A data quality study for the Positive science wearable eye tracker

Jan van den Heijkant – 5492564 Utrecht University

Master Thesis Applied Cognitive Psychology

Supervisor: Dr. I.T.C. Hooge

Second Assessor: Dr. J.S. Benjamins

Introduction

The field of eye tracking research has had a sizable impact on cognitive psychology as a whole. Eye tracking and the study of gaze directions hava a history stretching back over a century (Duchowski, 2002). The focus has shifted from eye movement facts, like saccadic suppression and angle velocity of the eye (Dodge & Cline, 1901), to a more applied focus in behaviourist experiments, like how people look at pictures (Buswell, 1935). This understanding of basic processes led to research into the more complex parts of the visual system, aided by the improving technology. Over sixty years ago the Russian scientist Alfred L. Yarbus described that saccadic eye-movements reflect cognitive processes and today the foundations his work laid are still being built on. Eye tracking is an objective way of studying looking behaviour. It has given psychologists a deeper insight into the way people interact with their environment and each other. There are different ways of eye tracking using different types of eye trackers. The two types I will go into further in this study are remote eye trackers and wearable eye trackers.

Remote eye trackers are a class of eye trackers that measure the subject's gaze direction from a distance. Stimuli are usually presented on a screen with a remote eye tracker under the screen facing the subject. The eye tracker contains at least one camera pointed at the subject's eye(s) and an infrared illuminator. The illuminator shines an infrared light on the subject's eyes which is reflected by the cornea and the eye tracker picks up the reflection of this light. The eye tracker can also register the location of the pupil. By combining information from the pupil with the location of the corneal reflection the gaze direction of the subject can be calculated.

Wearable eye trackers essentially function in the same way, with a few differences. There is at least one camera to film the eye and an infrared illuminator to create a corneal reflection. However, a wearable eye tracker also has a scene camera that films the surroundings of the subject. Since the entire setup is wearable, these eye trackers can be used without a measurement screen and digital stimuli.

These different types of eye tracker have been used in a wide variety of research. Remote eye trackers have been used in driving simulation tasks for example (Palinko, Kun, Shyrokov & Heeman, 2010). In this study the researchers tried to use remote eye trackers to measure cognitive load in a driving task. During the driving task the subject had to play two different word games to increase their cognitive load. In this study the eye tracker was used to measure pupil size instead of gaze direction. This is only one example of a remote eye tracker study, remote eye trackers are however not only useful for cognitive research. They have also been used in many other fields of research, such as neuroscience, industrial engineering, computer science and marketing research (Duchowski, 2002) as well as usability testing. Remote eye trackers have also been used for other visual research, like search behaviour and reading tasks (Mele & Federici, 2012) or music reading, typing and scene perception (Rayner, 1998).

Wearable eye trackers have been used to study many different things as well. Wearable eye trackers are especially fit for experiments where the subject needs to be able to move around. Because of the freedom of movement these types of eye trackers give to the subject, they can be used to study things like human crowd navigation (Hessels, van Doorn, Benjamins, Holleman & Hooge, 2020), they have even been used to study gaze behaviour in basketball shooting (Steciuk & Zwierko, 2015) as well as making tea in a kitchen environment. Making tea is a seemingly simple task, but it involves many complex coordinated movements that require a lot of different objects. By studying peoples fixations when they’re making tea, researchers discovered the close monitoring role of the eyes in automated activities (Land, Mennie, & Rusted 1999). The researchers found that the eyes always precede motor action within a task like making tea. The eyes seek out the places with relevant information to continue the task, independent of conscious attention. In the example of making tea, the eyes are already looking for the teabag while the cup is being filled with hot water. At the same time the eyes keep darting back to the teacup to check the water level. This sort of research allows for more insight into the underlying processes of everyday behaviour.

Problems with eye tracking

Eye tracking is a very useful method for studying the way people interact with their environment. However, it is also a very complicated process reliant on many factors to be conducted properly. Eye tracker manufacturers usually provide information about what kind of data quality is to be expected from their hardware. This information is usually based on optimal measurement conditions though, sometimes even using mechanical eyes to test accuracy. This of course doesn't necessarily correspond to real world testing conditions. There are a lot of things that can go wrong which can decrease data quality. Data quality, or poor data quality in this case, is operationalised here as a combination of systematic error, variable error and data loss.

Systematic error in the case of eye trackers is a measure of how far off the eye tracker can be. It can be seen as the inverse of the accuracy of the eye tracker, if systematic error is high then accuracy is low. The systematic error can increase or decrease under certain conditions. The amount of systematic error could for example differ throughout the measurement range as a result of bad calibration. This could mean that the reported gaze location is further off the true gaze position towards the edges of the measurement range than it is around the centre. Systematic error can be controlled for if the magnitude is known, but this is often not the case. If the maximum influence of the systematic error translates to a 10 centimetre shift in gaze location in the real world. This would mean that the measured gaze location could be up to 10 centimetres away from the actual gaze position of the subject. Making sure this 10 centimetre difference doesn't affect measurement within an experiment is possible. This could be done by making sure the objects that are relevant in the experiment, like cups in a search task, are twice as far apart as the maximum systematic error. If a subject is looking at cups that are over 20 centimetres apart, the maximum error of 10 centimetres shouldn't cause problems. Since in this case the measured gaze position would still be closest to the cup that the subject is actually looking at. If relevant objects are right next to each other this systematic error becomes a problematic error.

Variable error reflects the reproducibility of a measurement. in the context of eye tracking it is operationalised as the sample to sample variance of the reported gaze coordinates. If the variable error is low and a subject keeps their gaze at the same location, the reported gaze coordinates don't fluctuate very much. If on the other hand variable error is high, this would result in a signal that could be described as jittery or shaky. In an ideal situation this could be controlled for by measuring a lot, since random error should over time average out to zero. However, people usually don't stare at one place for a long time. As with the systematic error, it depends on the research setup and the magnitude of variable error whether the variable error is problematic or not (Orquin & Holmqvist, 2018).

Any time an eye tracker doesn't report gaze coordinates when it should, this is called data loss. An eye tracker should report as many gaze coordinates per second as its measurement frequency. The measurement frequency of an eye tracker is the number of times per second an eye tracker can report gaze coordinates. If an eye tracker has a measurement frequency of 30 Hz it should give 30 gaze coordinates every second. If only 15 gaze coordinates are reported for a given second that means 50% data loss in that time. The percentage of data loss over the entire period of measurement is the standard measure for overall data loss.

Data loss can happen for instance when the subject blinks or looks away. Data loss from blinking isn't problematic, since blinking is a part of normal looking behaviour. Data loss from other sources can be problematic though, since the missing data could have contained interesting information.

Data loss can also happen when a subject turns away from the screen and the eyes are lost to the eye tracker. This can happen when a subject gets distracted. It can also happen when the subjects that are studied don't follow instructions very well, with infants for example. A 2017 study focused on data quality has shown that the recovery time can differ between eye trackers (Niehorster, Cornelissen, Holmqvist, Hooge & Hessels, 2017). When the subject has looked away from the screen and turns back, the eye tracker needs to locate the eyes again and restart calculating gaze coordinates. This period is called the recovery time, until the eye tracker has recovered there will be data loss as well. Another source of data loss can be a non-optimal head orientation. Non-optimal in this case means anything other than horizontally straight ahead. This means no rotation in x, y or z axes. The amount of data loss is dependent on the type of orientation change with rolling along the x axis being most problematic (Hessels, Cornelissen, Kemner & Hooge, 2015). A change in head orientation can not only result in data loss, but also in an increased shift in gaze location. This means that under non-optimal head orientations where data can still be collected the systematic error could increase.

Data quality research

There are many factors that can influence data quality in eye tracking research. Thus, data quality itself has been the topic for many eye tracking studies. Strategies for improving data quality can be found by studying the factors that negatively influence data quality. A remote eye tracker works best when the eyes are in a certain location. This optimal location isn’t the same for every model eye tracker but the manufacturer usually provides this information. The subject’s eyes need to be as close to this location as possible, to ensure the best chance for accurate and reliable measurement. If the distance between the optimal location and the real location becomes too large, gaze coordinates might not be reliably reported or not reported at all. This virtual space where measurement should be most reliable and accurate is also referred to as the headbox. There have been studies that show there is an influence of subject head orientation on data quality, therefore the subject should be facing the screen as straight as possible. The optimal location and orientation can be difficult for a subject to maintain throughout measurement, therefore a chin rest is very useful, if the experiment allows it. There have also been studies that found that a subject wearing lenses can significantly offset measurement (Nyström, Andersson, Holmqvist & van de Weijer, 2012). Many different factors have been linked to reduced data quality, eye make-up, glasses or even eye colour (Hessels, Andersson, Nyström & Kemner, 2015). A schematic drawing of a remote eye tracker setup can be found in figure 1.

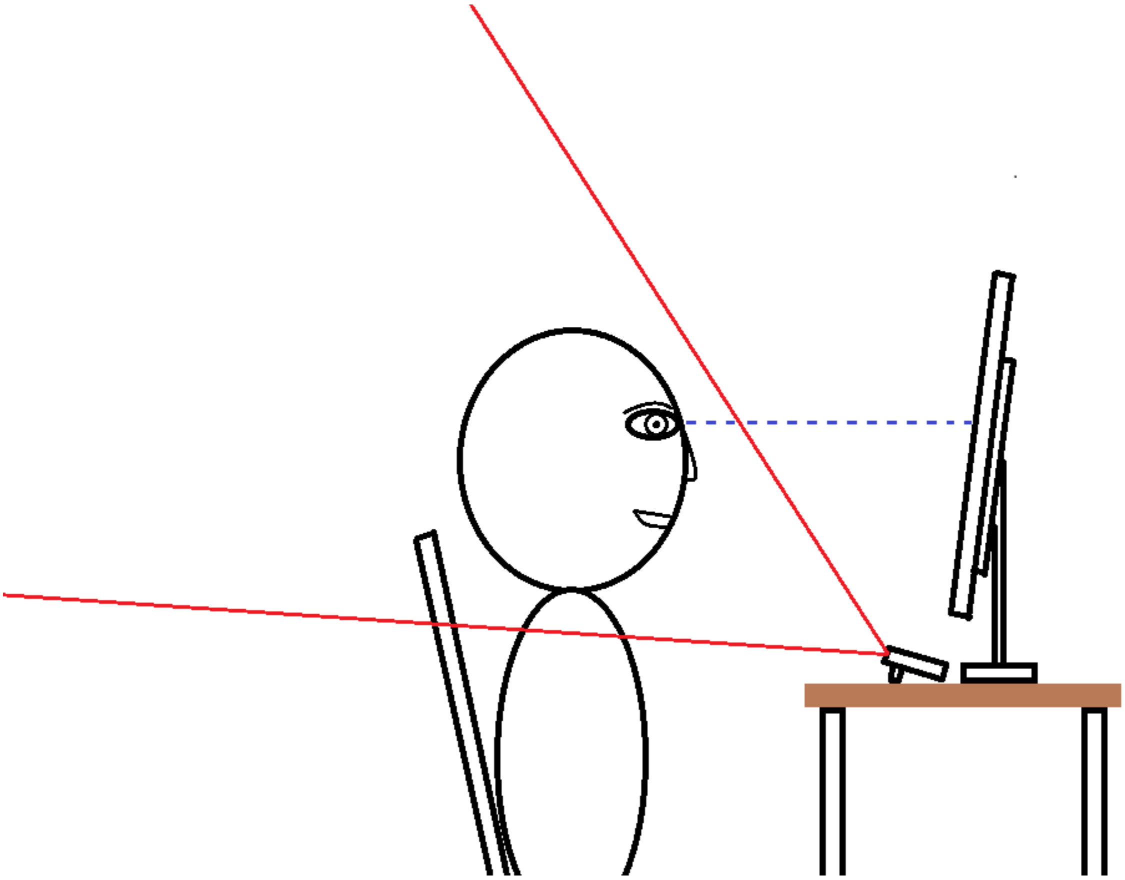


Figure 1 Schematic drawing of a remote eye tracker setup. A chin-rest could be added if the experiment requires it.

When using a wearable eye tracker there is no headbox to worry about, since the eye camera is attached to the subject's head. Instead, the relative placement of the cameras with regards to the subject's head is important. The eye camera needs to be placed in such a way that the eye is clearly visible throughout measurement, in all possible orientations. Care also needs to be taken that the camera doesn't obstruct the subject's view. The scene camera also needs proper placement, since the scene view image is where the subject's fixations will be coded. The scene camera should be properly positioned and oriented if the scene camera view roughly corresponds to the subject's field of view. Since wearable eye trackers are placed on subject's head, there is the possibility of slippage. Slippage is defined as a physical shift of the eye tracker with regards to the subject's head. This can happen as a result of shaking the head, making strong facial expressions or even talking. The effect of these actions on data quality was investigated by making subjects purposefully performing these actions to shift the eye tracker. The influence of slippage on data quality differs between eye trackers. Some eye trackers are hardly affected and others show increases in gaze deviation of up to 3.1 degrees (Niehorster, Santini, Hessels, Hooge, Kasneci & Nyström, 2020).

The use of eye trackers to study infants

Eye trackers have not only been used to study adults but they have been used in infant research as well. From as early on as the 1970s precursors of remote eye trackers were used to study infant looking behaviour (Aslin & McMurray, 2004). In the late 60s they used corneal reflection photography to determine the fixation direction of their subjects (Haith, 1969). This is somewhat similar to how modern remote eye trackers work. Such non-intrusive techniques to study infants are useful, since infants can't talk extensively about what they like or why they do certain things. Eye tracking is an objective non-intrusive way of quantifying attention, visual search capacity or looking behaviour in general. From the moment an infant first opens its eyes, looking behaviour plays an important role in their development. Remote eye tracking has been used from very early on to investigate infant looking behaviour (Gredebäck, Johnson & van Hofsten, 2010). Eye trackers have for example been used to study the influence of a mother's gaze on the infant's looking behaviour (Kochukhova & Gredebäck, 2010).

As I have mentioned earlier, the subject's location and orientation within the headbox can influence the eventual data quality. However, infants tend to move around a lot which can increase data loss and systematic error. Instructing the infant to sit still during measurement doesn't really work, since most infants can't follow instructions very well. Sometimes the infants are restrained in a car seat for example, so that they can't move around as much. This does improve data quality (Hessels, Andersson, Hooge, Nyström, & Kemner, 2015) but it can lead to infant distress, as they don't like being restrained for long periods of time. Using remote eye trackers to study infants also limits the kind of research that can be done. As I have mentioned before, to study more complex behaviours that involve moving around, a wearable eye tracker is better suited. In a wearable eye tracker study an infant can move around without causing the same difficulties it would cause in a remote eye tracker study. It also allows researchers to study their looking behaviour when movements are more free and self-motivated. Learning about the ways infants explore their environment is easier to achieve with a wearable eye tracking set-up than with a remote eye tracking set-up.

