Examensarbete

Eye Movement Event Detection for Wearable Eye Trackers

Akdas Hossain, Emma Miléus

LiTH-MAT-EX–2016/02–SE

Eye Movement Event Detection for Wearable Eye Trackers

Applied Mathematics, MAI, Linköpings Universitet

Akdas Hossain, Emma Miléus

LiTH-MAT-EX–2016/02–SE

Examensarbete: **30 hp**

Level: **A**

Supervisor: **Jonas Högström and Tobias Lindgren**,
Tobii Pro

Examiner: **Fredrik Berntsson**,
Applied Mathematics, MAI, Linköpings Universitet

Linköping: **June 2016**

Abstract

Eye tracking research is a growing area and the fields as where eye tracking could be used in research are large. To understand the eye tracking data different filters are used to classify the measured eye movements. To get accurate classification this thesis has investigated the possibility to measure both head movements and eye movements in order to improve the estimated gaze point.

The thesis investigates the difference in using head movement compensation with a velocity based filter, I-VT filter, to using the same filter without head movement compensation. Further on different velocity thresholds are tested to find where the performance of the filter is the best. The study is made with a mobile eye tracker, where this problem exist since you have no absolute frame of reference as opposed to when using remote eye trackers. The head movement compensation shows promising results with higher precision overall.

Keywords: Mobile Eye Tracking, I-VT Filter, MEMS Gyroscope

URL for electronic version:

http:// ????

Acknowledgements

We would like to express our gratitude to Tobii Pro who has been very welcoming showed big interest in our thesis. A special thanks to our supervisors Jonas Högström and Tobias Lindgren who we have had many long discussions with and guided us throughout the entire thesis.

Finally we would like to thank our examiner Fredrik Berntson for giving us valuable comments on the work and report.

Stockholm, 2016
Akdas Hossain and Emma Miléus

Contents

Chapter 1

Introduction

This thesis validates the possibility to improve eye movement event detection when using mobile eye trackers by compensating for head movements with gyroscope data.

1.1 Background

The first eye tracking research was done in late 1940s and eye tracking is today used in a wide range of applications. For example, in medical research eye tracking helps disabled people increase their independence and possibilities to communicate by using only their eyes[14]. Another growing area within eye tracking is gaming where some games now uses eye tracking[13]. These applications often use a remote eye tracker, i.e. eye trackers put on a computer screen. Mobile eye trackers are often used in market research[15] where the researcher wants to know where a customer directs his attention to products or advertising. Some other applications are clinical research, sports research, child research and for educational purposes.

To understand and visualize the data streams from a recording you need to filter and classify the data. There are different methods to do this, described in detail in Chapter 3, but they all have in common that they want to identify what kind of eye movements that was happening during the recording in order to identify where the gaze point was. There are three types of eye movements a person can do, fixations, saccades and smooth pursuits.

A saccade is a rapid eye movement, i.e. you change from looking at one object to another [3]. The velocity and duration of a saccade depends on which type of movement the person is doing. A short saccade can be when a person is reading and a long saccade when a person is looking around on a scene. For a velocity based filter, which is used in this thesis, a saccade is defined as when the velocity is above a certain threshold, see Figure 1.1.

Fixations are when a person keeps his eyes fixated on one point. In reality a person cannot fixate the eyes on one exact point over time due to constant

tremor, drift and micro-saccades [3]. These eye movements are however very small and results in small velocities. For a velocity based filter fixations are when the velocity is below the specified threshold, see Figure 1.1. Smooth pursuits occurs when a person fixates the gaze on a moving object.

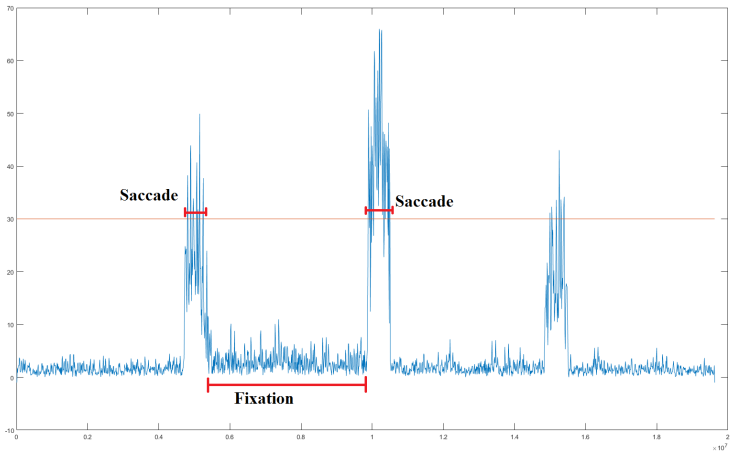


Figure 1.1: The definition of a saccade and fixation with a velocity based I-VT filter.

1.2 Tobii Pro

This thesis is performed in collaboration with Tobii Pro, which is a unit within the Tobii Group, that address's mainly researchers. Tobii is a market leading company within eye tracking located in Danderyd, Sweden and has during this thesis provided a desk, a computer with suitable software and necessary hardware to perform the study.

Tobii Pro suggested the thesis work after discovering that their velocity based fixation filter, which is an I-VT filter, did not perform as good on their mobile eye tracker as it did for their remote eye trackers when the recordings involved a lot of head movements since the velocities would be higher and many fixations would not be detected by the I-VT filter. The current solution for detecting this kind of fixations is making the velocity threshold higher on the mobile eye tracker. The remote eye trackers use a velocity threshold of 30 $°/s$ and the mobile eye trackers a velocity threshold of 100 $°/s$. By choosing a high velocity threshold the gaze point is allowed to change more within a classified fixation but this has on the other hand led to that the filter cannot detect short saccades.

1.3 Outline

The aim of this thesis was to improve the eye movement event detection filter for Tobii's mobile eye tracker by compensating for head movements using micro electrical mechanical systems (MEMS) data, i.e. gyroscope and accelerometer. The thesis also aimed to make it possible to lower the velocity threshold and still maintain the current precision by the I-VT filter. Further on an alternative, more straight forward, velocity calculation is proposed and compared to the existing velocity calculation. Mobile eye trackers are often suitable for researchers since they want to restrain the test subjects as little as possible. Mobile eye trackers allow the test subject to move around freely in, for example, a store.

One major difference between mobile and remote eye trackers is that mobile eye trackers measures the gaze direction relative to the head, giving no fix frame of reference in the room, while remote eye trackers always has a fix frame of reference since they are stationary. Existing eye movement event detection algorithms has difficulties to distinguish eye movements from head movements when the frame of reference moves which translates to incorrect classification of the eye movements by the I-VT filter. The implementation will be evaluated by recording four common types of head- and eye movements.

- **Fixations with Head Movements**
  Is the algorithm better of detecting fixations where the gaze is fixed on a target but the head moves around? With a mobile eye tracker the velocities can become high even when fixating on a target due to head movements. With the head movement compensation the algorithm should return lower velocities during a fixation with head movement.
- **Beginning and End of Fixations and Saccades**
  How well does the the algorithm detect the end and start of fixations and saccades? A saccade can start with a head movement before the eye actually stops fixating at one point and the head movement will then finish before the eyes have reached the new fixation point.
- **No Head Movements**
  A person will move the head slightly even when they do not intend to move the head. Is there are a difference for the new algorithm compared to the old when there are none or very small head movements?
- **Reading**
  Can the algorithm detect reading saccades and fixations? When reading the duration of fixations and saccades are short and the velocities of saccades quite low. With a high threshold these saccades can easily be undetected leading to inaccurate classification and inaccurate calculation of fixation point.

Smooth pursuits are not detected by the current fixation filter and will not be a priority in this thesis.

1.4 Disposition of the report

The outline of this report is as follows:

- In Chapter 1 an introduction of the area is given with background of the project
- In Chapter 2 the used hardware and technique is described
- In Chapter 3 the theory and related work is presented
- In Chapter 4 the implementation is described
- In Chapter 5 is the outcome of the project presented
- In Chapter 6 future work is discussed
- In Chapter 7 the conclusions of the project is presented

Chapter 2

Hardware and Data Recording

Recordings are made with Tobii's mobile eye tracker "Tobii Glasses 2", hereafter referred to as "glasses", see Figure 2.1. The glasses come together with a battery pack and different nose pads to fit all people.

Figure 2.1: The Tobii Glasses 2, from [15].

The glasses uses dark pupil tracking [15]. Dark pupil tracking is when the pupil appear dark by using an illuminator that is not near the optical axis of the imaging device, opposed to bright pupil tracking when the illuminator is close to the optical axis, see figure 2.2.

To measure the direction of the eye, four near infra-red cameras are used to identify the pupil and the reflections from the illuminators on the eye surface, so called glints. If the cameras can not find a glint no value will be recorded at that time. Failure to find a glint can occur for example during a blink or when the user is looking too far to one side causing the glint to end up outside of the cornea on the scalera.

Before each recording a calibration is needed to get good accuracy in the measurements. This is because the exact placement of the fovea in the eye is different for each person. The fovea is located near the center of the eye, see Figure 2.3, and the offset from the center is constant, the constant is different for each person. During the calibration procedure the person looks at the center of the

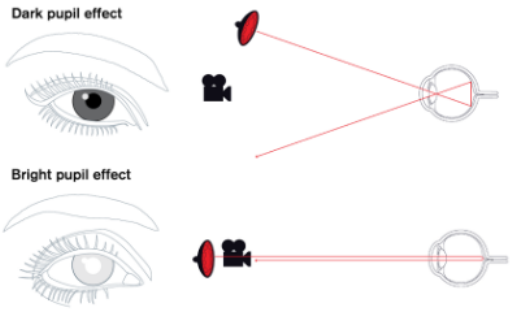


Figure 2.2: Dark pupil tracking vs bright pupil tracking below, from [15].

calibration marker. The calibration marker is a circle with a diameter of 46 mm.

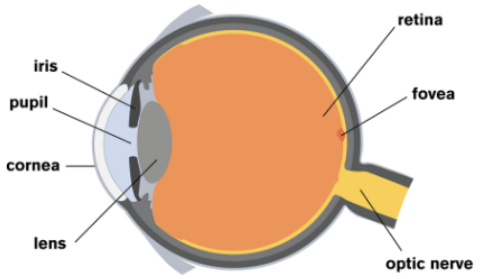


Figure 2.3: The structure of the eye, from [15].

