Gesture Based Input Method for Wearable Devices

Anders Beck

Kongens Lyngby 2018

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Richard Petersens Plads, building 324,
2800 Kongens Lyngby, Denmark
Phone +45 4525 3031
compute@compute.dtu.dk
www.compute.dtu.dk

Abstract

Modern technologies has allowed for the opportunity to collect personal data about human activities related to health. Some activities such as heart rate and physical activity can be collected passively and fully automated, but collecting psychological activities requires the user to manually log the data. Traditionally this has been achieved with pen and paper. While modern solution utilizing smartphones and smartwatches has greatly improved this, there still is a need to reduce the burden related to self tracking. Therefore this thesis proposes a solution using a wristband device together with arm gestures for self tracking. This will allow for tracking psychological activities in the moment, and allowing the user to rate the severeness on a continues scale similar to a Visual Analogue Scale (VAS). The wristband should be used together with a companion app that can import, display and export the data collected on the wristband. The thesis details the design and implementation of this solution. An experiment was designed and conducted to test the implemented solution compared to digital version of a VAS. In the experiment participants rated shades of grey and numbers using both the wristband with two different gestures and a digital VAS.

The results showed that the accuracy of the solution was not significantly different from the digital VAS, when rating shades of grey. When rating numbers the digital VAS performed significantly better. Mapping the results to a 0-10 scale (resembling a VAS), the mean absolute error using the wristband was within ± 1, about 50% of the data points were within ± 1 and about 80% of the data points were within ± 2, concluding that the solution is reliable. Comparing data entry time showed no significant difference between the wristband and digital VAS.

Preface

This thesis was prepared at DTU Compute in fulfillment of the requirements for acquiring an M.Sc. in Engineering.

The thesis detail the design and implementation of a wristband device and companion app for manual collection of data on subjective experiences using a gesture based system. An experiment was conducted in order to compare the performance of the wristband with a touch screen implementation of a Visual Analogue Scale.

Lyngby, 02-July-2018

Anders Beck

Acknowledgements

First I would like to thank my supervisor Jakob Eg Larsen for suggesting me the topic for this thesis. Furthermore I want to thank him for his guidance, feedback and engagement throughout this project.

Thanks to everyone that volunteered their time to participate in the experiment.

Thanks to my friends, family and former colleagues at IBM for support and their feedback to this project. A special thanks to my friends and former colleagues Morten Due Christiansen and Alexander Lillelund for all their help and feedback in the development process.

Last but not least I would like to thank my girlfriend Julia for all her help, support and patience with me throughout this project.

Contents

Abstract **i**

Preface **iii**

Acknowledgements **v**

1 Introduction **1**
1.1 Thesis Goals 3
1.2 Contributions 3
1.3 Thesis Structure 4

2 Related Work **5**
2.1 Reducing the Burden of Self Tracking 5
2.2 Long Term Use of Self Logging 8
2.3 Comparing Ratings of Behavior 9
2.4 Authors Previous Work 10
2.5 Relation to This Thesis 11

3 Analysis **13**
3.1 Self Tracking 13
3.2 Usage of VAS 14
3.3 Orientation in 3D Space 15
3.4 Inertial Measurement Unit 16

4 Design **19**
4.1 Overview 19
4.2 From Gesture to Value 20
4.2.1 Orientation of the Arm 20
4.2.2 Problems with Yaw 22

4.2.3 Mapping Euler Angles to a 0-10 Scale 22
4.3 Functionalities and Requirements 23

5 Implementation **25**
5.1 Proof of Concept 25
5.2 Prototype 27
5.3 Companion App 30
5.4 Performance Testing 31

6 Experiment **35**
6.1 Participants 36
6.2 Apparatus 37
6.2.1 Essential Hardware 37
6.2.2 Tablet App 38
6.2.3 Other 46
6.3 Procedure 47
6.4 Design 48
6.5 Pilot Tests 48

7 Results and Discussion **51**
7.1 Accuracy 54
7.1.1 Distribution and Scatters of Stimuli and Responses 54
7.1.2 Error Distribution and Variance 59
7.2 Task Completion Time 65
7.3 Post Experiment Survey 65

8 Future Work **69**
8.1 Improvement to the Experiment 69
8.2 Case Study 70
8.3 Improvements to Companion App 70
8.4 Investigating the Wrist's Critical Point 71

9 Conclusion **73**

A Survey **75**

B Outlier Visualization **79**

C Code **87**
C.1 Proof of Concept 87
C.2 Companion App 87
C.3 Experiment App 88
C.4 Data analysis 88

D Data **89**

E Experiment App Screens **91**

Bibliography **101**

Chapter 1

Introduction

Many patients suffer with a chronic illness or disorder (e.g. chronic pain[28], migraines[15], anxiety[7], panic attacks[8] or PTSD[34] just to name a few). To better treat these illnesses it is important to figure out what can trigger an episode (episode meaning the negative effect of the illness, e.g. a headache if suffering from migraines), this can be done by tracking when and where episodes happen in order to look for patterns. The problem with these illnesses is that their episodes require to be tracked manually by the patient, since no technology yet exist that can measure these on its own. Providing a way to accurately gather when patients are experiencing the effects of their conditions would be valuable information that could help treat their illness, or at least help to prevent scenarios that can trigger episodes.

Manual tracking has traditionally been done with pen and paper, either in the moment or as a daily diary where the patient by the end of the day will log their experiences for the whole day. The problem with daily diaries are that they introduce memory recall bias[35], which causes issues with the reliability of data reported by patients. Thus it is important to gather data on patients when they are exposed the effects of their illness. In recent times different different solutions have been created trying to improve this, using an app on either a smartphone or smartwatch eliminates memory recall bias by enabling the patient to log what they are experiencing in the moment. At first this seems like an ideal solution, but one problem becomes apparent and that is burden of use. Using a tracking app on a smartphone requires the user to get the smartphone out of their pocket, turn it on, unlock it, find the app, open it and then log the event. While this is a bit easier using a smartwatch since it in it natures is always at your wrist, it still requires at least a couple of interactions and requires the user's full attention while interacting with the smartwatch. It is of great importance to minimize the burden of logging experiences, and often the simplest solutions are shown to be the most effective.

An example has been a "smartbutton" which is a wristband device equipped with a single button, pressing it will log a single time stamp[21]. The interaction burden with such a device is very minimal, and it doesn't require the user's full attention, since it can be done whit out looking at the device. The downside of this solution is that it only allow to log when an event occurred and not the severeness. Another problem is how to measure and compare psychological events, in the medical field a Visual Analogue Scale (VAS)1.1 is used to measure such events by letting the patient rate the experience on a continuous scale from 0 to 10. The meaning of the scale is then described to fit the domain, most used is the pain scale as seen in Figure 1.1, where 0 is no pain and 10 is the worst imaginably levels of pain. The VAS can easily be used in many other domains simply by changing the explanation for the scale, for example to use with panic attacks 0 would be no panic attack and 10 would be worst imaginably panic attack.

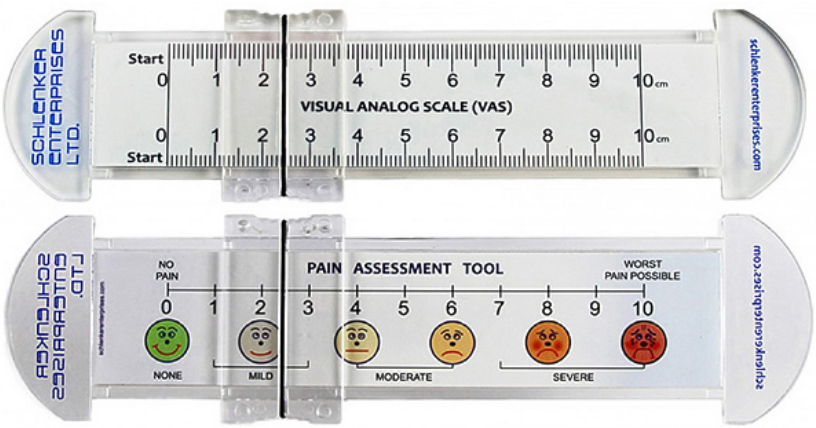

Figure 1.1: VAS Pain Scale Ruler[2]

The proposed solution shall allow users to track whenever they experience events related to their illness. It need to be simple to use in order to minimize the burden of use. It should be able to track the severeness of the events in a way that resembles a VAS. The solution proposed in this thesis will be to create a wristband device that using gestures will map the rotation of the wrist or the elevation of the arm to a 0-10 scale. When the button on the wristband is pressed it will log the information based on the gesture. A companion app will be created to collect the data from the device and display the results to the user.

1.1 Thesis Goals

In this thesis the design and implementation of a wristband device and companion app for manual collection of data on subjective experiences using a gesture based system will be described. It aims to minimize the burden compared to other solutions and it aims to be reliable. An experiment will be conducted in order to compare the performance of the wristband with a digital version of a VAS.

1.2 Contributions

The focus of this thesis has been to design, implement and test two gesture based input methods to track the severeness of psychological experiences. The work of this thesis has resulted in the following contributions:

1. Designed a system for tracking psychological experiences and their severeness on a continuous scale

2. Implemented said system using a wristband device and Android app

3. Created a method to map gestures of wrist and arm orientation to a continuous scale

4. Designed an experiment for comparing said gestures to a VAS

5. Conducted said experiment which results demonstrated the reliability of the implemented system

6. Stated further problems worth investigating and possible improvements to the gesture based input technique

1.3 Thesis Structure

This thesis is structured as follows. Chapter 2 detail the some of the most related work and it relates to this thesis. Chapter 3 analysis the key concepts for this thesis that will form the foundation for this thesis. Chapter 4 details the design process for creating the proposed solution. Chapter 5 details the implementation of the designed solution. Chapter 6 presents and details an experiment conducted in order to compare the proposed solution to a digital VAS. Chapter 7 presents the results collected results and discusses their meaning. Chapter 8 describes potential future work that could improve both the designed solution and experiment. Finally Chapter 9 concludes the thesis.

Chapter 2

Related Work

With the popularity of smartphones and smartwatches a lot of new possibilities has been made possible on the world of self tracking. This has in turn created a lot of research related to the topic. This Chapter will go through the most relevant work that related to this thesis, describing the work that has been done by others and how it relates to this thesis.

2.1 Reducing the Burden of Self Tracking

Self tracking can solve the problem of memory recall bias, by letting the user register observation in the moment (or very close to) they occur. But requiring the user to self register observations introduces a burden of the work involved in doing so. As mentioned traditionally this has been done with pen and paper, but with the invention of smartphones and smartwatches this burden can be reduces. The work of Ponnada et al., 2017[29] investigate the interruption burden with both smartphones and smartwatches when used with both regular EMA and μEMA. The difference between EMA and μEMA is that EMA will prompt the user less than μEMA, but requires the user to answer several questions back to back, where μEMA will instead prompt the user much more frequently, but only require the user to answer one question, which can be answered on the same screen showing the question.

In a prior 4 week pilot study Ponnada et al.[29] found that smartwatch μEMA demonstrated higher response rates and participants reported a lower perceived burden than smartphone EMA, even though the interruption rate for smartwatch μEMA was 8 times higher than smartphone EMA. In a new 4 week study they gathered data based on a smartwatch EMA in order to determine if the previous result where due to the fact that the data was gathered on a watch, or if it in fact was due to the μEMA. Their results showed that there were no statistically significant differences in compliance, completion, and first-prompt response rates observed between smartphone EMA and smartwatch EMA. But smartwatch μEMA response rates where significantly higher than smartwatch EMA. Their results suggest that the higher compliance and lower perceived burden is more likely to do with μEMA than the device itself and that compliance with EMA may not improve simply by changing the device.

The work of Ponnada et al. only looked into EMA where the device prompts the user for information, not situations where the user registers an observation. In his thesis Kamiński[20] investigates how to enable self logging with different forms of input: a binary sample, selection input form a list, VAS and a numerical scale. He designed and implemented all of the input methods onto a smartwatch, where on the watch face each method input method could be selected. In Figure 2.1 we can see his design, the watch face has a shortcut for each input method. In Figure 2.2 we see how the interaction occurs, and that it takes 1-3 touches to log the desired data depending on the input method.

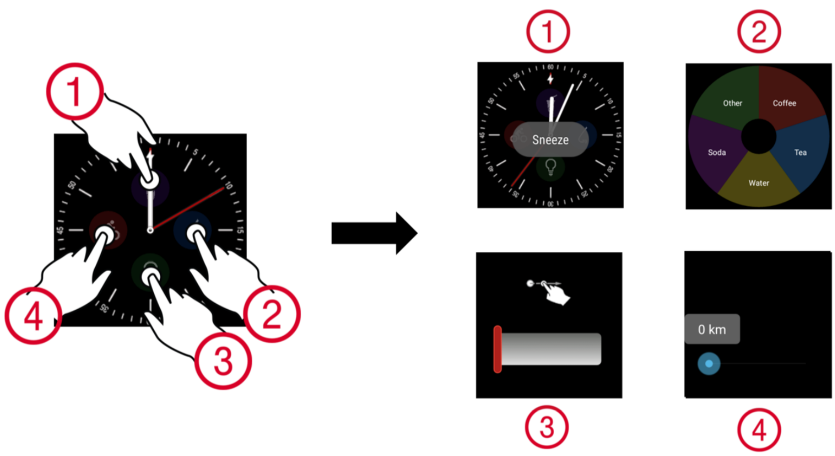

Figure 2.1: Design for self logging on a smartwatch by Kamiński[20, p.23]

For comparison Kamiński also implemented a similar design on a smartphone following the *Android material design guide*[17] and conducted an experiment

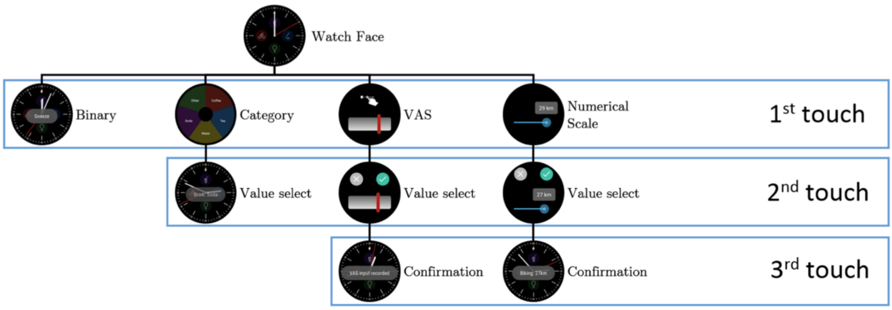

Figure 2.2: Interaction for self logging on a smartwatch by Kamiński[20, p.23]

where he compared the two devices performance (interaction time, lower is better). His result showed a reduction in the user interaction time by approximately 30% on average, thus successfully reducing the burden related to self logging. Beside his experiment Kamiński also had himself and to others use his design to collect data about their daily life (sneezing, skin itching and water consumption) over a 13-21 week period. What he found was a high compliance rate, all though it is a fairly biased test, since all three participants where involved in the project and are working with self logging technologies on a daily basis.

Even though Kamiński's design was a success, there persist an underlying problem with smartwatches; that they still require the user to user to activate them (often by tapping the screen several times) and then the user has to look at the screen to touch the right section to register the input. Dam-Jensen[16] investigates this issue in his thesis. He introduces three ways of performing self logging, all using a "smartbutton" (wireless button connected to smartphone/smartwatch) to achieve a display-less logging system. The three input method that Dam-Jensen designed was:

1. Hold Button: Measure the time the button is held down.

2. Rotate Lower Arm: Using the built in gyroscope in the smartbutton, the difference in rotation from when the button was pressed to it was released is measured. In order to achieve a measurement the user would hold the button in the hand, rotate the lower arm to a neutral position, hold down the button, rotate the lower arm to the end position and release the button.

3. Rotate Upper Arm Around Elbow: Similar to the previous method, but instead of rotating the lower arm, the user would raise/lower his/hers upper arm around the elbow.

For a comparison baseline method a standard VAS was used. In order to compare his three methods to the VAS Dam-Jensen conducted an experiment based on the work of Matejka et al.[23], where participants are asked to rate a color between white and black on a grey scale where the colors are equally spaced out on the CIELAB color space[14] (more on this later). His results showed that there where no significant difference in accuracy between his three designed input methods compared and to the baseline, thus meaning that all three methods are viable and could be used for self logging, and in their simple nature can reduce the burden on the user.

2.2 Long Term Use of Self Logging

As mentioned Ponnada et al.[29] conducted two four week studies with EMA and μEMA and Kamiński conducted a even longer 13-21 week use case where three people used his design to track daily activities. Both Ponnada et al. and Kamiński's work show great potential for real use, but none of them utilizes a device similar to that of Dam-Jensen's work, where there is no screen and only a single button for registering observations.

In a case study Larsen et al.[21] investigated the real life use case of a novel device for self logging. They had a Danish veteran suffering from PTSD equipped with a "smartbutton", that when pressed will store a time stamp. By connecting a smartphone the data can be transferred from the smartbutton to a data file for further analysis. The veteran was then instructed to press the button whenever he experienced a specific "trigger event" related to his PTSD (the "bodily sensation" was chosen after a two-hour assessment interview together with his therapist). The veteran wore the smartbutton for 100 days, pressing the button whenever he experienced the specific bodily sensation.

The results showed that the veteran had a high compliance, actively using the device throughout the 100 days (due to a technical issue no data was collected during the second week). The data collected revealed patterns that explained why/when the veteran's events would be triggered, which would not have revealed itself from the conversational assessment. In other words, not only was it a success from a technical and usability standpoint, but it also succeeded in bringing valuable new information to the treatment process.

2.3 Comparing Ratings of Behavior

When developing new methods for self logging it is important to access the feasibility of these methods. As mentioned Kamiński did this simply by comparing the user interaction time, while Dam-Jensen did a more comprehensive comparison by looking into how precise the data recorded with his designs compared to the VAS. He did so by having his participants rate shades of grey with the different input methods. That experiment was inspired by the work of Matejka et al.[23] where they investigate the bias of different slider designs. In order to do so they base their experiment on the work of Borg and Borg[12] that extensively studied using a “scale of blackness” as a stimulus and concluded that it serves as a good test of general rating behaviour, and further more it provides a truth value (the color displayed) which can be compared to the response rating from the participant. Matejka et al. designed their experiment to test severely different slider designs with various combinations of labels, tick marks and bands. In Figure 2.3 a design with 2 labels and 5 ticks is shown. Matejka et al. recruited participants via *Amazon Mechanical Turk*[6]. Participants where then showed one of the 50 shades of grey (as seen in Figure 2.4), then asked to rate the color from white to blakc using the slider. This process was repeated 50 times so each participants was exposed to all the shades of grey (in random order), this was then repeated until the participants had been exposed to each shade of grey 4 times, 200 trials in total. Each participant was only exposed to one slider design, and was afterwards excluded to participate further.

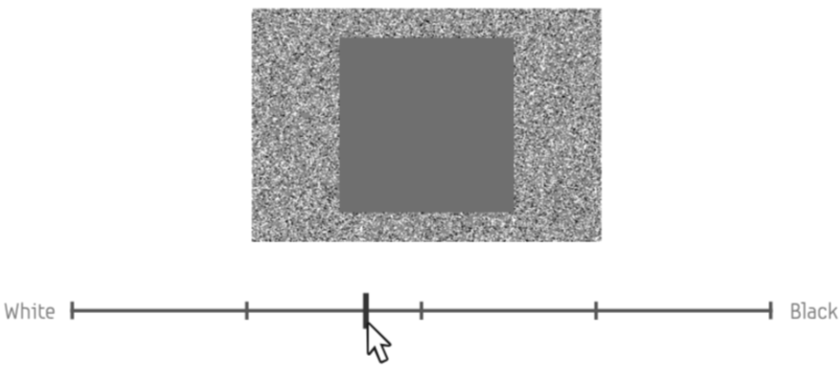

Figure 2.3: Slider bias experiment by Matejka et al.[23] with 2 labels and 5 ticks

With their results they investigated the bias for each slider design by looking at the distribution of responses and calculating a “smoothness” value representing the bias as well as precision and response time. Their results showed that sliders

Figure 2.4: 50 shades of grey used by Matejka et al.[23]

with ticks introduces the most bias towards the tick marks, followed by sliders with many labels and combination of the two. Banded sliders showed only small amount of bias and sliders with only two labels showed the least amount of bias. The higher amount of both ticks and labels resulted in higher precision and the average reaction dime only differed by one second between the fastest and slowest designs.

2.4 Authors Previous Work

In a previous project the author created an app for self tracking a person's mood[10]. The app's main screen had 11 buttons, labeled from 0-10, with the added text on "Very bad mood" on 0, "Neutral bad mood" on 5 and "Very good mood" on 10 and a button labeled "Mood history" as seen on the left in Figure 2.5. When one of the buttons from 0-10 is pressed the "mood value" and a time stamp is saved. Pressing the "Mood history" will take the user to the second screen (seen on the right in Figure 2.5) and a scroolable list of all recorded moods are shown, with the latest entry on top together with a button for sharing the mood history (export to a CSV file and send via email) and a button to delete the mood history.

The focus off that small project was on the implementation of the app, so no investigation went in to the actual use of the app (beside the author testing it on himself), but it was prioritized that the app was as quick and easy to use as possible. With the limitation of being a smartphone app (not a widget similar) this design seems to one of the most simple and efective solutions, it enables the user to quickly register events, since the first screen seen when the app is opened is the main screen. But it does illustrate the limitation of smartphone apps, comparing it to the designs of Dam-Jensen it is quite clear that the time and burden of getting your phone out of your pocket, turning it on/unlocking it, finding and opening the app before finally being apple to register a input is much greater than the time and burden involved with for example rotating your lower arm and pressing a button.

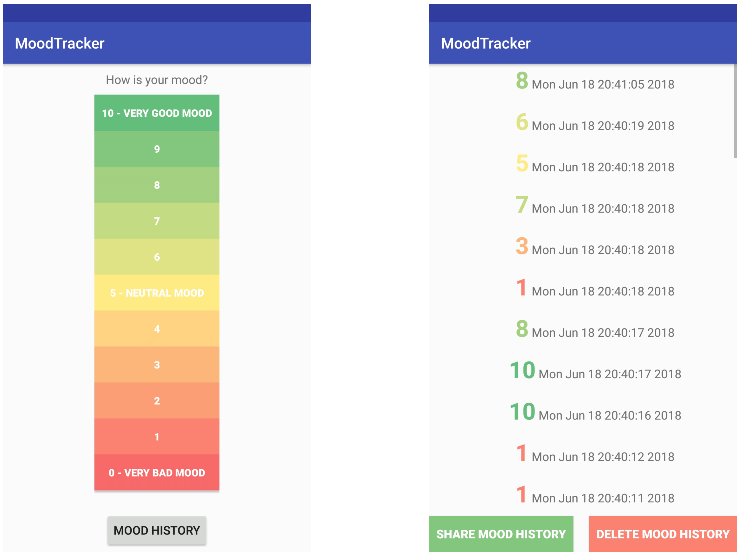

Figure 2.5: Mood Tracker app. Register a mood on the left & mood history on the right[10]

2.5 Relation to This Thesis

The work of Ponnada et al.[29], Kamiński[20] and Dam-Jensen[16] all point towards the benefit reducing the burdens. Their work has been a great inspiration for this thesis, together with Larsen et al. work showing the real world benefit of in situ measurements. This thesis will further investigate two of Dam-Jensen's designs, the "Rotate Lower Arm" and "Rotate Upper Arm Around Elbow" methods, developing these methods a bit further and then comparing them to a VAS by a "rating of blackness" experiment. And the authors previous work has shown that even a well designed smartphone app still creates a lot of burden for the user when comparing it to the more novel designs of Dam-Jensen[16].

Chapter 3

Analysis

The previous Chapter described the most relevant work related to this thesis. This Chapter will analyze and describe the key concepts that will be used in this thesis.

3.1 Self Tracking

Experience Sampling Method (ESM) and Ecological Momentary Assessment (EMA) are the research methodologies that refer to the usage of self tracking[22]. What defines these methods are the use of collecting real world data in the moment, whether that be psychological (thoughts, feelings, mood, pain etc) or physical (food/drinks consumed, sneezing, cramps etc). Traditionally this had been achieved with diaries, which could have the same set of questions printed on each page designed to collect the desired data. In modern times digital diaries have been created, for example as apps on smartphones and smartwatches. The benefit of these methods is that the prevent recall bias, since the data is collected in the moment (or shortly after) compared to daily diaries where a user by the end of the day has to think back and recall the data. The more time passed between the event and the time the data is collected, the larger the risk of recall bias will be, and the data can even be forgotten.

When collecting data that requires active self tracking, there will always be a burden involved. The simpler the data is to collect, the smaller the burden can be. For example tracking each time a person sneezes only requires a time stamp which can be collected by wearing a smartbutton like the one made by Larsen et al.[21]. On the other hand, if tracking the pain levels of a patient, the patient needs a way to input that on scale preferably a VAS. One solution has been to create a digital VAS on a smartphone or smartwatch, similar to the work of Kamiński[20]. Compared to the smartbutton, this solution has a much higher burden, the user must turn the display on, navigate the software to input the value and it requires a user's full attention to execute. With the improvements Dam-Jensen[16] introduced, the burden was lowered to only have the user press, hold and release a button while doing an arm gesture.

As a evolution to the design of Dam-Jensen[16], this thesis will introduce a self tracking method where the user will follow the same gestures (described later in this Chapter) and press a button only once (not press, hold and release). This new method will reduce the burden even further compared to the other mentioned methods.