A problem with using wearable eye trackers for infant research is that most wearable eye trackers aren't fit for use on infants. Wearable eye trackers can be relatively heavy and big for infants. There is only one eye tracker commercially available for infant research, the Positive science wearable eye tracker. This eye tracker has been used in a number of eye tracking studies with infants. In a 2016 study by Yu & Smith this eye tracker was used to study the origin of sustained attention in infants. Sustained attention is an interesting topic of research, since it can be an indicator of later mental development of infants. In a 2011 study by Franchak, Kretch, Soska and Adolph wearable eye tracking was applied to infant research. They tried to measure infant looking behaviour in a more unconstrained setting using wireless wearable eye trackers in a free-play environment. This environment consisted of a room with many objects and toys the infant could explore and examine, while being free to move around. In this study the researchers wanted to explore self-motivated infant behaviour by putting the infant in an interesting environment it could and would want to explore by itself, while gaze data was collected.

As stated earlier, infant research has particular difficulties, especially relating to the infants' (in)ability to follow instructions and stay focused on the task at hand. Plus, infants can get fussy and irritable if they are kept from moving for a long time. The article by Yu & Smith, 2016 mentioned earlier doesn't describe any of the problems they might have run into. The focus of their article is on the theoretical background and describing the experiment itself. This is necessary information for replication of their study but it's not complete. By omitting their strategies to improve data quality or their solutions for putting infants at ease during measurement they are leaving out important information. It seems as if there were no difficulties at all, which is hard to believe. They did have their infant subjects sitting in a high chair, so that they weren't completely unrestrained. However, beyond this there is no description of infant related details. This omission makes it very difficult to replicate their study in exactly the same way. In the article by Franchak et. al. (2011) infant related details are explained a little more. The researchers explain they used a wireless wearable eye tracking setup from Positive Science. They describe that everything about the setup was kept as small and light as possible to reduce infant fussiness. The eventual setup with eye tracker, cap, vest, battery pack and wireless transmitter weighing only 271 grams. Of their 44 infants only 4 refused to have the eye tracker put on them. The calibration procedure they used is also explained as well as how the infants were kept busy whilst installing the equipment. At first, the calibration sequence was run on a computer screen by individually showing 9 images in a 3 by 3 grid. The flashing image was coupled with sounds in an attempt to draw the infant's attention. However, the procedure was later changed to use a poster board with windows where a confederate could draw the infant's attention with a toy. No explanation is given as to why the procedure was changed. Maybe the computer screen didn't really work to get the infant's attention or maybe using a poster board was simply easier. The reason for this change would be useful information for researchers trying to replicate or expand on this research. If a poster board calibration with an infant's favourite toy works better than a digital calibration sequence this is very useful information. If this is the case there is no need for an extra screen for the calibration sequence. When calibration is done with a poster board the infant's favourite toy could be used for calibration, instead of digital stimuli.

The aim of this study

This thesis is an in depth look into the Positive Science wearable eye-tracker with regards to data quality. A few wearable eye-trackers have been investigated with respect to data quality, like in the study by Niehorster et al. (2020). However, the Positive Science wearable eye-tracker hasn't been studied like this. Yet, at the moment it is the only commercially available wearable eye tracker that can be used in infant research. Therefore I will put this eye tracker to the test with regards to data quality. Thereby I will attempt to clarify the possibilities and challenges of using the Positive Science wearable eye tracker in infant research. This study will consist of two experiments using adult subjects (figure 2).

First I will investigate the measurement range of the eye tracker and how it compares to the scene camera image.

Secondly I will investigate the effect of eye-in-head orientation on the systematic error.
The choice to use adults instead of infants was made because adults can be clearly instructed on the purpose of these tests. The purpose of these tests is to find out the influences of effective measurement range and eye-in-head orientation independent of infant related difficulties. Eye-in-head orientation refers to the relative orientation of the eyes in the head.

Figure 2 Participant wearing the positive science wearable eye tracker on a custom mount fit for adults

Methods

Subjects

Subjects were recruited at Utrecht University. All consented to their data being used for this thesis and any use of pictures of subjects has been authorized by the respective subject. A total of 7 people (1 female, 6 male) participated in the experiments in this study. There was some subject overlap between the experiments, 3 subjects participated in experiment 1 (1 female, 2 male) 5 subjects in experiment 2 (1 female, 4 male). All subjects had knowledge of the purpose of the tests they participated in and all participated voluntarily.

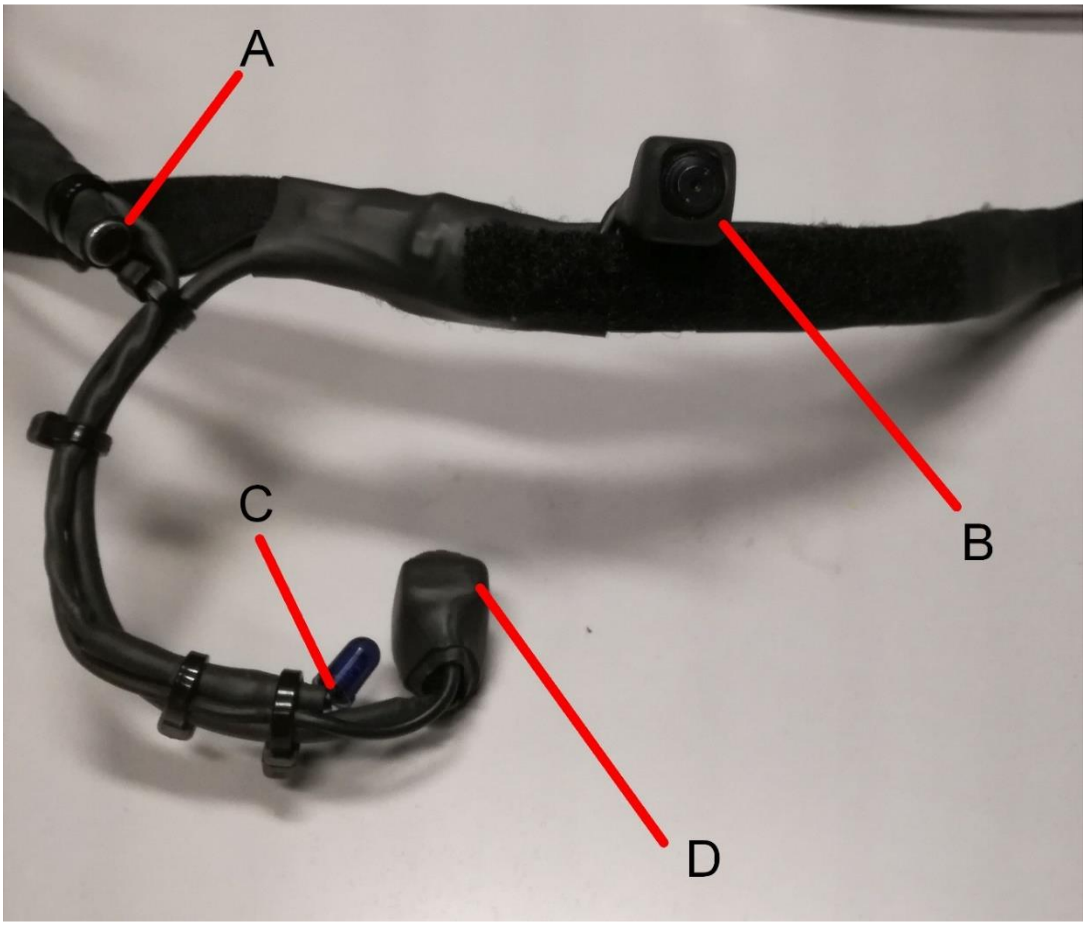


Figure 3: Positive science wearable eye tracker with parts highlighted. *A* = microphone *B* = scene-camera *C* = infrared illuminator *D* = eye-camera

Materials

The eye tracker

In this study, the Positive Science wearable eye tracker was used. This eye tracker (see Figure 3) is light, and it can be fitted to a flexible cap on a baby's head. The eye tracker consists of a semi flexible bar that holds a microphone (A) and the scene camera (B) as well as a flexible arm where the eye camera (D) and infrared illuminator (C) are attached. The eye tracker can be mounted using a strip of Velcro attached to the back of the semi flexible bar. For infants it is attached to a cap made of soft fabric that stretches to fit properly, however, this is too small to use for most adults. To use the eye tracker for adults it was attached to the inner fastenings of a safety helmet. The helmet's mounting system has adjustable bands so it can be properly fitted to most people.

The software

This eye tracker uses two different programs for data collection, Yarbus for calibration and PS live capture for recording. Using these two programs allows calibration after the fact which can be very useful. Yarbus can also be used on its own to make videos, if started in live capture mode. However, calibration is then done beforehand and can't be adjusted. This approach better suits quick insight instead of precise measurement. Both methods of calibration and data collection are described in the attachment "Calibration Manual".

IBM SPSS statistics 23 was used for all statistical analysis.

Procedure

Calibration

This study consisted of two separate experiments, for both of these the calibration procedure was done in the same way described in the "Calibration Manual" in attachment 1. There are two options for calibration and recording video. The first is calibration and recording using only Yarbus and the second is recording in PS Live Capture and calibration in Yarbus. For this experiment I used recording and calibration in Yarbus so that I could directly see any potential problems. To calibrate the Positive Science Wearable eye tacker in Yarbus a minimum of 4 points need to be defined in the scene camera view. The choice was made to use 5 calibration points. This choice was made because later on in the experiments to assess the effective measurement range it is useful to have a clearly defined centre of the screen. Calibration was done with a paper calibration grid (figure 4) on a white background, because using a black background proved to be problematic because of the contrast in the scene camera image. When the white paper was in front of a black background the markings on the paper became invisible on the scene camera view. The contrast was so low that the white paper seemed to be solid white. For the other experiments in this study, calibration was done with pieces of post-it paper stuck to a wall in a grid pattern. This choice was made because it allowed more freedom in setting up a calibration grid as they can be placed in any grid size. This allows the researcher more freedom to create a calibration area of relevant size on any surface, fit for the specific task that is to be performed.

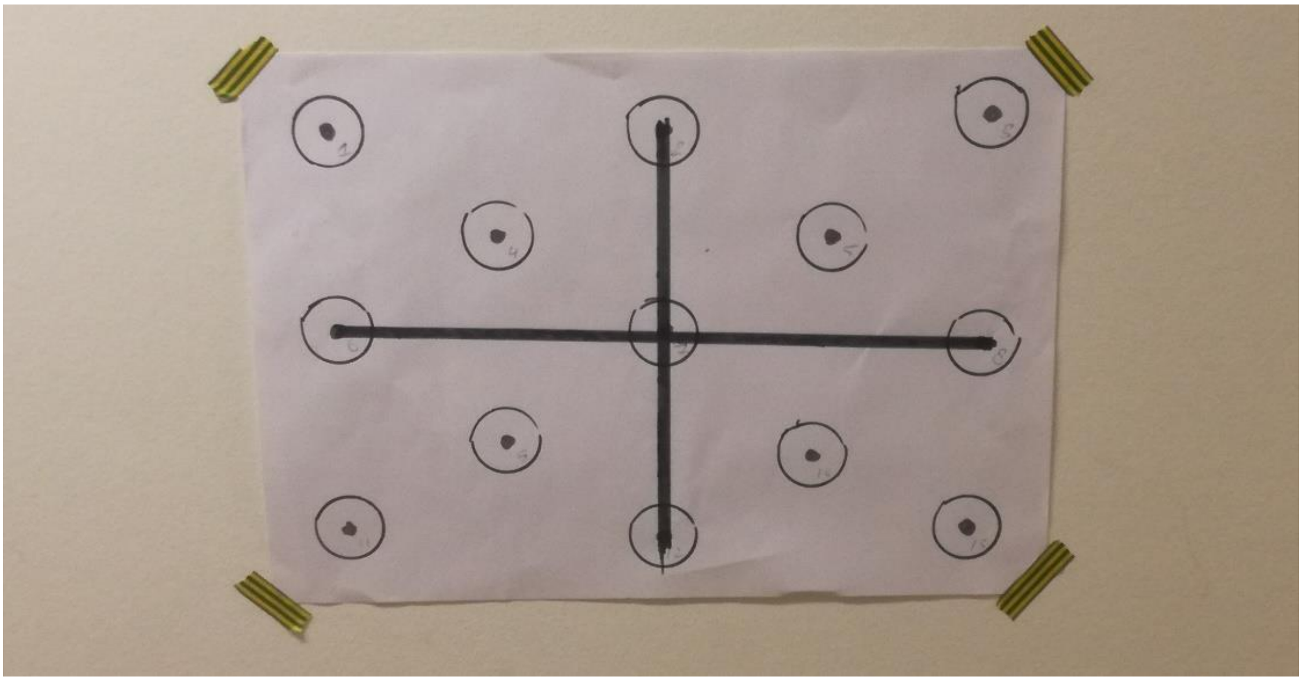


Figure 4 First version of calibration grid

Experiment 1

The aim of the first experiment was to find the extent to which the eye-tracker is capable of reporting gaze coordinates in the edges of the visual field. A test was devised to determine the area of the scene camera view wherein the gaze position could reliably be measured. The setup for this test, shown in figure 5, consisted of: a wall, a subject, the wearable eye tracking setup, a table with a chin rest to restrict movement and post-it notes to mark the relevant locations on the wall.

First of all the field of view of the scene camera was measured on the wall using a tape measure. To determine the field of view of the scene camera the corners of the area that was visible on the scene view were marked with post-it paper. The scene camera feed was monitored live as guidance. The subject was instructed to move as little as possible and keep their chin on the chin rest to keep the image steady. The visual angle of the camera was calculated by using the tangent formula.

