberg. *Using the ADXL202 Duty Cycle Output*. http://www.analog.com/UploadedFiles/Application_Notes/320058905AN604.pdf Last visited 05052005).

[70] Greg Welch and Eric Foxlin. Motion Tracking: No Silver Bullet, but a Respectable Arsenal. *IEEE Computer Graphics and Applications*, 22:24–38, 2002.

[71] Bernard Willers and Sep Vrba. Artificial Neural Networks—Emulating the Operation of the Human Brain.

[72] Kuan Zhang, Patricia Werner, Sun Ming, F. Xavier Pi-Sunyen, and Carol N. Boozer. Measurement of Human Daily Physical Activity. *Obesity Research*, 11(1), January 2003.

[73] Wen-Yi Zhao, Rama Chellappa, P. Jonathon Phillips, and Azriel Rosenfeld. Face Recognition: A Literature Survey. *ACM Computing Surveys*, 35(4):399–458, 2003.

[74] Qifeng Zhu and Abeer Alwan. On the use of Variable Frame Rate Analysis in Speech Recognition. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 2000.

A The Gait Collector's Software Design

The software which is programmed onto the Gait Collector is partly based on the pre-programmed demo-code embedded with the AVR Butterfly. It is written in C and compiled with WinAVR. The main purpose of this software is to collect data from the two ADXL 202 accelerometers connected to the Buttefly, store these data and transfer it using the USART[1] interface when the butterfly is connected to a computer.

A.1 Interacting with the Butterfly

The Butterfly is equipped with both an LCD display and a joystick. Though these features are good methods for interacting with the Butterfly, they was not suitable for this project, since it was necessary to incapsulate the entire butterfly into a casing to protect its circuits during usage. The only available interface with the users is therefore the USART interface. The code that listen for user interaction through this interface is fairly simple and shown in listing A.1. It is a simplification of the main-loop of the Gait Collector's software, where—among other, it is checked for activity on the USART interface.

Listing A.1: The main loop in the Gait Collector software.

```
for(;;) {
 CollectAcc();     // Collect acceleration data
 WriteToFlash();   // Write acceleration to dataflash
 CheckUSART();     // Listen for commands from USART
}
```

As shown, the Gait Collector continuously collects a serie of acceleration data from the two accelerometers and saves it on the dataflash. After this has been performed, it listens for commands from the USART interface, and interacts according to the command. The command might tell the Butterfly to create a new experiment, dump the dataflash to the USART interface or delete the dataflash. If the received command is not recognised by the system, a brief menu will be sent to the USART informing the user some information regarding the available commands. The detection of whether there has been received a character on the USART is performed using the code in listing A.2.

Listing A.2: How to listen for activity on the USART interface.

```
return ((UCSR0A & (1<<RXC0)) ) ? UDR0 : NULL;
```

This method checks whether there is an unread byte in the UDR0 register by checking whether or not the RXC0 register is 1. UCSR0A is a status register which also is useful to read.

[1] Universal Syncronous Asynchronous Receiver / Transmitter — The interface for transferring data between the Butterfly and a computer.

A.2 Detecting the Acceleration

There are two ADXL202 accelerometers connected to the AVR Butterfly, each of them capable of detecting accelerations in 2 axis. The output from the accelerometers are in duty cycles, where the duration of each duty cycle is relative to the acceleration.

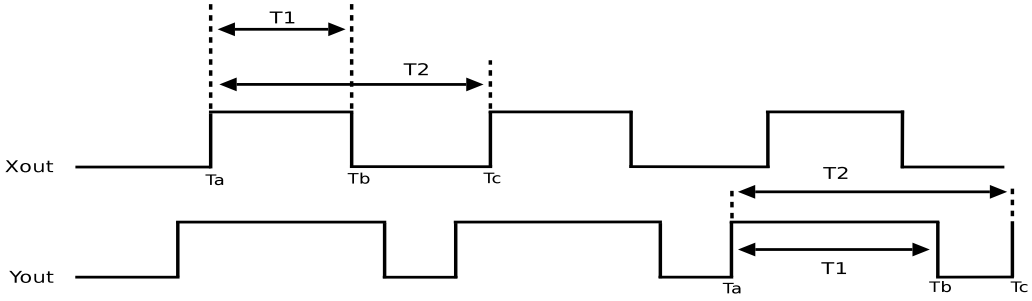


Figure A.1: The ADXL202 Duty Cycle Output.