3.2 Usage of VAS

In Figure 1.1 we see one of the VAS rulers that is sold to and used by hospitals to measure the perceived pain level of patients, while they come in different designs their principals are the same. On one side of the ruler we see 5-6 faces with expression going from happy to discomfort to sad and crying, these expression are meant to illustrate the amount of pain and are often displayed together with a text description explaining the feeling of pain as seen in Figure 3.1. On the other side of the VAS ruler there is regular ruler of 10cm (sometimes shown as 100mm instead). On the ruler there is a slider, the patient will be shown the side of the ruler with the faces, and asked to adjust the slider to fit the level of pain they experience, on the backside the slider will indicate a position on the ruler from 0 (no pain) to 10 (most pain), this position is then registered by a doctor or nurse as the pain level. The VAS is not limited to measuring pain, it can be used for any subjective characteristics or attitudes that cannot be directly measured, for example mood, annoyance, anxiety, but the measuring of pain is the most common use of VAS.

The VAS can be compared to the Likert scale and Borg scale[18], but what makes the VAS different is that it is continuous where the others are discrete. Evidence show that VAS have better metrical characteristics compared to discrete scales and therefore a wider range of statistical methods can be applied to

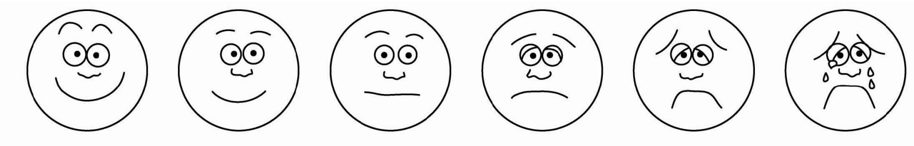

Figure 3.1: Example of faces and descriptions on VAS for measuring pain[27]

the measurements[31].

Despite its benefits VAS are a burden to use, even digital version since the require a screen and attention. It is therefore interesting to investigate if other solutions with less burden can perform equally or better than the VAS.

3.3 Orientation in 3D Space

Before creating a device that translates the orientation of the hand to a scale, one must first understand orientation in 3D space. We can rotate an object around each of its 3 axis, these actions are refereed to as pitch, roll and yaw. In Figure 3.2 we see an illustrations of these rotations performed on a airplane.

Pitch is rotation around the axis across the airplane, this rotation can be felt during takeoff and climbing (shortly after the wheels lift of the ground) of a airplane, where the pitch is dramatically changed from 0° (parallel to the ground) to $\sim$15-20°[9].

Roll is rotation around the axis along the airplane, this rotation can sometimes be felt when an airplane has to perform a sharp turn and then rolls a little to one side so one wing tip is higher in the air than the other (when airplanes turn the yaw will also change), this rotational movement where the airplane rolls over is the roll axis.

Yaw is the rotation around the last axis, the same axis you would measure the height of airplane on. This rotation is barely felt in a airplane, but it is this orientation that will change the direction of the airplane, from lets say facing north to facing west by rotating 90° to the left (from the pilots perspective).

Getting the orientation of an object is then a matter of measuring the pitch, roll and yaw, there a different ways to do this and they will be discusses a little later, for now we will measure each rotation in degrees and define the orientation of 0^o pitch, 0^o roll and 0^o yaw to be an object that is level in both directions and facing north. A full rotation in one axis is 360^o and would leave the object in the same position as it started out. This definition is very similar to Euler angles witch will be discussed further.

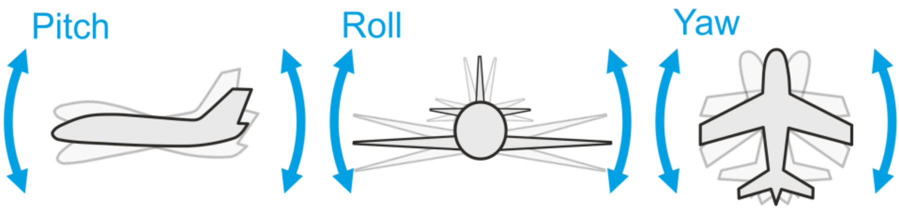

Figure 3.2: Pitch, roll and yaw of an airplane (edited version of figure found online[1])

3.4 Inertial Measurement Unit

In order to measure the orientation we will mount a device on the wrist which contains an Inertial Measurement Unit (IMU) from which the orientation can be read. Most IMUs are equipped with accelerometers, gyroscopes and magnetometers and offers either raw sensor values or the calculated Euler angles and quaternions for absolute orientation. In it self the raw sensor values can't provide the orientation, since it measures change for example the gyroscope alone will measure the rate of rotation, which in it self can't lead to an absolute orientation. But together the sensors provide enough data that this can be calculated, such an algorithm is called Sensor Fusion and is illustrated in Figure 3.3. Each IMU will have a different Sensor Fusion algorithm, and exactly was goes on is often a trade secret, but the pricnipal is as follows. The IMU will use its magnetometer to find north and use this to calibrate yaw, using its accelerometers the IMU can measure gravity and calibrate both pitch and roll. As a result of this yaw is measured from 0^o to 360^o independently (like a compass), pitch is measured from -180^o to 180^o degrees (0^o to 180^o is pointed towards the ground and -180^o to 0^o is pointed towards the sky) and roll is measured from -90^o to 90^o. Roll is dependent on pitch, which is why it only has a span of 180^o instead of the full 360^o. All together this lets the IMU calculate the absolute orientation. These calculations also requires the IMU to be calibrated and if not done correctly this can lead to issues, such as drift, where the orientation values will change even though the IMU sits completely still, this will be further investigated in Chapter 5.

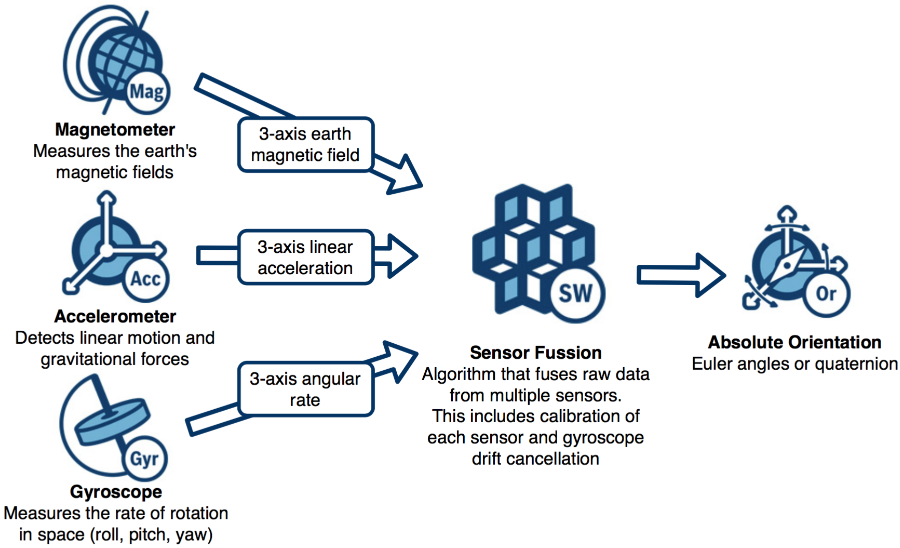

Figure 3.3: Model of Sensor Fusion algorithm (icons taken from Bosch[13])

The most common ways to represent absolute orientation is using either Euler angles[32] or quaternions[19]. Euler angles describes the orientation by angles of the 3 axis, this is the most intuitive approach but it has it downfall. When two of the gimbals in a gyroscope align into a parallel configuration, then one of the three degrees of freedom is lost, this is called Gimbal lock[36]. When Gimbal lock occurs changes in the locked axis cannot be measured, in order to fix it the axis which aligned itself with the other must be changed to get out of alignment and thus getting back the lost degree of freedom. This problem is often experienced in animation, where by defining movement only by total change in the three axis can cause strange and unintended results. The solution is to use quaternions, which is a mathematical number system which can represent the orientation of an object. Not getting into the math behind quaternions, but they solve the gimbal problem and quaternions can be transformed into Euler angles. Luckily when measuring the angles for the arm movements described gimbal lock does not occur, since the different axis does not come close to alignment in any position within the described movement.

Chapter 4

Design

This Chapter will describe the design for the proposed solution, in the end a list of functionalities and requirements will be created which will be the basic for the implementation in the next Chapter.

4.1 Overview

The solution that has been proposed is to use the gestures defined in Figure 4.2 and 4.3 instead of using a VAS. The orientation for each gesture should be translated to a value on a scale from zero to ten. To achieve this a wristband equipped with an IMU and a button will be used, the user will use one of the gestures and the press of the button to input the desired value. In order to view the data the wristband will be accompanied by a smartphone app. The companion app will display the logged data and offer the possibility to export it to a data file. This design is illustrated in Figure 4.1.

Since the main focus of this report is to compare the gesture based input method to a VAS, the approach to the design of the companion app will be to create a Minimum Viable Product (MVP)[30]. This means that it will focus a minimal set of features, just enough to be use able, like vise the user interface (UI) will be kept simple.

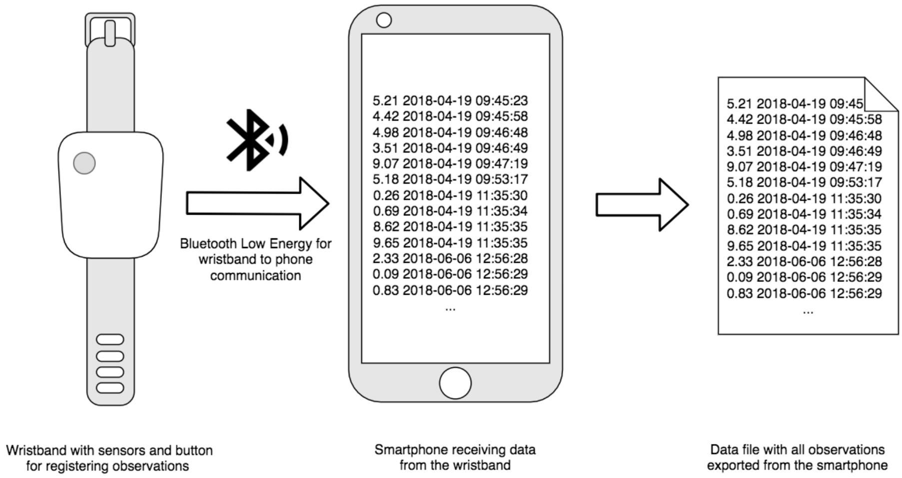

Figure 4.1: Wristband device with sensors and a button for registering observation. Smartphone to import, display and export the data

4.2 From Gesture to Value

In order to convert a gesture to a value on a scale we must first define said gestures and then create function to map their measurement to the preferred scale. We shall start by defining the gestures based on the orientation of the arm.

4.2.1 Orientation of the Arm

With the orientation defined as the rotation of the three axis, we need to translate this into movement of the arm. In Figure 4.2, 4.3 and 4.4 the movement to adjust pitch, roll and yaw is described.

Pitch is changed by rotating the arm around the elbow, resulting in the lower arm being raised/lowered. The angle between the lower arm and table (horizontal plane) is then the measurement for this rotation.

Roll is changed by rotating the wrist/lower arm, the angle between the back of the hand and table is then the measurement for this rotation.

Yaw is changed by bending the lower arm the elbow, keeping it level with the

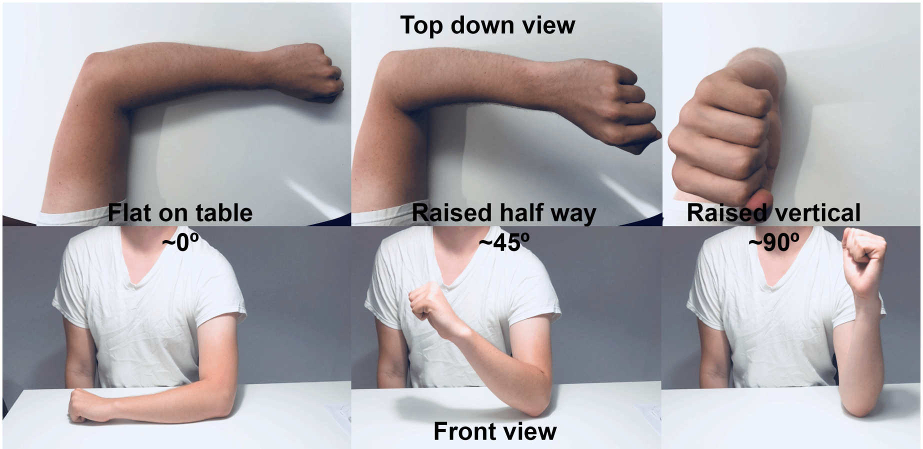

Figure 4.2: The movement for adjusting pitch by rotating the arm around the elbow

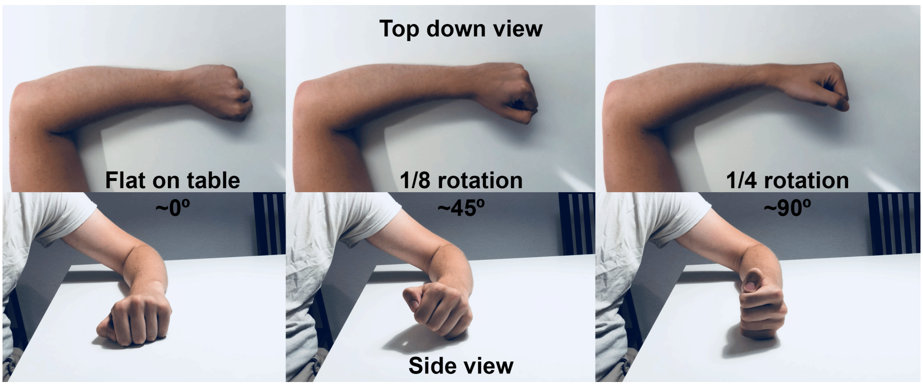

Figure 4.3: The movement for adjusting roll by rotating the wrist/lower arm

table, thus changing the direction of the lower arm. The angle of direction is then the measurement for this rotation.

Of course all three rotation can be achieved in other ways, for example yaw can be changed by turning the entire body to face another directions. But for the purpose of this thesis we look at these motions because they are done in front of one self, can be executed both while sitting at a table (for best reference) or while standing, and for all these motions the body itself also is a reference point, where turning your entire body would require another reference point to make sense of the scale.

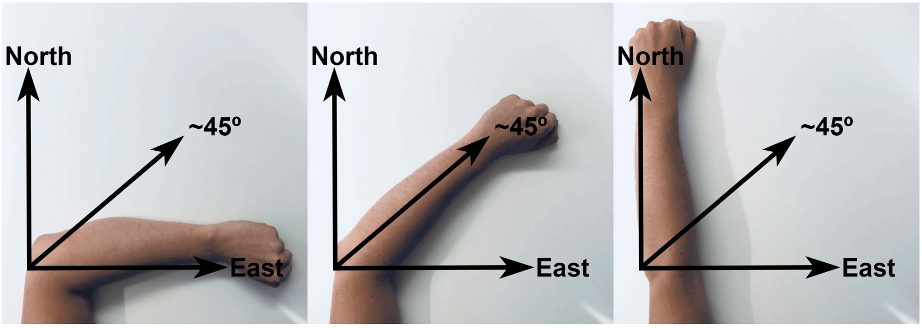

Figure 4.4: The movement for adjusting yaw by rotating the lower arm around the elbow (note the arm is flat/horizontal on the table)

4.2.2 Problems with Yaw

As described when measuring absolute orientation, then the magnetic north is used as the calibration point, meaning that 0° yaw will be pointing north, 90° pointing east and so on. Both pitch and roll is relative to the ground (horizontal plane), therefore no matter the yaw orientation you can get an absolute reading based on one point within the arm movement described earlier, e.g. you can rotate your wrist to a desired angle and take one measurement and know the orientation in that axis. This can't be done for yaw, it would require that you either first align yourself to a specific direction (north for example), and then bend the lower arm out in front to the desired angle, or it requires two measurement (one in the starting position, and one in the desired position) in order to calculate the difference, and achieve the desired angle. For this reason using the yaw orientation will not be investigated further, since it would introduce a higher burden for the user.

4.2.3 Mapping Euler Angles to a 0-10 Scale

We have now defined the two gestures as seen in Figure 4.3 and 4.2, we shall refer to them as wrist and arm gesture. We will assume that performing both gestures will have a linear relation between the movement and the Euler angle measured. As mentioned the roll angle will always be between -90° and 90° depending if the orientation clock or counter clock wise compared to 0° (horizontal). To map the values to a 0-10 scale the absolute value is divided by 9, e.g 90° and -90° will become 10, 45° and -45° will become 5, 25° and -25° will become 2.7778 etc. Mathematically this can be expressed with the function seen in Equation 4.1 where sv is the scaled value from 0-10 and ra is the Euler angle for the roll axis.

$$sv\,(ra) = \frac{abs\,(ra)}{9} \tag{4.1}$$

Equation 4.1: Mapping roll angle (ra) to scaled value (sv)

A consequence of this mapping is that if the wristband is turned pass the vertical position the scaled value will begin to decrease. This means that the user won't achieve a 10 rating simply by turning the wrist way beyond the vertical position. A different solution could be made to detect when the wristband pass the vertical position and set the scaled value at 10. It can be debated as to which solution is the greater, but by letting the scaled value decrease will force the user to get a better sense of the scale and won't allow "lazy" inputs.

Mapping the pitch angle needs to take into account that the roll value will be between -180° and 180°. The solution is to create a function that maps the value differently depending if the absolute pitch angle is greater than 90, basically first mapping the -180° to 180° range down to a -90° to 90° degree range, and then mapping it to the scaled value. The function in Equation 4.2 will express this where *sv* is the scaled value from 0-10 and *pa* is the Euler angle for the pitch axis.

$$sv\,(pa) = \begin{cases} \frac{90-(abs(pa)-90)}{9} & if\ abs(pa) > 90 \\ \frac{abs(pa)}{9} & otherwise \end{cases} \tag{4.2}$$

Equation 4.2: Mapping pitch angle (pa) to scaled value (sv)

4.3 Functionalities and Requirements

With the general outline for the wristband and companion app laid out, we can create a list of functionalities and requirements that shall be implemented in order to create a MVP.

- **Log data:** The user should be able to log observation using the wristband, this should be possible to do independently from the companion app
- **Import data:** The companion app shall import the logged data from the wristband

- **Display data:** The companion app shall display the logged data (scaled value and time stamp)
- **Export data:** The companion app shall be able to export the logged data to a data file for further investigation
- **Simple solution:** It should be simple and easy to perform all of the above actions.

Chapter 5

Implementation

This Chapter describes the implementation process of the design from the previous chapter. All non trivial issues will be described for both the wristband and companion app. The end of the Chapter will shorty describe the testing process that was done during the implementation.

5.1 Proof of Concept

Before the actual implementation of the design from the previous Chapter, a Proof of Concept (PoC) was created. This had several purposed, most importantly to see if it was at all possible to have at least some sense of the scale when performing the gestures. The component listed bellow was connected together as shown in Figure 5.1 and 5.2, the finished PoC can be seen in Figure 5.3.

- Adafruit Feather M0 Adalogger[4] with a 350mAh 3.7V battery
- Adafruit 9-DOF Absolute Orientation IMU Fusion Breakout - BNO055[3]
- Flora Wearable Bluefruit LE Module[5]
- A push button

- A 100kΩ resistor

The Feather M0 Adalogger was controlled by an Arduino script, that after establishing a bluetooth connection waits for the button to be pressed. When the button is pressed the Euler angles from the IMU is read, and transmitted over the bluetooth connection, the LED is connected to the switch so it will light up as long as the button is pressed. A simple GUI (as seen in Figure 5.6) was created using a Python script, that connected to the bluetooth module and when receiving data displayed the raw values and the mapped scaled value between zero and ten using the mapping function seen in Equation 4.1 (the scaled value was rounded for simplicity). Only the wrist gesture was tested using the PoC, if the PoC proved the wrist gesture to be possible the arm gesture would too, since the wrist gesture is much finer than the arm gesture. Both the Arduino and Python scrips are linked to in Appendix C.1.

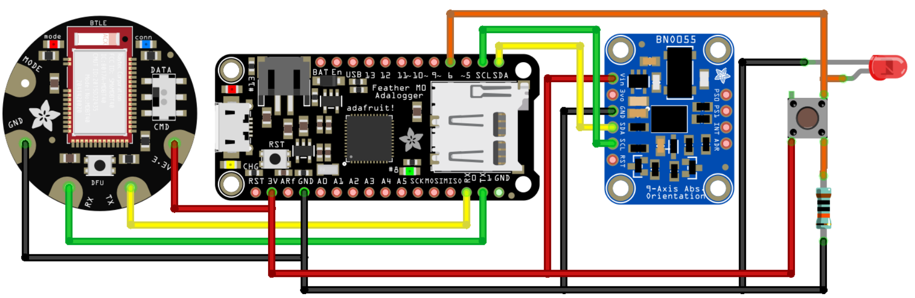

Figure 5.1: Wiring for of the proof of concept. Battery not shown

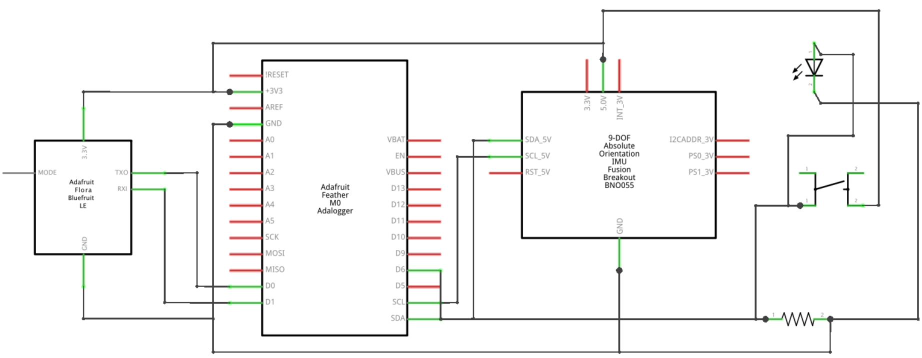

Figure 5.2: Schematic for the proof of concept. Battery not shown

The first thing that was tested using the PoC was sensor drift, since it was important to prevent this. A test was preformed logging the Euler angles with

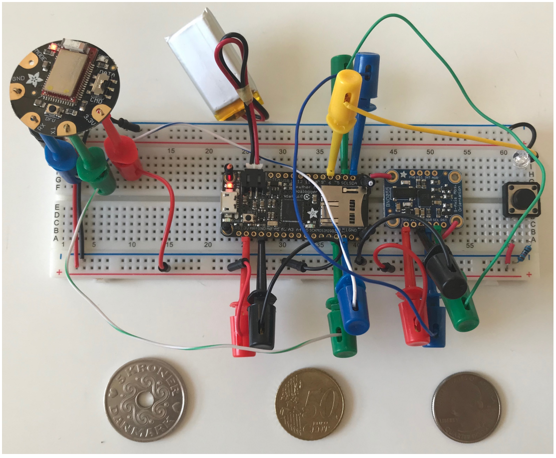

Figure 5.3: Proof of concept. Components from left to right on the breadboard: Bluefruit LE Module, Feather M0 Adalogger and battery, IMU Fusion Breakout - BNO055, push button and led. Coins below for scale (5dkk, 0.5€ & 0.25$)

and without calibration while the IMU sat completely still. Once every minute the Euler angles was read from the IMU and logged to a file. Two trials where run, one for nine hours and one for twelve hours without and with calibration respectively. The results can be seen in Figure 5.4 and 5.5, we see that without calibration the sensor drifts and the Euler angles change quite dramatically over time, while with calibration the Euler angles only had minor change. This concluded that with proper calibration sensor drift would not be an issue.

The PoC was then tested by strapping it to the back of the lower arm using rubber bands, then moving the arm to different positions using the wrist gesture and pressing the button. It was validated that the concept worked, and it was possible to input different values using the gesture. The question still remained how precise it was. At this point the author thought that this concept would only be precise enough to be used on a three point scale and would not be able to compete with a VAS.