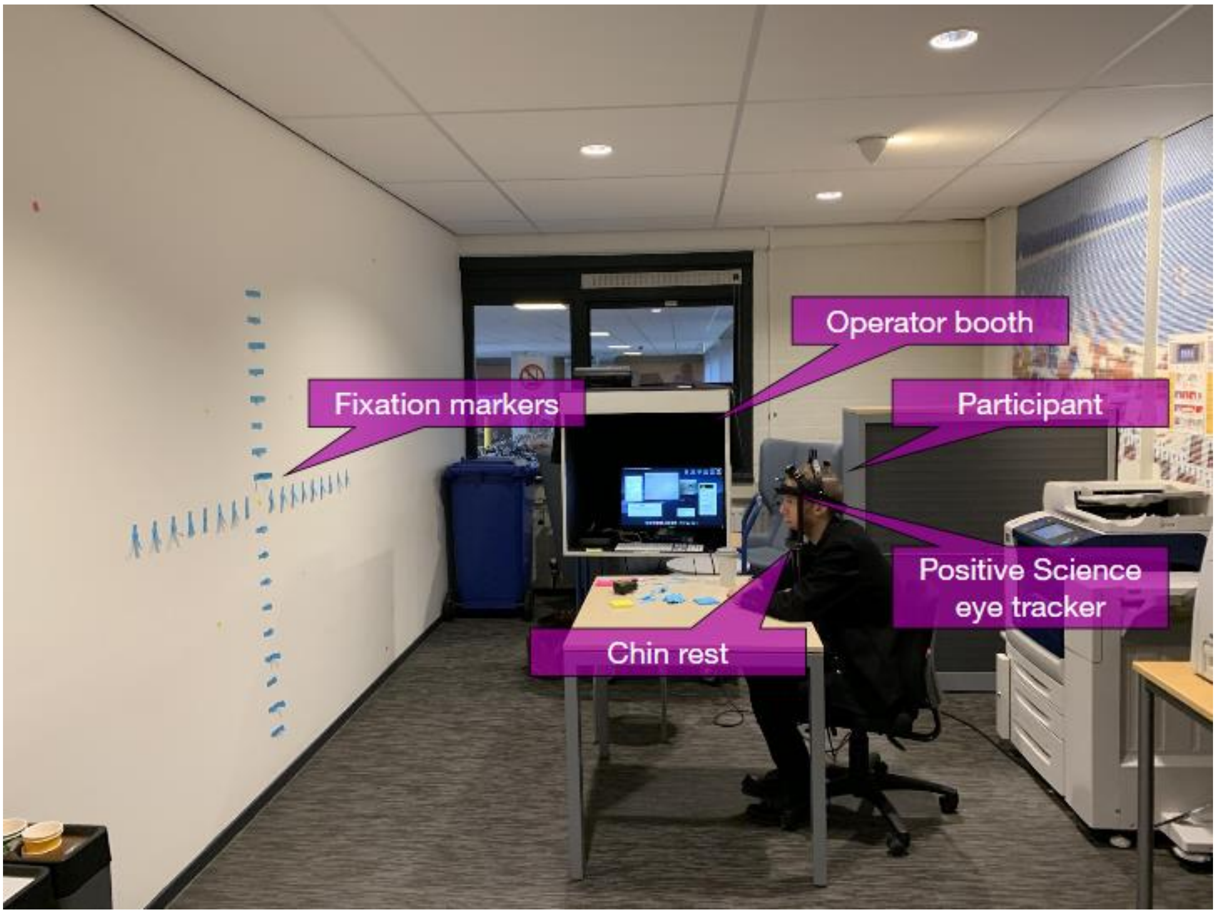


Figure 5 Test setup for assessing the range wherein measurement can reliably be taken

Secondly the effective range wherein measurement can be taken was determined. The subject was again instructed to move their head as little as possible while placed in the chin rest. The set up for this experiment can be seen in Figure 5. First a calibration was taken with the subject. The calibration frame was slightly reduced in size for one of the subjects since not all points could be reliably measured. The size of the calibration frame was 80 by 80 centimetres for 2 subjects and 60 by 60 centimetres for the other subject. After calibration the centremost point was taken as the centre of the effective field of view. Using different coloured post-its two axes were marked on the wall at 10 centimetre intervals. One of the axes horizontal and one vertical, intersecting in the centremost point of the calibration. The subject was then instructed to look at the point in the centre of the calibration frame (figure 5). Next the subject was instructed to subsequently fixate on the individual markers in order in one direction on one of the axes. Switching from marker to marker until gaze coordinates couldn't be measured anymore. This was done for all four axes in order, starting at the centre every time, thereby creating a visual representation of the axes of the effective measurement range.

Experiment 2

The effect of different or changing eye-in-head orientations on the systematic error was also investigated. The subjects were instructed to change their yaw (z axis), pitch (x axis) and roll (y axis) orientation as shown in figure 6. For this test, subjects were instructed to restrict their movements to rotation along the aforementioned axes, while not translating their head in any direction. This was done without a chinrest, because some of the rotations require more movement than the chinrest allowed. For this test 5 points were used for calibration while the subject was looking straight ahead at the calibration area. The subject was then instructed to pitch their head up until the calibration area was close to the edge of the scene view camera. In this position the subject looked at all 5 calibration points. This was done for changing orientations around all axes tilt up/down yaw right/left

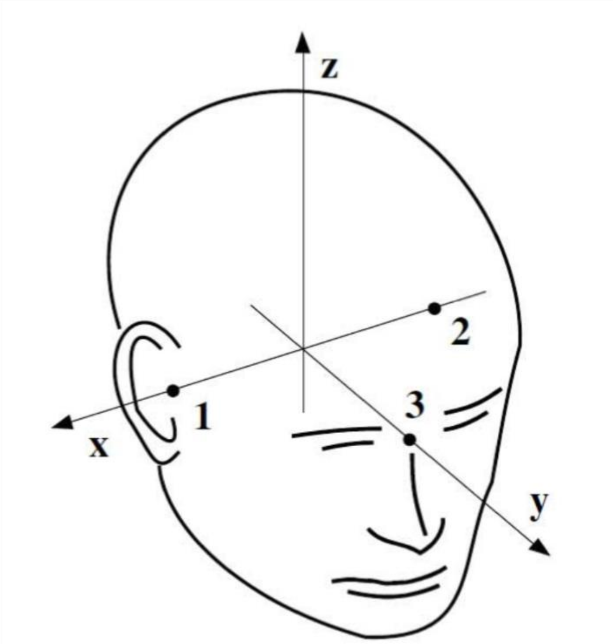


Figure 6 Head orientation axes

and roll right/left as demonstrated in figure 7. In every orientation the subject looked at all 5 points. This experiment was conducted with 5 subjects, measuring the 7 different head orientations. In all of these orientations the subject looked at 5 points. For each of these fixations 5 individual frames were captured and the distance between the measured gaze coordinates and the instructed fixation point was determined.

Straight ahead

Yaw left Yaw right Pitch up

Roll left Roll right Pitch down

Figure 7 subject demonstrating measured head orientations

This was done by measuring the number of horizontal and vertical pixels in the video frame. The absolute difference between the reported gaze position and the point the subject was instructed to look at was defined as the systematic error, measured in pixels. The number of pixels was chosen as the unit of distance since this can be directly measured from the frames that make up the individual trials in this experiment. On the surface of the wall at a distance of 2 metres 1 pixel roughly corresponds to .36 centimetres or 2.78 pixels correspond to 1 centimetre. In the centre of the screen 1 pixel roughly corresponds to 0.109 degrees.

Results

Experiment 1

The first experiment was designed to find the field of view of the scene camera and the effective measurement range. The results for these tests can be found in figure 8. This experiment was done with three different subjects. For all three subjects the effective measurement range is smaller than the field of view of the scene camera. There are differences between subjects in the size of the effective measurement range but the cause of these differences is unknown. The first subject that was measured had the largest effective measurement range. For the first and second subject the effective measurement range reached beyond the scene camera's field of view. This can be seen in the bottom area of the chart for subject 1 and at the top area of the chart for subject 2.

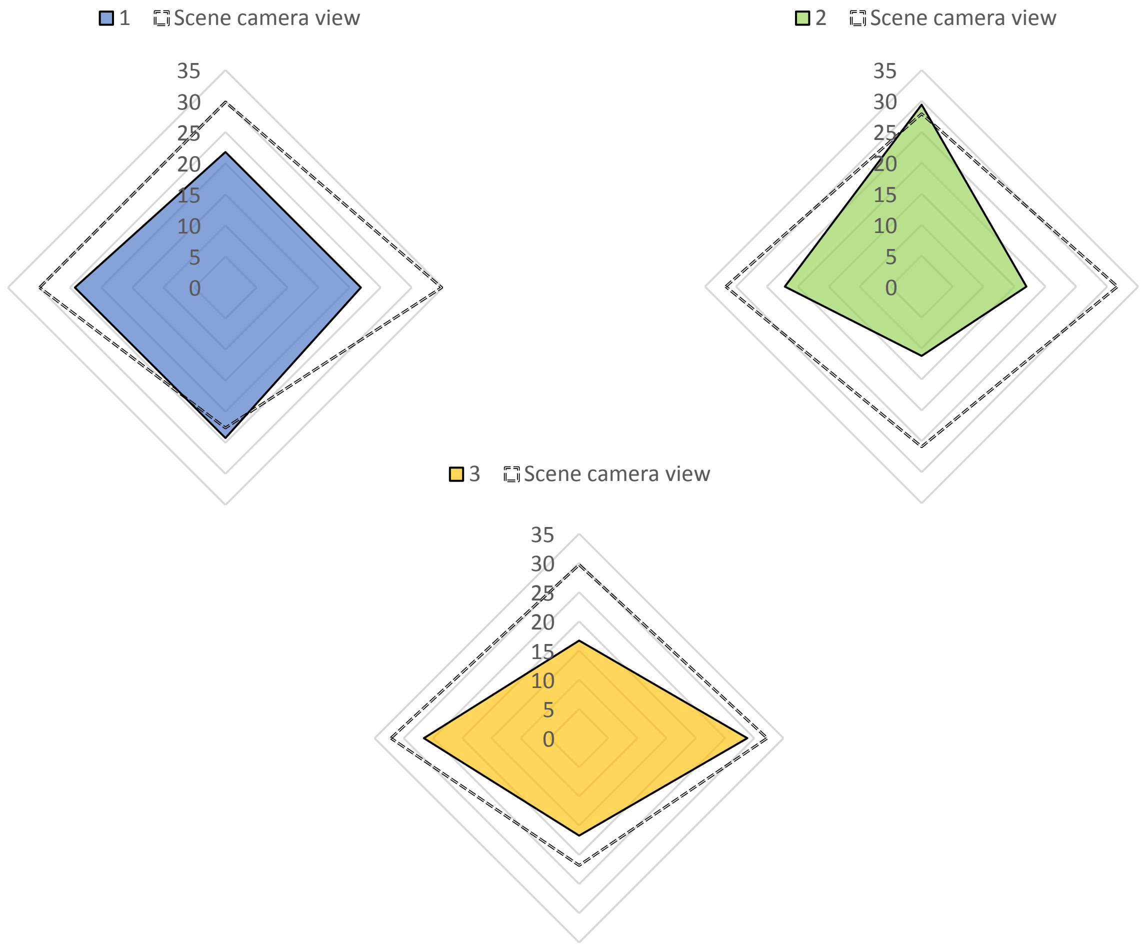


Figure 8 Effective measurement range for every subject is shown as the coloured box, calculated scene camera view per subject is shown as a dotted black line. Axis values are in degrees.

Experiment 2

First of all the average systematic error for every eye-in-head orientation was determined. The systematic error is defined as the absolute distance between the gaze coordinates measured by the eye tracker and the location that the subject was instructed to look at. The distance was measured in pixels in a horizontal and vertical component. Pythagoras theorem was used to calculate the Euclidean distance. The average systematic error for every eye-in-head orientation can be found in figure 9.

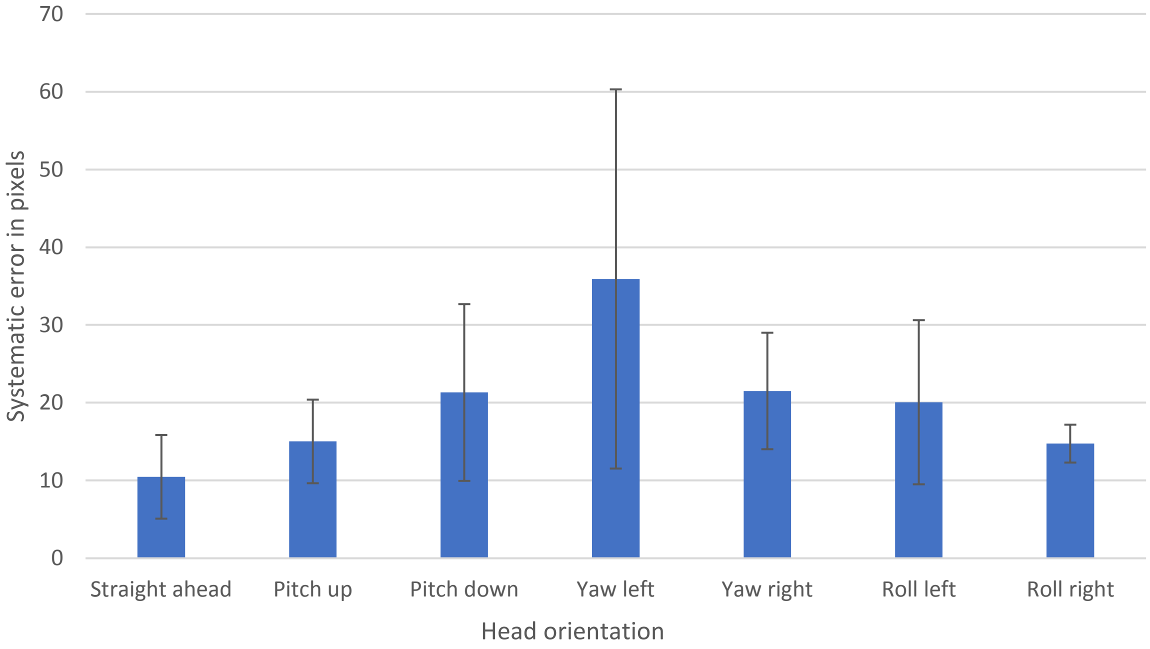


Figure 9 Average systematic error for every head orientation, values are given in pixels with standard deviation values as error bars

To calculate the significance of the differences seen in figure 9 a repeated measures ANOVA was executed in SPSS 23. Sphericity could not be checked using SPSS since the number of subjects was lower than the number of levels in the repeated measures factor. However, I decided to assume the data are spherical, since this allows me a more liberal criterium for rejecting the null-hypothesis.

Assuming the data are spherical the differences in systematic error between the individual head orientation are significant *F*(6,24)=3.60 p=.011. Therefore it is likely that specific head orientations influence the systematic error for the Positive Science wearable eye tracker. Post-hoc tests of individual differences weren't significant however. This is probably because the effect was small to begin with, and the Bonferroni correction is quite strong.

The average systematic error for every fixation point results can be found in figure 10. The same dataset is used for both analysis methods in experiment 2, they are grouped differently however. The first method groups the fixation points together for every different head orientation. The second method groups the head orientations together for every different fixation point.

The differences in systematic error for every fixation point were analysed in SPSS 23 using a repeated measures ANOVA.