Figure A.1 shows the duty cycle output from X_{out} and Y_{out}. The two accelerometers are mounted on the Buttefly at a straight angle on each other in order to capture acceleration in three axes. The function which performs the acceleration data collection makes use of a timer to detect the duty cycles from each of the accelerometers outputs simultaniously. This is done as shown in listing A.3.

Listing A.3: The code which registers the acceleration in X, Y and Z axis simultaneously.

```
ResetTimer();
do {
 Byte changePort = WaitForChange();
 if( changePort != 0b00000000 ) {
  time = GetTimer();
  if((changePort&0b00010000)==0b00010000)
    AccDataX[Xcounter++] = time;
  if((changePort&0b01000000)==0b01000000)
    AccDataY[Ycounter++] = time;
  if((changePort&0b10000000)==0b10000000)
    AccDataZ[Zcounter++] = time;
 } while( (Xcounter != 4) && (Ycounter != 4)
       && (Zcounter != 4) );
```

This is a simplified version of the code which collects the acceleration data from the accelerometers. As shown above, it makes a call to the function WaitForChange(), which looks like the code in listing A.4.

Listing A.4: The code which checks for changes in output from one of the accelerometers.

```
Byte res;
do {
 res = status^port;  // If status XOR port not is 0,
} while((res == 0)); // a change has occurred.
status = port;       // Res will tell which accelerometer
return res;          // has changed its output.
```

This function listens to the port where the two accelerometers are connected, performs a logical XOR operation to detect changes on the port, and loops while there are

no changes on these pins. It then updates the status byte with the new state, and returns the changes. It is then possible to perform a logical AND operation to determine whether or not there was a change at a given pin. The data will be put in the pin's array, and the array pointer will increase so the next value can be read. This way the values T_a, T_b and T_c from figure A.1 are read in all three axis, and it is possible to calculate T1 and T2, whereas T2 is the actual acceleration, and therefore the value that is written to the dataflash. This way, the accelerations in all three axis are measured simultaneously.

A.3 Storing the Gait Data

The AVR Butterfly is equipped with a 4Mbit dataflash [4] for storage purposes. The data collected from the accelerometers are stored on this dataflash until it is transferred to a computer.

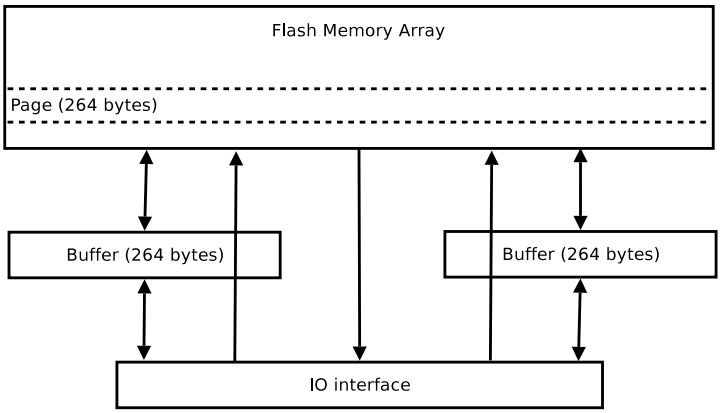


Figure A.2: The AT45DB041B dataflash arrangement.

The dataflash is arranged as shown in figure A.2. The dataflash consists of 2048 pages, each at the size of 264 bytes. The pages can be loaded into one out of two 264 byte buffers for reading and writing purposes, and the modified buffers can be sent back into one of the pages in the dataflash. This way it is possible to perform reading and writing operations on the dataflash much like you would read and write to a ordinary file. Using a serie of pointers, it is possible to determine which page in the dataflash should be written to and read from next. Similarly, two pointers can tell which byte in the buffer should be written to and read from. This way it is possible to implement simple read and write operations which just reads and writes data to and from the next position in the dataflash. Listing A.5 shows how it is possible to write one byte of data to the flash by monitoring where in the buffer and on the pages one are currently working.

Listing A.5: The code which writes a character to dataflash.

```
// If necessary, fetch a new page from the flash
if(BufferPointerW == 0)
 Page_To_Buffer(PagePointerW, 1);

// Write the next byte to the page
Buffer_Write_Byte(1, BufferPointerW, c);
BufferPointerW++;

// If exceeded the buffersize, add the page to the flash,
```

```
// and increase the pagepointer
if( BufferPointerW == BUFFERSIZE ) {
 Buffer_To_Page( 1, PagePointerW );
 PagePointerW++;
 Page_To_Buffer( PagePointerW, 1 );
 BufferPointerW = 0;
 // If exceeded the pagesize, increase the overflow pointer
 // and reset the two pointers
 if( PagePointerW == PAGECOUNT ) {
   Overflow++;
   BufferPointerW = 0;
 }
}
```

A.4 Transfering the Gait Data

The data stored on the Gait Collector is transferred to a computer using its USART interface. Any software capable of sending and receiving data from such an interface can be used for transferring the acceleration data. For this project, the program Terminal by Bray was used[2].