5.2 Prototype

With the PoC being a success the design from Chapter 4 could be implemented. The components used for the PoC could in theory be rearranged and soldered

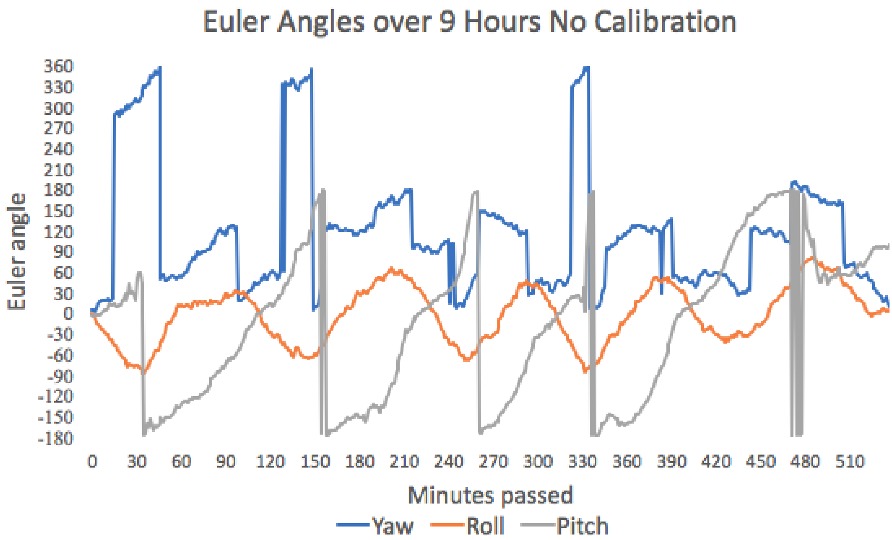

Figure 5.4: Reading the Euler angles from the BNO055 IMU[3] without calibration

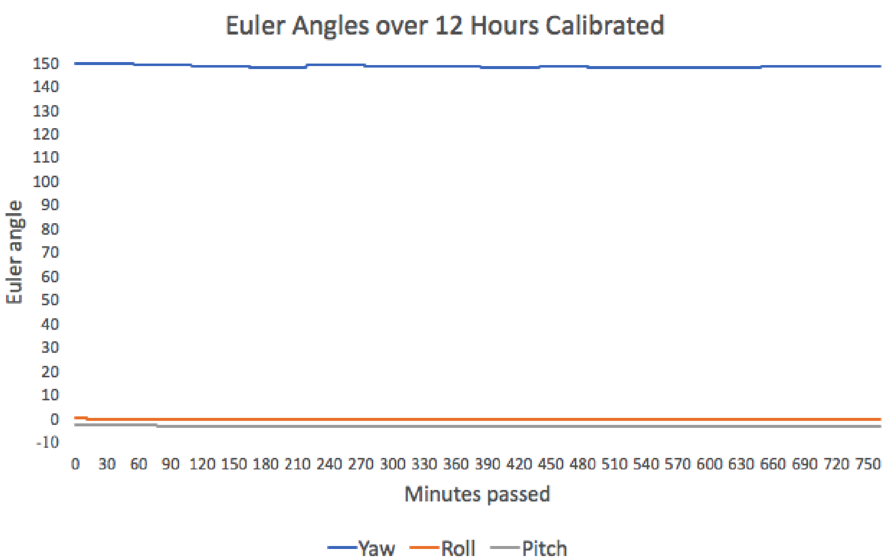

Figure 5.5: Reading the Euler angles from the BNO055 IMU[3] with calibration

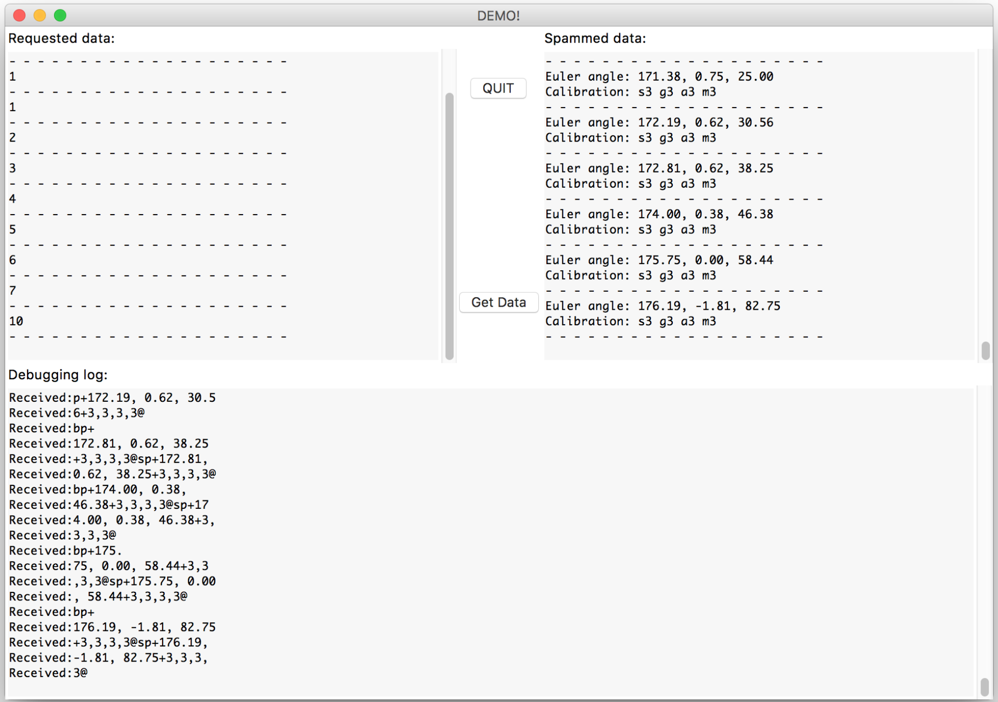

Figure 5.6: PoC GUI. Top left shows the data mapped to the 0-10 scale. Top right shows the raw data and calibration. Bottom shows a debug log

together into a more compact design, but in practice it would be quite bulky for a wristband device. Instead a small device called "MetaMotionR" created by mbientlab[26] was used. The device is roughly the same size as a regular wrist watch, although thicker, and is packed with sensors, most importantly an IMU with the same Sensor Fusion algorithm as the BNO055. The MetaMotionR is also equipped with a push button and has Low Energy Bluetooth built in, making it an excellent hardware choice for this implementation.

The MetaMotionR must be programmed using mbientlab's API, available for several platforms[25]. Since the design all ready requires a companion app for the wristband the API for Android will be used. The MetaMotionR has 8MB of flash memory which can be used to store commands and log data, it allows for more than 200.000 log entries when logging Euler angles, which should be

Figure 5.7: MetaMotionR by mbientlab[26]

more than enough for this use case. The MetaMotionR is programmed to run the Sensor Fusion algorithm and whenever the button us pressed it will log the current orientation, turn its LED on and vibrate to give user feedback. It was considered to use the vibration motor to give varied feedback dependent on the value logged, but the vibration motor doesn't offer much customization, only the time interval can be customized, and if the vibration is any shorter than half a second it might not vibrate, therefore this idea was discarded for this implementation.

When the companion app connects to the MetaMotionR the logged data can be imported over the app and is afterwards deleted from the MetaMotionR to free up space for future data entries. Beside storing the Euler angles each data entry has a time stamp. The time on the MetaMotionR is synchronized whenever it connects to a smartphone. The Euler angles are measured with three decimals precision, which is much finer than the precision that can be achieved by using the gestures. One flaw with the MetaMotionR is that logged entries and the commands are stored in the flash memory, which requires power, this means that if the MetaMotionR runs of power all logged entries will be lost, and it needs to connect to the companion app before it can be used to log entries again. The battery life of the MetaMotionR will be explored later in this Chapter.

5.3 Companion App

Following the requirements from Section 4.3 the companion was implemented. The app was build upon mbientlab's Android tutorial "starter app"[24]. The starter app handles the bluetooth connectivity (searching for devices and connected), beside this the rest was left to be implemented. As explained the app programs the MetaMotionR, this happens every time the device connects to the app. The MetaMotionR is programmed to turn on its LED, log the Euler angles (saved to on board memory), vibrate the motor and turn the LED off again every time the button is pressed. Once programmed the MetaMotionR can run independently.

The app itself consist of three buttons, one to import, share and delete data. The largest part of the screen is reserved for displaying the data in a scroolable list, with the latest entry on top. When the data is imported from the device the Euler angles get mapped to the scaled value, this value together with the Euler angles, time stamps and battery percentage of the MetaMotionR are saved to internal file (in CSV format). The scaled values are displayed in the app together with its time stamp and battery percentage. The scaled value is color coded from green (low values) to yellow (middle values) to red (high values) as seen in Figure 5.8. The color coding was implemented using HTML formatting and simply looking at a color table to map the scaled value to colors based so 0.0-0.99 gets the greenest color, 1.0-1.99 gets the next shade of green an so on (if no scaled value was recorded the color is assigned to blue).

After data has been imported it is deleted from the MetaMotionR. The “Delete Data” button lets the user delete the data that has been imported to the app, when pressed a confirmation dialogue appears to prevent accidental deletion. If confirmed the app will delete the internal file containing the data. The “Share Data” button lets the user export the data via. email. When pressed the app creates a copy of the internal file to external storage, an intent is then made to create an email with the file attached. The email can then be send and the attached data can be examined in a spreadsheet or other tool. Lastly the app has a toggle switch labeled “Time Only”. The default configuration is to have this toggle off, in this state the MetaMotionR behaves as described. If the toggle is on the MetaMotionR is programmed into a “Battery Saver” mode, where the Sensor Fusion will be turned off, and therefore Euler angles and scaled values wont be logged. Data entries made in this state will be listed with “–” instead of the scaled value. This mode was implemented after it was discovered that the battery life of the MetaMotionR didn’t live up to expectations, more on this later. The finished app can be seen in Figure 5.8. The code used to display, share and delete the data in the companion app was based on the code from the authors previous project “Mood Tracker”[10], but has been heavily modified.

5.4 Performance Testing

The performance of the MetaMotionR and the companion app was tested throughout the implementation process, this was done to ensure that it worked as intended. The first problem that was discovered was related to Sensor Fusion module of the MetaMotionR. Using it the first approach was to only start the algorithm when the button was pressed, measure the Euler angles and then

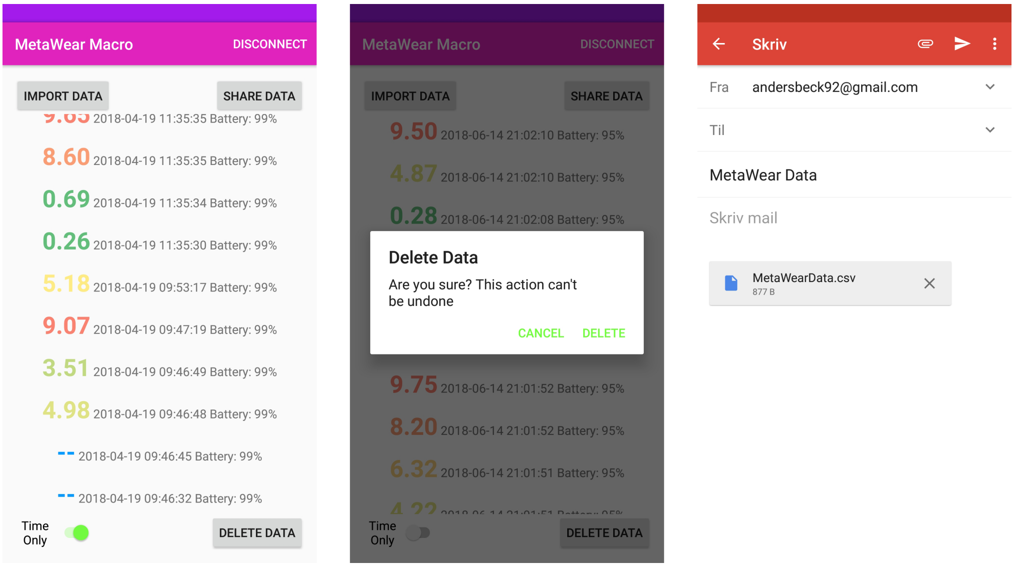

Figure 5.8: Companion app. Main screen on the left, Delete Data in the middle & Share Data on the right

turn the algorithm off waiting for the next button press. The problem turned out to be that the Sensor Fusion algorithm needed a short time to "catch up" before it could be used, which resulted in the readings not being accurate, but a little delayed. Solving this problem required the Sensor Fusion algorithm to be turned on at all times, the downside of this is a higher power consumption.

It was then discovered that the battery life of the MetaMotionR was quite limited, depended on the workload. A artificial test was setup to simulate a worst case scenario, the MetaMotionR was programmed to read and transmit the Euler angles to the companion app once every second. The battery percentage was noted at about every 15-30 minutes until the MetaMotionR ran out of battery. The test was run twice, and the results can be seen in Figure 5.9, the first test ran for a little less than five hours, while the second time only ran for a little over two hours. These results disappointing, but the test were extreme and real usage would not be that demanding. In order to get a feel of what could be expected by the MetaMotionR in real use, the author choose to wear the device every day for one week, logging entries at least 20 times a day at random times. What was found that with this demand the device could run for at least one day without a charge. It was discovered that the battery percentage would stay in the ~90-80% range through most of the day and night, and once the battery percentage came below ~70-60% it would dramatically drop to zero. Because of this, it was decided to implement the option to disable the logging of orientation, so if this device and app were to be used in real life, one could still get the time stamp recorded which is still valuable information to collect as shown by Larsen et al.[21]. When used in this mode the battery life is much longer, how long is unknown but it was able to hold a comfortable battery percentage for over a week. One benefit to the battery is that it will fully charge in less than 2 minutes.

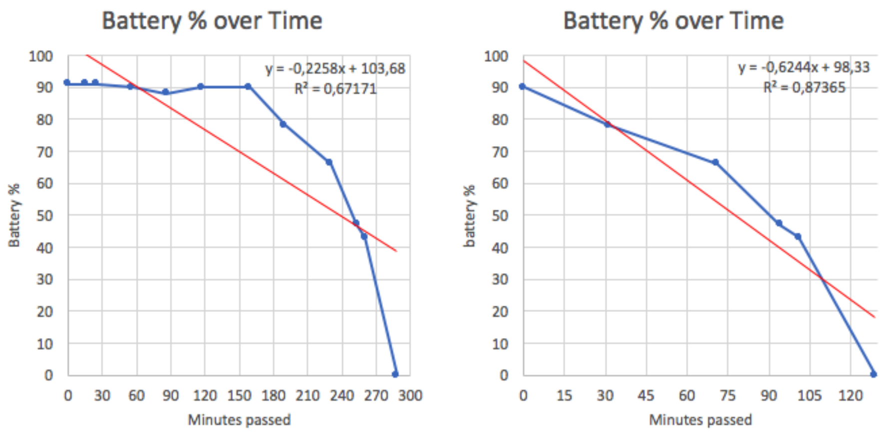

Figure 5.9: Battery Tests

Chapter 6

Experiment

This chapter will describe the method of the experiment conducted in order to compare the gestures using the wristband to a VAS. For comparison a test was designed where the participants would be shown a stimuli and asked to input it using the gestures or a digital VAS. Two kinds of stimuli was used, an integer between zero and one hundred, or a box with a shade of grey between white and black based on the experiment by Matejka et al.[23]. Before describing the experiment further lets declare some terms for better clarification. As stated, there are three input methods:

1. The digital VAS. This is using a slider on the tablet. This input method will be referred to as “slider”
2. Using the wristband and the gesture of turning the wrist as described in Figure 4.3. This input method will be referred to as “wrist gesture”
3. Using the wristband and the gesture of raising the lower arm as described in Figure 4.2. This input method will be referred to as “arm gesture”

The two types of stimuli will be referred to as “number scale” and “grey scale” for the integer and shade of grey respectfully. With these three input methods and two stimuli a total of six different exercises must be conducted. The naming for each of these exercises is a combination of the input method and stimuli, e.g. turning your wrist to input a shade of grey will be referred to as “wrist gesture on grey scale”. All of the exercise’s names can be seen in the top of Table 6.2. The rest of this section will cover every aspect of this in more detail.

In Figure 6.1 we see a participant during the training before the experiment, is shows how the experiment was setup and provides a good basic understanding of how the experiment was executed.

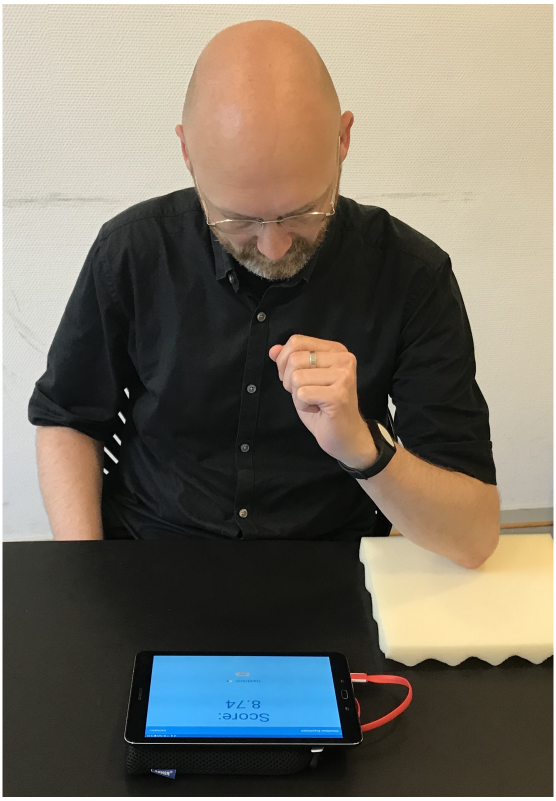

Figure 6.1: Participant during training with the wristband before starting the experiment

6.1 Participants

Twenty four (24) participants where chosen for this experiment. Since the experiment consist of six exercises a multiple of six was needed to ensure a minimal bias as explained later. Then based on the time frame for the experiment twenty four seemed to be a reasonable amount while being large enough to collect enough data. The participants volunteered their time and where not paid or in other ways obligated to participate. All of the participants were either from The Department of Applied Mathematics and Computer Science at DTU or the authors colleagues at IBM, thus most of the participants had above average technical experience, but based on the survey conducted after the experiment (Appendix A) 42% had not any prior experience with wearable devices.

Out of the twenty four participants 18 where men and 6 where female. The mean age was 27.83 and the median age was 26.5, both age and gender distribution can be seen in Figure 6.2.

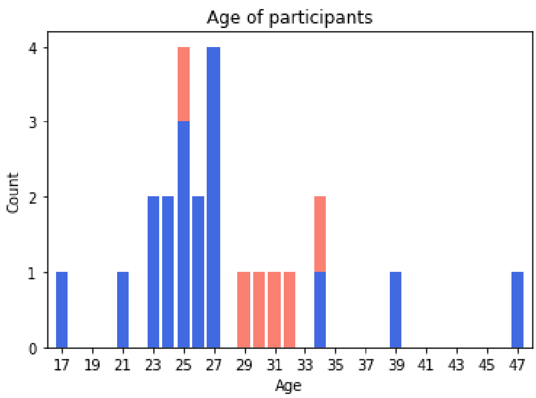

Figure 6.2: Age and gender distribution of the 24 participants: 18 males in blue and 6 females in red

6.2 Apparatus

6.2.1 Essential Hardware

For the experiment three pieces of hardware was used:

1. MetaMotionR by mbientlab[26] as the wristband device for registering input from the participant using the gestures
2. Samsung Galaxy Tab S2[33], this tablet run the app developed for the experimenting, keeping track of exercises, showing stimuli, collecting input

from the participant, saving the data etc. Much more detail about this is this section

3. A computer used to fill in the survey after the experiment. In the experiment a MacBook Pro 13 inch model was used, although the computer used isn't important

6.2.2 Tablet App

The app created for the experiment is built upon the same code as the companion app described in Chapter 5, but changed to fit the experiment. Has three main functions: Settings, Train and Experiment as seen in Figure 6.3. The settings is for configuring the experiment and exporting or deleting the collected data as seen in Figure 6.4. Two variables can be changed that has influence on the experiment, the first is the "Number of stimuli", this value will change the number of stimuli the participant has to rate for each exercise, the number of stimuli can be changed in steps of five. The "Next Subject ID" will change the ID used to identify each subject in the data. The order of exercises for the participant is bases on this value. Both of these settings should not be changed though out the experiment, so that the participant's IDs will be in sequential order and that they have the same amount of stimuli for each exercise. For this experiment the number of stimuli was set to 20 based on pilots tests. Beside these two settings there is the option to "Toggle back button", when this option is enabled a "Back" button will appear on any screen in the app, and when pressed will go back to the main screen. This option is purely for debugging allowing one two go back to the main screen at any point, during the experiment this options should be disabled.

The last actions on this screen is to export and delete data. There are two options to export the date: "Move data" and "Email data", the first will simply copy the collected data from the internal storage of the app to the external storage on the tablet so it can be accessed by any file manager. The later will also copy the date to external storage, but it will create an intent to email the data which will start the default email app and attach the data. "Delete data" will delete the collected data from internal storage, a confirmation prompt will appear first to prevent accidental deletion. Whenever the main or setting screen is accessed, the current battery percentage from the MetaMotionR is displayed on the screen, keeping an eye on this prevent the wristband to run out of power in the middle of the experiment.

The train screen is made to let the participant practice using the wristband device and get a feeling of the scale. When this screen is activated the wrist-

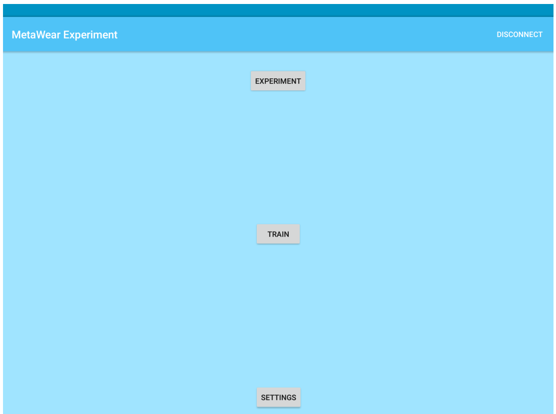

Figure 6.3: Experiment app main screen

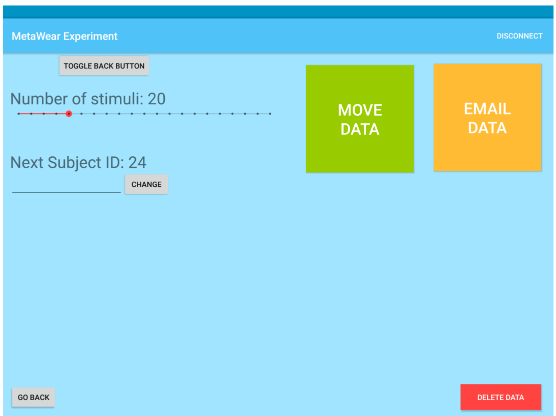

Figure 6.4: Settings screen

band will be programmed similar as explained in Chapter 5, when the button is pressed it will read the orientation from its IMU and transmit it to the tablet. The light will flash and vibration motor will do a short pulse to provide feedback. The difference between the previous described implementation and the programming for the experiment is that instead of logging the data for later export, it will transmit it instantly. When the orientation data is received by the app it will map the Euler angles to a 0-10 scale and display the result. The mapping is dependant on the toggle switch, if the switch is in the "Hand" position as in Figure 6.5 the mapping will depend on the roll angle and if the switch is in the "Arm" position as in Figure 6.6 the mapping will depend on the pitch angle. This of course correspond to the gestures described in Figure 4.3

and 4.2. Mapping the Euler angles to the scaled value is done in the same way as described in Chapter 4 using the simple mathematical functions as seen in Equation 4.1 and 4.2. In order to map the two functions from a 0-10 scale to the grey scale (0.49) and number scale (0-100), the result of the the functions are multiplied by 4.9 and 10 for each scale receptively.

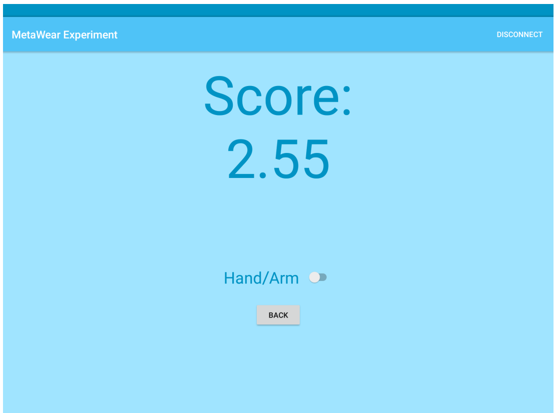

Figure 6.5: Train screen Hand (wrist gesture) mode

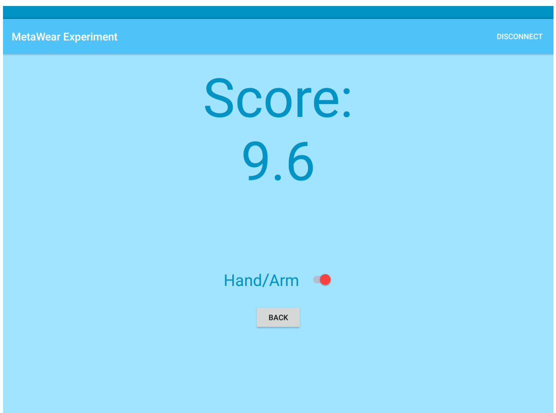

Figure 6.6: Train screen Arm (arm gesture) mode

The last screen is the experiment itself, this is the most complicated screen and consist of many "sub screens". Before the experiment begin a screen for inputting the gender and age of the participant is presented as seen in Figure 6.7. In the top right the assigned ID for the participant is showed, this is just for confirmation and to be used in the survey after the experiment (so the survey and experiment data can be synchronized if needed). Once the "Start" button is pressed the experiment will begin. As mentioned the experiment consist of six exercises and the order of these is dependent on Table 6.2. Before each exercise an explanation screen will show, this screen will have some text to explain the task of the exercise and a short GIF that illustrates it. The text will stay on top during the whole exercise. This screen purpose is to make sure the participant is sure of what task they are required to do, the participant will have gotten instructions beforehand, so this screen is just reassure them. Three of the six explanation screens can be seen in Figure 6.8, 6.9 and 6.10.