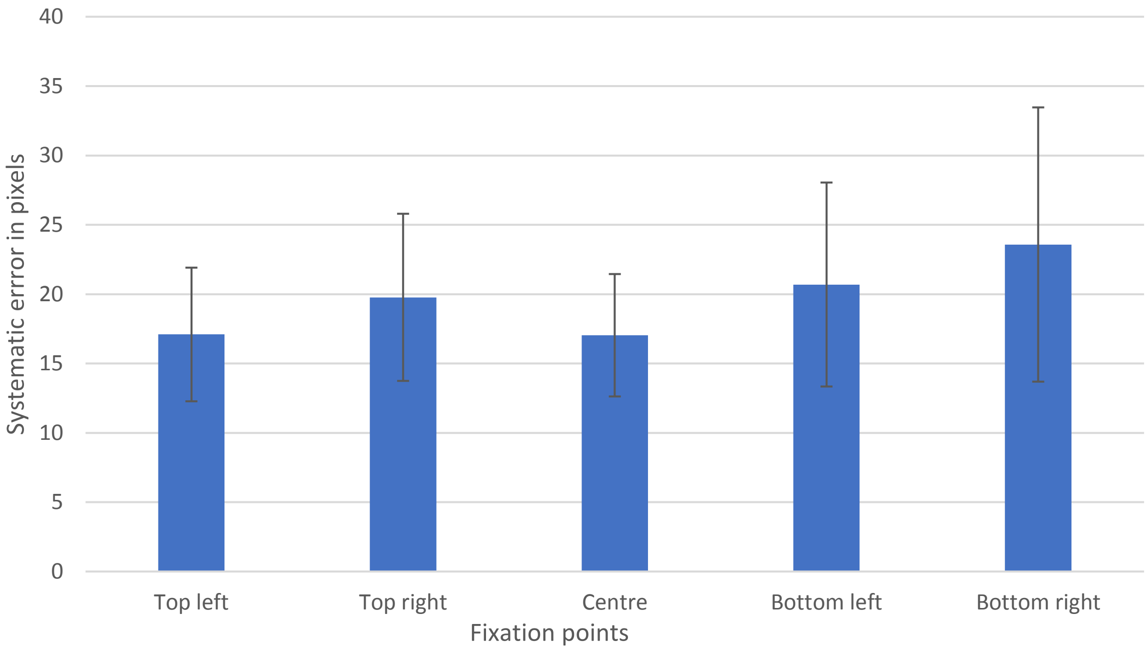


Figure 10 average systematic error for every fixation point, values are given in pixels with standard deviation as error bar values.

The sphericity of the systematic error for every fixation point was checked with Mauchly's test of sphericity. This differences in variance were significant p<.001, therefore sphericity was violated and the Greenhouse-Geisser correction was used for the analysis. The differences in systematic error between specific fixation points weren't significant *F*(1.42,567)=1.20, p=.34. This means that there is no significant influence of looking at specific fixation points independent of head orientation on systematic error.

Discussion

In this study I looked into the influence of and eye-in-head orientation on the data quality of the Positive Science wearable eye tracker. I also measured the field of view and determined the effective measurement range of the Positive Science wearable eye tracker. The experiment designed to calculate the effective measurement range made it apparent that the effective measurement range is smaller than the scene camera image. This means that gaze data can't be collected for every location that is visible on the scene camera image. If it is visible it isn't necessarily measurable. There were however unexplained interpersonal differences.

Two of the three subjects that were measured had an effective measurement range that partly extended outside the scene camera image. The first subject's effective measurement range extended slightly below the scene camera image while the top of the scene camera image couldn't be measured. For the second subject this was exactly the other way around. This could mean that if the cameras were set up slightly different the entire height of the scene camera view might have been measured. This can't be said for sure though. However, this finding shows the independence of the effective measurement range and the scene camera image. This is important to keep in mind when setting up an experiment with this eye tracker. Especially since there are interpersonal differences in the size of the effective measurement range. If the cameras are properly setup for one subject, that isn't a guarantee for another subject, both cameras should be carefully positioned for every subject.

The second experiment was designed to determine the influence of specific eye in head orientations on the systematic error of the eye tracker. The data for this experiment was grouped in two different ways. These two groups are head orientations averaged over fixation points and fixation points averaged over head orientations. There was an influence of specific head orientations on the systematic error of measurement. Post-hoc tests didn't show any significant differences between individual groups however. Sphericity of the data was assumed. This choice was made in order to have a more liberal criterion for rejecting the null hypothesis. The possibility of incorrectly rejecting the null hypothesis seems like the lesser of two evils. Ignoring an important effect could lead to more problems with research findings than overestimating a negative influence on data quality. The post-hoc tests didn't show any significant individual differences between head-orientations. However, numerically the smallest systematic error was found in the "straight ahead" head orientation. This was to be expected since this condition is closest to the calibration conditions. Data quality focused research with remote eye trackers has shown that the systematic error is smallest when the measurement conditions are closest to the calibration conditions (Niehorster et. al., 2018). The effect of head orientation on systematic error should be explored further in the future in order to confirm or disprove these findings.

Besides effective measurement range and eye-in-head orientation other relevant influences on data quality were discovered for the Positive Science wearable eye tracker. Specifically, calibration, camera positioning and contrast in the scene camera image. Next, these influences will be individually elaborated on. First, for successful calibration with the positive science wearable eye tracker 4 calibration points need to be defined. These calibration points need to be created with the eye being in different orientations relative to the eye camera. This is necessary for the eye tracker to be able to correctly estimate the coordinates of all other possible gaze locations. This might prove difficult in infant studies however. Since infants usually look more with their head than with their eyes. By this I mean that they turn their entire head towards the thing they are looking at, more than adults (Daniel & Lee, 1990). If this happens, the calibration will be unreliable since the eye-in-head orientation doesn't change. Therefore this could pose problems in infant research. The second possible problem with calibration is the fact that infants often grab at the eye tracker pulling it off their head or moving it out of place, as mentioned in a recent article (Mulder, van Houdt, van der Ham, van der Stigchel en Oudgenoeg-Paz, 2020). The problem with this is that the calibration procedure needs to be repeated if this happens. If the eye tracker is moved with respect to the subject's head, the calibration would become invalid. Every time the infant touches the device in such a way that one of the cameras moves or the entire thing shifts on his head the procedure needs to be repeated. This is a really big problem for research with the positive science wearable eye tracker. Even if calibration can be done quickly and reliably, this will still happen during measurement. There is a wearable eye tracker for which removing and replacing it isn’t really a problem, the Tobii glasses pro. This is because the aforementioned eye tracker uses a digital model of the eye and automatic calibration. If the subject removes this during measurement and places it back in a slightly different orientation the eye-tracker automatically adjusts for this shift. However, the Tobii glasses are too big to be used in infant research, so at the moment this isn’t a solution.

Camera positioning was challenging with the positive science wearable eye tracker. This is because the eye camera is attached to a very flexible arm, which makes it difficult to move precisely. If the eye camera is pushed into the desired position it has a slight elasticity that moves it out of alignment again. Therefore it is important to be careful when positioning the eye camera and to make sure you check the position when not touching the camera.

Another issue I found with the positive science wearable eye tracker was the contrast in the scene camera image. For the first subject that was measured in the eye-in-head orientation task the

Figure 11 example of bad scene camera contrast, the centre piece of post-it paper is visible, but only just, the four corners are almost impossible to distinguish calibration points were pieces of yellow post-it paper. These yellow pieces of paper on the white wall were clearly visible to the subject. In analysis however this wasn't the case. Since the contrast of the scene camera was quite poor the yellow post-it pieces were almost indistinguishable from the background, as can be seen in figure 11.

This could cause problems in analysis, but it is prevented easily enough by making sure overlapping or touching objects have high enough contrast. In this study the choice was made to use pink pieces of post-it paper instead going forward.

Conclusion

The positive science wearable eye tracker is the only commercially available wearable eye tracker fit for infant research. It has some limitations, like any other eye tracker. The effective measurement range isn't necessarily the same size as the eye camera view. The subject's head orientation can influence the amount of systematic error that occurs in measurement. Calibration can be difficult with infant subjects and the camera's need to be carefully positioned. All of these challenges should be kept in mind when planning an experiment with this eye tracker in an attempt to keep data quality as high as possible.

References

Aslin, R. N. & McMurray, B. (2004). Automated corneal-reflection eye tracking in infancy: Methodological developments and applications to cognition. *Infancy,* 6(2), 155-163.

Buswell, G. T. (1935). How people look at pictures: a study of the psychology and perception in art.

Daniel, B. M. & Lee, D. N. (1900). Development of looking with head and eyes. *Journal of experimental child psychology,* 50, 200-216

Dodge, R., & Cline, T. S. (1901). The angle velocity of eye movements. *Psychological Review*, *8*(2), 145.

Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers*, 34(4), 455-470.

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-mounted eye tracking: A new method to describe infant looking. *Child development*, *82*(6), 1738-1750.

Gredebäck, G., Johnson, S., & von Hofsten, C. (2009). Eye tracking in infancy research. *Developmental neuropsychology*, *35*(1), 1-19.

Haith, M. M. (1969). Infrared television recording and measurement of ocular behavior in the human infant. *American Psychologist*, *24*(3), 279.

Hessels, R. S., Andersson, R., Hooge, I. T., Nyström, M., & Kemner, C. (2015). Consequences of eye color, positioning, and head movement for eye-tracking data quality in infant research. *Infancy*, *20*(6), 601-633.

Hessels, R. S., Cornelissen, T. H., Kemner, C., & Hooge, I. T. (2015). Qualitative tests of remote eyetracker recovery and performance during head rotation. *Behavior Research Methods*, *47*(3), 848-859.

Hessels, R. S., van Doorn, A. J., Benjamins, J. S., Holleman, G. A., & Hooge, I. T. (2020). Task-related gaze control in human crowd navigation. *Attention, Perception, & Psychophysics*, 1-20.

Kochukhova, O., & Gredebäck, G. (2010). Preverbal infants anticipate that food will be brought to the mouth: an eye tracking study of manual feeding and flying spoons. *Child development*, *81*(6), 1729-1738.

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(11), 1311-1328.

Mele, M. L., & Federici, S. (2012). Gaze and eye-tracking solutions for psychological research. *Cognitive processing*, *13*(1), 261-265

Mulder, H., van Houdt, C. A., van der Ham, I. J. M., van der Stigchel, S., Oudgenoeg-Paz, O. (2020). Attentional flexibility predicts a-not-b task performance in 14 month old infants: a head-mounted eye tracking study. *Brain Sciences*, 10(279)

Niehorster, D. C., Cornelissen, T. H., Holmqvist, K., Hooge, I. T., & Hessels, R. S. (2018). What to expect from your remote eye-tracker when participants are unrestrained. *Behavior research methods*, *50*(1), 213-227.

Niehorster, D. C., Santini, T., Hessels, R. S., Hooge, I. T., Kasneci, E., & Nyström, M. (2020). The impact of slippage on the data quality of head-worn eye trackers. *Behavior Research Methods*, 1-21.

Nyström, M., Andersson, R., Holmqvist, K., & Van De Weijer, J. (2013). The influence of calibration method and eye physiology on eyetracking data quality. *Behavior research methods*, *45*(1), 272-288.

Orquin, J. L., & Holmqvist, K. (2018). Threats to the validity of eye-movement research in psychology. *Behavior research methods*, *50*(4), 1645-1656.

Palinko, O., Kun, A. L., Shyrokov, A., & Heeman, P. (2010, March). Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 symposium on eye-tracking research & applications* (pp. 141-144)

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3), 372.

Steciuk, H., & Zwierko, T. (2015). Gaze behavior in basketball shooting: Preliminary investigations. *Trends in Sport Sciences*, *22*(2).

Yu, C., & Smith, L. B. (2016). The social origins of sustained attention in one-year-old human infants. *Current Biology*, *26*(9), 1235-1240.

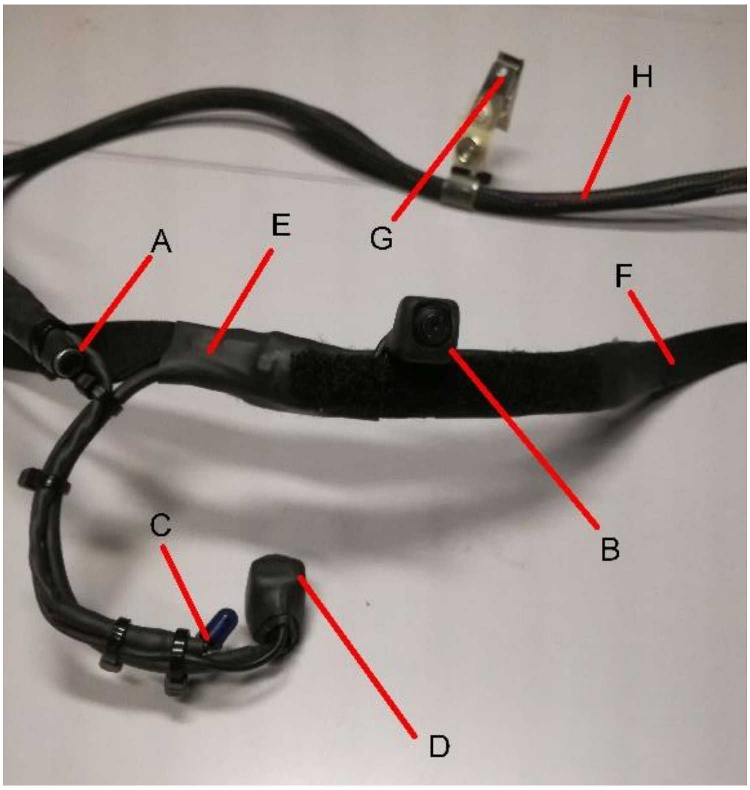
Attachment 1: Positive Science wearable eye-tracker Calibration and start-up procedure

- 1. Components
- 2. Software
- 3. Step-by-step calibration procedure

1 Components

The Positive Science wearable eye-tracker has a very simple lightweight design.

A: the microphone, to record sound alongside the video
B: the scene-camera, to record the surrounding scene
C: an infrared light, to illuminate the eye
D: the eye-camera, to record eye-movements
E: the frame, a semi-flexible bar with Velcro attached to the back, for easy and adjustable mounting
F: extra Velcro strip, to mount the eye-tracker more securely
G: clip to attach the cable to a participants clothing, to keep it out of the way
H: data and power cable, to transmit data to the computer



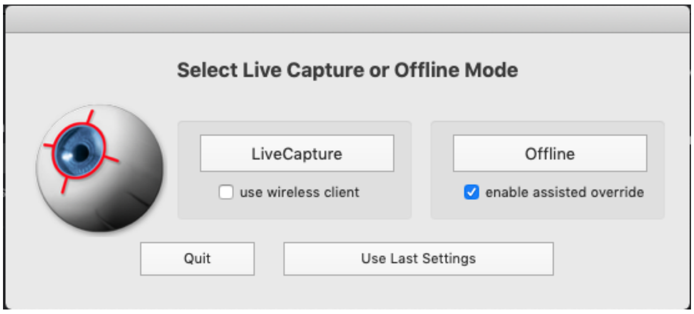
2 Software

The programs used for this eye-tracker are PS live capture and Yarbus. PS live capture is useful for the actual recording and Yarbus is used for calibrating the eye-tracker after the fact. Yarbus can also be used to create videos but these can't be calibrated afterwards. If this eye-tracker is used in the field and there isn't a lot of time to take measurements it is very useful to quickly check the quality of the setup in Yarbus and then record the video in PS live capture. This allows you to take as much time as needed to properly calibrate the video later on as long as the calibration sequence has been properly recorded. This can be achieved by making sure that while shooting the video there are clear visual or auditory ques when the subject looks at one of the calibration points. In infant research for example this can be very useful, since the time spent with the subject is reduced, reducing the chances of the participant becoming fussy or difficult without losing accuracy of measurement. Assuming that this is the goal the combining of these programs is very useful, for simple visualisation of the capabilities of this eye-tracker it might be easier to just use Yarbus instead, since the eventual calibration is done in this program anyway and it can also record video. This might take more time while taking measurement, but it will take less time to process the videos and you don't have to calibrate after the fact. This can be useful for instance when using the eye-tracker in adult research, these participants can usually be instructed more clearly and follow instructions more clearly than infants. This also allows you to immediately show the recorded video back to the participant, which they might find very interesting and gives them a reason to help you out with research.