The glasses has a built-in gyroscope of model L3GD20 and an accelerometer of model LIS3DH. They sample in 95Hz and 100Hz respectively. The gyroscope measures rotational velocities around the x-, y- and z-axis and the accelerometer measures proper acceleration, "g-force", in x-, y- and z-axis. Further technical information of the glasses, accelerometer and gyroscope is found in appendix C.

During recordings Tobii's own software "Tobii Pro Glasses Controller" is used, hereafter referred to as "Controller". To use the glasses you connect them to the computer wirelessly using the built-in WLAN. The calibration and recording is then performed in Controller.

The recordings and data streams can be replayed and visualised using "Tobii Pro Glasses Analyzer", hereafter referred to as "Analyzer". Analyzer was used to export the raw data to a tab-delimited file. The implementation is done completely in MATLAB. A list of the raw data sets that was exported from Analyzer is found in appendix A. A plot of how the gyroscope data can look after export and filtering can be found in appendix D. The parameter settings

of the I-VT filter is found in appendix B.

An old version of Tobii's former software "Tobii Studio" was given in MATLAB and was modified to work more or less as Analyzer.

Chapter 3

Theory and Related Work

In this thesis a velocity based I-VT filter is used. Other common fixation filters are I-DT, I-HMM and I-AOI. I-DT is a dispersion filter, using the measurement of dispersion or spread distance to classify fixations and saccades. The user must specify the dispersion threshold and the cluster size or the duration threshold. Due to fixations generally lasting at least 100ms the duration threshold is set to values between 100 and 200 ms [1]. The I-DT takes all the points within the specified duration threshold, starting from the first point, and calculates the dispersion, if the dispersion for these points are lower than the threshold more points are added until the dispersion value exceeds the dispersion threshold. When exceeded, a fixation is recorded at the centroid of the points and the points are excluded from future calculations. Figure 3.1 shows the steps of the I-DT.

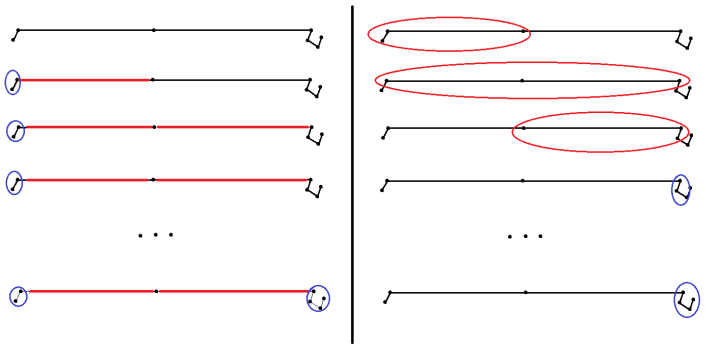


Figure 3.1: Left: example of the I-VT algorithm. Right: example of the I-DT algorithm.

Hidden markov model filter, I-HMM, is a filter based on probability [9]. Through probability it finds the most likely identification for a given protocol. An obser-

vation probability shows the expected velocity within the state and a transition probability how likely it is to stay in the present state and how likely it is to change state, see Figure 3.2. The saccade has a distribution around high velocities and fixations a distribution for lower velocities. The parameter estimation is complex and chosen with a method called reestimation. Reestimation learns the probabilistic values by training on given sets.

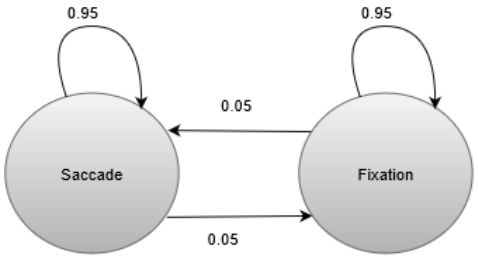


Figure 3.2: An example of the probability states of an I-HMM model.

Area of interest filter, I-AOI, is not a real fixation filter due to the fact that it does not find all fixations within a data set [1]. The algorithm takes a predefined rectangular window and defines all data points within the window as fixation points and the rest of the data set as saccades. Within the window, all consecutive fixation points redefines as one single fixation. If the fixations doesn't span longer than the defined minimum duration threshold the fixation is removed. According to [1] the results of using this method is similar to the I-DT filters result, using the same data set. Although to get good result with this method the window must be chosen carefully, otherwise fixations will not be detected.

The head movement compensation can be done in many different ways. Most articles suggests adding a IMU, i.e. gyroscope, accelerometer and magnetometer, to the eye tracker, as in [6]. The magnetometer gives an absolute orientation in the coordinate system and and has shown good results. The Direction Cosine Matrix method[2] establish an absolute coordinate system using accelerometers and gyroscopes. To reduce risk of drift and noise an Proportional Integral(PI) controller is used based on The Good Gain method for PI controller tuning. The Good Gain Method is an experimental method for tuning PI controllers. One benefit of using this method is that you do not need any prior knowledge about the process model [4].

Reducing noise can also be done with a Kalman filter. One way is to model a virtual gyroscope which includes the two most common errors of a gyroscope, i.e. angular random walk (ARW) and rate random walk (RRW),[10]. The Allan variance method is used to quantify the noise terms. It is simple to implement and calculates the variance by dividing the samples in clusters and calculating an averaging factor. Another way is to model the gyroscope as a first order Markov process [7] if knowledge about the process model exists. A statistical error model is used and the acceleration is modelled as a low pass filter.

3.1 Velocity calculation

The angular velocity can be calculated in different ways. The already implemented angular calculation, introduced by Tobii, is done by using the law of cosines. To calculate the velocity for sample t, the angle is calculated between the samples before and after sample t, t_1 and t_2. The velocity is then given by dividing the angle between t_1 and t_2 with the time between the two samples. To do this the vectors a, b and c are calculated as

$$a(t) = GazePosition3D(t1) - EyePosition3D(t1), \tag{3.1}$$
$$b(t) = GazePosition3D(t2) - EyePosition3D(t2), \tag{3.2}$$
$$c(t) = GazePosition3D(t2) - GazePoisition3D(t1). \tag{3.3}$$

The vectors a, b and c is visualised in Figure 3.3. The law of cosines gives the angle α according to

$$c^2 = a^2 + b^2 - 2abcos(\alpha) \ . \tag{3.4}$$

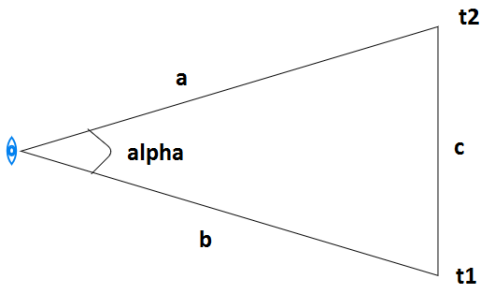


Figure 3.3: Angle calculation between 2 sample points, s(t_1) and s(t_2).

When the angle is known the velocity, $v(t)$, is easily calculated with the angle for that time since the sampling time is known,

$$v(t) = \frac{|\alpha(t)|}{|t_2 - t_1|} \ . \tag{3.5}$$

The velocities are then used by the I-VT filter to classify the sample points. If the velocity is higher than the specified threshold the point is defined as a saccade and if the velocity is lower it is a fixation.

3.2 Alternative Velocity Calculation

The alternative velocity calculation, proposed in this thesis, calculates velocity similarly as the method above in Chapter 3.1 but with one major difference. Instead of defining the vectors $a(t)$ and $b(t)$ as in equation 3.1-3.2 for a sample t the recorded data GazeDirection is used directly with the samples before and after t, t_1 and t_2, giving

$$a(t) = GazeDirection3D(t1) \tag{3.6}$$

and

$$b(t) = GazeDirection3D(t2). \tag{3.7}$$

The angle $\alpha(t)$ is calculated as

$$\alpha(t) = atan2d(|a(t) \times b(t)|, a(t) \cdot b(t)) \ . \tag{3.8}$$

This is a special command in MATLAB that calculates the four quadrant arctangent of the $a(t)$ and $b(t)$, given angles between $-\pi$ to π.

3.3 Head compensation

The head movement compensation is done with the gyroscope data. The gyroscope records the rotational velocities around x-, y- and z-axis, i.e. pitch, yaw and roll respectively, see Figure 3.4.

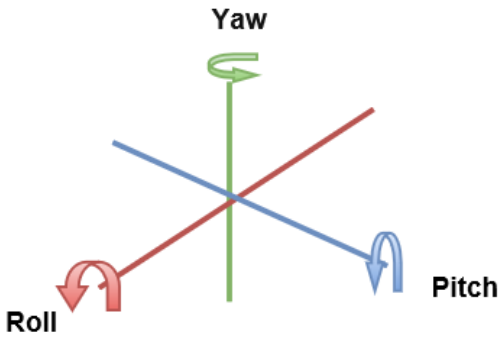


Figure 3.4: Definition of pitch, yaw and roll.

The rotational velocities found in the vectors $GyroX(t)$, $GyroY(t)$ and $GyroZ(t)$ can be integrated over time which will give corresponding Euler angles

$$\phi = \int_t GyroX(t)\mathrm{d}t, \tag{3.9}$$

$$\beta = \int_t GyroY(t)\mathrm{d}t, \tag{3.10}$$

and

$$\gamma = \int_t GyroZ(t)\mathrm{d}t, \tag{3.11}$$

where ϕ corresponds to the pitch rotation, β corresponds to yaw rotation, γ corresponds to roll rotation. With the Euler angles a rotational matrix, $\mathbf{R}$, can be calculated as

$$\mathbf{R} = R(\phi)R(\beta)R(\gamma), \tag{3.12}$$

where

$$R(\phi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & cos(\phi) & -sin(\phi) \\ 0 & sin(\phi) & cos(\phi) \end{pmatrix}, \tag{3.13}$$

$$R(\beta) = \begin{pmatrix} cos(\beta) & 0 & sin(\beta) \\ 0 & 1 & 0 \\ -sin(\beta) & 0 & cos(\beta) \end{pmatrix}, \tag{3.14}$$

and

$$R(\gamma) = \begin{pmatrix} cos(\gamma) & -sin(\gamma) & 0 \\ sin(\gamma) & cos(\gamma) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3.15}$$

Using $\mathbf{R}$, the vector $\alpha(t)$ then rotated in accordance with how much the head has moved between $t1$ and $t2$,

$$\alpha_{hc}(t) = \mathbf{R}\alpha(t), \tag{3.16}$$

where a_{hc} is the head compensated visual angle vector. Using $\alpha_{hc}(t)$ new velocities are calculated again using (3.4) and (3.5).