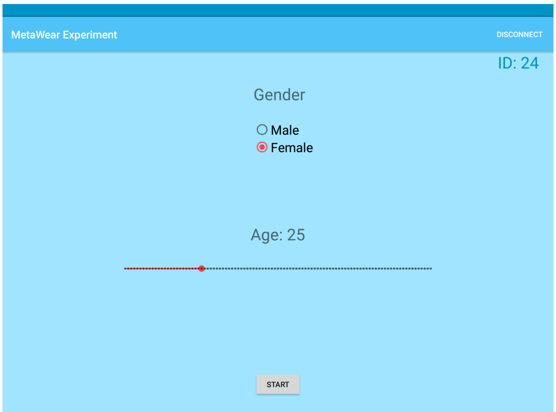

Figure 6.7: Experiment start screen

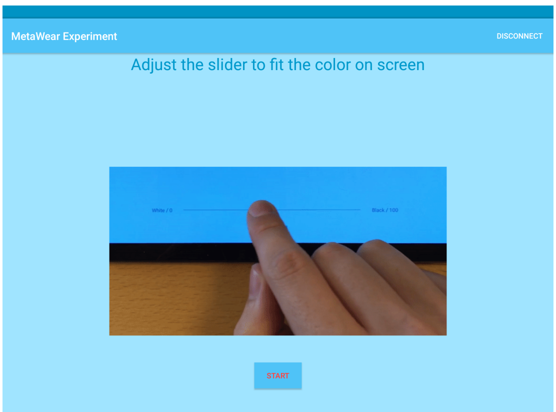

Figure 6.8: Slider on grey scale explanation screen

After the explanation screen each exercise follow the same pattern: show stimuli, wait for input, refresh input (optional), register input, next input. This pattern is repeated until the set number of stimuli has been shown and then the next exercise is started, after all six exercises are completed a “finish” screen is shown. The slider on grey scale exercise will be used to explain this pattern in more

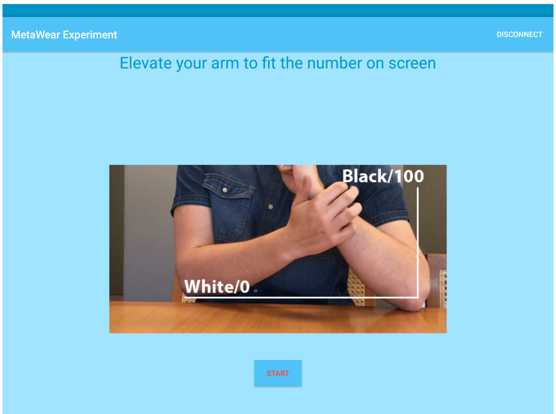

Figure 6.9: Arm gesture on number scale explanation screen

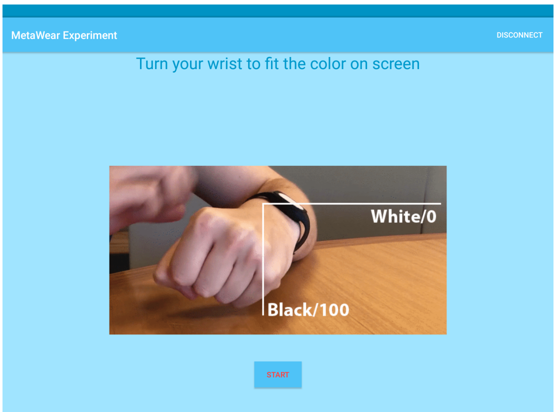

Figure 6.10: Wrist gesture on grey scale explanation screen

detail. After the exercise explanation the first stimuli is shown as seen in Figure 6.11, then the participant registers a input using the slider as seen in Figure 6.12, a small vertical bar indicate the input. The input can be changed simply by adjusting the slider (for arm and wrist gesture exercises the participant can just press the button on the wristband again), to register the input the participant has to press the “Submit” button as seen in Figure 6.12. Once the input has been registered the stimuli will disappear and the participant has to press the “Next” button as seen in Figure 6.13, this button will alternate to appear on the left and right side of the screen (this will be explained later), once the button has been pressed the next stimuli will appear (as seen in Figure 6.14)and the process will repeat until the participant has been through the set amount of stimuli and moves on to the next exercise. In Figure 6.15 we see a screen from the wrist gesture on number scale exercise where the participant has pressed the button on the wristband, beside the feedback from the wristband (light and vibration) the screen also show that the button has been pressed, but no feedback is given on what value was registered nor how well they performed. This was done on purpose to reflect the real life use case where the user wouldn't know the registered value before importing it to the companion app. In Figure 6.16 we see the "Finish" screen after the participant has completed the experiment. All app screens shown here plus a couple extra can be seen in larger scale in Appendix E.

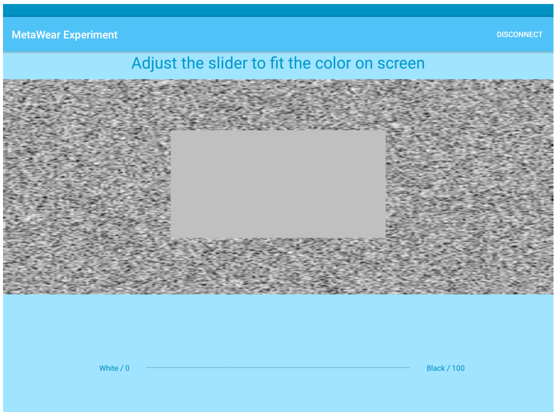

Figure 6.11: Show stimuli screen with shade of grey

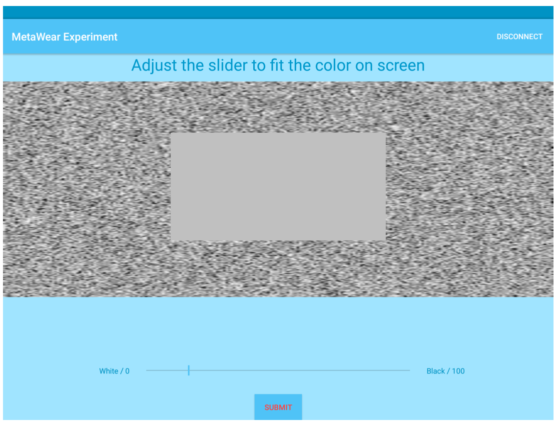

Figure 6.12: Participant has given an input using the slider

The stimuli for each exercise follow a specific set of rules. First of all the stimuli for each exercise is independent. For number scale exercises the stimuli

Figure 6.13: Next stimuli screen

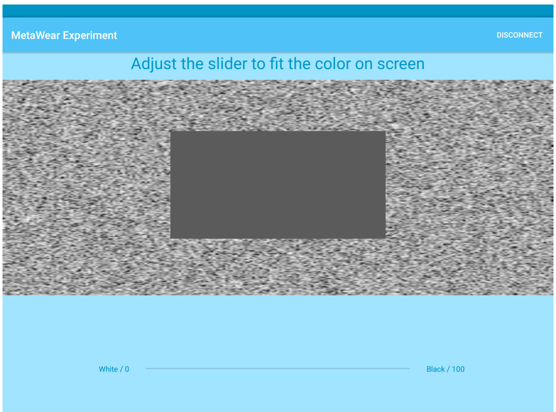

Figure 6.14: Pattern is repeated

is generated by creating an array of integers from 0-100, e.g. [0, 1, 2 .. 98, 99, 100]. This array is shuffled so the order is random. When a stimuli is needed the first element in the array is used, once used it is removed from the array. Once the array is empty the process is repeated. For grey scale exercises the process is the same, but the array consist of integers from 0-49, here 0 correspond to pure white, 49 is pure black and the rest are the shaded of grey in between. The shaded of grey are the same as used by Matejka et al.[23] as seen in Figure 2.4, where each shade is of equal distance from each other in the CIELAB color space[14]. What this means is that the order of stimuli is completely random, but the distribution of stimuli is completely even, this means that for each participant the stimuli is random but not repeated, while keeping the overall

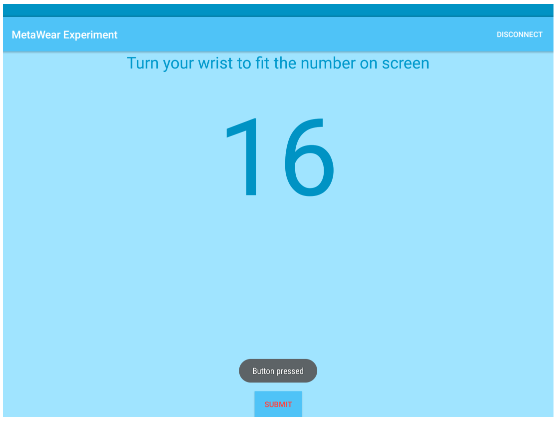

Figure 6.15: Input registered from wristband with integer stimuli

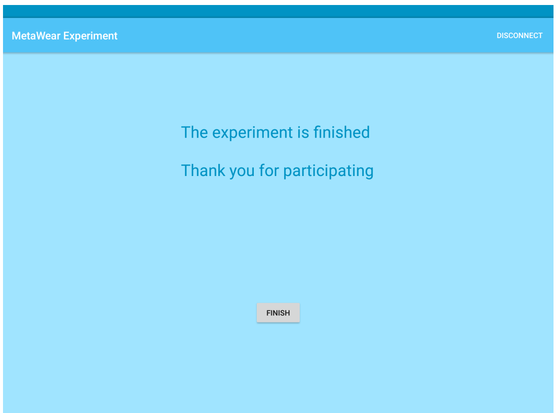

Figure 6.16: Finished experiment screen

even distribution between participants. Again emphasizing on the fact that the stimuli for each exercise is independent.

The data collected from the experiment is described in Table 6.1. Data is recorded whenever the participant pressed the submit button for each trial. With the six exercises each having 20 trials, this results in 120 data points for each participant and a total of 2880 data points for the complete experiment. Most of the data is self explanatory and doesn't need further explanation than the table. The "type" refers to the exercise, the keyword for each exercise should be self explanatory. "type_order" refers to the order of the exercises, this data is redundant since it can be deducted based on the "id" and Table 6.2.

"stim_order" is the order the stimuli had and is also redundant since it can be deducted based on the time stamp, but if any of these were to be investigated this would greatly reduce the work. The "pitch", "roll" and "yaw" are the Euler angles recorded, these could be used to change the mapping functions from Euler angles to scaled value without needing to redo experiment (these values are only recorded for wrist and arm gesture exercises). The "reaction_time" is recorded from when the stimuli was shown on screen until the participant pressed the submit button. The data collected is saved to a CSV file in the apps internal storage, and can be exported from the settings screen as described earlier. Beside this data, a separate data file was created where to total time taken to complete the experiment (from pressing start in Figugre 6.7 to the finsih screen was shown).

column name	type	descriptiom
id	int	Uniqe id for each particpant
gender	String	Participant gender: "Male" or "Female"**
age	int	Participant age**
time	Date object	Timestamp when input was registered
type	String	Type of exercise: "slider_grey", "slider_num", "wrist_grey", "wrist_num", "arm_grey" or "arm_num"
type_order	int	Order for the type of exercise: 0-5*
stim_order	int	Order for the stimuli: 0-49 or 0-100*
stim	int	Stimuli: 0-49 or 0-100
response	double	Response: 0.0-49.0 or 0.0-100.0
pitch	double or String	Value of pitch for exercises using the wristband, else "NA"**
roll	double or String	Value of roll for exercises using the wristband, else "NA"**
yaw	double or String	Value of yaw for exercises using the wristband, else "NA"**
reaction_time	long	Time in milliseconds from the stimuli was shown until a response was registred

*Redundant data. **Data not investigated

Table 6.1: Structure of Collected Data

6.2.3 Other

Beside the described hardware and software a small foam pad was used for the participant to place their elbow on, this reduced fatigue. A power bank was connected to the tablet to keep the it from running out of power and to keep the angle between the display and table at 15°, this reducing the glance in the display from the ceiling lights and gave a better viewing angle.

6.3 Procedure

Each participant completed the experiment in one continuous sitting, the whole process took approximately 15 minutes from the entered the room until they left. First participant was seated in front of the tablet and assisted to wear the wristband correctly. They were told to wear it on the arm the felt most comfortable with, but all participants choose their left arm. Then the participants was informed of how the device worked, they where shown the same gestures described in Figure 4.2 and 4.3 (the gestured was performed on them, they weren't shown the Figures). It was made clear that the horizontal position would register as a minimum value, that the horizontal position would register a maximal value. They were instructed about how the scaled value would decrease if they performed the gesture beyond the vertical point.

They were then taken to the "Train" screen of the app, where they were given 1-2 minutes with each gesture to get a sense of the scale. They were asked to produce an input as close to zero and ten as possible. Again they were told about the what would happen if the went beyond the gesture, and they could see the result. They where then instructed to after each input with the wristband to lay their hand flat on the table. This was to have the same baseline for each input.

After training they were instructed on the structure of the experiment. They were told about the three input methods and two kinds of stimuli, and that there would be six exercises in total. They were told that each exercise would start with an explanation screen, then show a stimuli that they had to rate according to the instructions. They were told that each exercise had several stimuli, but they weren't told the exact amount, only that they had to continue the exercises until the saw the finish screen. They where told that if they where unhappy with an input, they could just press the button on the wristband again or adjust the slider before they pressed submit, but after pressing submit the input was final. Once they confirmed that they understood the instructions (any questions about the procedure was answered) they where taken to the experiment screen, where they registered their age and gender and then started the experiment. Most participants had no problems following the on screen instructions, but some participants asked questions during the experiment to clarify the task they had to perform.

After completing the experiment they were handed the survey (Appendix A) on a computer and asked to answer it.

Exercise	slider on grey scale	slider on number scale	wrist gesture on grey scale	wrist gesture on number scale	arm gesture on grey scale	arm gesture on number scale
ID	**0**	**1**	**2**	**3**	**4**	**5**

	Exercise Order					
Participant id	First	Second	Third	Fourth	Fifth	Sixth
0, 6, 12, 18	**0**	**1**	**5**	**2**	**4**	**3**
1, 7, 13, 17	**1**	**2**	**0**	**3**	**5**	**4**
2, 8, 14, 20	**2**	**3**	**1**	**4**	**0**	**5**
3, 9, 15, 21	**3**	**4**	**2**	**5**	**1**	**0**
4, 10, 16, 22	**4**	**5**	**3**	**0**	**2**	**1**
5, 11, 17, 23	**5**	**0**	**4**	**1**	**3**	**2**

Table 6.2: Exercise ordering based on a Balanced Latin Square

6.4 Design

This experiment was a 3 $\times$ 2 within-subject design. There was three input methods: the slider, wrist gesture and arm gesture. Te input types was shades of grey and numbers. The independent variables has the stimuli value associated to it, likewise the dependent variables has the response (input value) associated to it. The goal of the experiment is to compare the accuracy of the dependent variables to each other with the independent variables being the ground truth.

The within-subject design was chosen in order to achieve more data with less participants, this also balance any variance due to participants' predispositions and it will be approximately the same across test conditions. Unfortunately this can introduce bias because of the practice effect. To prevent this as much as possible counterbalancing has been introduced using the balanced Latin square seen in Table 6.2. This requires a multiple of six participants and it was decided to have 24 participants providing a good amount of data while still being achievable within the given time frame.

Each of the six exercises consisted of 20 trials, with 24 participants that totaled to 2880 trials/data points in total.

6.5 Pilot Tests

Small pilot test was ran throughout the development of the experiment and app. This ensured that the experiment stayed as streamlined as possible. When the implementation of the experiment app was completed some pilots test was ran in order to determine how long it took to complete the exercises. Figure 6.17 shows the results of seven pilots with varying numbers of stimuli for each exercise. Not surprisingly there was a linear correlation between number of stimuli and completion time. It was a goal to only have the experiment take about 15 minutes for each participant, with about 10 minutes for the experiment and 5 minutes for introduction, training, survey etc. Therefore based on the results 20 stimuli per exercise was chosen, since it took about 10 minutes to complete.

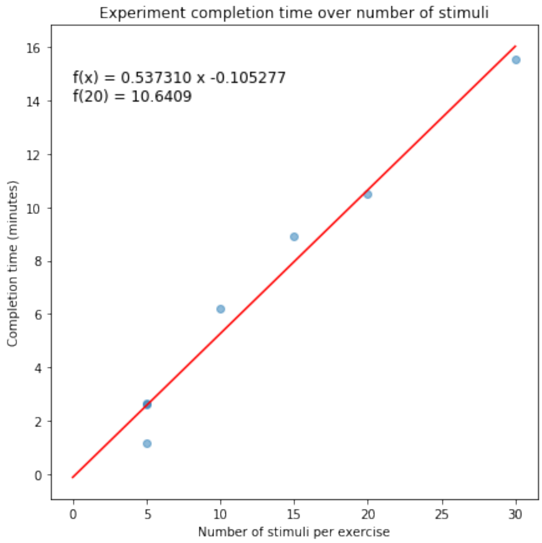

Figure 6.17: Completion time of pilot tests, with the regression line in red

After the initial seven pilots tests, another four pilot tests was conducted, this time with the form of the experiment in mind, looking to see if anything needed changing e.g. the instructions given, the survey etc. Based on these four tests the only thing that was changed was to include age and gender in the survey (even though it could already be conducted by cross referencing participant ID in the survey with the data collected). Since this was the only change to the execution of the experiment it was decided to include these four pilot test in the results of the experiment. No factors surrounding the experiment itself had been changed, only the survey.

CHAPTER 7

Results and Discussion

The data collected was described in the previous Chapter, in Figure 7.1 we see a small snippet of the actual data collected. The whole data set and code for analysis can be found in Appendix D and C.4.

In the work of Matejka et al.[23] they define any response with an error greater than half the scale to an outlier, debating that an response this far from the stimuli must be seen as a human error. Using the same argument we will remove any data point where the difference between the stimuli and response was greater than half the scale. With the grey scale exercises 50 shades of greys where shown, with a corresponding value from 0 (pure white) to 49 (pure black), therefore we will remove any data point with a difference between stimuli and response that is more than 25. Likewise with the number scale exercises where the scale is from 0 to 100 any data point with a difference between the stimuli and response greater than 50 is removed. This results in the removal of four data points, which can be seen in Figure 7.2.

It seems reasonable to remove these four data points, but another approach is visual inspection of the data. This can be done by plotting each participants response over stimuli for each exercise, in Figure 7.3 we the result of this for the slider on grey scale exercise, the plot has been divided into four plots with six participants on each. It then becomes easy to see by looking at each line if a point should be considered an outlier. If a line is fairly straight, but a single

	id	gender	age	time	type	type_order	stim_order	stim	response	pitch	roll	yaw	reaction_time
38	0	Male	24	Sun May 20 18:44:35 GMT+02:00 2018	slider_num	1	18	50	51.17	NaN	NaN	NaN	5742
39	0	Male	24	Sun May 20 18:44:44 GMT+02:00 2018	slider_num	1	19	65	61.88	NaN	NaN	NaN	8075
40	0	Male	24	Sun May 20 18:45:07 GMT+02:00 2018	arm_num	2	0	63	46.30	41.633152	20.192804	304.02070	17045
41	0	Male	24	Sun May 20 18:45:14 GMT+02:00 2018	arm_num	2	1	12	23.30	20.955986	21.641567	290.03125	4639
42	0	Male	24	Sun May 20 18:45:22 GMT+02:00 2018	arm_num	2	2	79	48.70	43.794170	18.117966	299.70346	6171

Figure 7.1: Snippet of the collected data

	id	gender	age	time	type	type_order	stim_order	stim	response	pitch	roll	yaw	reaction_time
77	0	Male	24	Sun May 20 18:49:02 GMT+02:00 2018	wrist_grey	3	17	3	28.1750	-1.684786	51.767326	285.30933	3785
478	3	Male	34	Wed May 23 15:07:19 GMT+02:00 2018	slider_grey	5	28	10	38.9109	NaN	NaN	NaN	2004
1080	9	Male	25	Fri May 25 11:09:31 GMT+02:00 2018	wrist_num	0	79	6	79.0000	-17.046492	71.109856	74.16516	12244
1620	13	Female	34	Mon May 28 09:36:11 GMT+02:00 2018	wrist_num	3	58	77	18.2000	143.890780	16.414879	238.16380	4039

Figure 7.2: Four outliers that was removed from the data set

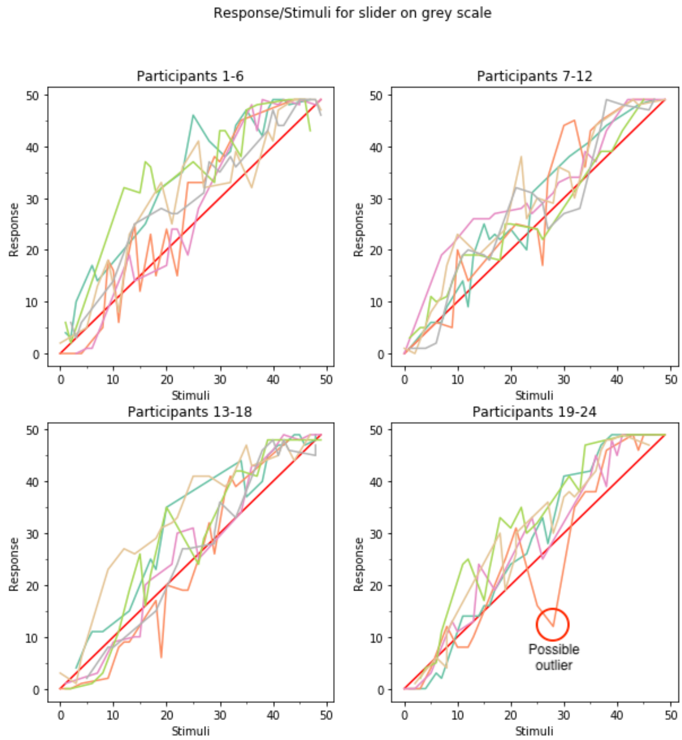

Figure 7.3: Visualization of outliers for the slider on grey scale exercise. Each line correspond to the response from a single participant. The red diagonal would be a perfect response

point is way off it can be considered an outlier, but if a line is fluctuating a lot, then it is more likely that the participant just wasn't very consistent, and therefore the a point on that line can't be considered an outlier. If we look to the lower right plot in Figure 7.3 an follow the orange line, we see that it is fairly straight, with the one exception at about (28, 12), since this is a single point that is way out line, this could be considered an outlier that must be due to user error. While this method can be effective at removing outliers, it can't be known for curtain if these point in fact are outliers, therefore this method has been discarded. The plots for the other exercises can be found in Appendix B.

In order to compare the wristband to the digital VAS we will look into how they performed in two aspects, one is the accuracy which will be greatly investigated as this is the most important aspect to validate the use ability of the wristband. Beside the accuracy a short analysis of the task completion time of the different methods will be discussed.

7.1 Accuracy

In this experiment error is measured as the absolute difference between the stimuli and response, the lower the error the better the accuracy. Beside the error we will look into the the distribution and variance in order to describe the accuracy and compare it between the input methods.

7.1.1 Distribution and Scatters of Stimuli and Responses

Looking at the distribution of stimuli and responses will reveal if there is any bias using any of the input methods. Since the stimuli was evenly distributed we would expect to see the same of the responses if they don't have any bias, in the following Figures the expected amount of is marked with a red line. A scatter plot of response over stimuli will visualize the errors, a perfect response will fall along the diagonal (marked with a red line), the closer the scatter is to the diagonal the smaller errors will be. We see both the distributions and scatter plots of stimuli and responses for each exercise in Figure 7.4, 7.5, 7.6, 7.7, 7.8 and 7.9.

First lets investigate the distributions of stimuli and responses. First of all we see that the stimuli is evenly distributed between either 0-49 for the grey scale exercises or 0-100 for the number scale exercises, this of course is no surprise since the experiment was designed this way. We do notice that there is a little unevenness in the stimuli distribution, this is due to that each exercise had 24 participants that each performed 20 trials for a total of 480 trials. Neither 50 nor 101 is evenly dividable by this, resulting in a few stimuli being shown one time less than the others. The four outliers removed also influence this, but the overall distribution is as close to even as possible. With the distributions of the responses we don't see a perfect distribution, e.i. completely bias less, but neither do we see a largely skewed distribution meaning there isn't a strong bias. That said there are some interesting observations.