The two different ways you can use the program are both in the step-by-step procedure, From step 18 it splits into two

3 Step-by-step calibration procedure

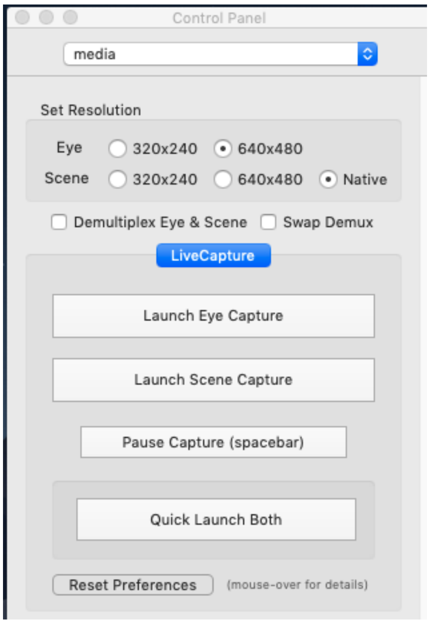
- 1: Start up the computer
- 2: attach the hardware to your subject
- 3: connect the cables to the computer, while making sure they can't get tangled or caught between objects to easily
- 4: find Yarbus and start it up



Select Live Capture or Offline Mode

- 5: start up in LiveCapture

- 6: This gives you the start-up screen
- 7: Launch eye-capture, in the left control panel



Control Panel

media

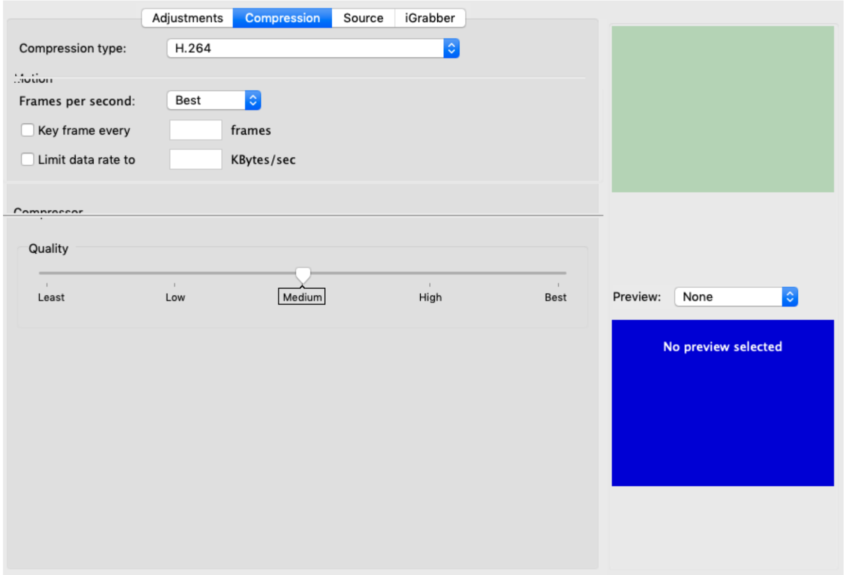
Set Resolution

Eye ○ 320x240 ● 640x480

Scene ○ 320x240 ○ 640x480 ● Native

☐ Demultiplex Eye & Scene ☐ Swap Demux

- 8: if the eye-tracker isn't (properly) connected it looks like the screen below



Compression type: H.264

Frames per second: Best

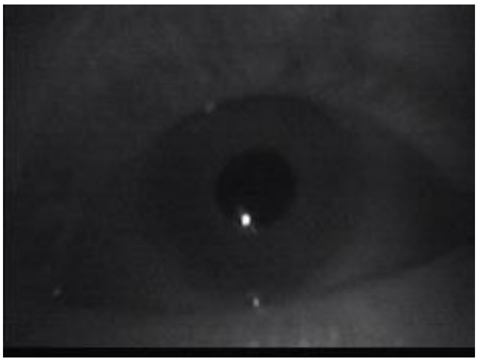
☐ Key frame every frames

☐ Limit data rate to KBytes/sec

Preview: None

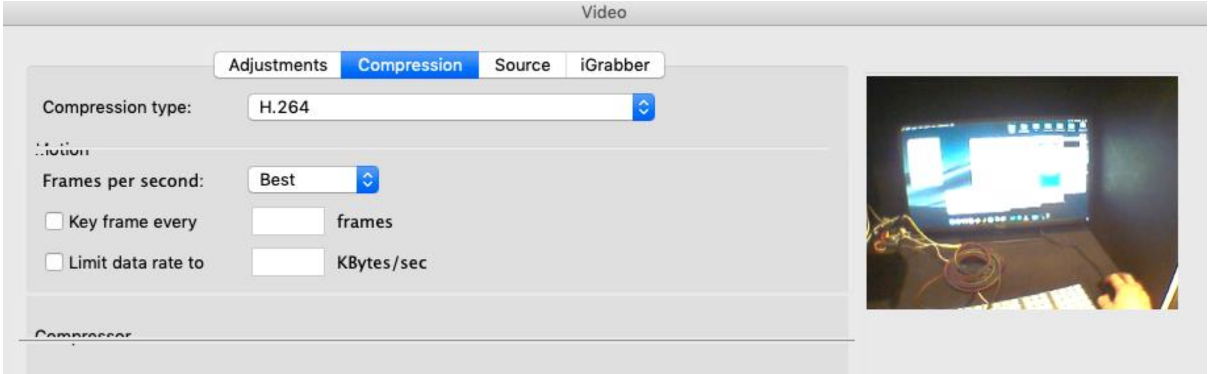
No preview selected

- 9: if the cables are properly connected it will show a preview of the eye like this



- 10: if the eye isn’t properly visible the camera needs adjusting, but not yet.
- 11: after the input has been selected for the eye camera you get back to the start-up screen, with the addition of the eye-camera view

- 12: launch scene capture in the left control panel (see step 7)the scene capture launcher is basically the same as the eye-camera launcher, with different video input



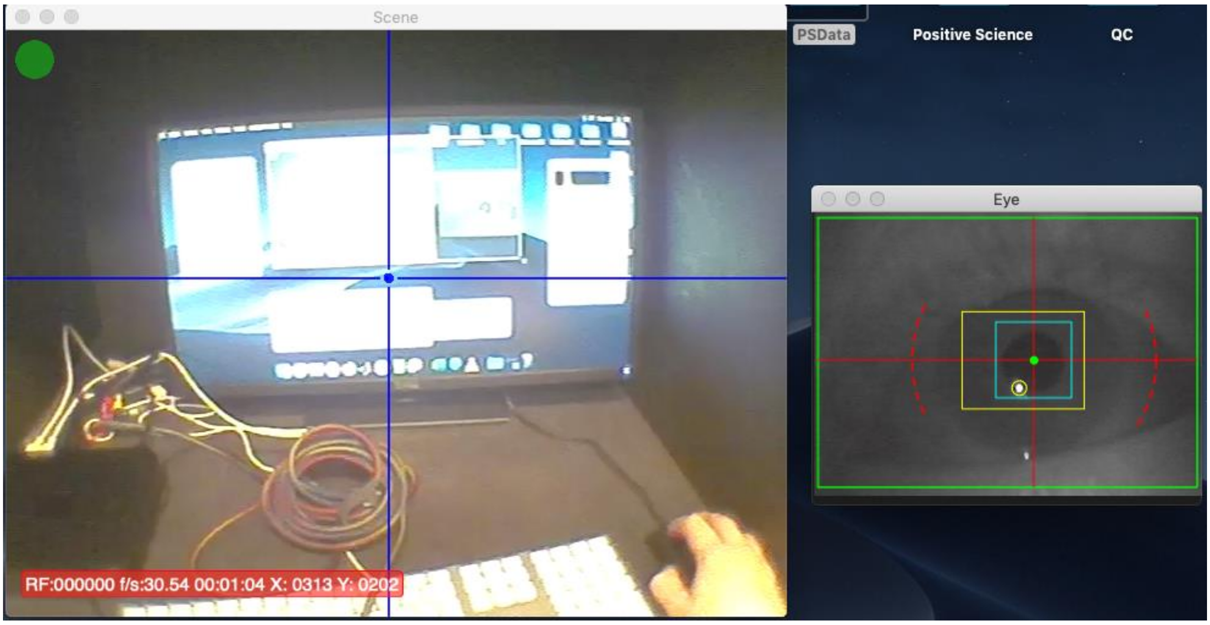
Compression type: H.264

Frames per second: Best

☐ Key frame every frames

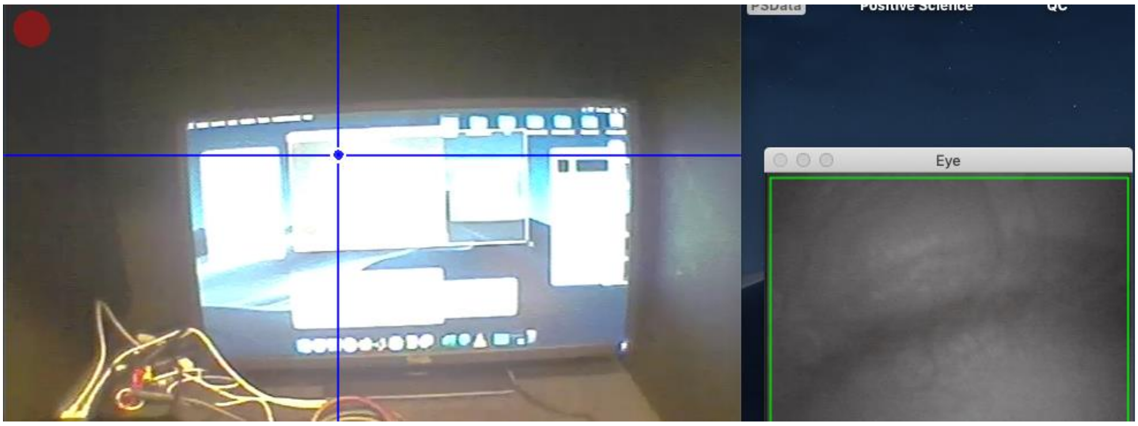
☐ Limit data rate to KBytes/sec

- 13: after launching both scene and eye cameras the input should look something like this:

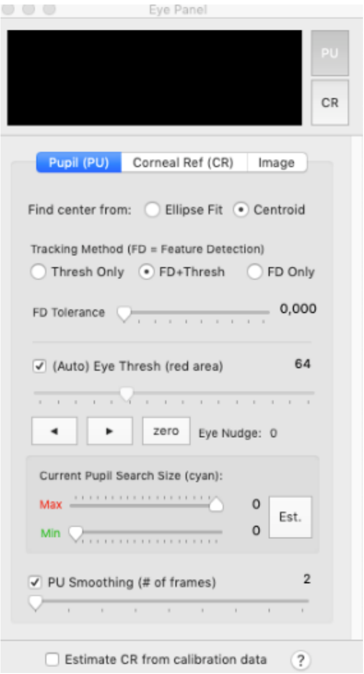


RF:000000 f/s:30.54 00:01:04 X: 0313 Y: 0202

- 14: note the green dot in the top left corner and the markers in the eye view camera on the right. These are automatically generated when the program properly detects everything. The yellow circle is the corneal reflection, the red crosshairs shows the centre of the pupil and the blue square show the estimated pupil size. In the scene view the blue crosshairs correspond to the calculated view of the participant and the dot in the top left corner indicates the “confidence” the computer has that it can properly “see” the subjects eye, this can be either green, certain, yellow, not certain or red, certainly not. As seen below when the subject closes his eyes.



- 15: as stated before, the blue crosshairs shows where the computer has calculated the subject is looking, but in this case it can be misleading, since calibration data is saved by Yarbus, meaning that it isn’t right until you calibrate with the current subject.
- 16: check whether the scene camera is set-up properly, are all relevant things visible on the scene camera or does it need to be adjusted (adjusting the scene camera first is important, because if the entire eye-tracker needs to be moved on the subjects head the eye-camera will automatically be moved)
- 17: check the eye-camera view, are all parts of the eye properly visible?
- 18: have the subject look around, can you collect data for all relevant parts of the screen? Check whether or not the dot in the top left corner stays green throughout.
- 18.2: if this is not the case but the eye is properly visible you could adjust the pupil tracking settings, in the right control panel. you can change the tolerance for pupil tracking or the way it is tracked over time, try to make sure the dot is green most of the time