The process of dumping the gait data from the Gait Collector to USART is reading one and one byte from dataflash until it reaches an EOF marker. Listing A.6 shows the code for writing the entire dataflash to USART.

Listing A.6: The code for dumping the dataflash to USART.

```
// Dump dataflash
FlashCloseWrite();       // Write EOF to the end of flash
FlashOpenRead();         // Reset reading pointers
for (;FlashRead(&c,1);) // Read one byte until EOF
 sput(c);                // Send the byte to USART
FlashCloseRead();        // Reset reading pointers
FlashOpenAppend();       // Remove EOF marker
```

The sput() function is in reality an output stream to the USART interface, which was opened using the fdevopen() command.

The code for reading data from flash is very similar to the one for writing to flash, and is shown in A.7.

Listing A.7: The code which reads a character from dataflash.

```
char c;

// Fetch the current page into the buffer
Page_To_Buffer( PagePointerR, 1 );

// Read the next byte from the buffer
c = Buffer_Read_Byte( 1, BufferPointerR );
BufferPointerR ++;

// If the end of the buffer is reached,
// increase the page pointer
if( BufferPointerR == BUFFERSIZE ) {
```

[2]Available from `http://bray.velenje.cx/avr/terminal/` (Last visited 08052005).

```
   BufferPointerR = 0;
   PagePointerR ++;
   // If the end of the dataflash is reached,
   // reset the pointers
   if( PagePointerR == PAGECOUNT ) {
    PagePointerR = 0;
   }
  }
return c;
```

B Matlab Code for Local Minima Detection

This function was used to compare two datasets by looking for spikes in the results from the spectrum subtraction. The whole Matlab code is not listed, but enough of the code is provided to understand the local minima detection process.

Listing B.1: The code for detecting local minimas in a dataset.

```
% Apply polygamma function and smooth the dataset
for(i=1:15);
    comp(:,i) = smooth(psi(comp(:,i)));
end;

filter = 35; % The filter strength to use
spikes = zeros(15,size); % The array containing the spikes

% Iterate through all 15 comparisons
for(i=1:15);
% Detect the spikes using Serge Koptenko's function
 [l1,l2] = lmin(comp(:,i),filter);

 % Find the start and end position of each spike
 % and put it in the spikes array
 for(j=1:length(l2));
  min = l2(j);
  max = l2(j);
  while((min>2)&&(comp((min-1),i)>comp(min,i)))
   spikes(i,min)=1;min=min-1;
  end;
  while((max<size-1)&&(comp((max+1),i)>comp(max,i)))
   spikes(i,max)=1;max=max+1;
  end;
 end;
end;
```

Listing B.1 shows how the code for detecting local minimas work. As shown, it starts by applying the polygamma function and a smooth filter to all the datasets that should be compared. This enhances the local minimas which is desirable to detect, and removes the smallest minimas that should not be included in this detection process. Then Serge Koptenkol's local minima function[1] is applied to perform the local minima detection

[1] `http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=3170&objectType=file` (Last visited 24052005)

itself. The results from this detection is then used to try detecting the size of each minima, and the results is being placed in the spikes array.

Unfortunately, this algorithm does not perform particularly well compared with the other analysis methods that were performed. This is probably due to the fact that the local minima detection algorithm does not manage to detect the spikes in the datasets that are characteristic for a comparison of two datasets from the same person. Therefore, this algorithm must be improved further before it can be used in an authentication system.

C Statistical Tests on the Results from the Large Scale Experiment

After the large scale experiment had been performed, a serie of methods for statistical analysis were performed on the results. Since the results from these tests were not analysed or discussed in a great extent, they are presented in this appendix. Only a brief summary of the results from these tests are listed here. The full output of the tests in Microsoft Excel format can be downloaded from `http://www.torkjel.com/results/SPSS_largescale.xls`.

The statistical analysis was performed in SPSS[1], a widely used analysis software that can test statistically subjects, like correlation and hypotheses.

C.1 How the Results were Analysed

Three different statistical tests where performed on the results from the experiment to detect whether there is any kind of relation between persons compared with themselves and persons compared with each other.