In Matejka et al.[23] work they find that when using a slider similar to the one used in this experiment, that there was a bias towards the two extremes (min and max) and the middle. In Figure 7.4 we see a similar bias towards the extremes, and arguably towards the middle, although not as strong a bias as Matejka et al.'s results. They explain that the bias observed was due to the

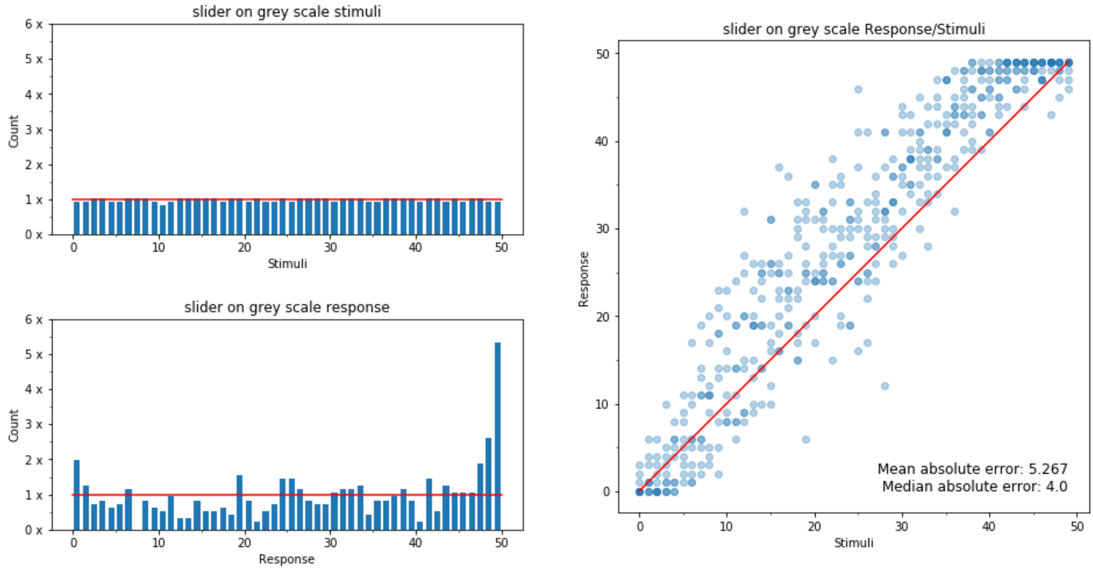

Figure 7.4: Distributions and scatter plot of stimuli and responses for slider on grey scale

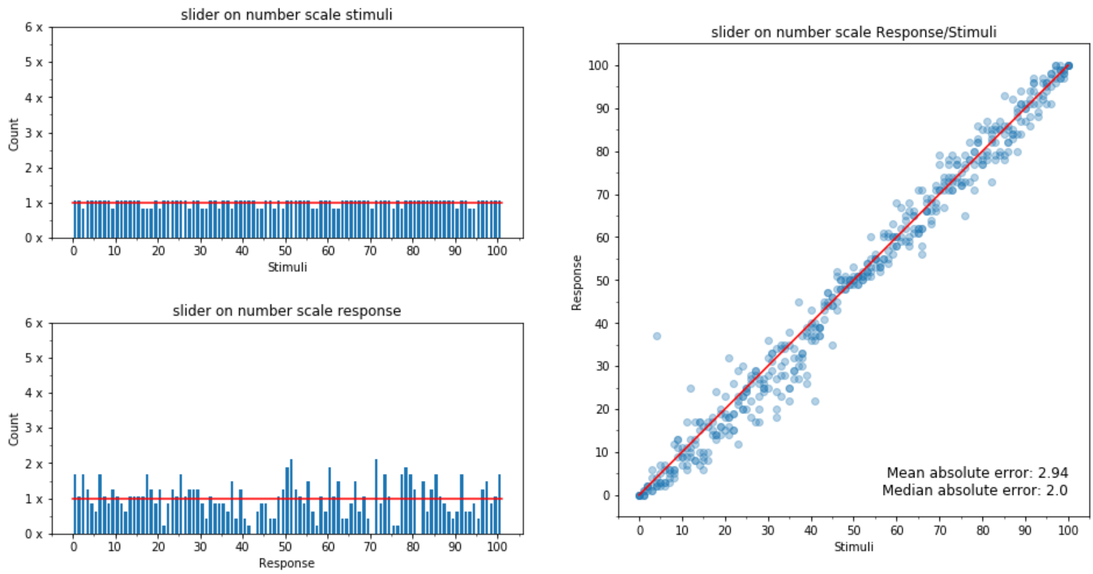

Figure 7.5: Distributions and scatter plot of stimuli and responses for slider on number scale

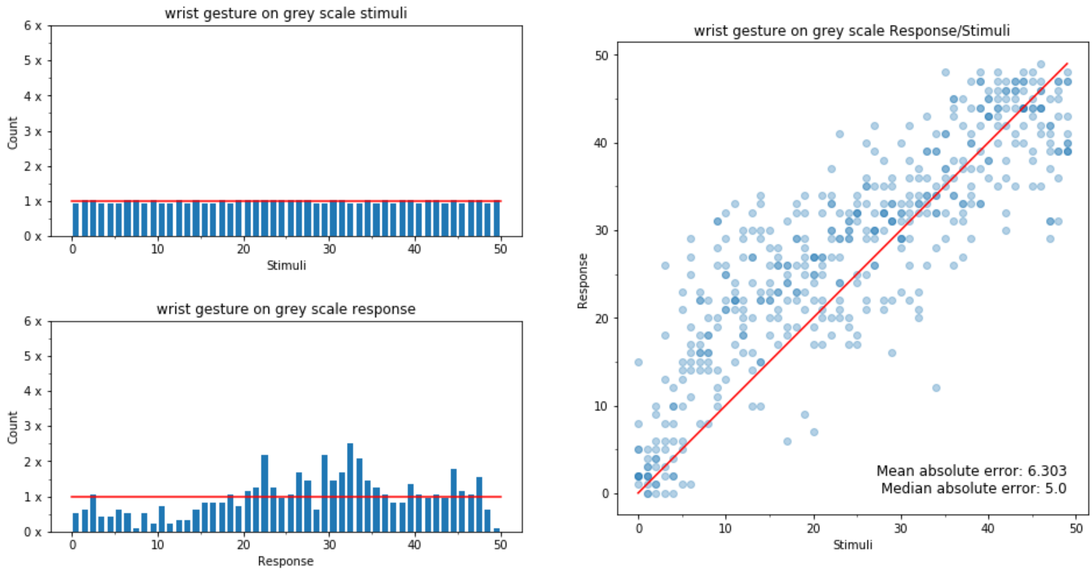

Figure 7.6: Distributions and scatter plot of stimuli and responses for wrist gesture on grey scale

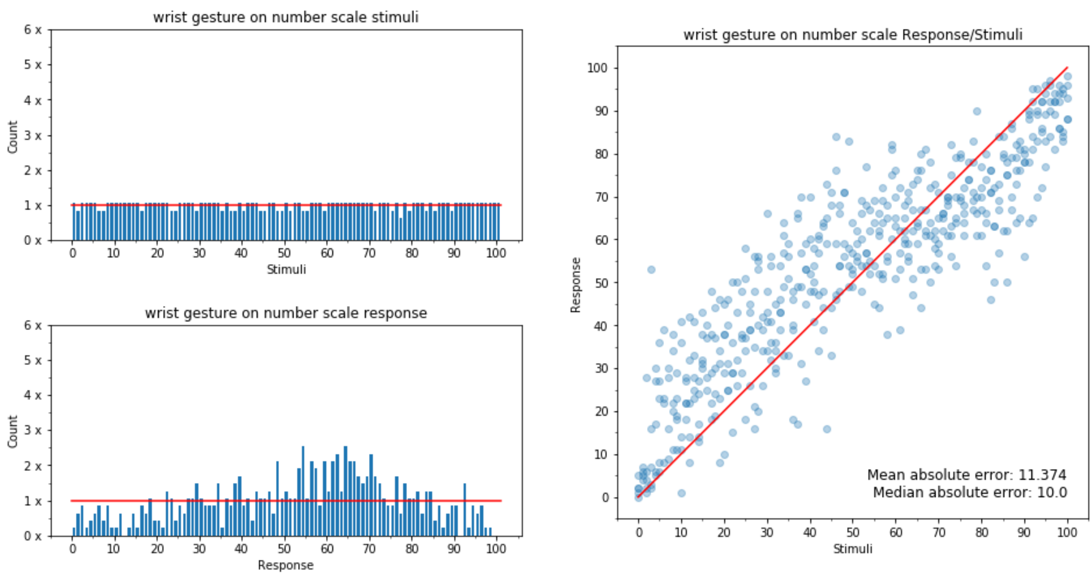

Figure 7.7: Distributions and scatter plot of stimuli and responses for wrist gesture on number scale

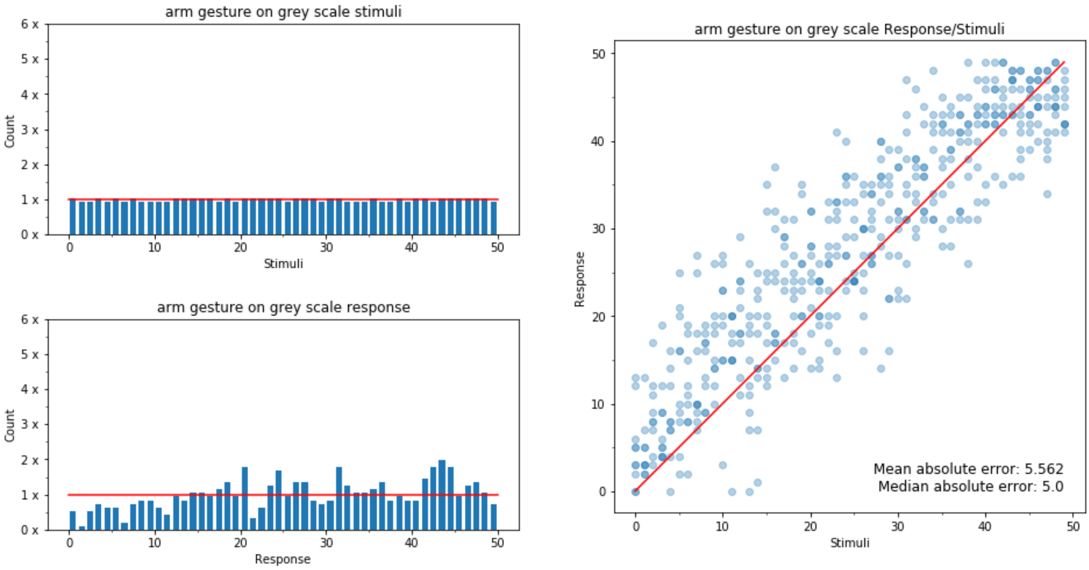

Figure 7.8: Distributions and scatter plot of stimuli and responses for arm gesture on grey scale

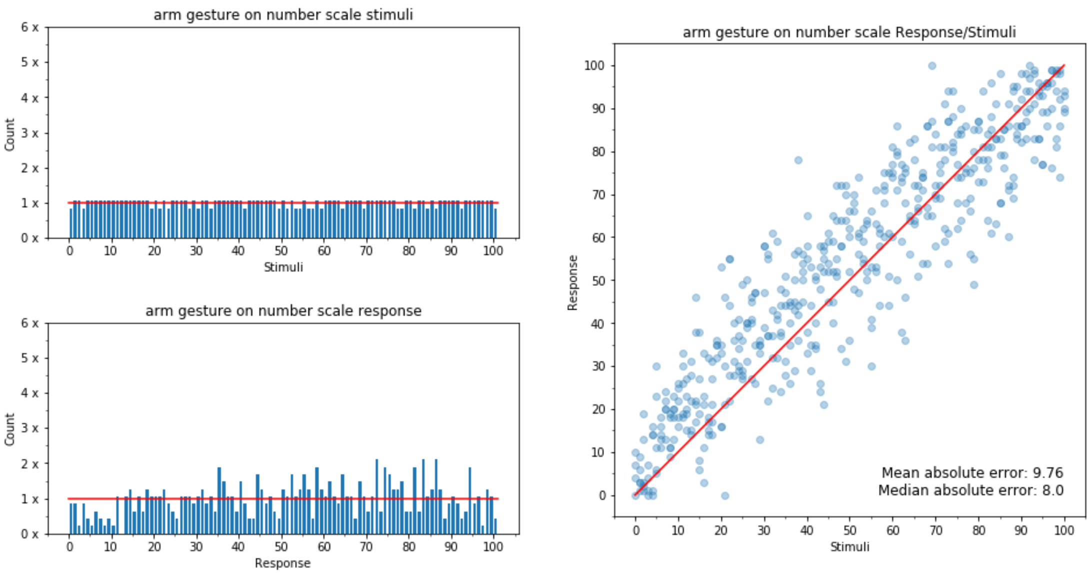

Figure 7.9: Distributions and scatter plot of stimuli and responses for arm gesture on number scale

use of a slider (easy for humans to find the middle, and we tend to choose the extremes). But the distribution seen in Figure 7.5 doesn't show any strong bias, since the only difference of these two exercises is the stimuli (shades of grey vs integers) then these results suggest that the bias observed in both Figure 7.4 and observed by Matejka et al. is in fact not caused by the slider, but caused by the shades of grey. Looking at the shades of grey in Figure 2.4 it becomes quite apparent, since it is almost impossible to differentiate pure black from the darkest shades. It is a bit easier to tell pure white apart from the light shades, this also explains why the bias toward zero is less significant. Dam-Jensen[16] found in his results that participants where only able to differentiate around 11 shades of grey, this again explain why the darker shades of grey get registered as pure black by the participants.

In Figure 7.6 and 7.7 we see greater amount of responses around the 32 and 65 mark respectively. We will come back to this observation soon.

Looking at the scatter plots we see can see the errors. First of all it is quite clear that the slider on number scale exercise outperforms all of the others, the scatter is very close to the diagonal. The other scatters also show a clear correlation around the diagonal, but not as close to it. In Figure 7.6 and 7.7 we see a very interesting observation. Looking at the scatters we see that there is a trend, with stimuli at the lower half of the scale the responses are to high and with stimuli in the upper half responses are to low. This is especially clear in Figure 7.7. It seems that there is a turning point at approximately 32 and 65 for wrist gesture on grey scale and wrist gesture on number scale respectively. These turning points are the same as we observed earlier in the distributions. A reasonable explanation for this observation can be found in biology, when investigating the wrist gesture (Figure 4.3) it becomes quite apparent, because of the bone structure in the lower arm (as seen in Figure 7.10) the gesture isn't linear. In the starting position of the gesture the bones are crossed and in the vertical position the bones are parallel. Since the bones aren't the same size and aren't straight this movement becomes non linear, meaning that the amount of movement required to turn the wrist from the starting position till about the halfway point is much greater than the movement needed to turn the wrist from the halfway point to the vertical position. In other words, we have a finer adjustment for the first half of the rotation than we do for the last half. It is plausible that a mathematical model could be created to compensate for this, thus improving this method, but that is beyond the scope of this thesis.

In Table 7.1 we see the mean, median and scaled absolute error for each exercise, these are the same values as displayed on the scatter plots. We also see see the amount of data points within ± 1 and ± 2 on a 0-10 scale, more on this later. The mean and median absolute errors are calculated across all trials for each exercise. The scaled values is the value mapped to a 0-10 scale, this is simply

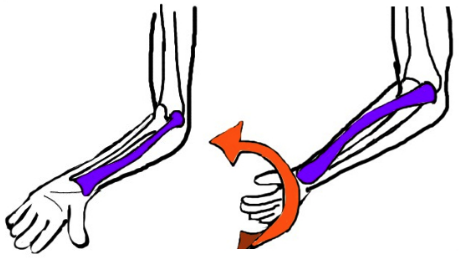

Figure 7.10: Bone Structure of Lower Arm[11]

achieved by dividing the value by either 4.9 or 10 respectively for grey scale and number scale exercises. We see that for all exercises that the mean and median absolute error is relatively low, when scaled they are all within close to one, meaning on a 0-10 scale the responses fall close to being withing ± 1 of the stimuli, of course the exception being the slider on number scale which performs even better than the others.

To further illustrate how the point in the scatter plots fall based on the 0-10 scale we will add some colored bands to the plots, the green area is where the points fall within ± 1, together with the yellow area this is where the points fall within ± 2 on the 0-10 scale. On the plots the percentage of point that fall within each area is also displayed (this is also in Table 7.1). This is seen in Figure 7.11, 7.12 and 7.13. We see that with the exeption of slider on number scale that all the other exercises performed about the same, with arm gesture on number scale performing just a bit better than the other (with the exception of slider on number scale). Based on the results so far we see no significant different between the three input methods when it comes to rating shades of grey. When rating integers from 0-100 slider on number scale strongly outperforms the other two while arm gesture on number scale is slightly better than wrist gesture on number scale.

7.1.2 Error Distribution and Variance

To further investigate the difference in performance of the different input methods we will look into the error distribution. They will be compared visually and using t-test to see if the are statistically similar. In Figure 7.14, 7.15 and 7.16 we see the distribution of errors for the six exercises. As we would expect the errors follow a normal distribution centered around zero. Again it is clear to see that the slider on number scale performs way better than the other number scale exercises, the variance is smaller resulting in a much steeper curve meaning

	Mean abs. error	Median abs. error	Scaled mean abs. error	Scaled median abs. error	Points within ±1	Points within ±2
slider on grey scale	5.267	4.0	1.075	0.8	51.98%	79.96%
slider on number scale	2.940	2.0	0.294	0.2	96.04%	99.58%
wrist gesture on grey scale	6.303	5.0	1.286	1.0	43.22%	72.03%
wrist gesture on number scale	11.374	10.0	1.137	1.0	48.54%	79.29%
arm gesture on grey scale	5.563	5.0	1.135	1.0	48.54%	77.92%
arm gesture on number scale	9.760	8.0	0.976	0.8	55.21%	88.33%

Table 7.1: Mean, median and scaled absolute error together with the percentage of points within ±1 and ±2 on the 0-10 scale for each exercise

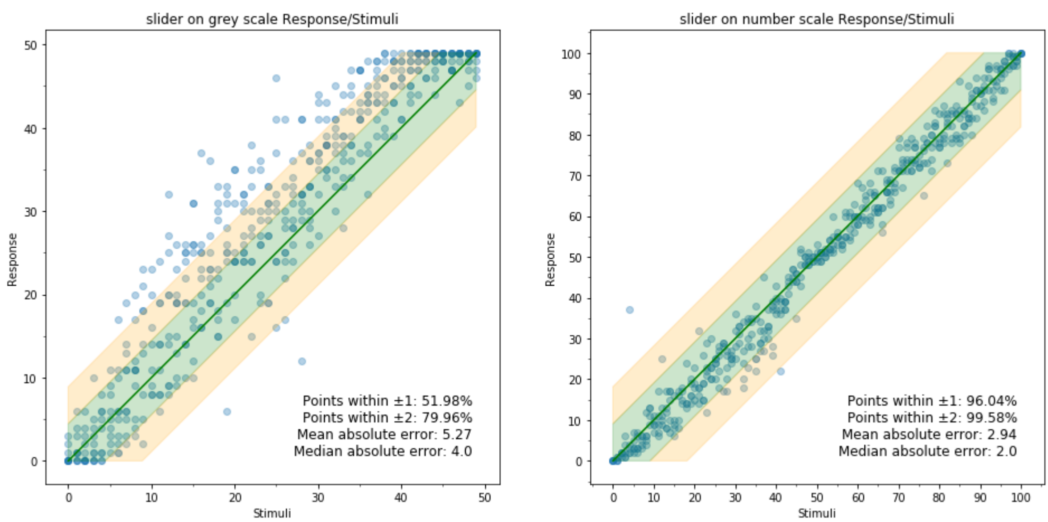

Figure 7.11: Scatter plot with color bands for slider

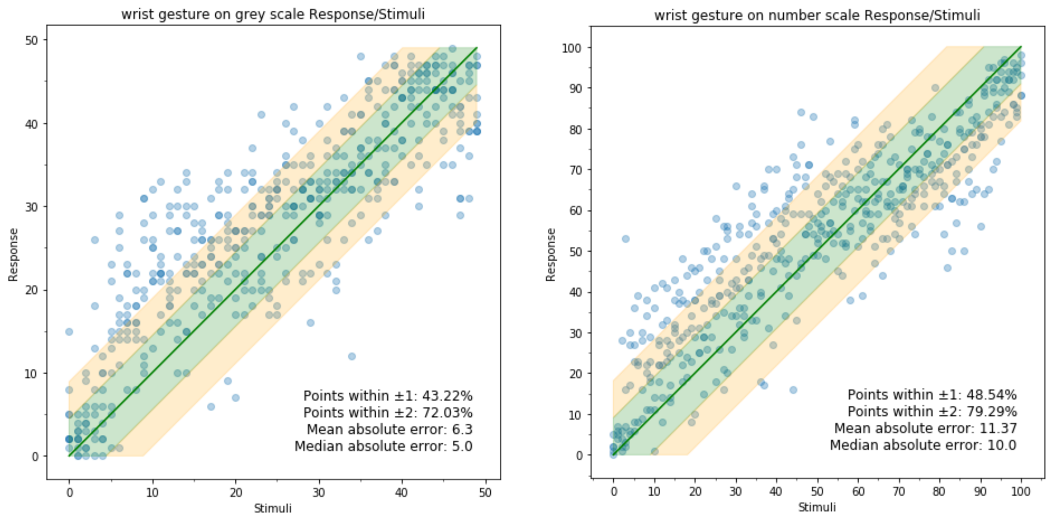

Figure 7.12: Scatter plot with color bands for wrist gesture

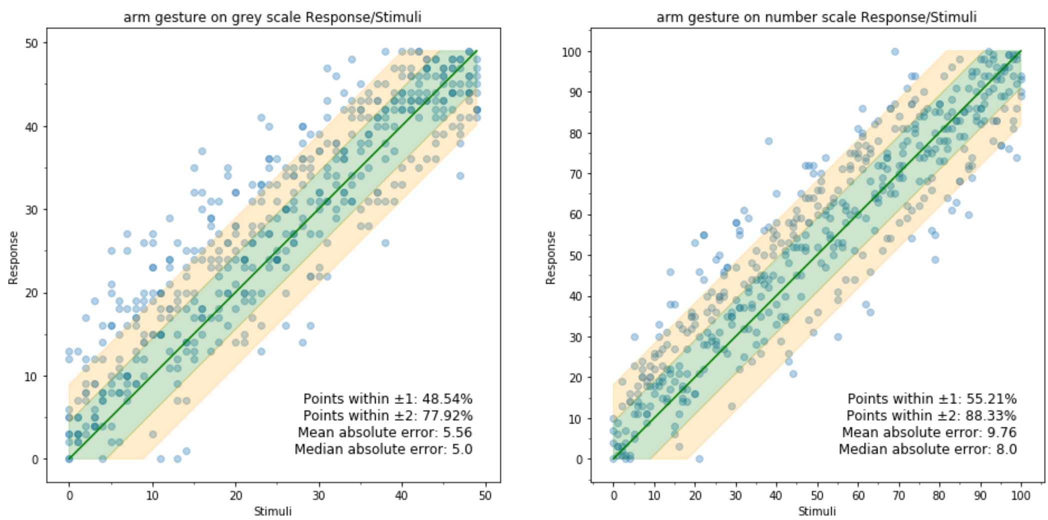

Figure 7.13: Scatter plot with color bands for arm gesture

that the response are much closer to the stimuli. The other two number scale exercises (wrist gesture on number scale and arm gesture on number scale) seem to perform about the same. Likewise all the grey scale exercises seem to perform about the same, no clear difference in the variance or steepness of the curves.

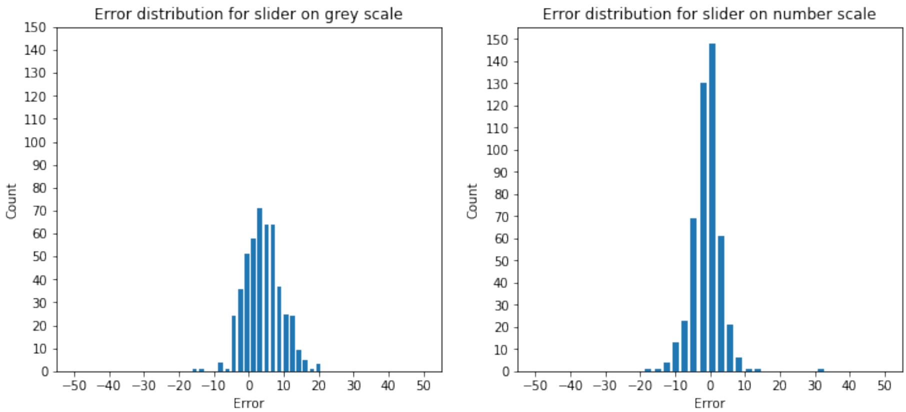

Figure 7.14: Error distribution for slider

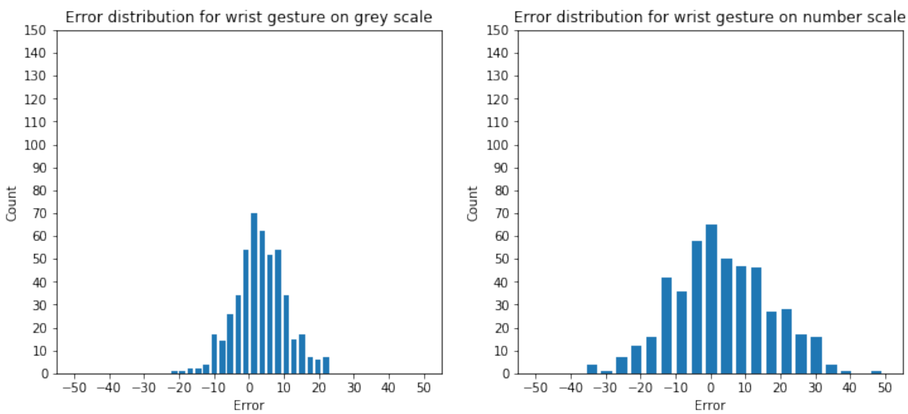

Figure 7.15: Error distribution for wrist gesture

T-test where performed between the grey scale exercises and between the number scale exercises to see if the are statistically alike. The results of the t-test can be seen in Table 7.2, for each comparison the errors from all trials of two exercises where used.

Lets start by comparing the grey scale exercises to each other. We see that wrist gesture on grey scale and arm gesture on grey scale are quite likely not

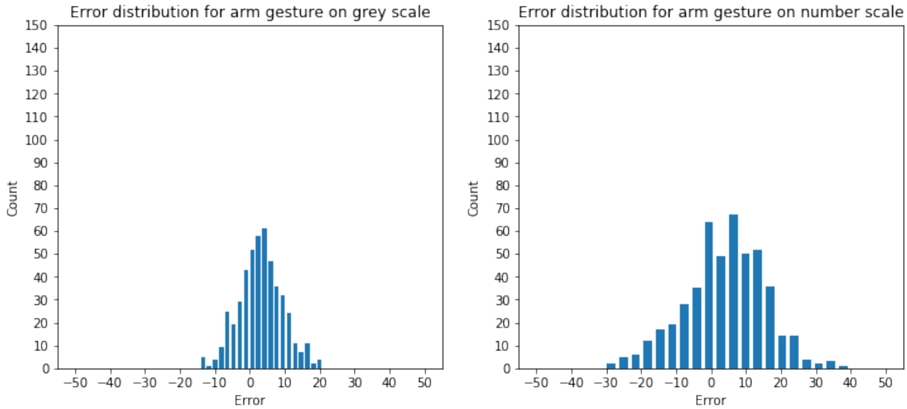

Figure 7.16: Error distribution for arm gesture

to have a significant difference. From the results of comparing slider on grey scale with wrist gesture on grey scale and arm gesture on grey scale it is hard to determine weather or not they are similar. On one hand slider on grey scale and wrist gesture on grey scale is somewhat likely not to be significant different, while slider on grey scale and arm gesture on grey scale on the other hand are likely to have significant difference. These results are a bit strange, since one would assume if A and B are similar and A and C are different, then B and C must be different too, but here we see that B and C are quite similar (A being slider on grey scale, B being wrist gesture on grey scale and C being arm gesture on grey scale). That being said, the p value for wrist gesture on grey scale and arm gesture on grey scale is rather high and the calculated t-statistic is close to zero, indicating that this is the most likely case to be true out of the three.

Comparing the number scale exercises we again see that there is a clear difference between slide on number scale and the other two number scale exercises. Comparing wrist gesture on number scale and arm gesture on number scale we see that they are quite similar, meaning that statistically the change of the data being part of the same distribution is high.