Find center from: ○ Ellipse Fit ● Centroid

Tracking Method (FD = Feature Detection)
○ Thresh Only ● FD+Thresh ○ FD Only

FD Tolerance 0,000

☑ (Auto) Eye Thresh (red area) 64

Eye Nudge: 0

Current Pupil Search Size (cyan):
Max 0
Min 0

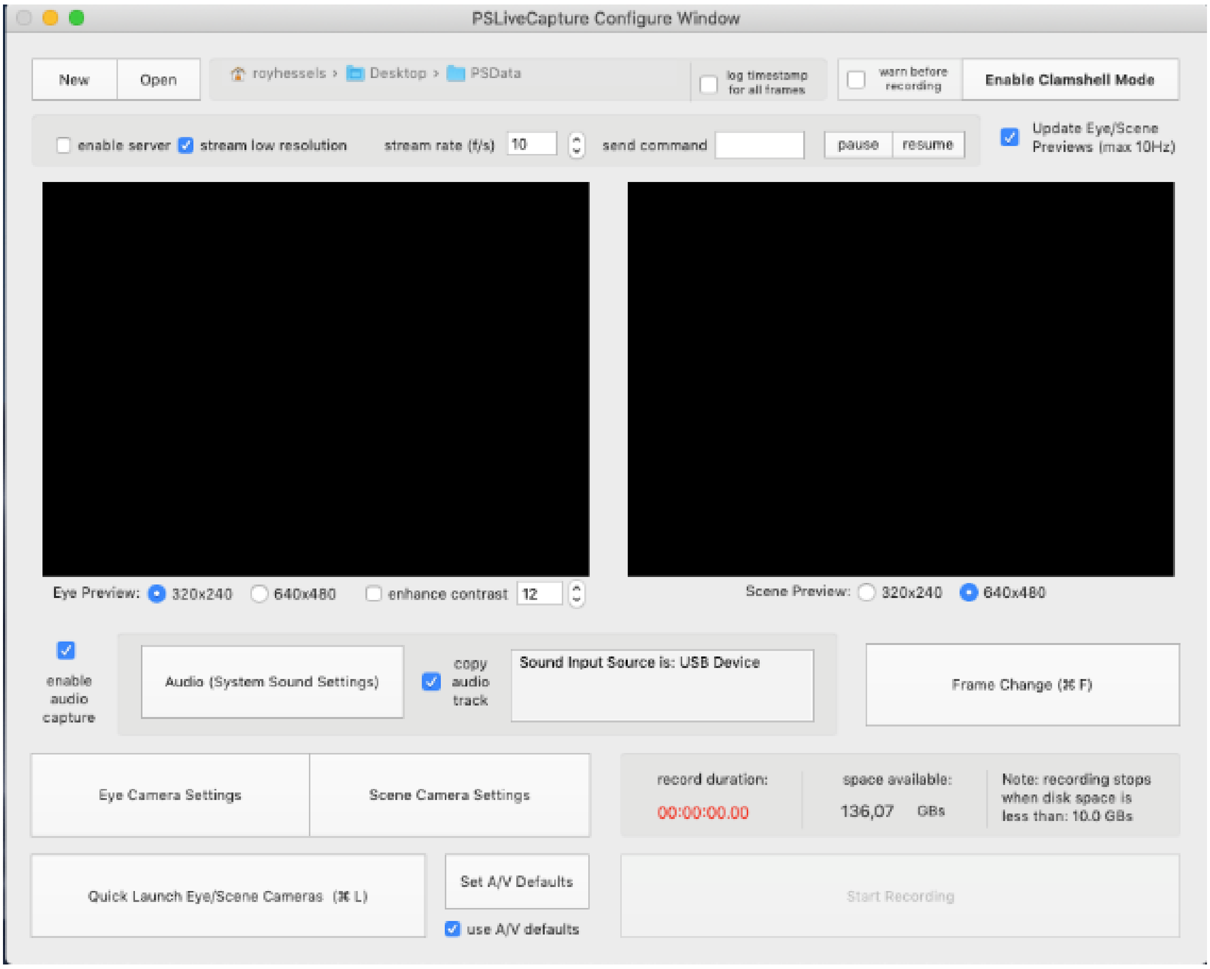
☑ PU Smoothing (# of frames) 2

☐ Estimate CR from calibration data

Choose your option

From here on the step-by-step splits into two options, option 1 is calibration after the fact using both Yarbus and PS live capture, option 2 is calibration and filming using only Yarbus. Option 1 continues here, option 2 continues on page 13.

- 19: After checking if everything is set up properly, you close Yarbus and start up PS live capture, where you'll see this start-up screen



☐ enable server ☑ stream low resolution stream rate (f/s) 10

☑ Update Eye/Scene Previews (max 10Hz)

Eye Preview: ● 320x240 ○ 640x480 ☐ enhance contrast 12

Scene Preview: ○ 320x240 ● 640x480

☑ enable audio capture

☑ copy audio track

Sound Input Source is: USB Device

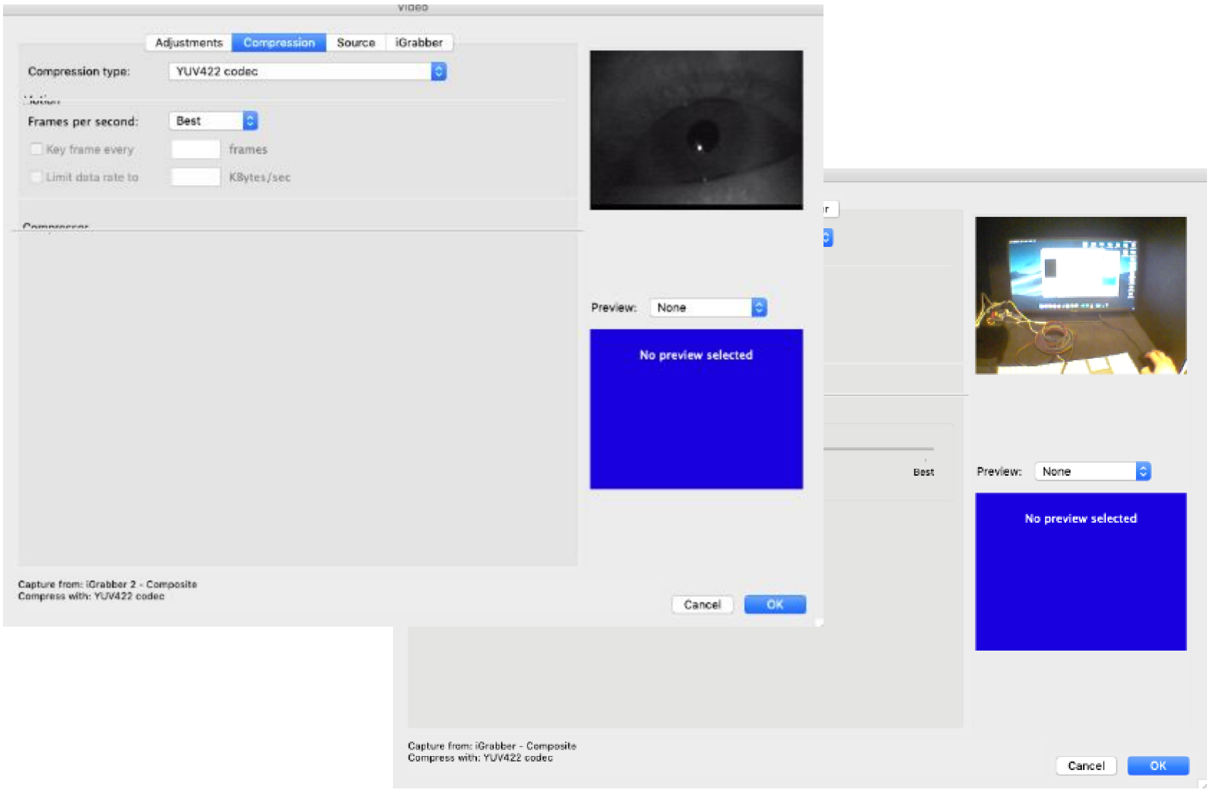
record duration: 00:00:00.00

space available: 136,07 GBs

Note: recording stops when disk space is less than: 10.0 GBs

☑ use A/V defaults

- 20: now you have to choose the inputs for your recording with the quick launch Eye/Scene cameras
- 21: this opens both cameras in separate screens, you can check the settings and quality again but it should be the same as in Yarbus so you don't need to do anything besides start the feed from both cameras by pressing OK.



Compression type: YUV422 codec

Frames per second: Best

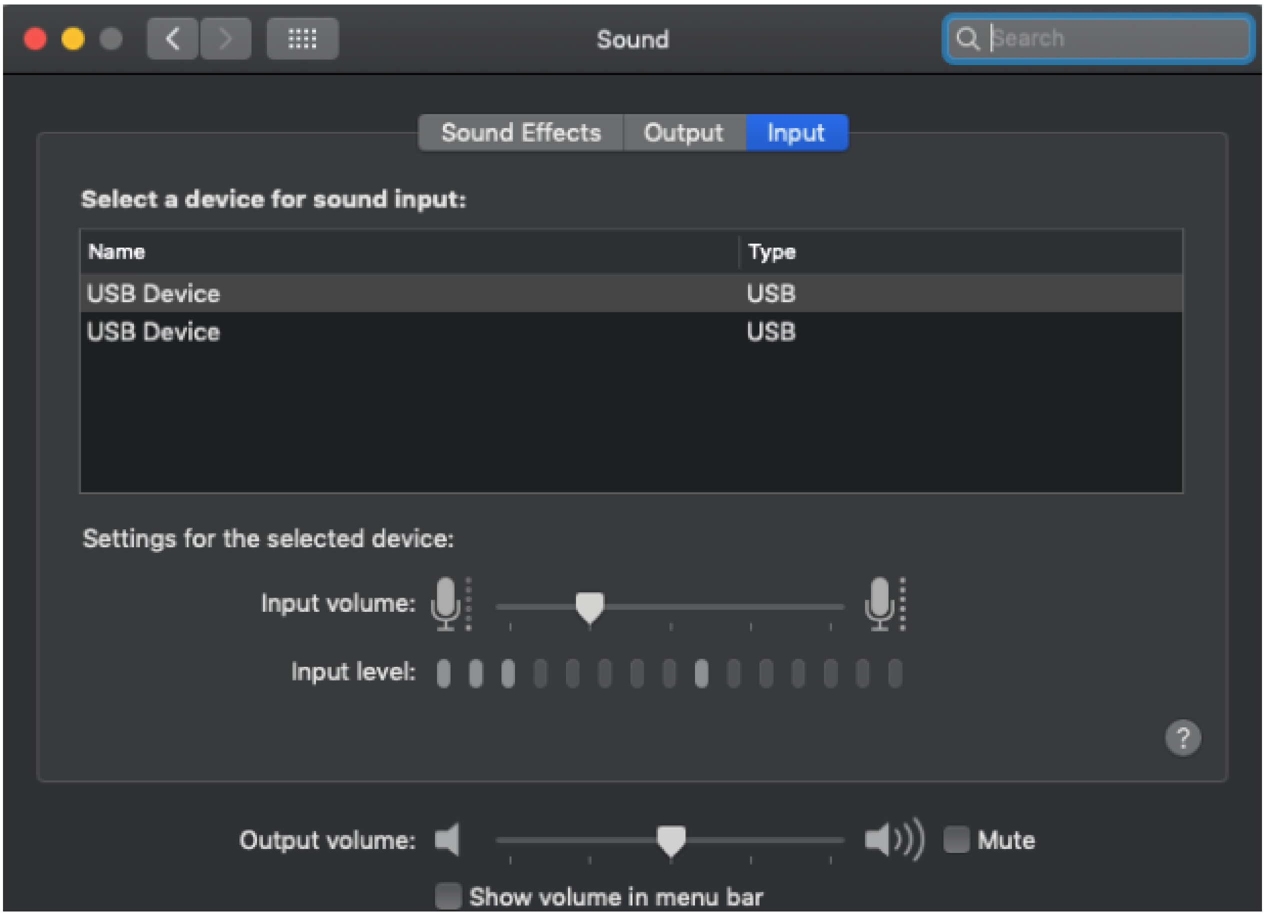
☐ Key frame every frames

☐ Limit data rate to KBytes/sec

Capture from: iGrabber 2 - Composite
Compress with: YUV422 codec

Capture from: iGrabber - Composite
Compress with: YUV422 codec

- 22: the next step is opening the sound input, this is done by selecting a USB device in the setup screen shown below. Make sure the device you choose captures sound by checking the bar at the bottom of the screen for input



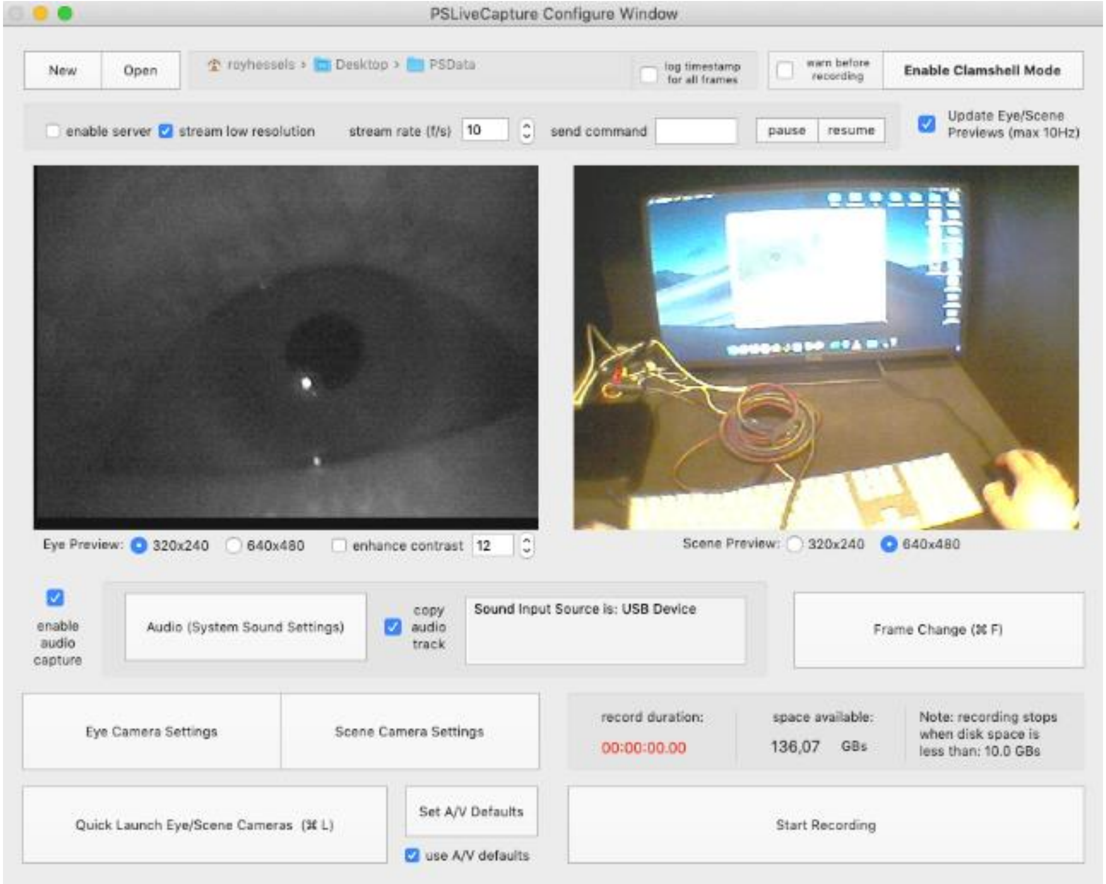
Select a device for sound input:

Name	Type
USB Device	USB
USB Device	USB

☐ Mute

☐ Show volume in menu bar

- 23: after you've done this the main screen of the program should look like the picture below. If everything checks out you can start recording by pressing the button at the bottom right of the screen.



roreal royhessels › Desktop › PSData

☐ log timestamp for all frames

☐ warn before recording

☐ enable server ☑ stream low resolution

stream rate (f/s) 10

send command

☑ Update Eye/Scene Previews (max 10Hz)