C.1.1 Student's T-test

This test checks whether two normally distributed populations are significantly different, or only differs due to chance. The latter case is often referred to as the null hypothesis. The t-test results in a significance value, called the P-value. If this value is less than 0.05 (5% probability), it is reasonable to believe the two groups are significantly different [50, 47, 22].

C.1.2 Mann-Whitney U Test

The Mann-Whitney test is another version of the t-test, and compares two groups testing whether the degree of overlap among these two groups is less than would be expected by chance. This is indicated using a ranking score and a U-value. During this analysis the U value can range between 0 and $20 \times 760 = 15200$. The U-value tells how many times a value in the first bin precedes a value in the second. Like with the t-test, the Mann-Whitney also return a P-value indicating the significance of the differences in the tested populations, but while the t-test is testing the means of the populations, the Mann-Whitney test looks at their medians [26, 22, 3]. The null hypothesis of a Mann-Whitney test is therefore that the median of the two bins are equal.

[1] `http://www.spss.com`.

C.1.3 Correlation

Correlation is a method for detecting linear relationships between two or more variables [22]. The correlation can be positive, negative or non-existent. This can be determined using Pearson's correlation coefficient, which is a value ranging from -1 to 1 where -1 indicates a perfectly negative correlation, 1 indicates perfectly positive correlation, and 0 indicates no correlation between the two variables.

C.2 The Results from the Statistical Analysis

The results from the Student's t-test and the Mann-Whitney test are summarised in Table C.3. The table shows whether it is reasonable to believe that the comparison of data from the same person and data from different persons will form two distinctive groups. As shown in this table almost all analysis methods get a P-value far below 0.05, which indicates that the way they grouped was not due to chance.

As shown in Table C.3, all except one method gets a P-value below 0.05 after the t-test, which indicates that the groups they form are not due to chance. The Mann-Whitney test gets P-values below 0.05 for all the analysis methods. This indicates that the overlaps that occur between the groups in each test are not due to chance. The obtained P-values in this test are being confirmed by the U-values, since hey are all outside the rank values.

C.2.1 Results from the Correlation Test on the Physical Characteristics of the Participants

A correlation test was performed to test whether there were any correlation between the participants' physical characteristics. The results are shown in Table C.1. In this table, the first value for all of the analysis methods should be close to 1 or -1 for a correlation to occur, while the second value should be less than 0.05 to indicate a significance. The table shows that there are some extent of correlation in the persons' weight and foot size, while there are almost no correlation in their height and leg length. However, the obtained correlations are very low.

C.2.2 Results from Student's T-test on the Genders

The last test being performed on the results from the large scale experiment was a Student's t-test on how differences between genders was compared to differences within the genders. The results from this test is shown in Table C.2. As this table shows the obtained P-values are in all except one case much larger than 0.05, which indicates that a person's gender does not effect the results of the data analysis in any extent. However, why analysing the horizontal alignment of the resultant vector gets a large differ within the genders are uncertain.

Method	Differences			
	Weight	Height	Leg length	Foot size
Comparing FFT of X axis data	-0.029	-0.009	**0.198**	-0.033
P-value	0.426	0.798	**0.000**	0.355
Comparing FFT of Y axis data	**0.212**	0.005	0.088	**0.188**
P-value	**0.000**	0.886	0.014	**0.000**
Comparing FFT of Z axis data	0.067	**0.098**	-0.018	0.057
P-value	0.069	**0.006**	0.616	0.113
Comparing FFT of the XYZ Resultant Vector	**0.208**	-0.006	0.035	**0.108**
P-value	**0.000**	0.867	0.324	**0.002**
Comparing FFT of YZ Resultant Vector	**0.216**	-0.005	0.042	**0.116**
P-value	**0.000**	0.887	0.246	**0.001**
Local Minima Detection on the FFT of XYZ Resultant Vector	**-0.148**	0.016	0.006	0.011
P-value	**0.000**	0.655	0.868	0.752
Local Minima Detection on the FFT of YZ Resultant Vector	**-0.157**	-0.013	**-0.076**	-0.064
P-value	**0.000**	0.715	**0.033**	0.073
Comparing the FFT of the Horizontal Alignment of XYZ Resultant Vector	0.001	0.045	0.052	**0.238**
P-value	0.984	0.207	0.148	**0.000**
Comparing the FFT of the Vertical Alignment of XYZ Resultant Vector	0.066	-0.030	**0.175**	0.068
P-value	0.072	0.400	**0.000**	0.057

Table C.1: The obtained data when performing a correlation test on the results from the large scale experiment. **The first values** shown for each of the methods are the correlations between differences in the compared persons' physics and the analysis method used. A value close to -1 and 1 indicates there is a correlation, while a value close to 0 indicates no correlation. **The second values** show the corresponding P-values, where a value lower than 0.05 indicates the correlations where not obtained due to chance. The situations with high correlation is emphasised.