Comparing grey scale exercises to number scale exercises only makes sence if comparing them with the same input method. Doing so we see the lowest p value and highest calculated t-statistic across the board when comparing slider on grey scale to slider on number scale, this is not surprising when we know how well slider on number scale performed. This indicates that there is a large difference when asked to rate shades of grey using the slider and rate integers using the slider. Both wrist and arm gesture shows to be similar when comparing the grey scale exercise to the number scale exercise, and it is quite interesting

Exercises compared		p value	Calculated t-statistic
slider on grey scale	wrist gesture on grey scale	0.10045	1.64426
slider on grey scale	arm gesture on grey scale	0.01598	2.41378
wrist gesture on grey scale	arm gesture on grey scale	0.61034	0.50976
slider on number scale	wrist gesture on number scale	1.29468e-10	-6.49960
slider on number scale	arm gesture on number scale	9.35024e-19	-9.02859
wrist gesture on number scale	arm gesture on number scale	0.39349	-0.85370
slider on grey scale	slider on number scale	6.82003e-55	16.65620
wrist gesture on grey scale	wrist gesture on number scale	0.93712	0.07892
arm gesture on grey scale	arm gesture on number scale	0.14421	-1.46148

Table 7.2: T-test between exercises

to see that wrist gesture on number scale and wrist gesture on grey scale has the highest change of being significant similar across the board.

With these results we must conclude that the two input methods arm and wrist gesture are quite alike, while we can't say for sure if they are similar to slider when rating shades of grey we can say for curtain that arm and wrist gesture is quite different from slider when rating integers. While the t-test lest us examine the likely hood of two data sets to be similar or not, it can't tell us which performs better.

Summarizing on the accuracy performance of the three input methods it is quite clear when it comes to rating integers that the slider outperforms both arm and wrist gesture. When it comes to rating shades of grey there seems to be no clear significant difference between the three inputs methods. And with the input methods arm and wrist gesture there doesn't seem to be a significant difference in rating shades of grey to rating integers.

	Mean	Median	Standard deviation
slider on grey scale	3.589	2.672	2.981
slider on number scale	4.552	3.942	2.436
wrist gesture on grey scale	4.089	3.195	2.588
wrist gesture on number scale	4.376	3.614	2.553
arm gesture on grey scale	4.338	3.563	2.522
arm gesture on number scale	4.618	3.823	2.633

Table 7.3: Task completion time (seconds) for each exercise. The time from stimuli being shown until data was registered by the participants in each trial

7.2 Task Completion Time

The task completion time for each trial was recorded, this was the time span from the stimuli was showed to the participant registered the data. A short analysis based on the results for all the trials showed that there were no significant difference between the six exercises as shown in Table 7.3. There is only a little more than a second difference between the mean task completion time of the fastest and slowest exercises. Because of this no further investigation of the task completion time was conducted. The results mean that it doesn't take a significantly longer time to use the wristband than to use the slider, keep in mind that the task completion time was recorded from when stimuli was showed on screen until the participant pressed the submit button. For real life use this gives the wristband a significant advantage, since it wouldn't take much longer to use the wristband in real life than it did in the experiment, but if one where to use a digital VAS on a smartwatch the person would first have to activate the smartwatch and go through the interactions required to input the value This would involve more steps than was present in the experiment, meaning a longer total interaction time.

7.3 Post Experiment Survey

After the experiment the participants were asked to fill in a survey, the full survey and all the answers can be found in Appendix A. Unfortunately an unknown error occurred while collecting the answers, resulting in three lost answers. This section will show the results of the most relevant questions.

Given the option between the arm and wrist gesture method the participants

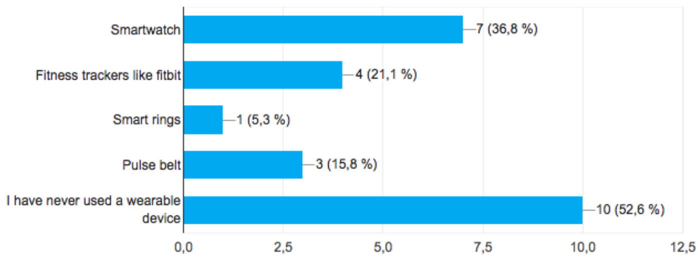

Figure 7.17: Participants prior experience with wearable devices

where asked which method they preferred, which they felt was more accurate and which they felt was most comfortable. For each question they were also given the option to answer that they felt the same for both methods. The results can be seen in Table 7.4. Among the participants that preferred the arm gesture method they the common answer to why was that the greater range of motion allowed them to have higher precision. For the participant preferring the wrist gesture the most common reason was that it was more comfortable. For the participant that felt that the arm gesture method was more accurate their reason was that it had more precision, making it easier to make small adjustment. The two participants that felt the wrist gesture was more accurate said that they just felt like it. One participant that felt the methods to be equally accurate explained that the task felt difficult in both cases, but with practice and feedback that one could become accurate when using either method. The many participants that felt that the wrist gesture method was more comfortable explained that it was less hassle, less awkward movement and that they didn't get tired. Among the participant rating the two method equally comfortable their comments was that neither of the methods were uncomfortable to use.

	arm gesture	wrist gesture	equal
Preferred	38.1%	42.9%	19%
Accurate	81%	9.5%	9.5%
Comfort	9.5%	61.9%	28.6%

Table 7.4: Distribution of Answers about arm gesture vs wrist gesture (Appendix A)

The participants was asked if they had any prior experience using any kind of wearable devices, and with which devices (it was possible to check multiple devices). The results can be seen in Figure 7.17. This showed that participants represented a good mixture of experience with wearable devices.

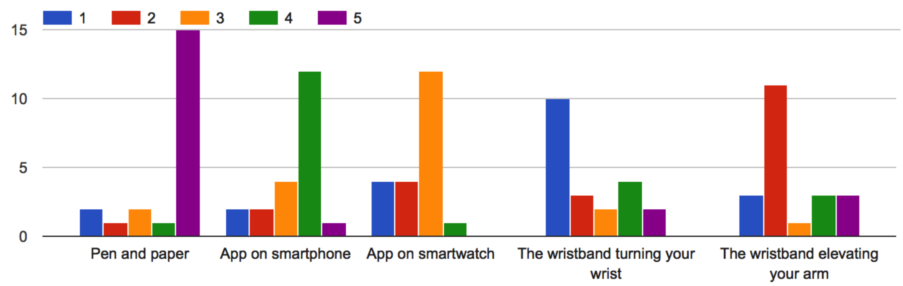

Figure 7.18: Rating of preferred method (1 is most preferred, 5 is least preferred)

Lastly the participant had a use case explained, where the had to imagine that they suffered from migraines and that their doctor had asked them to log whenever they had a headache and how severe it was. They where then presented with five different methods to log the headaches (which was described in detail) pen and paper, app on smartphone, app on smartwatch, the arm gesture method or the wrist gesture method. They where then asked to order all the method from most preferred to least preferred, the result can be seen in Figure 7.18. The results clearly show that the participants mostly preferred the wristband over the other methods, with pen and paper being the most disliked method. This question should only be seen as an indication, the question required the participants to think of a complex use case and make decisions on things where they probably don't have every piece of information to in order to take an informed decision. That being said the results shows promise for wristband device using the arm and wrist gestures.

Based on all of the results, both from the experiment and the survey, is shows great promise for a use case for the wristband. It was shown that the performance of the wristband did just as good for rating shades of grey which argues it would perform well for recording psychological events, where their isn't a specific value to record, but instead what the user feels. The arm and wrist gesture methods where shown to perform about the same, therefore either could be recommended. Based on the survey it seems to be a trade off between the feeling of accuracy and comfort, but since it has been shown that their isn't a significant difference in accuracy the wrist gesture method would be preferred to achieve a higher level of comfort.

Chapter 8

Future Work

The work done with the two input gestures is only the beginning. This Chapter will suggest what could be of interest if more work were to be done to this project.

8.1 Improvement to the Experiment

The experiment conducted was a success, but if anyone was to conduct the experiment again a few improvements could be made. With the design of the slider and submit button the button is positioned in the center below the slider. This had the unintended effect of serving as a reference point for the middle of the slider. Simply changing the position of the submit button to either side of the slider would resolve this. As discussed in the results, we humans are quite good at finding the center of a slider anyway, so this is smaller detail that most likely won't have a significant impact on the outcome.

It would be interesting to see if the results of the experiment is representative of a more general population, so recruiting participants with a broader demographic would give an idea of this. Likewise it would be interesting to see if the observed results will stay the same with a larger data set using more participants. In would also be optimal to increase the number of stimuli for each exercise, although it is a trade off since it could lead to fatigue.

Conducting another experiment where only 11 shades of grey and numbers between 0 and 10 would be of interest, to see how the three input methods would compare in a setting closer to the real life use case. Likewise an experiment without a ground truth could be of interest to compare the three input methods, e.g. where participants are asked to rate pain levels, mood or other psychological events. Such an experiment might be harder to conduct since you would need to provide a stimuli for the psychological events, for pain that could be easy, but might not be ethical. It is also harder to compare the results since there won't exist a ground truth, in this case it would be interesting to see how similar the results would be. Such an experiment could be conducted over a longer period of time, where the participants are asked to use both the digital VAS and one or both of the gestures to rate psychological events.

8.2 Case Study

To further develop the work of this thesis, the next step would be to conduct a case study. This could be done similar to the work of Larsen et al.[21], where a patient suffering from PTSD was asked to log whenever he experienced a specific event related to his illness using the smartbutton. The difference here would be that the patient would be asked to rate whenever he/she using the wristband and one of the two gestures (arm or wrist). PTSD is just one example that could be used in a case study, it would be equally interesting to conduct a case study on a patient suffering from migraines, chronic pain, anxiety, panic attacks or any similar condition where the common thing is that it will create value to track whenever the events of their conditions occurred and how severe the events are.

With such a case study it would be interesting to see if the tracking of severeness provides additional value over just tracking when events occur. It would be interesting to investigate the compliance of using the wristband with the hand or wrist gestures compared to the simpler smartbutton.

8.3 Improvements to Companion App

The companion app designed and implemented in this thesis was developed with the intentions to be a MVP, therefore some features where left out, but given more time it would create value to improve the companion app. Visualization of the data collected could make it easier to spot patterns. A simple proposal would be bar plot with days along x-axis and the number of tracked events on the y-axis, the bars could be colored to represent the severeness logged. If a given day had 5 data points with severeness rated at 1, 3, 6, 8 and 9 the bar could be colored with multiple color bands based on a color scale going from green, to yellow to red (like in the Mood tracker 2.5). This is just one suggestion, there exist many papers on how to visualize such data, it is important that the visualization is thoroughly tested on a wide demographic to ensure that it can be understood by the users. Beside a better visual presentation of the collected data, it would be interesting to the location of the observation, while this isn't possible with the wristband device, this can be achieved by synchronizing the time stamps of the logged data to the smartphone location history. This could provide additional value to the treatment process.

While these are the largest improvement that can be made to the companion some other minor adjustments could be made. In order to use companion app the wristband has to be connected first, this is of course important to program the wristband and import the data from it, but that means the user can't aces the collected data without first connecting the wristband. This is of course a inconvenience, the app should let the user see the collected data and only require the wristband to be connected when importing data or programming the wristband. Improvements could be made to the UI of the app to make it more visually pleasing, but that hasn't been the focus of this thesis. Features such as deleting individual data points would also improve the user experience of the app.

8.4 Investigating the Wrist's Critical Point

In connection to conducting a case study and improving the companion app it would be reasonable to investigate the conditions surrounding the "critical point" of the wrist gesture. As mentioned in the Chapter Results & Discussions7, the "critical point" is the condition observed that caused the participant's responses to over estimated both the shades of grey and numbers when they where on the lower half of the scale, and under estimate the responses for the upper half of the scale. As mentioned this is likely do the bone structure of the lower arm causing the movement of the gesture to be non linear. It would be interesting to investigate this further and try to develop a mathematical model to compensate for this. Another approach could investigate the use of machine learning and see if that could compensate for this condition, all though for this to be successful a larger data set would most likely be required.

Chapter 9

Conclusion

This thesis cowered the design and implementation of a wristband device and companion app for manual collection of data on subjective experiences using a gesture based system. An experiment was conducted in order to compare the performance of the wristband with a touch screen implementation of a Visual Analogue Scale (VAS). The results showed the designed solution to be feasible.

The thesis analyzed the problem and the key concepts that would form the foundation for the proposed solution. The two gestures of rotating the wrist and raising the arm was chosen since they corresponds to the roll and pitch axis.

The design of and implementation of the solution was described in detail. The wristband device would have a companion app to collect, display and export the data from the wristband. First a proof of concept was developed using off the shelf components, which was a success and laid the groundwork for a prototype. The prototype worked and was able to log values on a continuous 0-10 scale based on the gestures together with a time stamp.

In order to validate the performance of the solution an experiment was designed and conducted. Participants where tasked to rate different shades of grey and integers (stimuli), using both the gesture based methods and touch based VAS input methods. The collected data was analyzed for performance in accuracy and task completion time. The results showed that both of the designed input methods (arm and wrist gestures) had no significant difference in accuracy to the touch based VAS when rating shades of grey. When rating numbers the touch based VAS showed to have higher accuracy, but when mapping the results to a 0-10 scale, both of the gesture based inputs had e mean absolute within ± 1, and about 50% of the results was within this range and about 80% within ± 2. Thus it must be concluded that both methods are reliable as a replacement for a VAS. When comparing task completion times there was no significant difference between the methods. Based on user feedback the wrist gesture was a little more preferred and they felt it to be more comfortable, although they perceived the arm gesture to be more accurate. Since both gestures performed equally well and are equally preferred, it would be recommended to use the wrist gesture since it provide more comfort.

Finally, improvements where suggested that could improve the companion app and the experiment. In case the experiment was to be replicated it would be of interest to see if the results applies to a more general population and if the result will stay the same with a larger data set. A case study was suggested in order to investigate the real life performance of the solution.

Appendix A

Survey

Questions for the survey answered by the participants after the experiment. Unfortunately an unknown error occurred while collecting the answers, resulting in three lost answers.
The answers can be viewed on Google with charst following this link:
`https://docs.google.com/forms/d/e/1FAIpQLSfnv9BTGRywQlKB7XcPGaZfxnaf-ZsvQwbjxm_Wyicw4aScaQ/viewanalytics`
A CSV file is also available at this link: `https://github.com/TheBatAss/Gesture-Based-Input-Method-for-Wearable/blob/master/Survey/Post%20experiment%20survey.csv`

Questions:
ID (from the experiment):

Gender:
Male
Female

Age:

Which method did you prefer to use?
Arm
Wrist

Perferred equally

Why?

Which method did you feel was more accurate?
Wrist
Arm
The same

Why?

Which method did you feel was most comfortable
Wrist
Arm The same

Why?

Do you have experience using any kind of wearable devices?
Smartwatch
Fitness trackers like fitbit
Smart rings
Pulse belt
I have never used a wearable device
Other

Real usage
Imagine that you suffer from migraines and get headaches at irregular times throughout the day.

Your doctor ask you to log whenever you have a headache and how severe it is on a scale from 1-10.

You are given 5 options to do this:

Pen and paper You carry around a small notebook and a pen, and you have to write down the date, time and severeness each time you experience a headache.

App on smartphone You install an app on your smartphone. Each time you experience a headache you have to open the app which will ask you to select the severeness. The app automatically collects the time and date.

App on smartwatch You install an app on your smartwatch. Each time you experience a headache you have to open the app which will ask you to select the severeness. The app automatically collects the time and date.

The wristband turning your wrist Each time you experience a headache, you turn your wrist to match the severeness and press the button. Time and date is collected automatically.

The wristband elevating your arm Each time you experience a headache, you elevate your arm to match the severeness and press the button. Time and date is collected automatically.
Which method would you prefer? Order them from 1 to 5 (1 is most preferred, 5 is least preferred)
Pen and paper
App on smartphone
App on smartwatch
The wristband using your wrist
The wristband using your arm