Eye Preview: ● 320x240 ○ 640x480 ☐ enhance contrast 12

Scene Preview: ○ 320x240 ● 640x480

☑ enable audio capture

☑ copy audio track

Sound Input Source is: USB Device

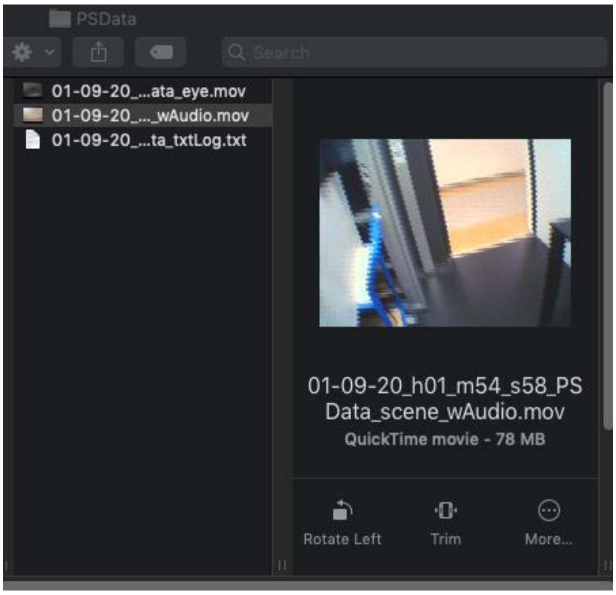
record duration: 00:00:00.00

space available: 136,07 GBs

Note: recording stops when disk space is less than: 10.0 GBs

☑ use A/V defaults

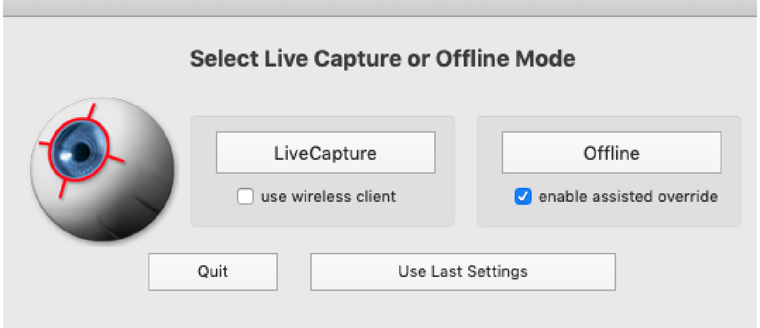
- 24: after you've collected your data and stopped recording you get three different save files in the target folder. One of the filenames ends in eye, this is the video file from the eye-camera. Another file ends in scenewAudio, this is the video file from the scene camera, including the audio from the video.



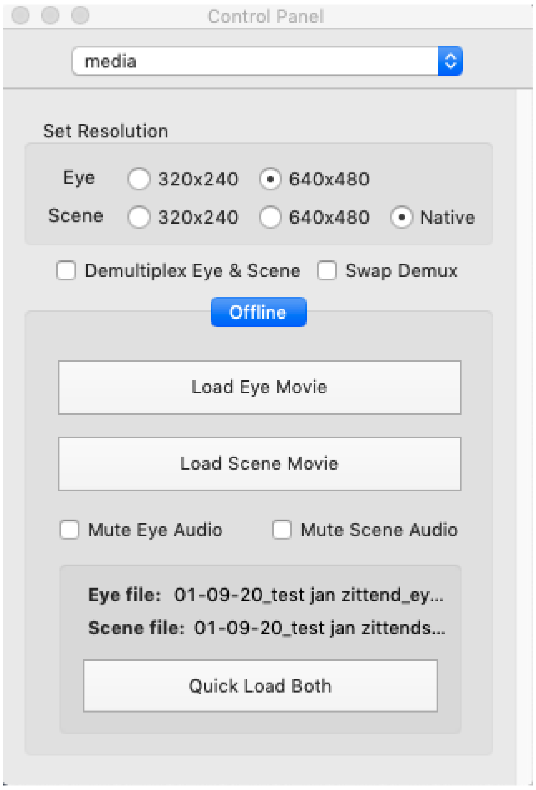
01-09-20_...ata_eye.mov
01-09-20_..._wAudio.mov
01-09-20_...ta_txtLog.txt

01-09-20_h01_m54_s58_PSData_scene_wAudio.mov
QuickTime movie - 78 MB

- 25: next up is the calibration procedure, you have to once again start up Yarbus but this time in the offline mode.

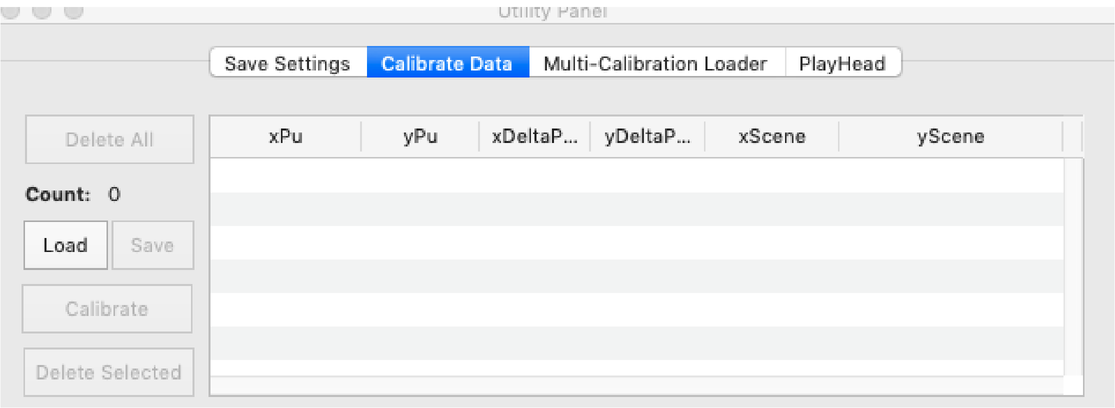


☐ use wireless client
☑ enable assisted override



Eye ○ 320x240 ● 640x480
Scene ○ 320x240 ○ 640x480 ● Native
☐ Demultiplex Eye & Scene ☐ Swap Demux
☐ Mute Eye Audio ☐ Mute Scene Audio
Eye file: 01-09-20_test jan zittend_ey...
Scene file: 01-09-20_test jan zittends...

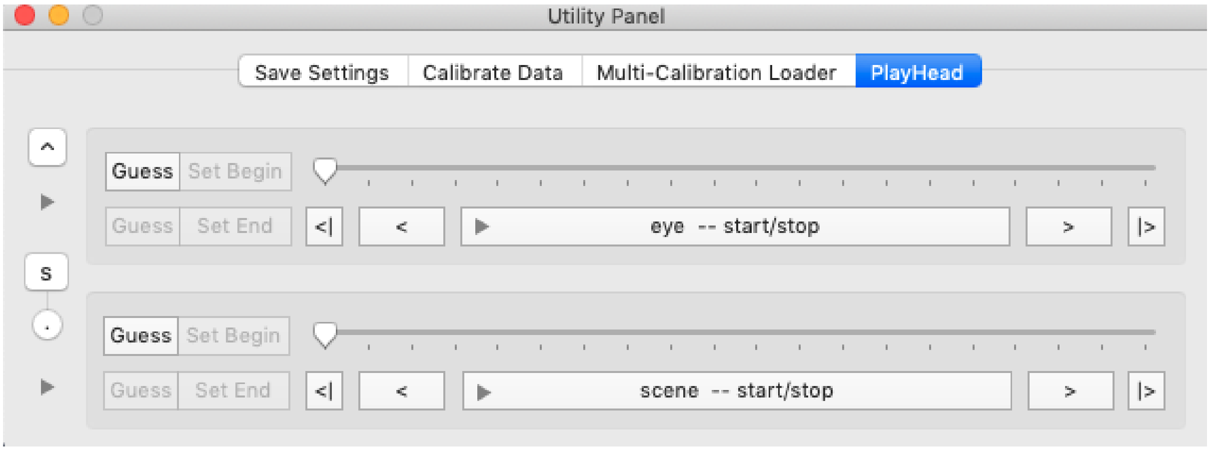
- 26: if Yarbus is started in offline mode you can load the different files into the program so you can calibrate them.
- 27: if all files have been loaded you need to make sure the files are synchronized in the utility panel, this can be done in the PlayHead tab on the right of utility panel



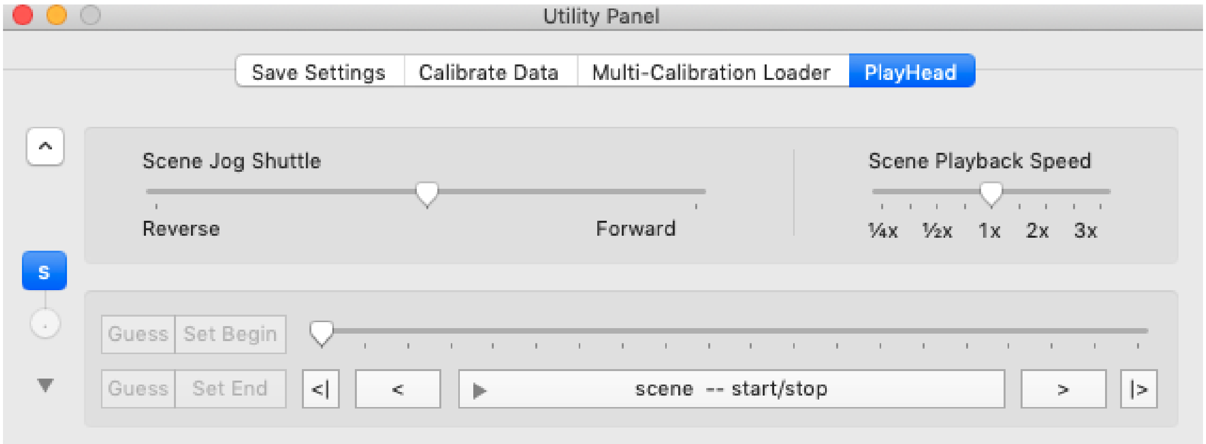
Count: 0

xPu	yPu	xDeltaP...	yDeltaP...	xScene	yScene

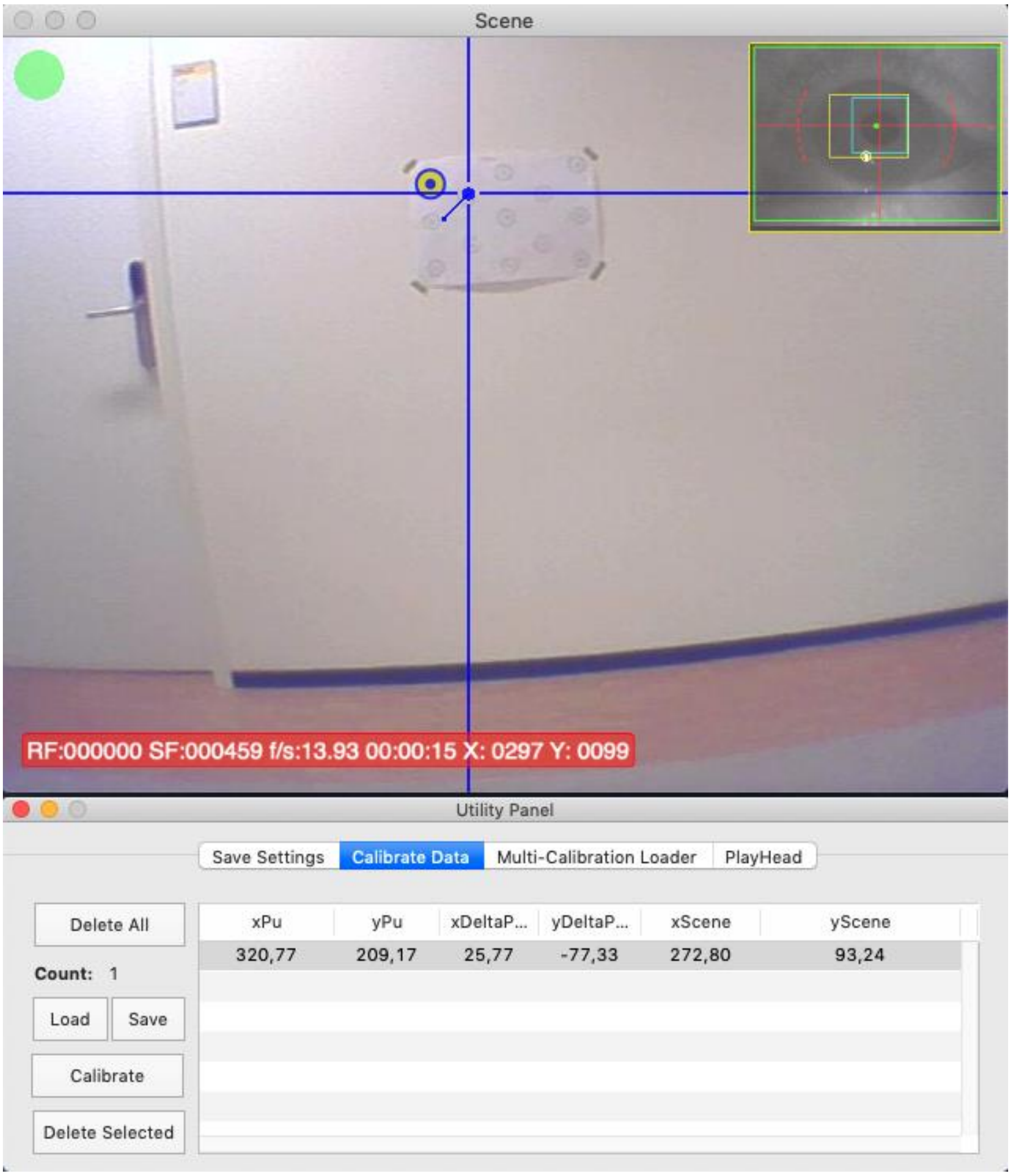
- 28: to synchronize the players you simply press the S button on the left of the screen. After you've done this the screen should go from looking like this



- To this



- 29: in the Calibrate Data tab of the utility panel you can now start putting in your calibration points, this is done by double clicking the location where you know the subject is looking. This is also why the audio input is very useful, you can provide auditory cues while recording the data to make putting in the calibration points easier. For example: having your subject verbally confirm they are looking at a specific point. If the data coming in is clear and the dot in the top left corner is green when you create the point, you should get a working calibration point with the coordinates listed in the utility panel below.