Chapter 4

Implementation

A schematic overview of full system is found in appendix E. The files "Import file" and "Time split" are functions for converting the tsv-file into readable format for MATLAB giving the struct ET with all variables in appendix A.

The gyroscope has a noticeable offset and after trying several devices the detected offset was shown to be unique for each device. Generally, the gyroscope gives quick and stable response but is prone to drift over time, meanwhile an accelerometer are more prone to noise but stable over time. Therefore both the gyroscope and accelerometer data was used to reduce noise and drift for the gyroscope using a Kalman filter [10].

A calibration is required to determine the glasses default values. The default values for the glasses used in this thesis is found in appendix D. The offset is dependent on external conditions which means that if external conditions change drastically, such as temperature, the calibration should be redone. The calibration is done with a recording of the glasses lying still on a flat surface. The recording is preferable at least a few minutes long but even a short calibration of only a few seconds decreases the noise significantly.

The built-in gyroscope and accelerometer used has different sampling frequencies, 95 Hz and 100 Hz, and therefore is the gyroscope data linearly interpolated to 100 Hz. Afterwards, the median offset of all the elements in each of the three gyroscope signals are subtracted from each signal to center the signal around zero,

$$O_{median} = \begin{cases} x_{(n+1)/2}, & \text{if n is odd ,} \\ \frac{x_{n+1}}{2}, & \text{if n is even ,} \end{cases} \tag{4.1}$$

where O_{median} is the offset and x is a vector of the sorted gyroscope values in ascending order. A virtual gyroscope is commonly modeled as

$$Z(t) = H\omega(t) + m(t), \tag{4.2}$$

where $Z(t)$ is an array of the outputs of the gyroscope and the accelerometer, H

is a vector of ones, $\omega(t)$ is the true rate signal of the gyroscope and accelerometer and m is a vector of the estimated white noise [10].

The axis output of the gyroscope is assumed to have a constant cross-correlation ρ and therefore the covariance matrix R is,

$$R = q_n \begin{pmatrix} 1 & \rho & \rho & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho & \rho & \rho \\ \rho & \rho & 1 & \rho & \rho & \rho \\ \rho & \rho & \rho & 1 & \rho & \rho \\ \rho & \rho & \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & \rho & \rho & 1 \end{pmatrix}, \tag{4.3}$$

where q_n is the calculated Allan Variance. The steady state vector, $\dot{\hat{X}}$, can be computed by the continuous-time Kalman filter,

$$\dot{\hat{X}} = -\sqrt{Cq_\omega}\hat{X}(t) + \sqrt{q_\omega/C}H^T R^{-1} Z(t), \tag{4.4}$$

where $C = H^T R^{-1} H$, q_ω is a variance determined by the noise level of the gyroscope and T is the sampling period.

The output of the virtual gyroscope, $\hat{X}$, is obtained by discretization of the continuous Kalman filter with a zero-order approximation,

$$\hat{X}_{k+1} = e^{-\sqrt{Cq_\omega T}}\hat{X}_k + \frac{1}{C}(1 - e^{-\sqrt{Cq_\omega T}})H^T R^{-1} Z_{k+1}. \tag{4.5}$$

Another modification is required because the eye position is sometimes not found, due to a number of reasons such as blinking or looking too far to the sides. "Eye position fill in" assumes that the eye position does not vary a lot between different time stamps meaning that it will add the position of the last found eye position where there is data loss. If there is lost data in the beginning of the recording the first found eye position is assumed to have been the starting eye position.

To reduce noise in the recorded gaze point and gaze direction data a median filter of order 3 is performed on the gaze data if the norm of the gyroscope values for that timestamp exceeds 5°/s,

$$|| < GyroX, GyroY, GyroZ > || < 5\ . \tag{4.6}$$

The reason for this is that early observations indicated that there is more noise in the gaze data when the recording contained a lot of eye- and head movements. To rectify this a median filter was chosen to filter the signal during head movements, the median filter was chosen so that the filtering would not

ruin the characteristics of the signal but still provide a cleaner, smoother signal than before filtering.

After the data has been filtered the data is ready to be sent into the I-VT filter. The first thing the I-VT filter does is gap fill in and eye position fill in. "Gap fill in" recovers occasional loss of data if the tracking should have failed for only a short moment. It linearly interpolates missing data from valid data in the neighbourhood of the missing data [8]. "Eye selection" is used by the original velocity calculation and takes the average of the two eyes and gaze positions to use in the velocity calculation [8].

The next steps are to calculate the visual angle and velocity, as previously mentioned the visual angles are calculated in two ways. The visual angles are calculated according to equations 3.1- 3.4 and 3.6- 3.8. It is here, after the calculation of visual angle that the values are compensated for head movements,

$$\alpha_{hc}(t) = \mathcal{F}\alpha(t),$$

giving the vector y of head compensated angles, where $\mathcal{F}$ is the *Head Compensation* described in Chapter 3.3 and $\alpha(t)$ are the visual angles prior head compensation. The velocities are then calculated using using 3.5.

After the velocity has been calculated each sample point are defined as either saccade, fixation or invalid by "classification",

$$EventType = \mathcal{A}v(t),$$

where $v(t)$ are the calculated velocities for each time stamp, $\mathcal{A}$ is the classification function and *EventType* is a vector containing 2,1 or -1 for saccade, fixation and invalid data respectively.

Afterwards, "merge adjacent fixations" merges fixations located close to each other in time and space,[8]. The angle between two adjacent fixations is calculated by taking the fixation points of the fixations and compensate with the head movement, in accordance with Chapther 3.3, from the middle of the first fixation to the middle of the second fixation. The middle was chosen in order to find a representative head movement during the fixations and is discussed further in Chapter 6.

If a fixation does not belong to a set of consecutive fixation points with the minimum length of 60 ms is the fixation discarded by "discard short fixations". 60 ms is the threshold because fixations that short are not meaningful studying user behaviour due to the processing time between eye and brain[8].

Chapter 5

Results and Discussion

The methods have been evaluated with the results from [5] as base. The following formulas for precisions are used,

$$PrecisionF = \frac{TF}{TF + FF}\,, \tag{5.1}$$

$$PrecisionS = \frac{TS}{TS + FS}\,, \tag{5.2}$$

where TF are the true fixations, FF are the false fixations, TS are the true saccades and FS are the false saccades. $PrecisionF$, as seen in Figure 5.1 is measuring the precision of fixations. In other words, $PrecisionF$ is a measurement of how many gaze points were correctly classified as fixations by the method and $PrecisionS$ the measurement of how many gaze points were correctly classified as saccades. This evaluation methodology only regards fixations and saccades, in reality in a recording there will be some data loss and the I-VT filter may also return an unknown eye movement. These points are not represented in the evaluation methodology.

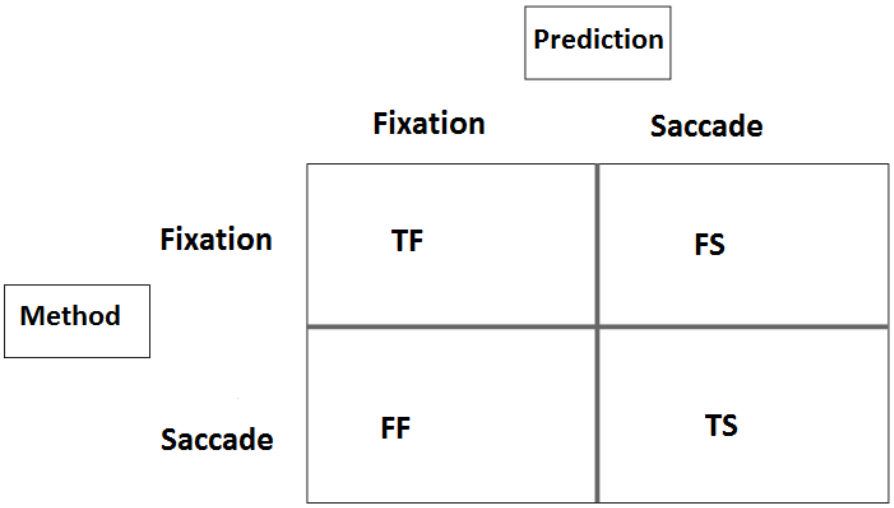


Figure 5.1: The four different labels used when calculating. A true fixation/saccade is a fixation/saccade which is both predicted to be one and also been classified as one by the method. A false fixation/saccade are points where the method wrongly classifies them as such.

Constructing the prediction, or gold standard, of how the method should correctly classify the data proved to be difficult as the recordings needed for this thesis requires natural head movements. In this thesis the test person received a description of the task but was not limited to how or when to perform it in order to encourage natural head movements. The one restriction was that the person had to look within a specified area during the whole recording. A snapshot of the area was taken and the gaze was manually mapped onto the snapshot, see Figure 5.3.

To make it easier to create the set of data which would serve as the prediction the videos were created where it is always known where the test person is trying to look. This is done by recording 3 videos where the test person look at different coloured dots on a white board in different ways, the idea is that by telling the test person to only look at the dots it will be easier to keep track on where the person is supposedly looking which is valuable information when creating a gold standard. The white board with coloured dots can be seen in Figure 5.2. An additional video of the test person reading a text was recorded to see how the algorithm fared in finding reading fixations/saccades. The test person who recorded the videos was familiar with eye tracking and the quality of the recordings are high, meaning that there is little to no loss of data during the recording.

The following is an example of how the task was described to the test person.