Method	P-value
Comparing FFT of X axis data	0.475
Comparing FFT of Y axis data	0.774
Comparing FFT of Z axis data	0.516
Comparing FFT of the XYZ Resultant Vector	0.487
Comparing FFT of YZ Resultant Vector	0.526
Local Minima Detection on the FFT of the XYZ Resultant Vector	0.360
Local Minima Detection on the FFT of the YZ Resultant Vector	0.183
Comparing the FFT of the Horizontal Alignment of the XYZ Resultant Vector	0.011
Comparing the FFT of the Vertical Alignment of the XYZ Resultant Vector	0.191

Table C.2: Shows the obtained P-values after applying Student's t-test to detect how differences in genders compares with differences in the two bins. **P-values** less than 0.05 indicates differences in genders are larger than differences in the bins for other reasons than chance.

Method	t-test p-value	Mann-Whitney test p-value	U-value	Sum ranks		Mean ranks	
				Group 1	Group 2	Group 1	Group 2
Comparing FFT of X axis data	0.000	0.000	2,249	2,459	302,131	123.0	379.5
Comparing FFT of Y axis data	0.000	0.000	3,872	4,082	300,508	204.1	395.4
Comparing FFT of Z axis data	0.003	0.005	4,798	5,008	299,582	250.4	394.2
Comparing FFT of the XYZ Resultant Vector	0.025	0.000	4,113	4,323	300,267	216.2	395.1
Comparing FFT of YZ Resultant Vector	0.027	0.001	4,216	4,426	300,164	221.3	395.0
Local Minima Detection on the FFT of XYZ Resultant Vector	0.005	0.005	4,835	10,574.5	294,015.5	528.7	386.9
Local Minima Detection on the FFT of YZ Resultant Vector	0.052	0.036	5,520	9,890	294,700	494.5	387.8
Comparing the FFT of the Horizontal Alignment of the XYZ Resultant Vector	0.000	0.000	3,322	3,532	301,058	176.6	396.1
Comparing the FFT of the Vertical Alignment of the XYZ Resultant Vector	0.000	0.000	2,769	2,979	301,611	149.0	396.9

Table C.3: A summary of the results from the statistical Student's t-test and Mann-Whitney tests using a 95% confidentiality interval. If the **p-values** are less than 0.005, one can conclude that the probability of this is statistically unlikely. An **U-value** outside the range of the two values in the **Sum of Ranks** column confirms the obtained P-values. The **Mean Rank** columns shows whether there are in some extent a gap between the two bins, indicating the extent of overlapping in the two groups.

D Statistical Tests on the Results from the Long Term Experiment

As with the large scale experiment in Appendix C, a serie of methods for statistical analysis were performed on the results from the long term experiment using SPSS[1]. The complete results from this analysis are available from `http://www.torkjel.com/results/SPSS_longterm.xls`. The data was analysed using the Student's t-test.

D.1 The Results from the Statistical Analysis

The results from the Student's t-test are summarised in Table D.1. The table shows whether it is reasonable to believe that the results obtained from the experiments are representative for how comparisons of data from one person and comparison of data from two different persons will form two distinctive groups. As the table shows, almost all methods show a statistically significance in their results indicating that the two groups they form are not due to chance. Only two of the methods receives a p-value above 0.05. Why they not have statistically significance is unknown.

Method	P-value
Comparing FFT of X axis data	0.000
Comparing FFT of Y axis data	0.000
Comparing FFT of Z axis data	0.078
Comparing FFT of the XYZ Resultant Vector	0.000
Comparing FFT of YZ Resultant Vector	0.000
Local Minima Detection on the FFT of the XYZ Resultant Vector	0.087
Local Minima Detection on the FFT of the YZ Resultant Vector	0.000
Comparing the FFT of the Horizontal Alignment of the XYZ Resultant Vector	0.000
Comparing the FFT of the Vertical Alignment of the XYZ Resultant Vector	0.000

Table D.1: Shows the obtained P-values after applying Student's t-test to detect whether the results from comparing data from the same person and data from two different persons forms two groups the way they do due to chance. **P-values** less than 0.05 indicates the differences are not due to chance.

[1] `http://www.spss.com`.