Appendix B

Outlier Visualization

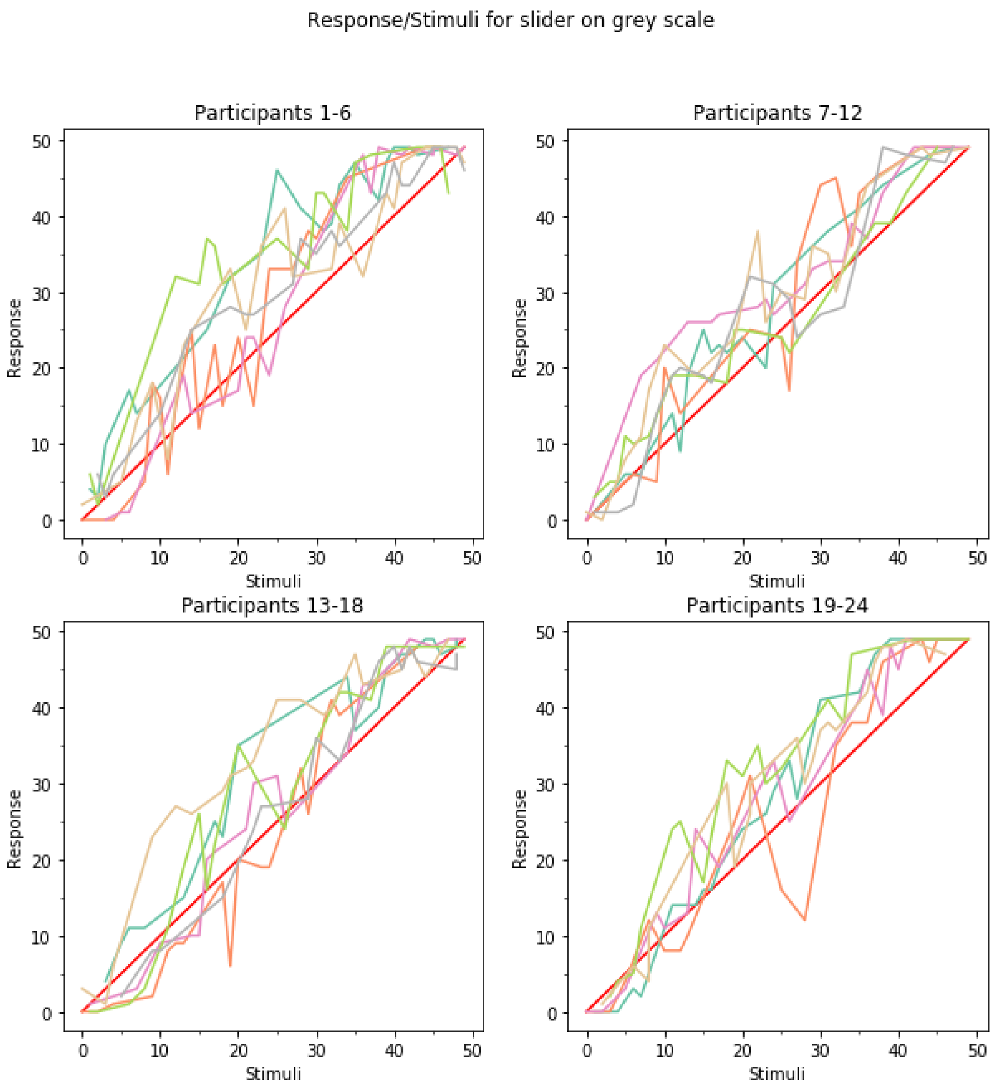

Figure B.1: Visualization of outliers for the arm_grey exercise

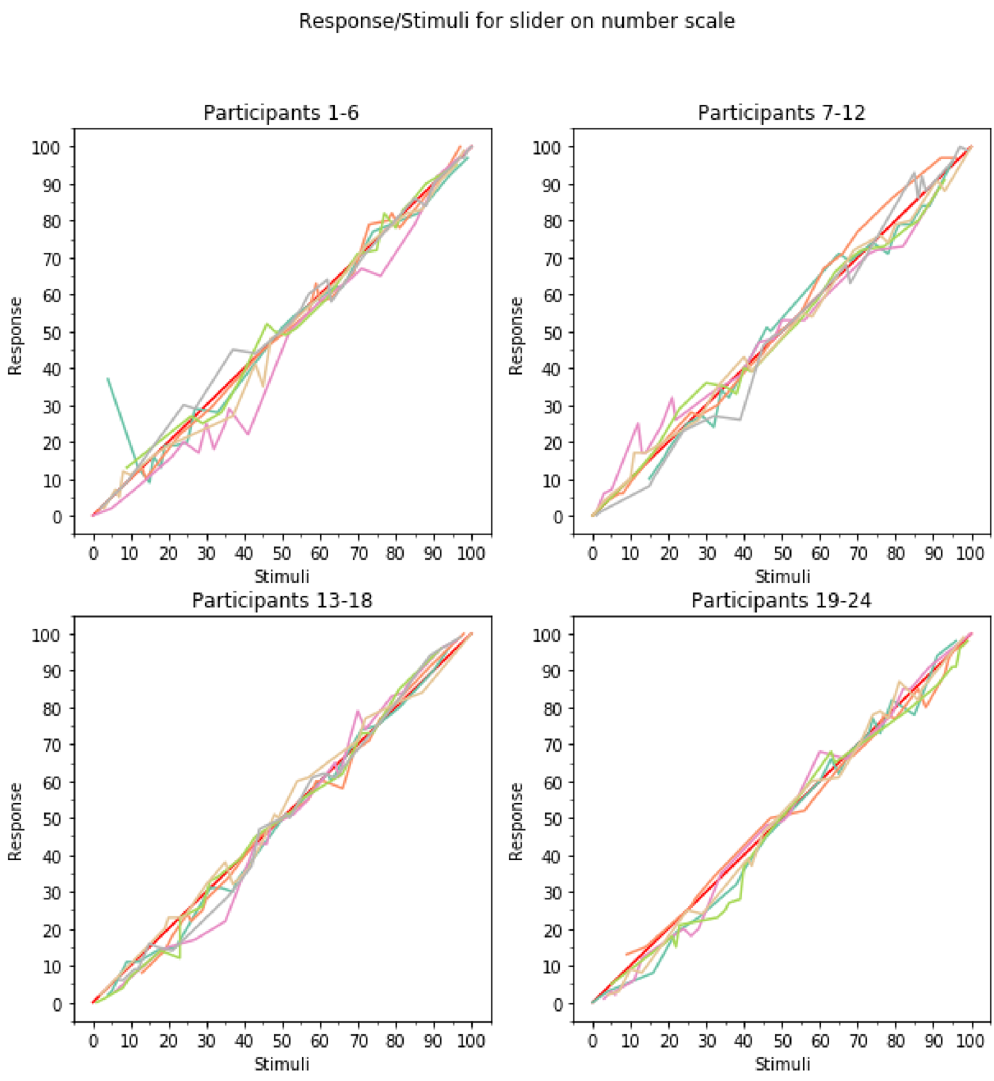

Figure B.2: Visualization of outliers for the arm_grey exercise

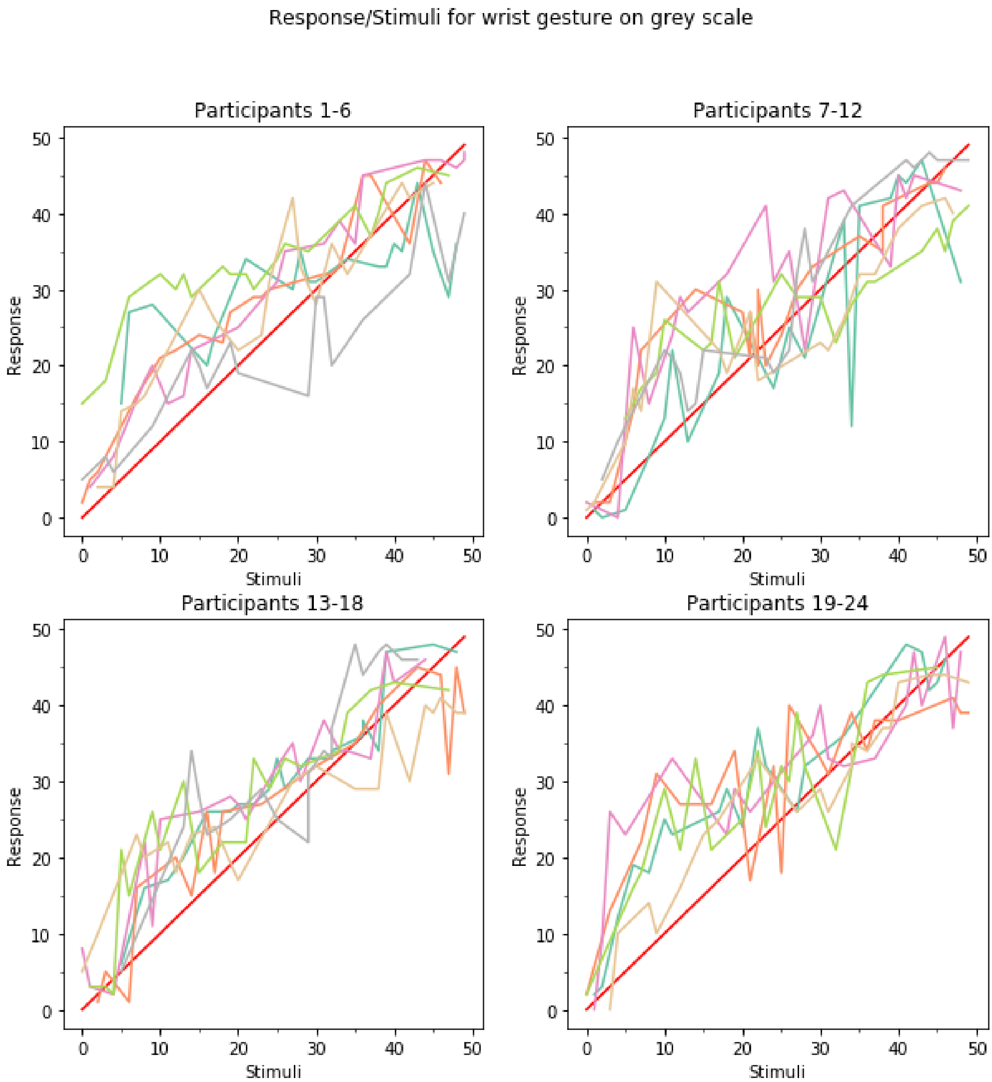

Figure B.3: Visualization of outliers for the arm_grey exercise

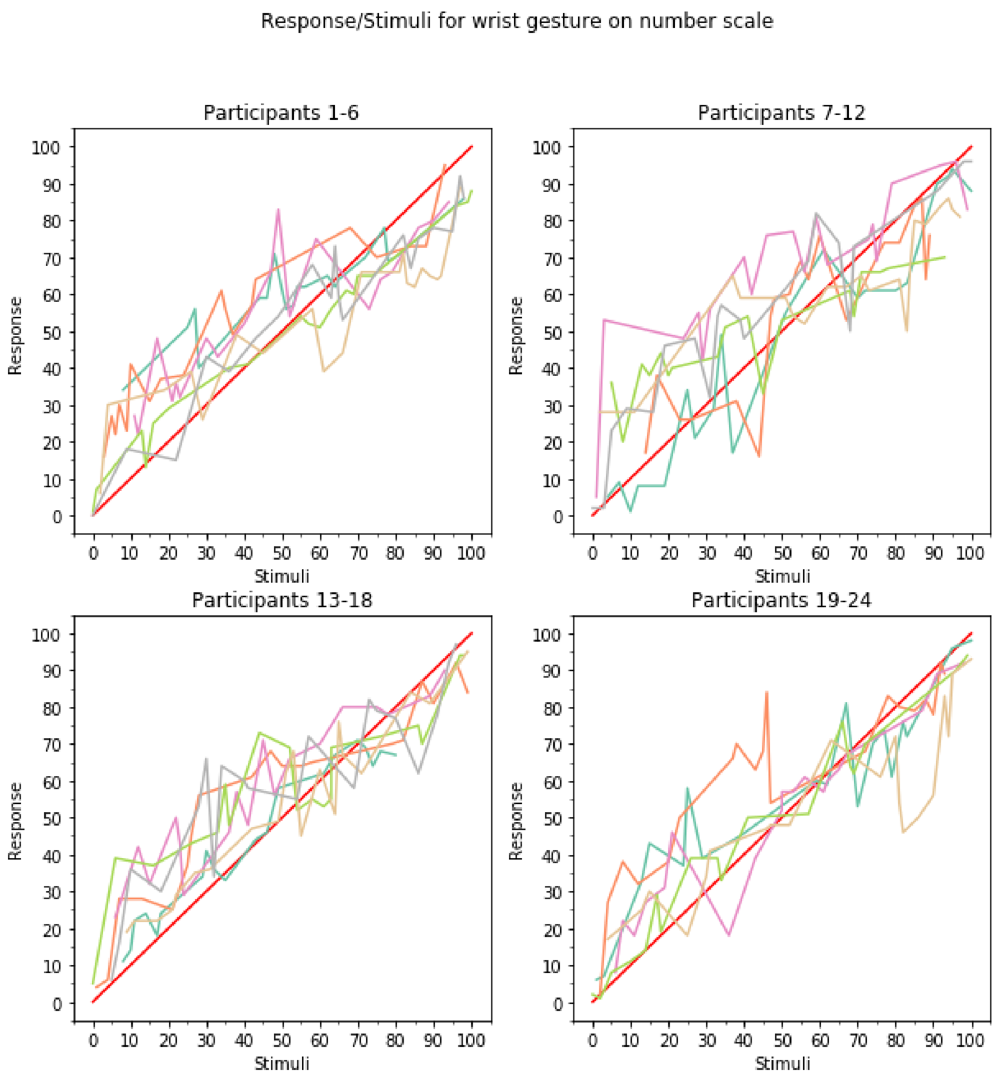

Figure B.4: Visualization of outliers for the arm_grey exercise

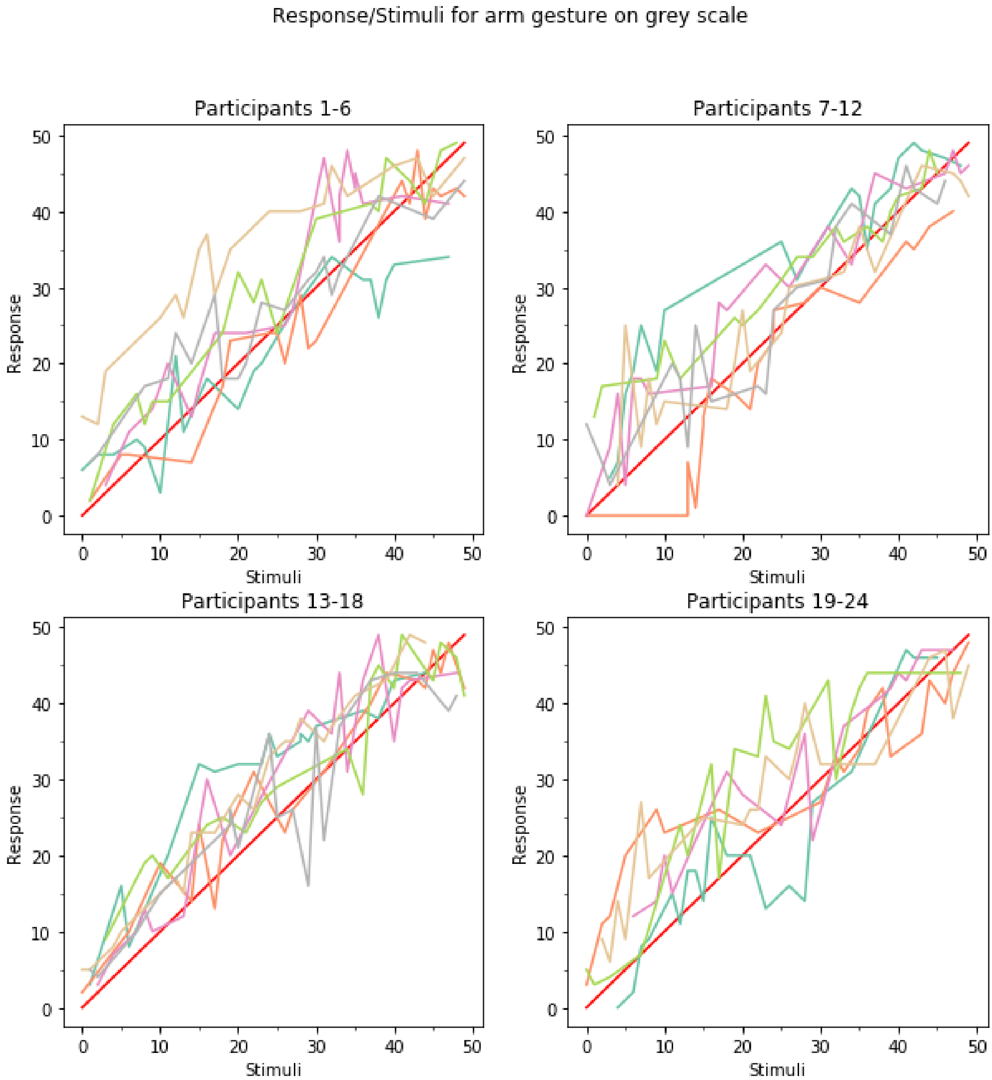

Figure B.5: Visualization of outliers for the arm_grey exercise

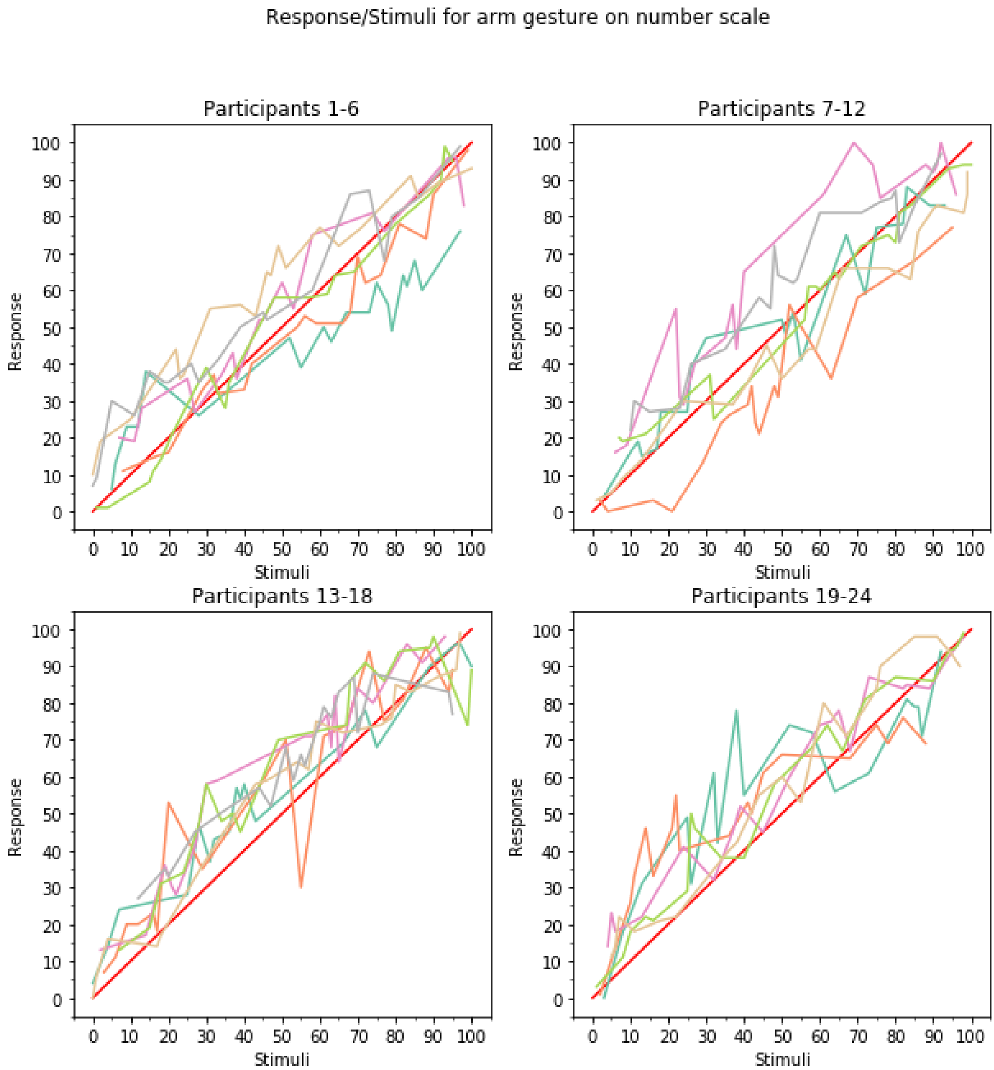

Figure B.6: Visualization of outliers for the arm_grey exercise

Appendix C

Code

All the relevant code for this thesis is hosted on GitHub at the following URL:
`https://github.com/TheBatAss/Gesture-Based-Input-Method-for-Wearable`

C.1 Proof of Concept

The code for the PoC can be found here:
`https://github.com/TheBatAss/Gesture-Based-Input-Method-for-Wearable/tree/master/proofOfConcept`

It consist of Python script "demo.py" which creates the GUI and communicates with the Adafruit Feather M0 Adalogger[4] which is controlled by the Arduino script "demo.ino".

C.2 Companion App

The code for the prototype's companion app "MetaWear Macro" can be found here:

https://github.com/TheBatAss/Gesture-Based-Input-Method-for-Wearable/tree/master/MetaWear-Macro

C.3 Experiment App

The code for the app "MetaWear Experiment" conduction the experiment can be found here:
https://github.com/TheBatAss/Gesture-Based-Input-Method-for-Wearable/tree/master/MetaWear-Experiment

C.4 Data analysis

The Python code for the data analysis can be found at this link:
https://github.com/TheBatAss/Gesture-Based-Input-Method-for-Wearable/blob/master/Data%20analysis/Data%20analysis.ipynb

Appendix D

Data

The data collected from the experiment can be found at this link:
`https://github.com/TheBatAss/Gesture-Based-Input-Method-for-Wearable/blob/master/Data%20analysis/ExperimentData.csv`

The completion times of the experiement and pilot test can be found at this link:
`https://github.com/TheBatAss/Gesture-Based-Input-Method-for-Wearable/blob/master/Data%20analysis/completionTime.csv`