> *"Try to always keep your eyes fixated on any of the five coloured dots. Feel free to move between the dots as you wish but try to make sure that when you fixate your gaze you do it on one of the coloured dots."*

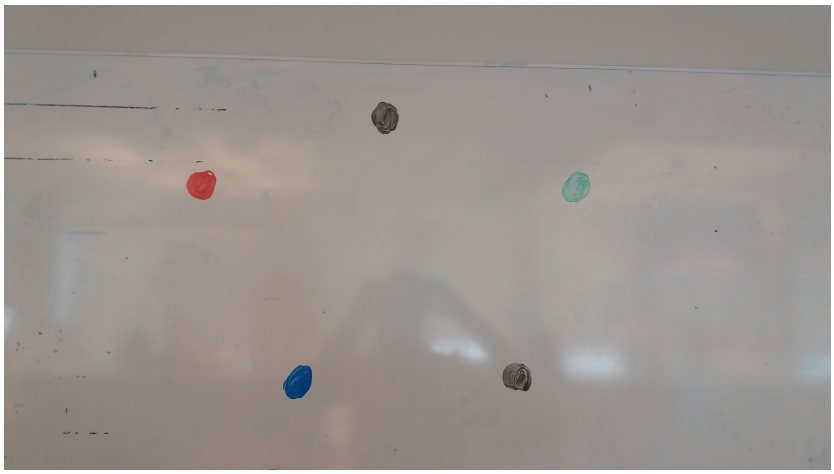
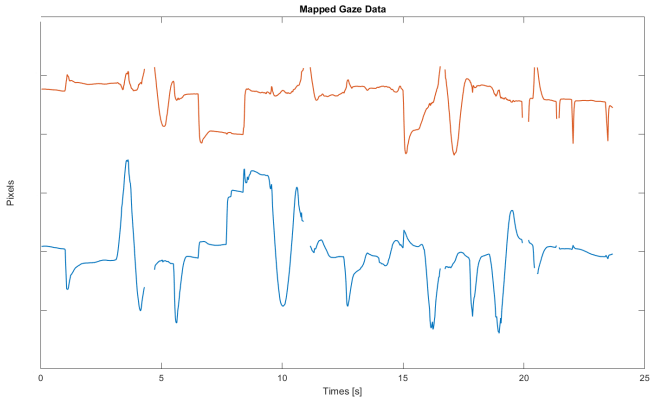


Figure 5.2: The white board with dots the test subject was asked to look at.

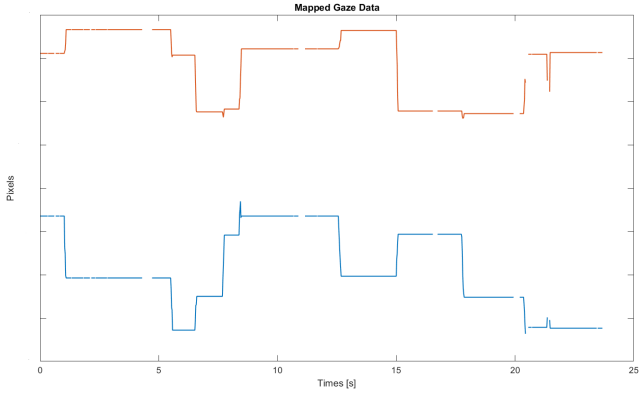
For each fixation a fixation point is calculated, which is the mean of all sample points in the fixation. If the fixation is correctly classified, the fixation point will be the place where the person was looking and the variance of the gaze points should only be artifacts from noise and tremor in the eyes. To evaluate how close the calculated fixation point is to the mapped gaze data, a histogram is used to illustrate the difference in pixels. It only takes points defined as TF in account. A big pixel difference indicates that saccades where missed.

In this chapter and in the figures below the names "method 1", "method 2" and method 3" are used, below is a description of the three methods:

- method 1 = original velocity calculation without head compensation, i.e. the starting point
- method 2 = original velocity calculation with head compensation
- method 3 = new velocity calculation with head compensation



(a) Data of x- and y coordinate of raw gaze data in 2D.



(b) Data of x- and y coordinate of gaze data mapped onto a snapshot in 2D.

Figure 5.3: The difference between raw gaze data and manually mapped gaze data.

The results the videos are presented by using several graphs. The first two graphs will show the *PrecisionS* and *PrecisionF* for different velocity thresholds, for both *PrecisionS* and *PrecisionF* higher values are desired. Next are graphs showing where the methods have classified the points as fixations, this is visualized by plotting mapped data and overlaying it with thick lines where fixations are found, the ideal results are thick lines whenever both the mapped data lines are horizontally straight. The last graphs shows the euclidean distances between the fixations and the actual gaze point for that time stamp measured in pixels, smaller difference is better.

5.1 Fixations with Head Movements

This recording measures the performance of the head movement compensation. Different head movements in pitch, roll and yaw, both combined and separately was done in the recording. The figures 5.4 and 5.5 shows the precision fixations and saccades. The precision of saccades is better for method 2 and 3 until the 100 $°/s$ threshold were they drop. Looking at method 2 compared to 1 in Figure 5.4 it is seen that there is a significant improvement using method 2 in correctly classify fixations. Method 3 slightly better than method 2 for lower thresholds and then quite equally to method 2 here.

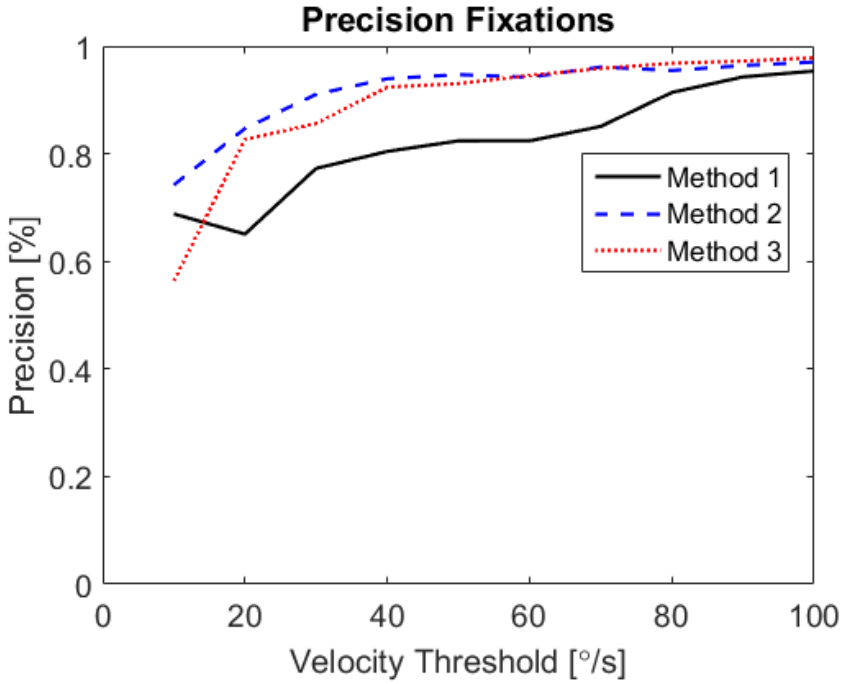


Figure 5.4: The *PrecisionF* for all three methods with different thresholds, higher value is better.

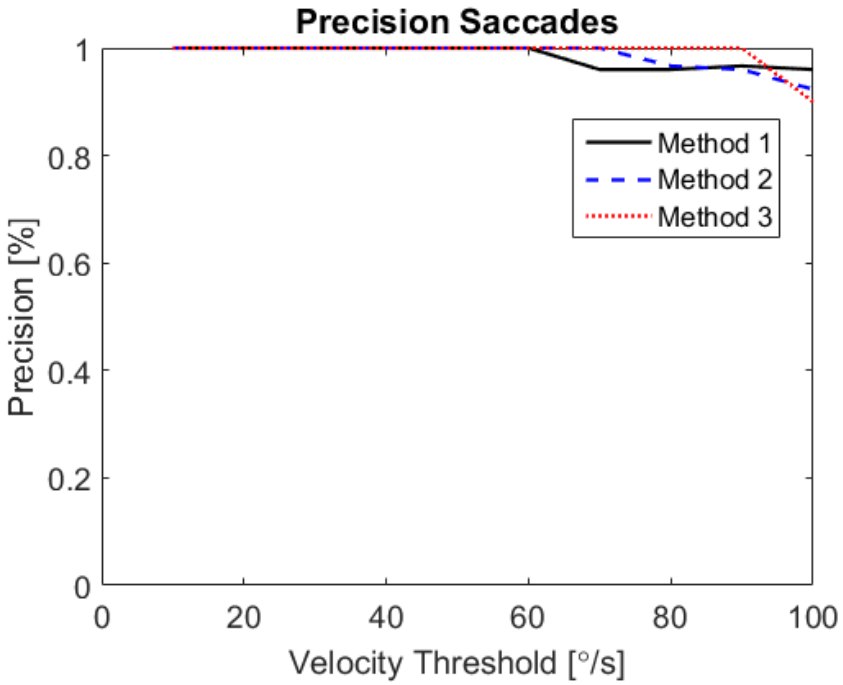


Figure 5.5: The *PrecisionS* for all three methods with different thresholds, higher value is better.

Figure 5.6 show where the fixations are found on mapped gaze data. Method 1 with a velocity threshold of 100 $°/s$ do perform quite well and it is seen that for both method 2 and 3 a velocity threshold of 30 $°/s$ is not high enough since many fixations are still wrongly classified and the fixations are split up in many short ones. The precision of fixations gets higher with higher velocity thresholds and by looking at Figure 5.5 the threshold can be as high as 70 $°/s$ for method 2 and 90 $°/s$ for method 3 without loosing precision in detecting saccades. Figure 5.7 shows the fixations with those thresholds. Using the 70 $°/s$ threshold for method 2 shows big improvement compared to the 30 $°/s$ threshold and it also looks like more of the fixations is correctly classified. Method 3 seem to perform better than both method2 and 2 with a threshold of 90 $°/s$ though. It has found most of the fixations and the fixations are less cut up into shorter fixations than for the other methods. Figure 5.8 and Figure 5.9 show that the fixation point is closer to all gaze samples for lower thresholds.

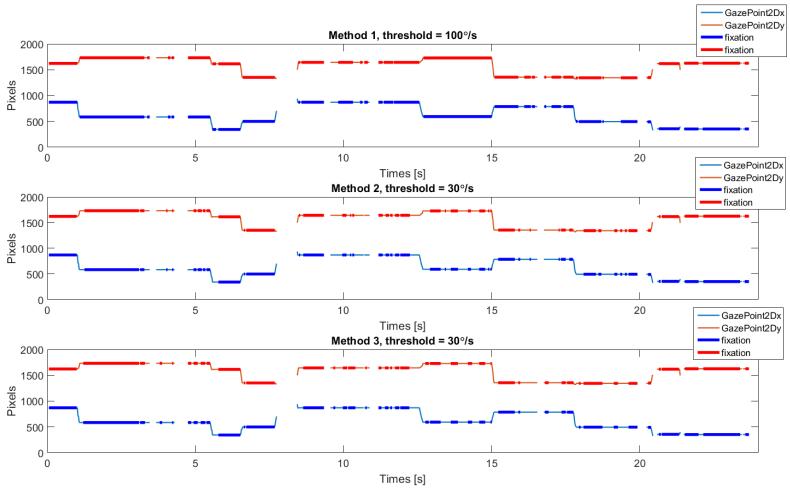


Figure 5.6: Mapped gaze data, thick lines represent where the method has defined the sample points as fixation.

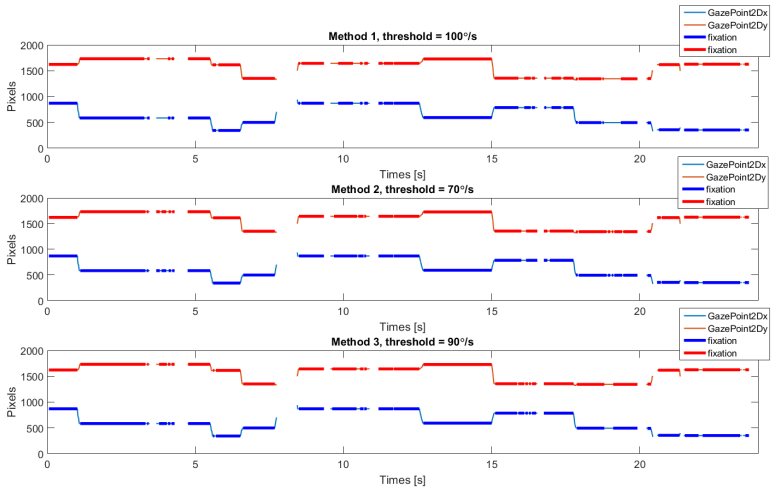
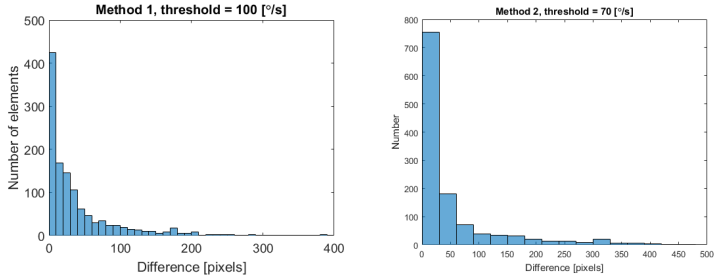


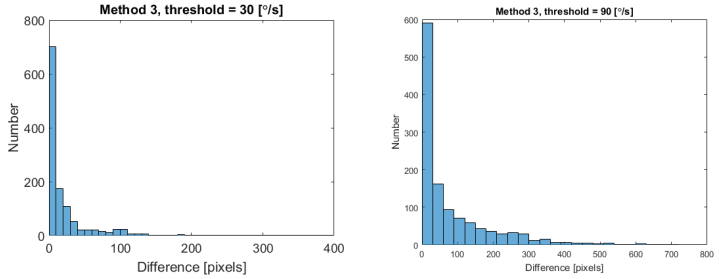
Figure 5.7: Mapped gaze data, thick lines represent where the method has defined the sample points as fixation.



(a) Method 1 with velocity threshold 100 °/s.

(b) Method 2 with velocity threshold 70 °/s.

Figure 5.8: Histogram of the euclidean distance between actual gaze point and calculated fixation point. Small values are desired.



(a) Method 3 with velocity threshold 30 °/s.

(b) Method 3 with velocity threshold 90 °/s.

Figure 5.9: Histogram of the euclidean distance between actual gaze point and calculated fixation point. Small values are desired.

5.2 Beginning and End of Fixations and Saccades

The purpose of this test case was to see how well the algorithm performs in the beginning and end of fixations/beginning and end of saccades. In many natural scenarios a person start moving their head while they they still are in a fixation, so the eyes moves in opposite direction of the head to still be looking at same thing. Shortly after the head and eyes moves in the same direction in the saccade. This kind of head/eye movement is done naturally often but is very hard to do when you're thinking about it, therefore many saccades was done in the recording to catch this typical behaviour. This was done by asking the test person to quickly move the gaze between two points back and forth, see Figure 5.2.

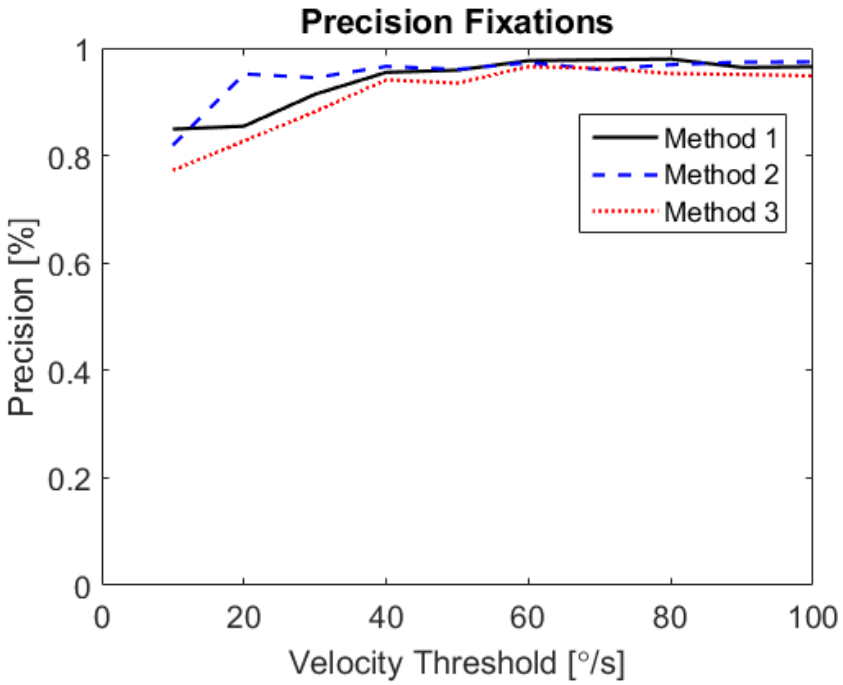


Figure 5.10: The precision of fixationsin percentage for all three methods with different thresholds.

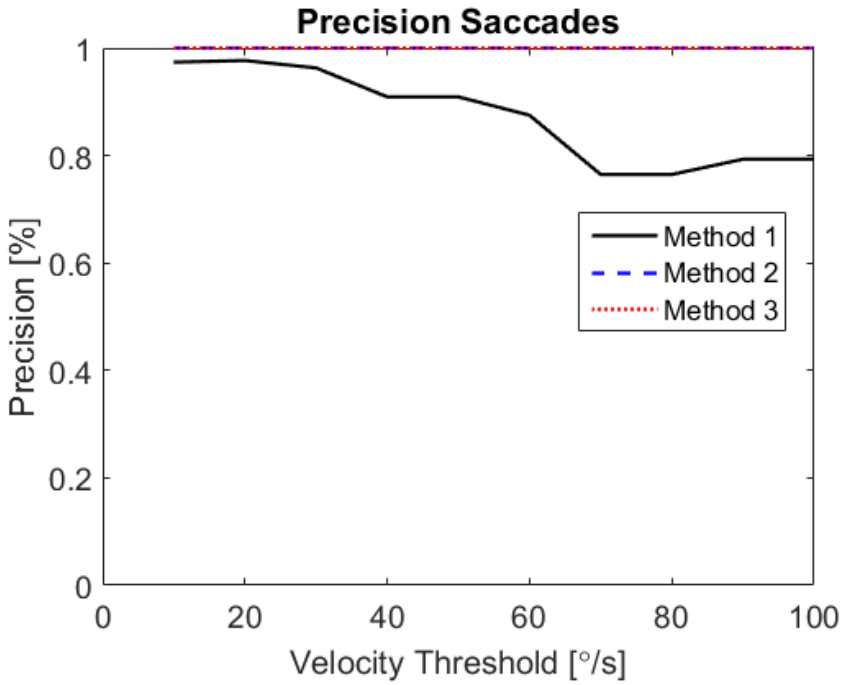


Figure 5.11: The precision of saccadesin percentage for all three methods with different thresholds.

From Figure 5.10 and Figure 5.11 we can see that the precision of fixations is very close to each other but the precision in saccades is a lot better for method 2 and 3. In Figure 5.12 we can see that it is particularly in the beginning of fixations that all three methods fail to correctly classify the data. One reason for this could be due to the difficulty of manually mapping the data. As small overshot can be seen in the recording which has been ignored in the mapping of the data since it is assumed that the fixation starts when the test person first reaches the point.

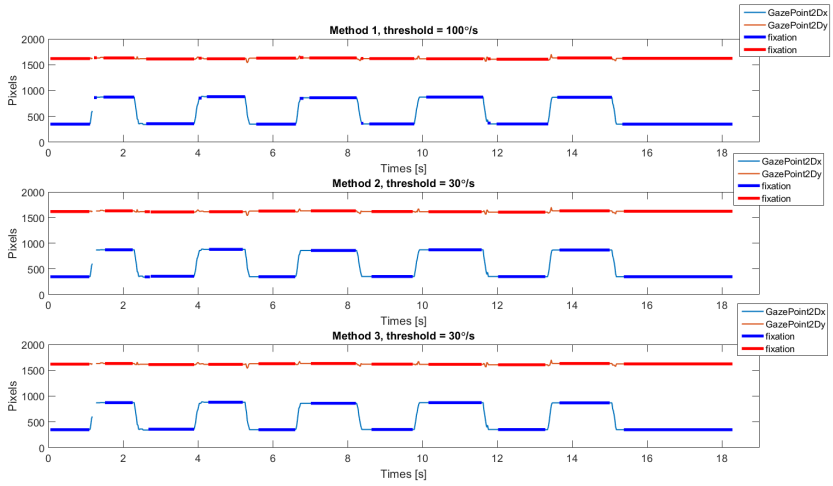
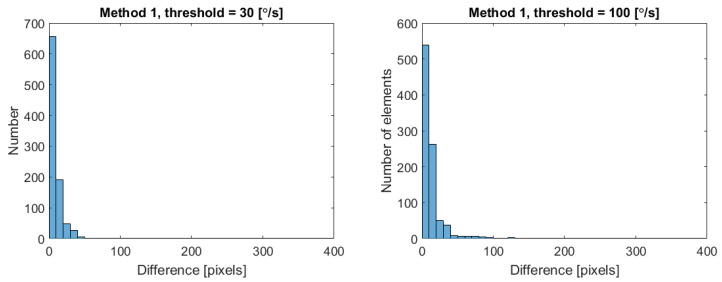


Figure 5.12: Mapped gaze data, thick lines represent where the method has defined the sample points as a fixation.