Appendix E

Experiment App Screens

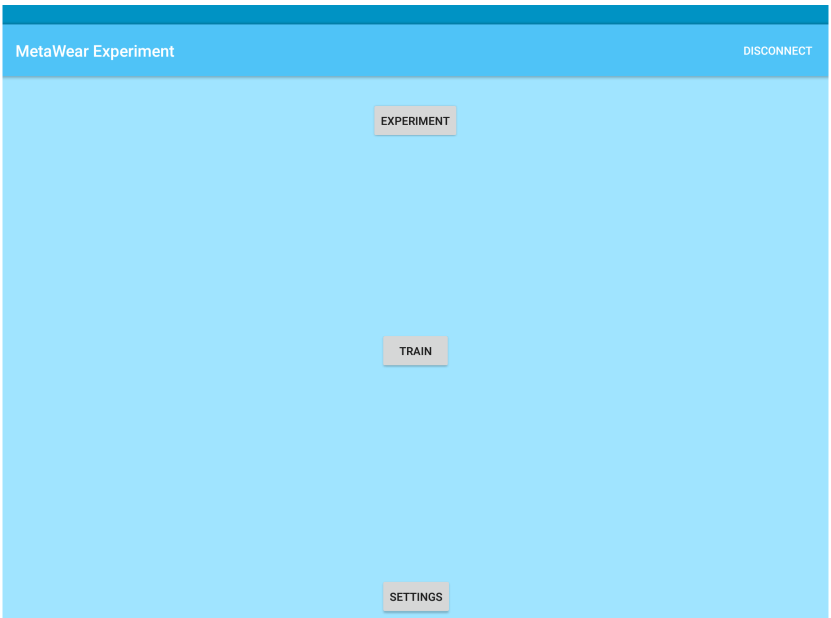

Figure E.1: Main screen

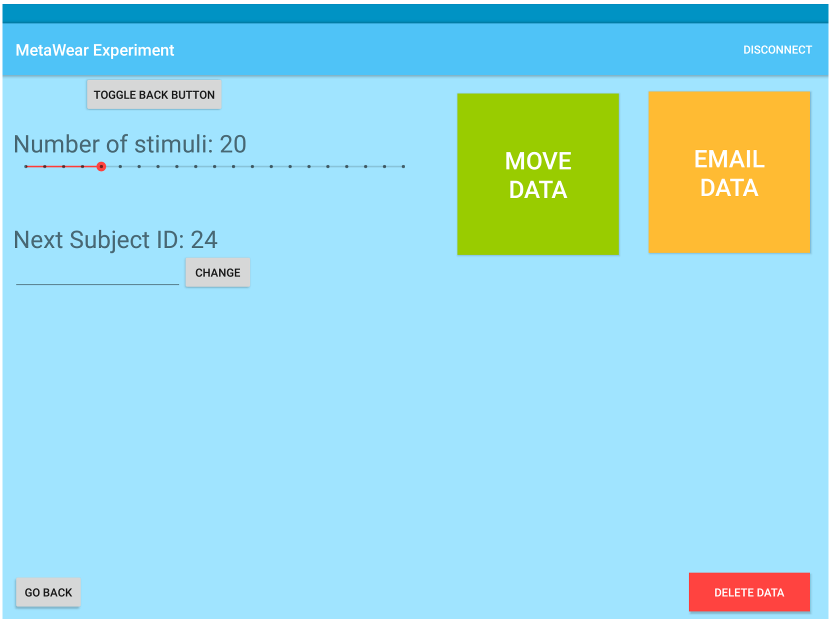

Figure E.2: Settings screen

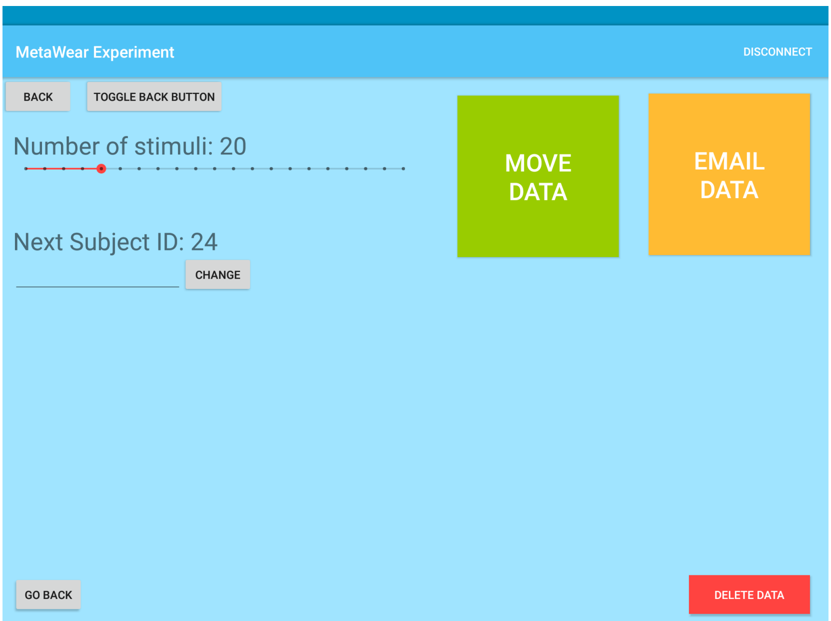

Figure E.3: Settings screen

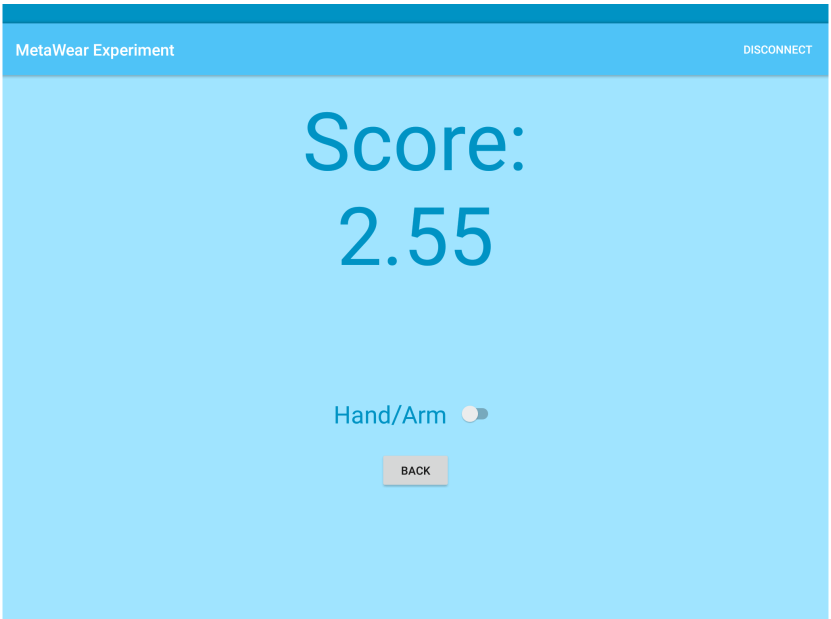

Figure E.4: Training screen

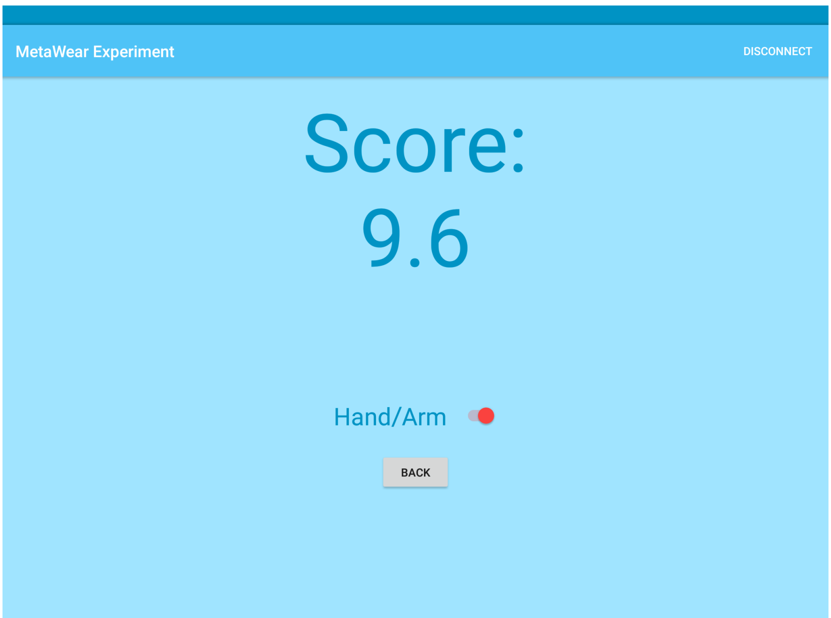

Figure E.5: Training screen

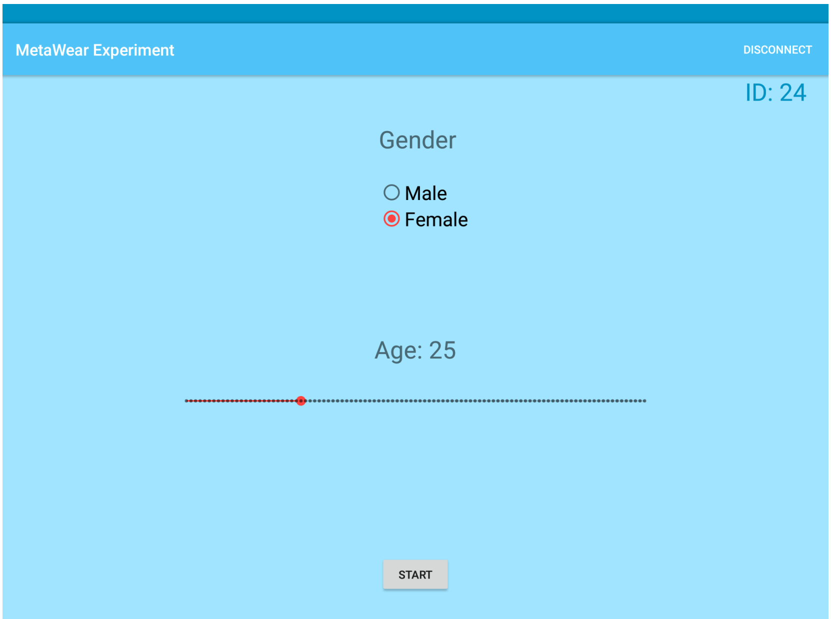

Figure E.6: Experiment start screen

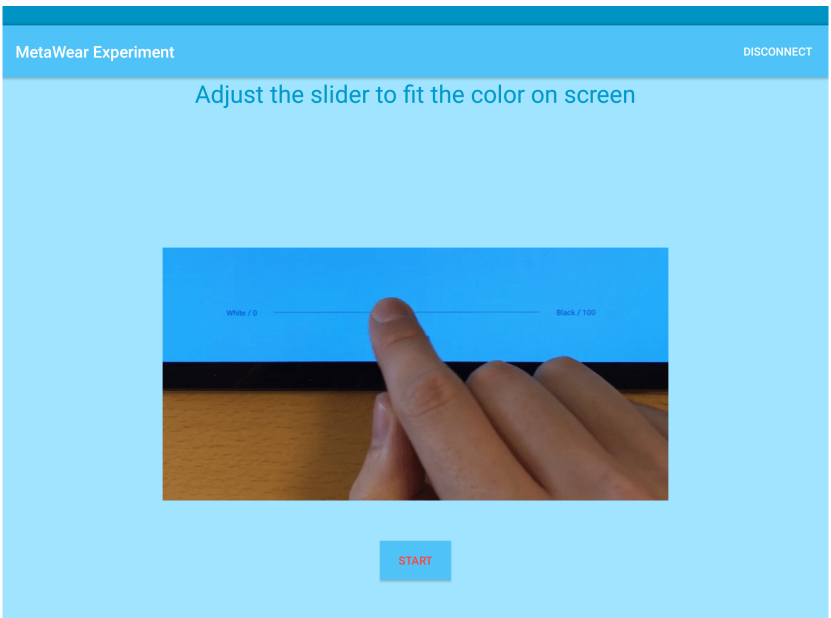

Figure E.7: slider_grey explanation screen

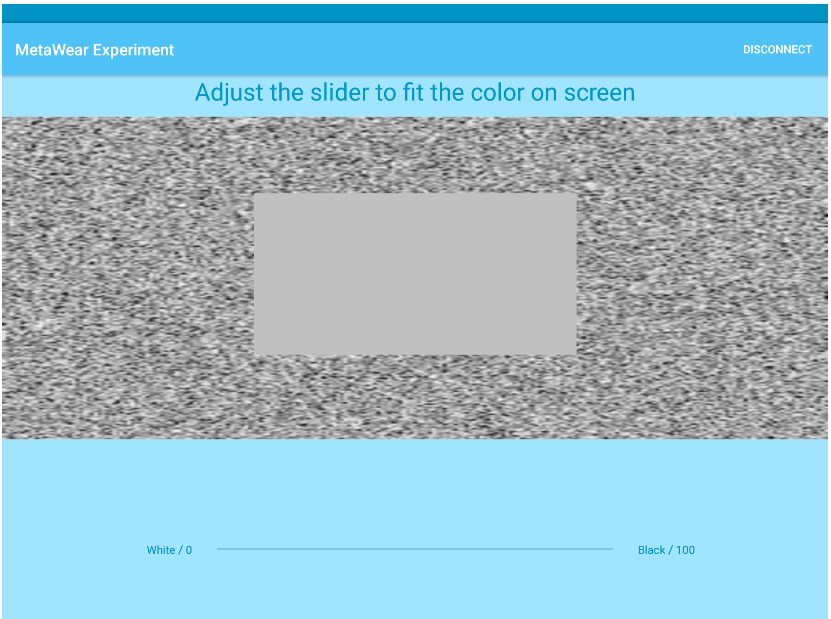

Figure E.8: slider_grey exercise screen

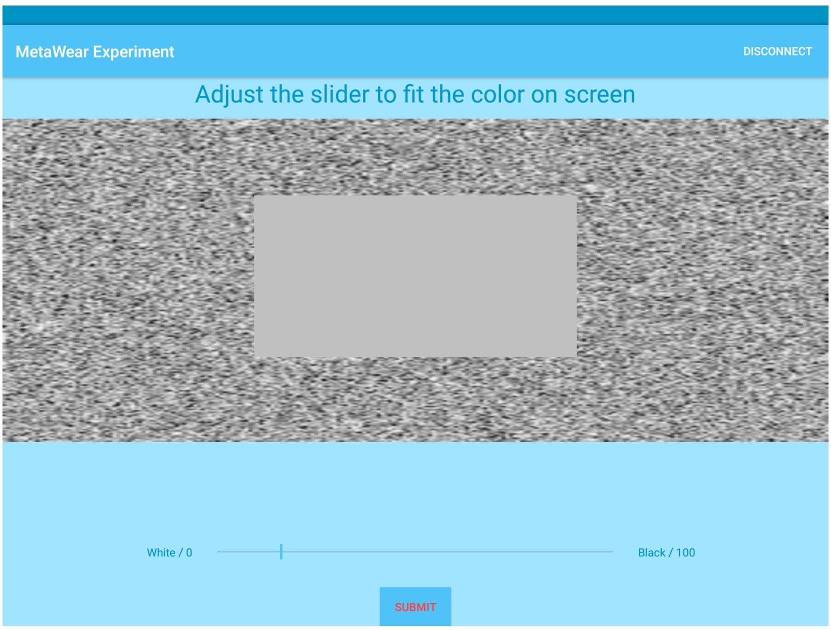

Figure E.9: slider_grey exercise screen

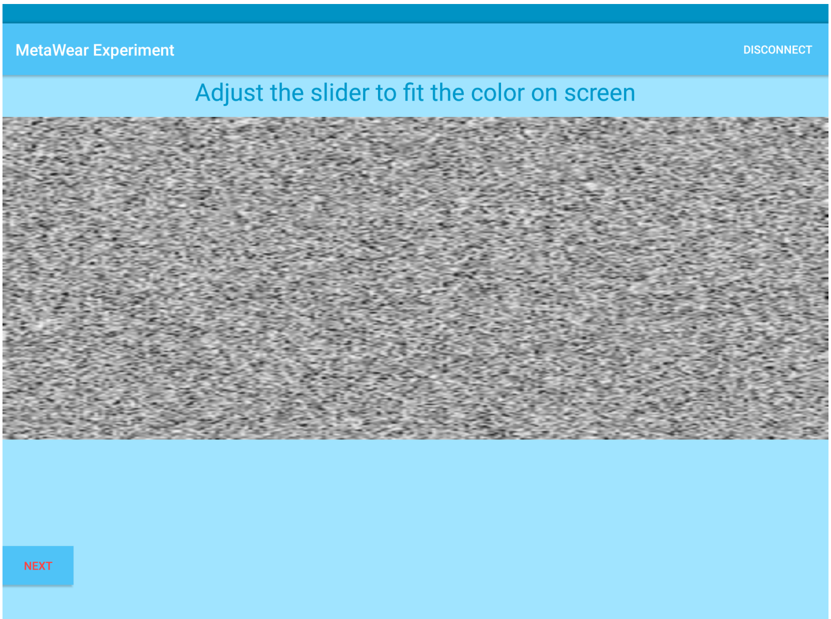

Figure E.10: slider_grey exercise screen

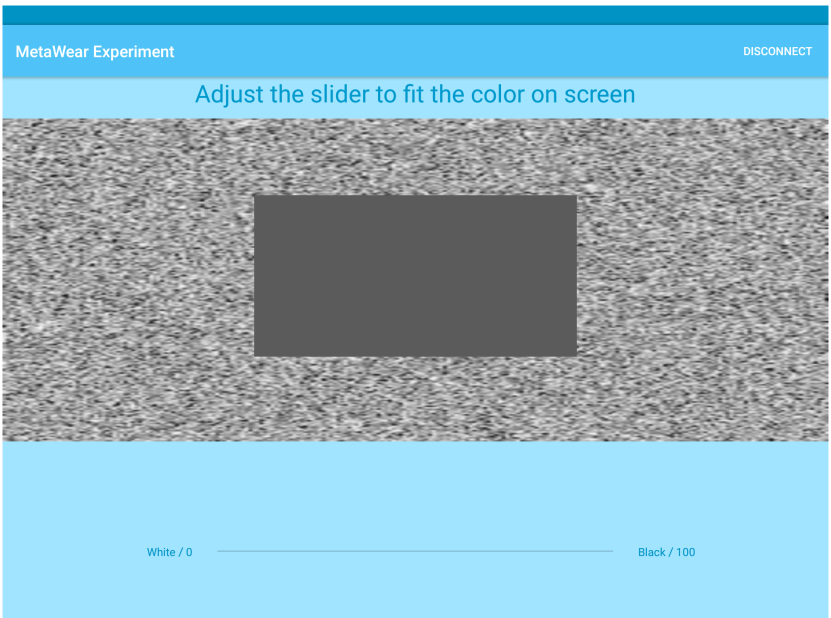

Figure E.11: slider_grey exercise screen

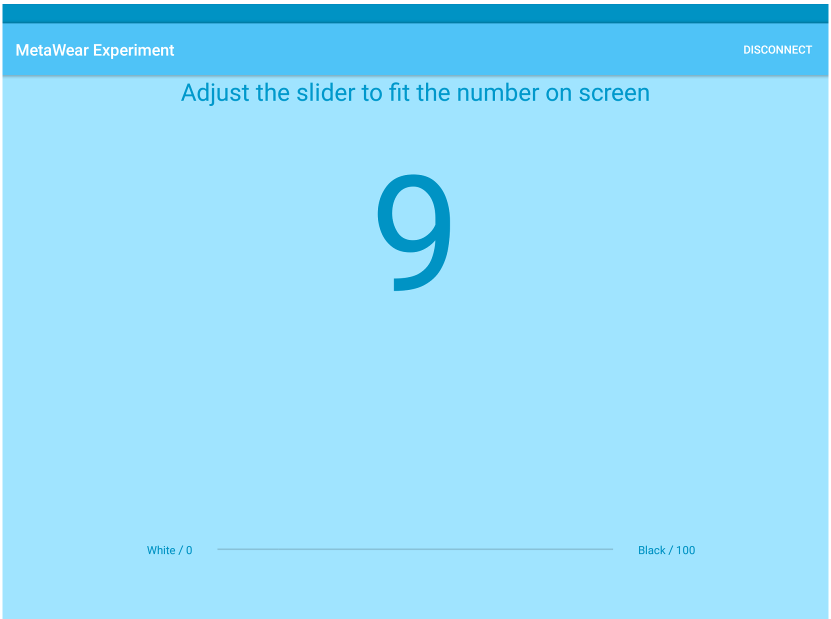

Figure E.12: slider_num exercise screen

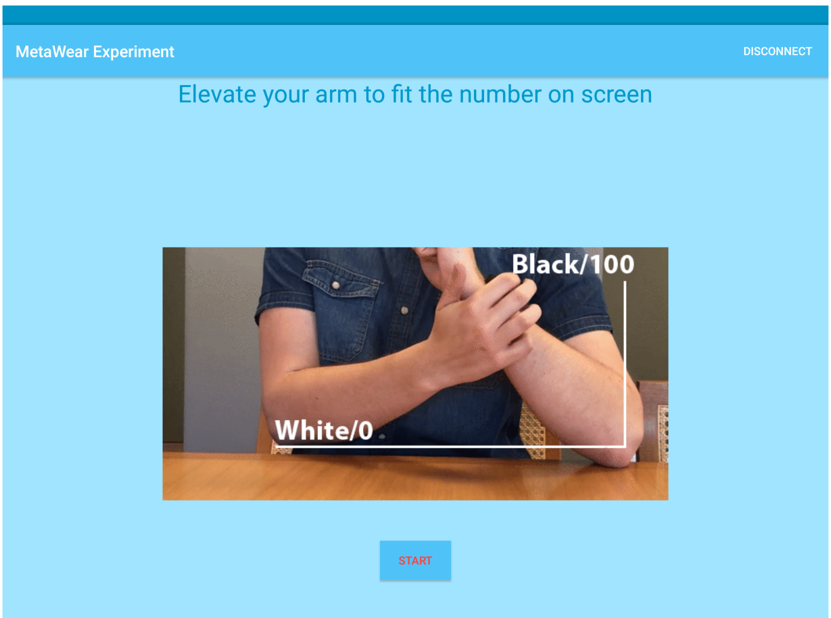

Figure E.13: arm_num explanation screen

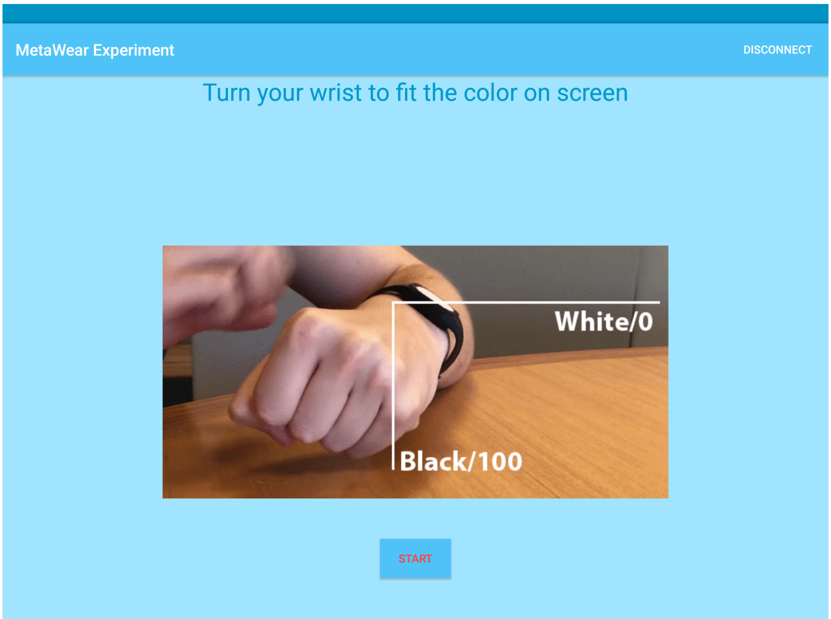

Figure E.14: wrist_grey explanation screen

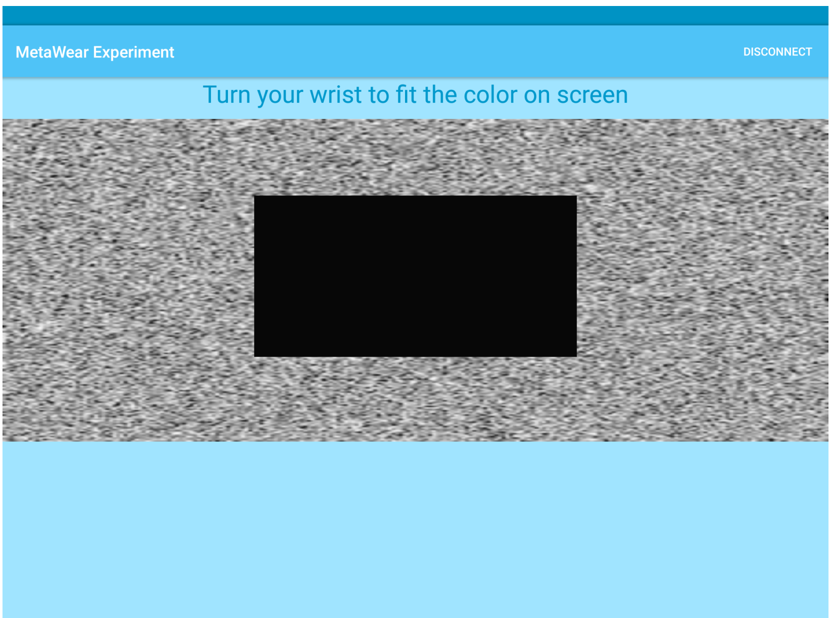

Figure E.15: wrist_grey exercise screen

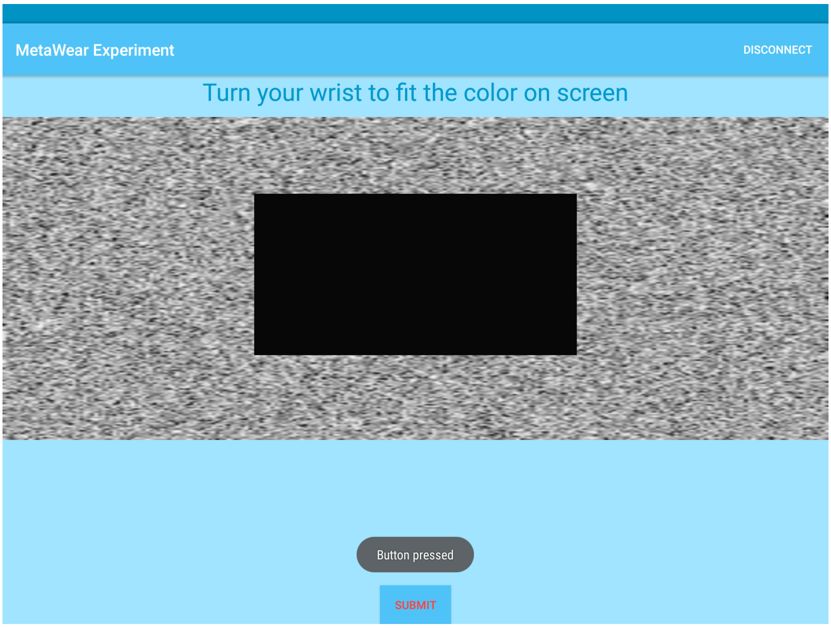

Figure E.16: wrist_grey exercise screen

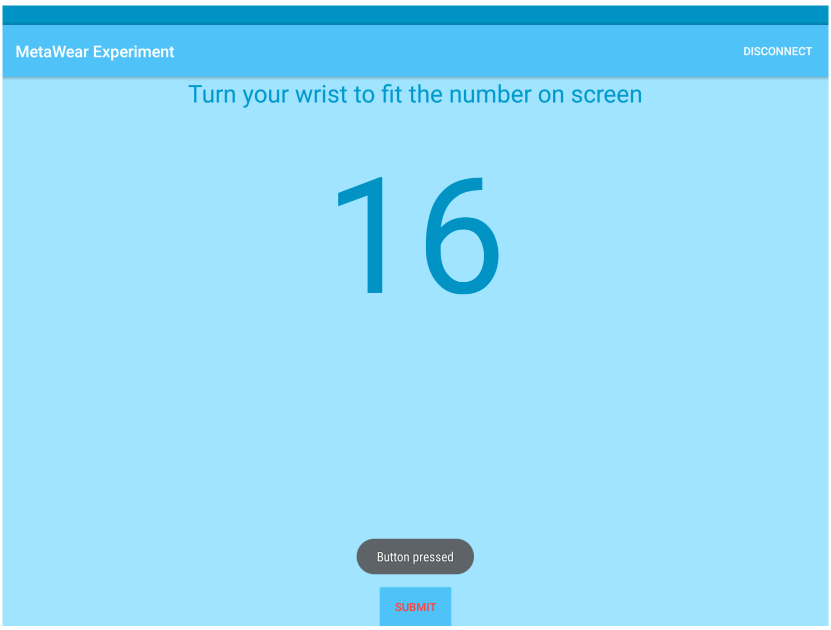

Figure E.17: wrist_num exercise screen

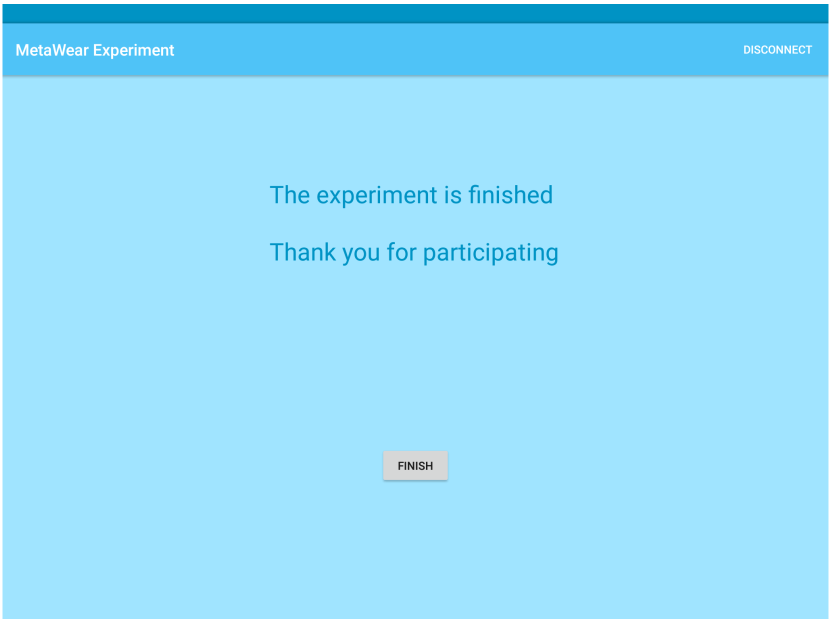

Figure E.18: Experiment finish screen

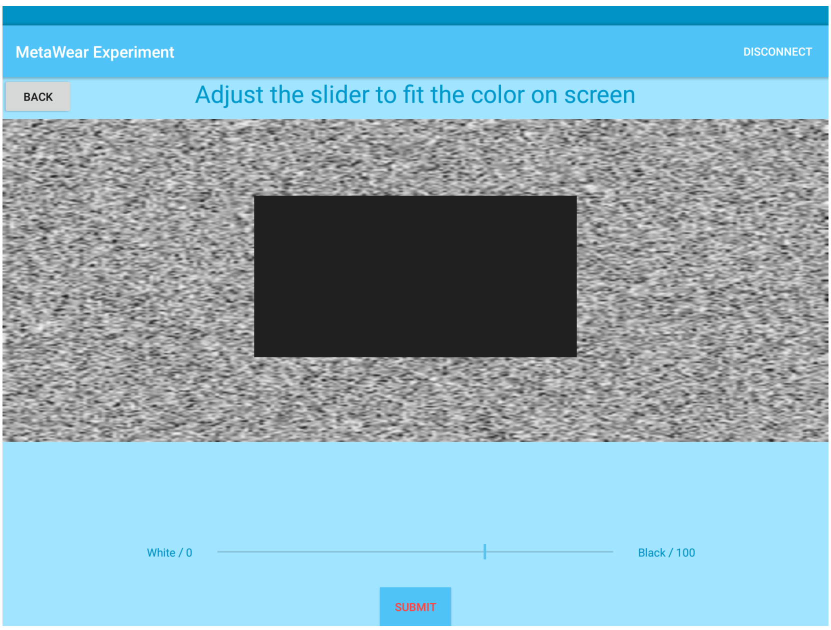

Figure E.19: slider_grey exercise screen

Bibliography

[1] Pitch, roll and yaw of an airplane, non edited figure. https://pic2.zhimg.com/2f32b07106e9c111da384b1afb7f2f12_r.jpg, 2018. [Online; accessed 18-June-2018].

[2] VAS Pain Scale Rulers 0-10 cm w/ Slider. http://www.custompromotionalrulers.com/visual-analog-scale-vas-rulers/vas-pain-scale-rulers-0-10-cm-w/slider/, 2018. [Online; accessed 16-June-2018].

[3] Adafruit. Adafruit 9-dof absolute orientation imu fusion breakout - bno055. https://www.adafruit.com/product/2472. [Online; accessed 18-June-2018].

[4] Adafruit. Adafruit feather m0 adalogger. https://www.adafruit.com/product/2796. [Online; accessed 18-June-2018].

[5] Adafruit. Flora wearable bluefruit le module. https://www.adafruit.com/product/2487. [Online; accessed 18-June-2018].

[6] Amazon. Amazon Mechanical Turk. https://www.mturk.com/, 2018. [Online; accessed 16-June-2018].

[7] Anxiety and Depression Association of America. Facts statistics. https://adaa.org/about-adaa/press-room/facts-statistics. [Online; accessed 2-July-2018].

[8] Anxiety and Depression Association of America. Panic disorder. https://adaa.org/understanding-anxiety/panic-disorder. [Online; accessed 2-July-2018].

[9] Bangalore Aviation. The typical takeoff and climb angles of all boeing planes. `https://www.bangaloreaviation.com/2009/05/typical-takeoff-and-climb-angles-of-all.html`. [Online; accessed 18-June-2018].

[10] Anders Beck. Mood Tracker. `http://science.beckup.dk/Mood%20Tracker%20-%20Anders%20Beck%20s134540.pdf`, 2017. [Online; accessed 18-June-2018].

[11] bonetalks.com. Broken forearm. `http://www.bonetalks.com/elbowforearmrad/`. [Online; accessed 24-June-2018].

[12] Gunnar Borg and Elisabet Borg. A general psychophysical scale of blackness and its possibilities as a test of rating behaviour. 1991.

[13] Bosch. Bosch sensortec. `https://www.bosch-sensortec.com/bst/products/all_products/bno055`. [Online; accessed 2-July-2018].

[14] U. Bosch and C. H. Stapfer. The cielab color space revisited: Surprise! *Spe Cad Retec 2004, Society of Plastic Engineers, Color and Appearance Division, Regional Technical Conference: Hot Colors - Cool Plastics - Conference Proceedings*, pages 71–77, 2004.

[15] Rebecca C. Burch, Stephen Loder, Elizabeth Loder, and Todd A. Smitherman. The prevalence and burden of migraine and severe headache in the united states: Updated statistics from government health surveillance studies. *Headache*, 55(1):21–34, 2015.

[16] Kristian Ohrt Dam-Jensen. Exploring interaction methods for self- logging with a wearable smartbutton, undersøgelse af interaktionsmetoder til selvregistrering med en 'wearable smartbutton', 2018.

[17] Android Developers. Material Design for Android. `https://developer.android.com/guide/topics/ui/look-and-feel/`, 2018. [Online; accessed 16-June-2018].

[18] S Grant, T Aitchison, E Henderson, J Christie, S Zare, J McMurray, and H Dargie. A comparison of the reproducibility and the sensitivity to change of visual analogue scales, borg scales, and likert scales in normal subjects during submaximal exercise. *Chest*, 116(5):1208–1217, 1999.

[19] Arthur S. Hathaway. Quaternion space. *Transactions of the American Mathematical Society*, 3(1):46–46, 1902.

[20] Thomasz Rafał Kamiński. A wearable system for acquiring data on subjective experiences, 2016.

[21] Jakob Eg Larsen, Kasper Eskelund, and Thomas Blomseth Christiansen. Active self-tracking of subjective experience with a one-button wearable: A case study in military ptsd. *Proceedings of the 2nd Computing and Mental Health Workshop at Acm Chi 2017*, 2017.

[22] Reed Larson and Mihaly Csikszentmihalyi. The experience sampling method. *Flow and the Foundations of Positive Psychology: the Collected Works of Mihaly Csikszentmihalyi*, pages 21–34, 2014.

[23] Justin Matejka, Michael Glueck, Tovi Grossman, and George Fitzmaurice. The effect of visual appearance on the performance of continuous sliders and visual analogue scales. *34th Annual Chi Conference on Human Factors in Computing Systems, Chi 2016*, pages 5421–5432, 2016.

[24] mbientlab. Android starter app. `https://github.com/mbientlab/MetaWear-Tutorial-Android`. [Online; accessed 18-June-2018].

[25] mbientlab. mbientlab apis. `https://mbientlab.com/developers/`. [Online; accessed 18-June-2018].

[26] mbientlab. Metamotionr developer kit. `https://mbientlab.com/product/metamotionr-dev-kit/`. [Online; accessed 18-June-2018].

[27] Operative Neurosurgery. Visual analog scale (VAS). `http://operativeneurosurgery.com/doku.php?id=visual_analog_scale`, 2017. [Online; accessed 18-June-2018].

[28] National Institutes of Health. Nih analysis shows americans are in pain. `https://www.nih.gov/news-events/news-releases/nih-analysis-shows-americans-are-pain`. [Online; accessed 2-July-2018].

[29] Aditya Ponnada, Caitlin Haynes, Dharam Maniar, Justin Manjourides, and Stephen Intille. Microinteraction ecological momentary assessment response rates. *Proceedings of the Acm on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):1–16, 2017.

[30] Dobrila Rancic Moogk. Minimum viable product and the importance of experimentation in technology startups. *Technology Innovation Management Review*, 2012.

[31] Ulf-Dietrich Reips and Frederik Funke. Interval-level measurement with visual analogue scales in internet-based research: Vas generator. *Behavior Research Methods*, 40(3):699–704, 2008.

[32] Aleix Rull and Federico Thomas. On generalized euler angles. *Mechanisms and Machine Science*, 24:61–68, 2015.

[33] Samsung. Samsung galaxy tab s2. `https://www.samsung.com/global/galaxy/galaxy-tab-s2/`. [Online; accessed 18-June-2018].

[34] PTSD United. Ptsd statistics. `http://www.ptsdunited.org/ptsd-statistics-2/`. [Online; accessed 2-July-2018].

[35] M. Van Den Brink, E. N G Bandell-Hoekstra, and H. Huijer Abu-Saad. The occurrence of recall bias in pediatric headache: A comparison of questionnaire and diary data. *Headache*, 41(1):11–20, 2001.

[36] Gergely Vass. Avoiding gimbal lock. *Computer Graphics World*, 32(6):10–11, 2009.