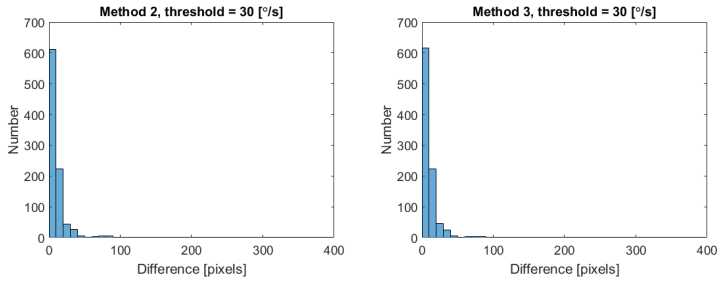
Figure 5.13 and Figure 5.14 shows that method 2 and 3 gives better results than method 1. Method 2 and 3 perform equally in this case.



(a) Method 1 with velocity threshold 30 °/s.

(b) Method 1 with velocity threshold 100 °/s.

Figure 5.13: Histogram of the euclidean distance between actual gaze point and calculated fixation point.



(a) Method 2 with velocity threshold 30 °/s.

(b) Method 2 with velocity threshold 30 °/s.

Figure 5.14: Histogram of the euclidean distance between actual gaze point and calculated fixation point and the calculated velocities.

In this case no distinct difference between method 2 and 3 can be seen, although both of them perform better than method 1. All methods seem to have difficulty in the beginning of fixations, a higher threshold than 30 °/s would improve this. Looking at Figure 5.10 the precision seem to be better and better for having a high velocity threshold without losing any saccade precision. One reason for that it seems like the higher threshold the better, is because all saccades in the video are very distinct ones and therefore they do not risk being missed by a high threshold. In a video with shorter saccades the precision of saccades might would have gone down.

5.3 No Head Movements

In this recording the test subject was told to move the head as little as possible during the whole recording, preferably no movement at all . As can be seen in Figure 5.15 and Figure 5.16 the head movement compensation with the original velocity calculation and the one without head movement compensation, i.e. the starting point of the project, is the same. This is of course good results since we do not want the head movement compensation to do anything if no head movements are done.

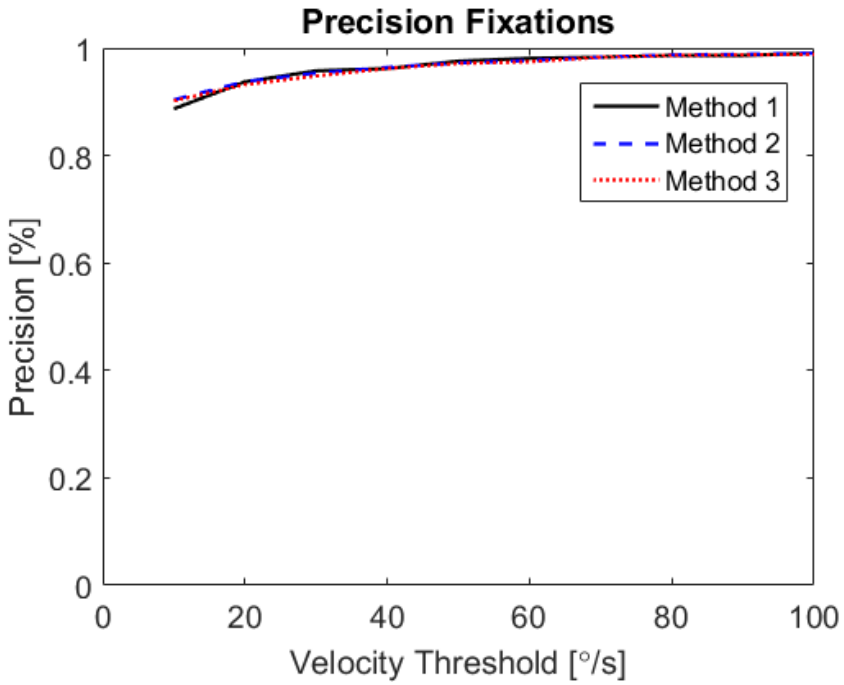


Figure 5.15: The precision of fixations in percentage for all three methods with different thresholds.

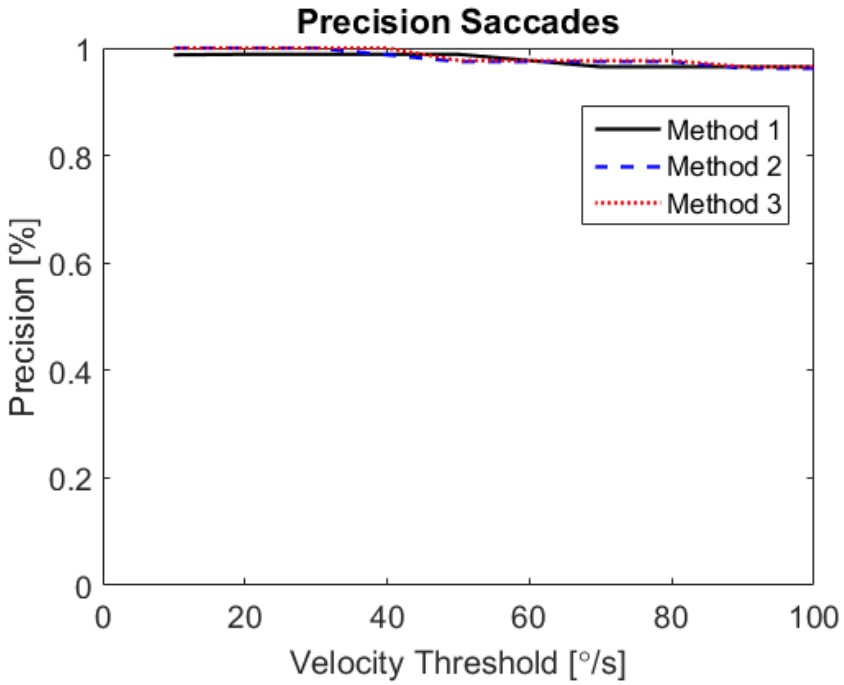


Figure 5.16: The precision of saccades in percentage for all three methods with different thresholds.

Looking at Figure 5.15 and Figure 5.16, the three methods score the same in precision of fixations and method 1 slightly higher in precision of saccades. Notable is that all methods give very high precision. As expected, the head compensation does not do anything when the head is almost still during the recording. It also seems that no additional noise is noticeable from adding the head compensation. The difference in classification is seen in Figure 5.17.

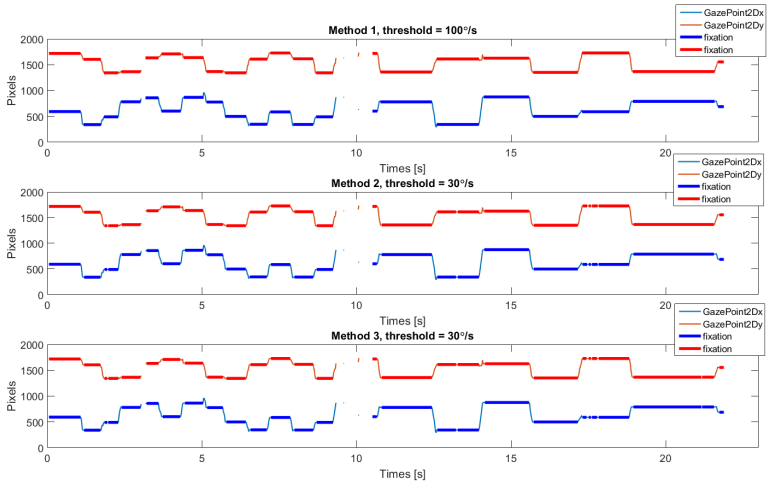
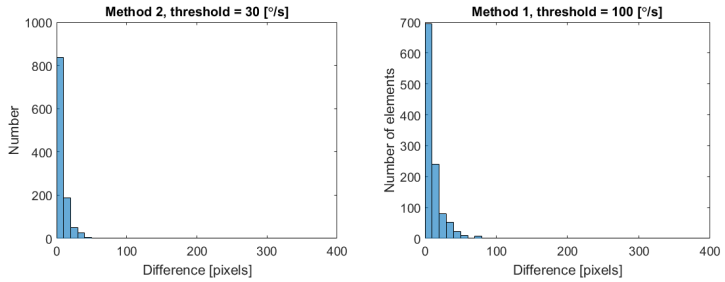


Figure 5.17: Mapped gaze data, thick lines represent where the method has defined the sample points as a fixation.

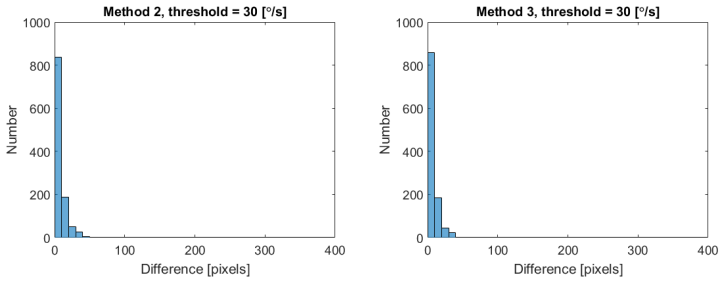
The difference between the calculated fixation point and actual gaze point should be small if the method has classified the fixations correctly. The histogram of the difference is shown in Figure 5.18 and Figure 5.19. The head compensation seems here to improve the results since the maximal difference is almost half of method 1 and more elements has a really small difference. Method 3 seems to improve it slightly but the difference between method 3 and method 2 are small.



(a) Method 1 with velocity threshold 30 °/s.

(b) Method 1 with velocity threshold 100 °/s.

Figure 5.18: Histogram of the euclidean distance between actual gaze point and calculated fixation point.



(a) Method 2 with velocity threshold 30 °/s.

(b) Method 3 with velocity threshold 30 °/s.

Figure 5.19: Histogram of the euclidean distance between actual gaze point and calculated fixation point and the calculated velocities.

All together, the three methods seem to perform almost identical in identifying saccades and fixations. Method 3 and 2 seems to be almost the same even when comparing the fixation point but a bigger difference is seen for method 1. Using a velocity of 30 °/s for this recording seems viable.

5.4 Reading

In this recording the test person was given a paragraph to read from a paper. Looking at the figures 5.20-5.22 method 1 performs slightly worse in both the precision of fixations and the precision of saccades at lower velocity thresholds but the differences are negligible.

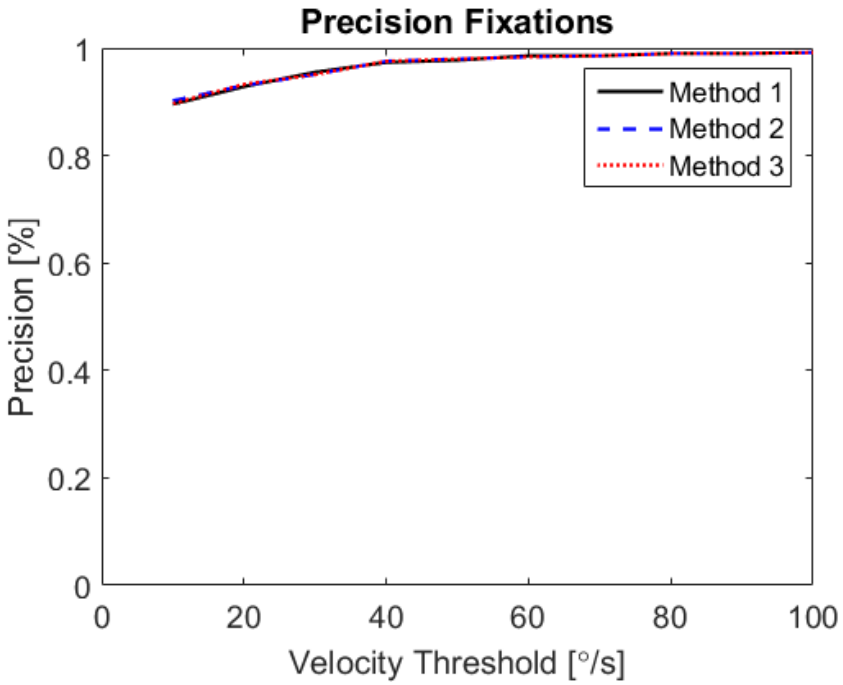


Figure 5.20: The precision of fixations in percentage for all three methods with different thresholds.

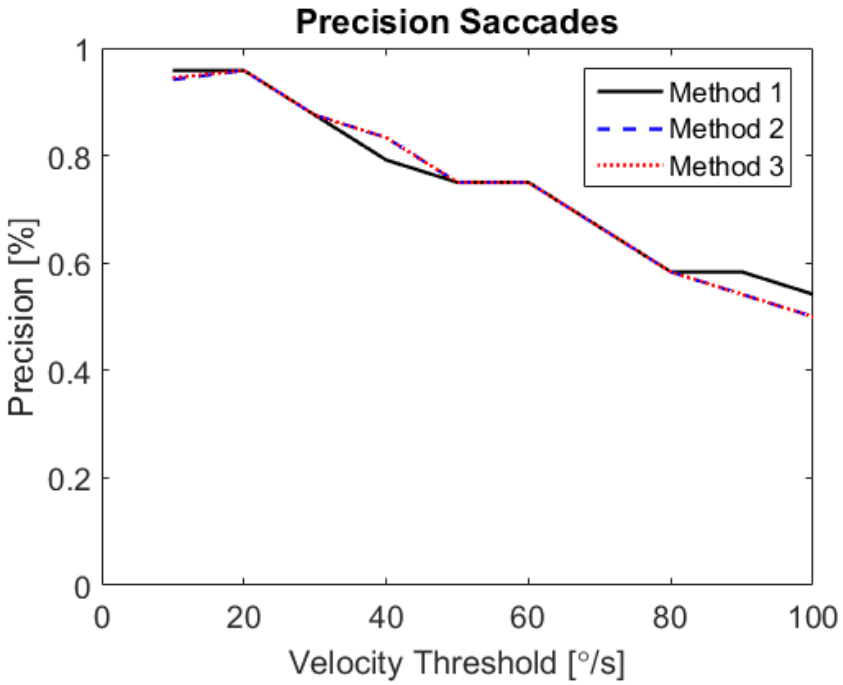


Figure 5.21: The precision of saccades in percentage for all three methods with different thresholds.

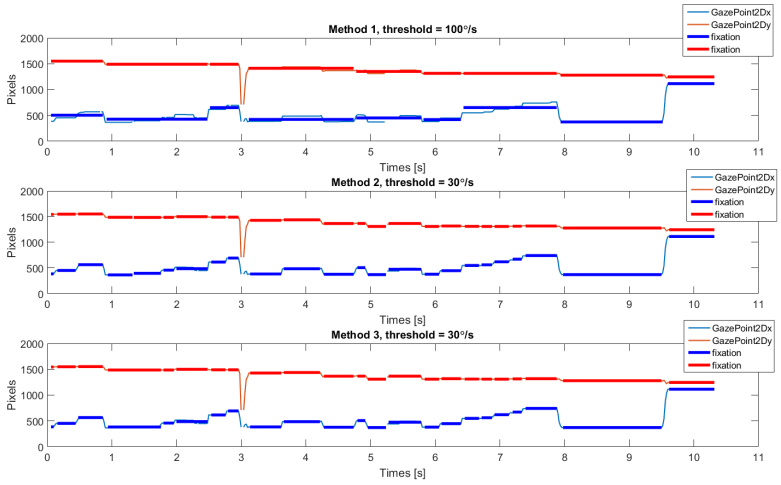
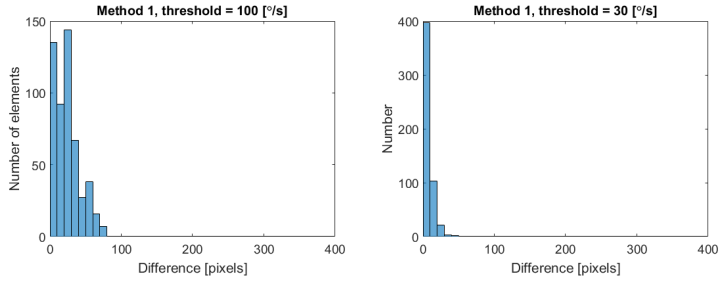


Figure 5.22: Mapped gaze data, thick lines represent where the method has defined the sample points as a fixation.

Looking at Figure 5.22 it is evident that using a threshold of 100°/s without head compensation is far too high. It is clearly seen that the method misses several saccades which is the behavior that want to be avoided. The saccades

between reading fixations are below 100°/s which means that it is impossible for the algorithm to correctly identify reading saccades/fixations by setting the threshold at 100°/s.

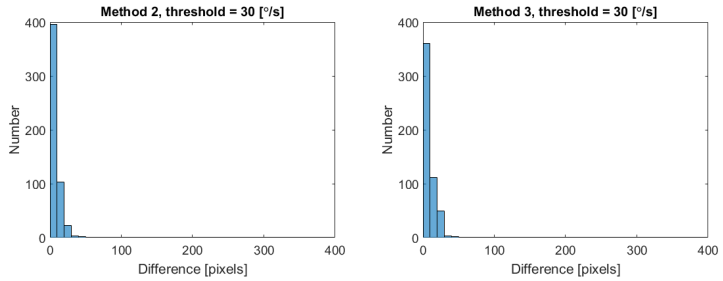
Setting a lower threshold at 30°/s shows improvement in finding reading saccades/fixations compared to 100°/s. Looking at the histograms 5.23-5.24 it is obvious that the best performing algorithms are the ones at 30°/s since the euclidian distance between the actual gaze point and the calculated fixation point is smallest for these methods. Furthermore there is no big difference between method 1 and method 2 The reason for this can be due to the fact that the recorded video did not contain a lot of head movements, the test person read the text without having to move his head too much. Method 3 at 30°/s performs about the same as method 1 and 2 with a threshold of 30°/s.



(a) Method 1 with velocity threshold 100 °/s.

(b) Method 1 with velocity threshold 30 °/s.

Figure 5.23: Histogram of the euclidean distance between actual gaze point and calculated fixation point.



(a) Method 2 with velocity threshold 30 °/s.

(b) Method 3 with velocity threshold 30 °/s.

Figure 5.24: Histogram of the euclidean distance between actual gaze point and calculated fixation point and the calculated velocities.

Chapter 6

Further Development Possibilities

During the thesis many possibilities for further work has presented themselves. Other than implementing the head compensation we have researched and implemented many smaller functions and algorithms which all could be further developed. To score our algorithm we used the results from an earlier thesis work but we feel that our method can be further developed with the hope of finding a scoring method which will work for all eye tracking algorithms. The goal would be to find a scoring algorithm which gives an objective result and leaves little room for interpretation.

We suspect that more work can be done using the accelerometer. At the moment we did not fully utilize the accelerometer in our head compensation but we think the accelerometer can be used more, especially when it comes to head movements in the pitch direction. It is also indicated in related work that expanding the head compensation algorithm with a magnetometer may improve the drift reduction in the gyroscope. Most research suggest that a magnetometer improves the accuracy in the measurement. During the tests it could be seen that there was noise that we did not manage to filter away, further work could focus on making the data sets 'cleaner' with better filtering algorithms.

The implemented filter does not detect smooth pursuits and to improve the filter further this could be added. Adding the functionality of smooth pursuit detection would make the eye movement detection complete and probably decrease inaccurate classification. To achieve good results in detecting smooth pursuits the I-VT filter could be combined with another filter.

According to the four different videos different thresholds should be used depending on which kind of recording is being studied. It is reasonable to have a very low velocity threshold for recordings of reading where only a few head movements will be done whereas it will be wise to raise the threshold some in other cases. Finding the optimal thresholds for the different cases is a task in itself. In the best case scenario the filter performs exceptionally well even at low thresholds which would allow the user to always use a low threshold

regardless of what kind of video.

In general we found that there are many small improvements that can be made in several of the steps in the I-VT filter. For example it would be interesting to look into the conditions set for merge fixations.

Chapter 7

Conclusions

All recordings show that head compensation improve the saccade and fixation identification when there are head movements. In cases where there are few head movements the head movement compensation does not impact the results. The new velocity calculation, method 3, perform about as good or better in all of the recordings studied in this thesis than the previous previous velocity calculations, method 2. A velocity threshold of 100 $^\circ/s$ is not necessary with head movement compensation. The results indicate that a threshold between 30- 90 $^\circ$ is appropriate to use depending on which type of movements that dominate during the recording.

Bibliography

[1] M. Alt, C-T. Nguyen, P. Tobien *A Quantitative Comparison of Fixation Filters* 2015

[2] J.A. Barraza-Madrigal et. al. Instantaneous Position and Orientation of the Body Segments as an Arbitrary Object in 3D Space by Merging Gyroscope and Accelerometer Information 2014.

[3] P. Blignaut *Fixation identification: The optimum threshold for a dispersion algorithm* 2009

[4] F. Haugen The Good Gain method for simple experimental tuning of PI controllers 2012.

[5] G. Larsson *Evaluation Methodology of Eye Movement Classification Algorithms* 2010

[6] L. Larsson et al. *Compensation of Head Movements in Mobile Eye-Tracking Data Using an Interial Measurement Unit* 2014.

[7] H.J. Luinge, P.H. Veltink Measuring orientation of human body segments using miniature gyroscopes and accelerometers 2005.

[8] A. Olsen *The Tobii I-VT Fixation Filter* 2012

[9] D. D. Salvucci J.H. Goldberg *Identifying Fixations and Saccades in Eye-Tracking Protocols* 2015

[10] L. Xue et al. A novel Kalman filter for combining outputs of MEMS gyroscope array 2012.

[11] *www.adafruit.com* Last checked: 2016-02-28

[12] *www.pololu.com* Last checked: 2016-02-28

[13] *www.tobii.com/tech* Last checked: 2016-05-02

[14] *www.tobiidynavox.com* Last checked: 2016-05-02

[15] *www.tobiipro.com* Last checked: 2016-05-02

Appendix A

Export Variables to MATLAB import file

- Recording timestamp
  - Timestamps of all data
- Gaze Point X
- Gaze Point Y
  - 2D gaze coordinates, data given in pixels
- Gaze 3D position left X
- Gaze 3D position left Y
- Gaze 3D position left Z
- Gaze 3D position right X
- Gaze 3D position right Y
- Gaze 3D position right Z
  - Calculated 3D coordinates given in millimeter, calculated as the point closest to both direction vectors, left and right.
- Gaze 3D direction left X
- Gaze 3D direction left Y
- Gaze 3D direction left Z
- Gaze 3D direction right X
- Gaze 3D direction right Y
- Gaze 3D direction right Z
  - Direction vector with origin at pupil, normalized.
- Pupil position left X

- Pupil position left Y
- Pupil position left Z
- Pupil position right X
- Pupil position right Y
- Pupil position right Z
  - Pupil position, left and right, given in millimeter.
- Event
  - Name of Event. Costumed events of type Fixation Begin, Fixation-Start, FixationEnd and Fixation End is used. These costumed events is used if data is manually labeled and mapped.
- Mapped gaze data X
- Mapped gaze data Y
  - Mapped gaze coordinates on snapshot, data given in pixels. Optional export variables. Limited to one snapshot.
- Gyro X
- Gyro Y
- Gyro Z
  - Gyroscope measurements measured in °/s. Rotation around X-, Y-, Z-axis corresponds to pitch, yaw and roll respectively.
- Accelerometer X
- Accelerometer Y
- Accelerometer Z
  - Accelerometer measurements measures proper acceleration, "g-force", in $°/s^2$ in the three directions.

The parameters are exported from Analyzer as tabdelimited .tsv-file.

Appendix B

Default Values for I-VT filter

Table B.1: Default Values for I-VT filter

Velocity Window length	20 ms
Max Gap Fill In Length	75 ms
Merge Angle between Fixations	0.5°
Merge Gap between Fixations	75 ms
Discard Fixation Duration	60 ms

Appendix C

Technical Specification

In the table C.1-C.3 the technical specifications of the used glasses, gyroscope and accelerometer is presented.

Table C.1: Tobii Glasses 2, from [15].

Gaze sampling frequncy	50 Hz
Number of eye cameras	4
Camera recording angle horizontal	82 °
Camera recording angle vertical	52 °
Frame dimensions	179x159x57 mm
Dimensions	130x85x27
Battery	Li-on
Weight incl. battery	312 g
Storage media	SD card
Connectors	eternet,WLAN, 3.5mm jack
Wireless	2.4 GHz and 5 GHz

Table C.2: Gyroscope L3GD20, from [12].

Supply voltage	2.4 to 3.6 V
Temperature range	-40 to +85°C
Digital rate output	95 Hz
Temperature sensitivity	$\pm$ 2%

Table C.3: Accelerometer LIS3DH, from [11].

Supply voltage	1.71 to 3.6 V
Temperature range	-40 to +85°C
Rate noise density	220 $\mu g/\sqrt{Hz}$
Digital rate output	100 Hz
Temperature sensitivity	$\pm$ 0.01 %/°C %

Appendix D

Gyroscope Default Values

The set of experimentally chosen default values for the glasses used in this thesis is shown in table D.1.

Table D.1: Default values for filtration of gyroscope.

Parameter	Default value
fs	95 Hz
q_w	0.3
ρ	-0.3
q_n	0.0182

In Figure D.1 and Figure D.2 is the gyroscope signals shown, before and after calibration respectively.

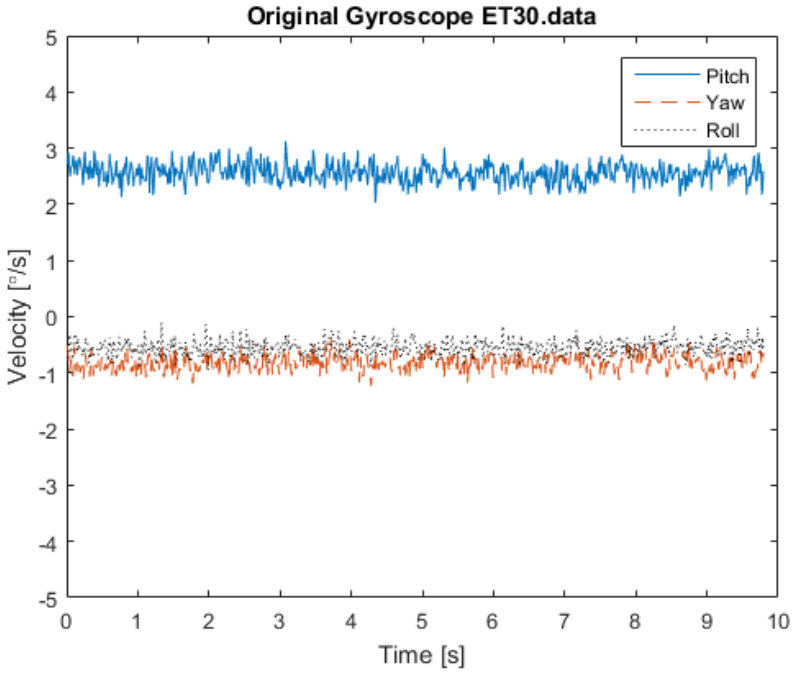


Figure D.1: The unfiltered, original gyroscope data

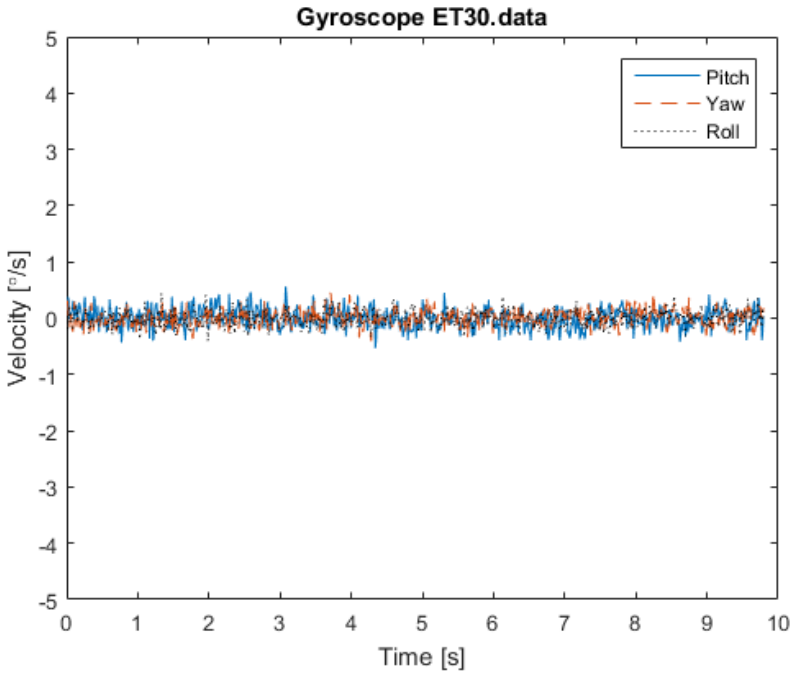


Figure D.2: The filtered gyroscope data after calibration

Appendix E

Implementation Overview

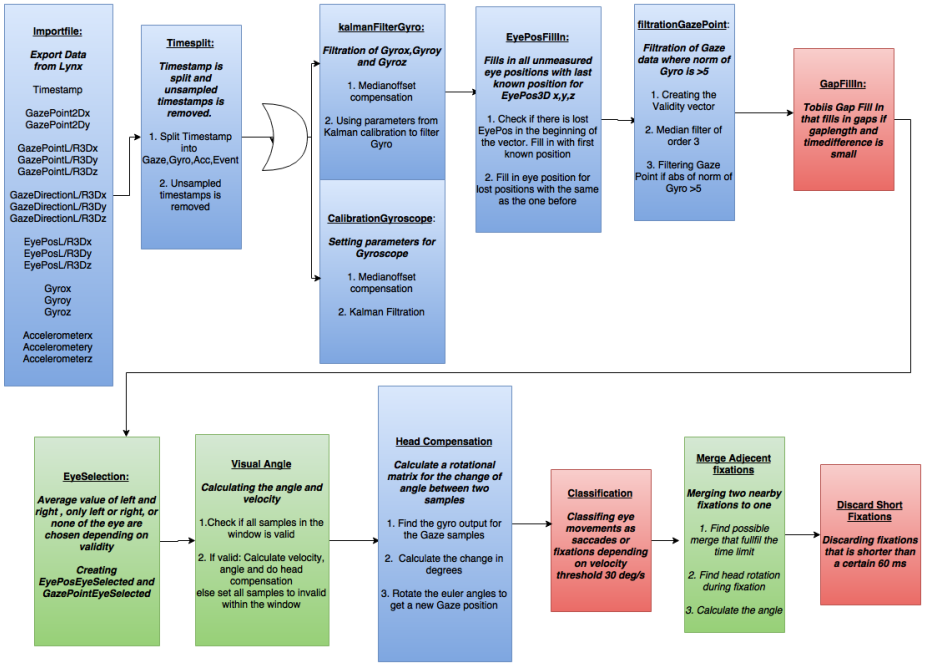


Figure E.1: An overview of the implemented system. Blue colour represents new functionality that is added, green given functions that has been modified and red colour given functions that is kept as they were.