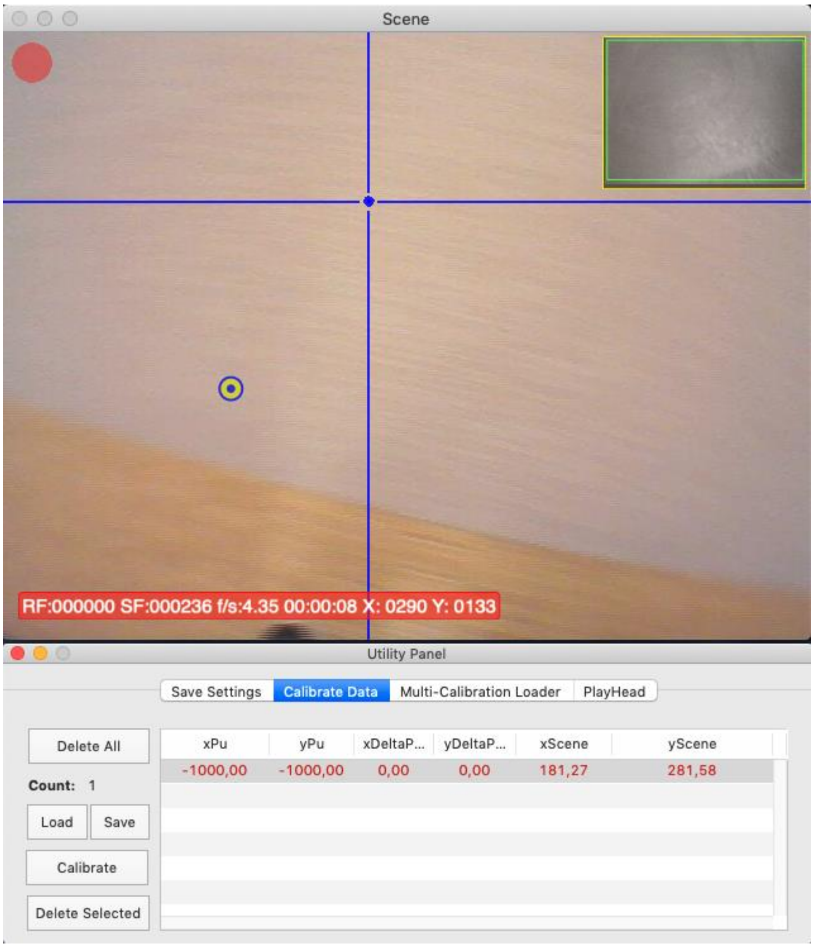


RF:000000 SF:000459 f/s:13.93 00:00:15 X: 0297 Y: 0099

Count: 1

xPu	yPu	xDeltaP...	yDeltaP...	xScene	yScene
320,77	209,17	25,77	-77,33	272,80	93,24

- 30: if however the dot isn't green when you create the calibration point it won't work but the point still ends up in the list of calibration points as seen below, note the red numbers indicating coordinates which means the calibration point doesn't work. If this happens you should select the point and delete it.

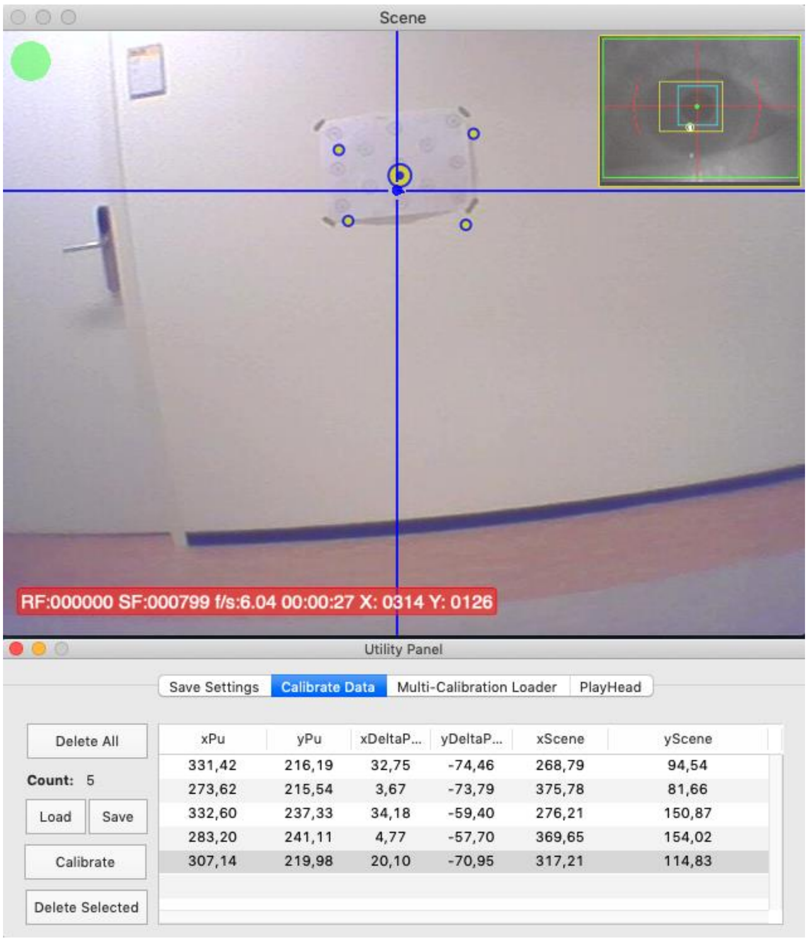


RF:000000 SF:000236 f/s:4.35 00:00:08 X: 0290 Y: 0133

Count: 1

xPu	yPu	xDeltaP...	yDeltaP...	xScene	yScene
-1000,00	-1000,00	0,00	0,00	181,27	281,58

- 31: if all points are properly added the list should look like it does in the screen below.

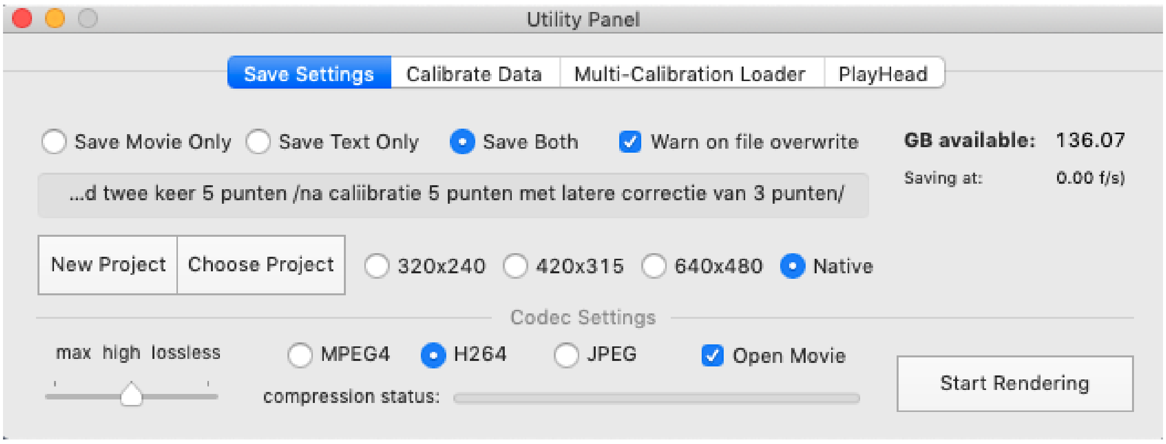


RF:000000 SF:000799 f/s:6.04 00:00:27 X: 0314 Y: 0126

Count: 5

xPu	yPu	xDeltaP...	yDeltaP...	xScene	yScene
331,42	216,19	32,75	-74,46	268,79	94,54
273,62	215,54	3,67	-73,79	375,78	81,66
332,60	237,33	34,18	-59,40	276,21	150,87
283,20	241,11	4,77	-57,70	369,65	154,02
307,14	219,98	20,10	-70,95	317,21	114,83

- 32: after this you're almost done, now all you have to do is go to the Save Settings tab in the Utility Panel as seen below, make sure the output location is what you want it to be and click the Start Rendering button. After this you'll have a video file with the eye-tracking in the same file and you can start analysing.



○ Save Movie Only ○ Save Text Only ● Save Both ☑ Warn on file overwrite

GB available: 136.07

Saving at: 0.00 f/s)

...d twee keer 5 punten /na caliibratie 5 punten met latere correctie van 3 punten/

○ 320x240 ○ 420x315 ○ 640x480 ● Native

○ MPEG4 ● H264 ○ JPEG ☑ Open Movie

compression status:

Option 2 continues here

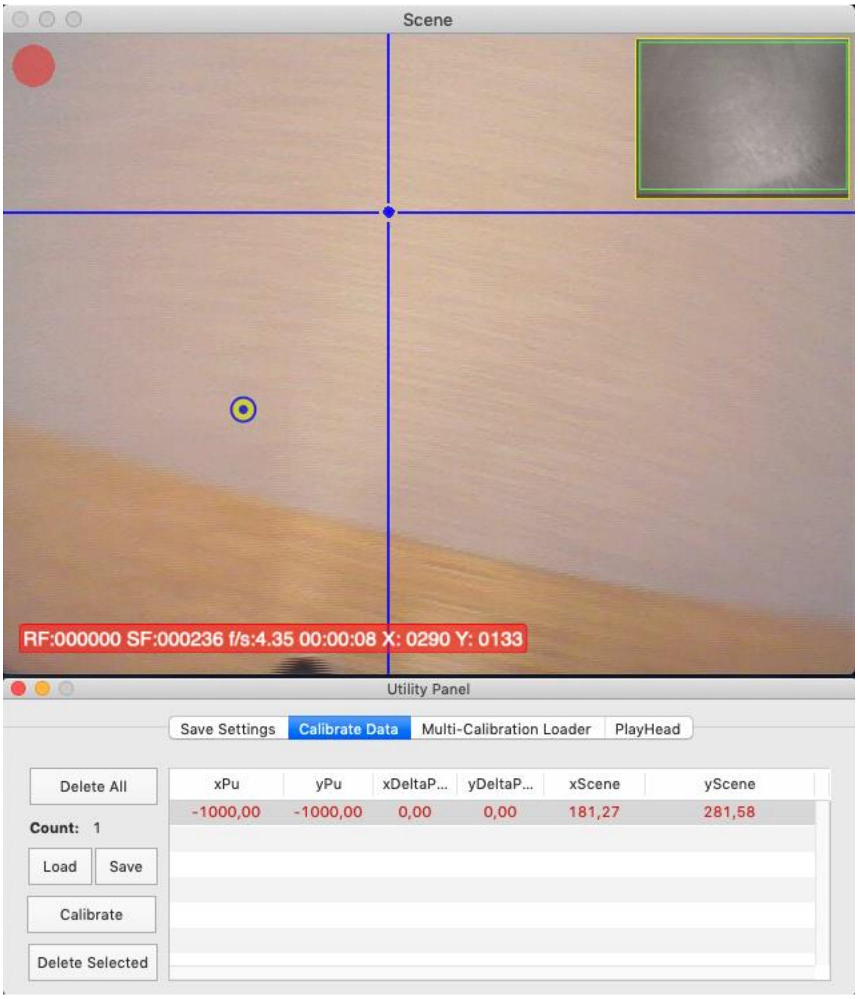
19: If the data coming in is clear the dot should be green most of the time, except when the subject is blinking or looking very far into their peripheral vision.

RF:000000 SF:000459 f/s:13.93 00:00:15 X: 0297 Y: 0099

Count: 1

xPu	yPu	xDeltaP...	yDeltaP...	xScene	yScene
320,77	209,17	25,77	-77,33	272,80	93,24

- 20: after this you can start the calibration, have your subject look at specific points on the screen and double click the corresponding part of the scene view in the scene panel. If done right it should look like this. → if however you try to make a calibration point and the dot is red at that moment, it won't register properly and in the calibration points list the coordinates will appear in red, as shown below. If this happens you should select the point in the list and delete it.

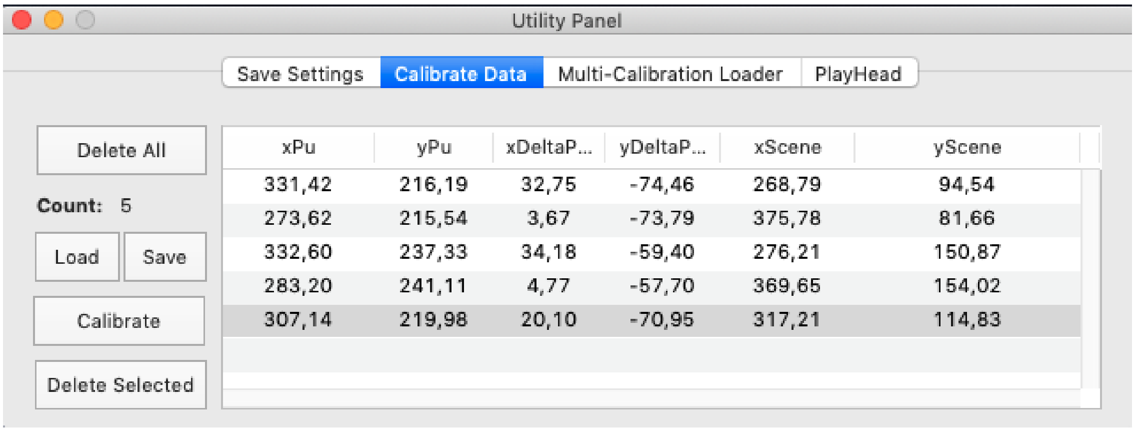


RF:000000 SF:000236 f/s:4.35 00:00:08 X: 0290 Y: 0133

Count: 1

xPu	yPu	xDeltaP...	yDeltaP...	xScene	yScene
-1000,00	-1000,00	0,00	0,00	181,27	281,58

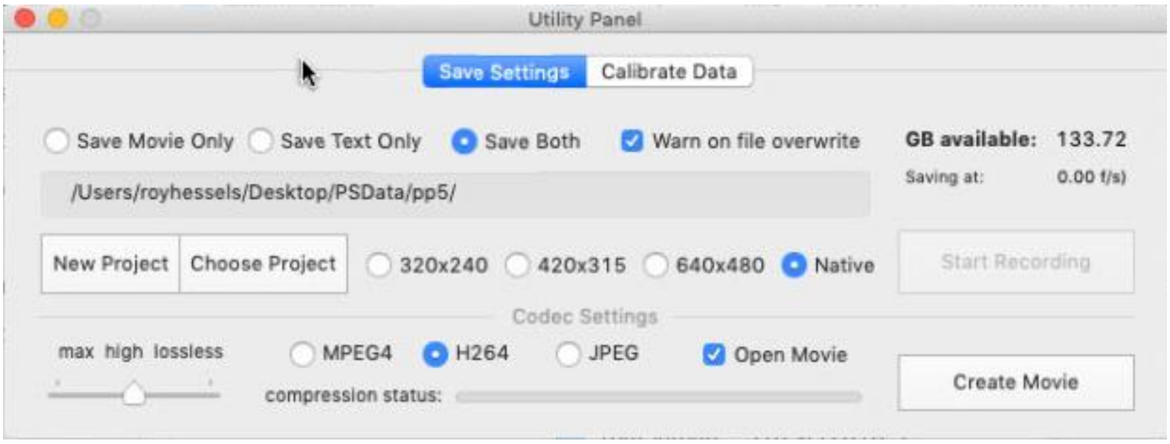
- 21: if all point have been properly added you should end up with a calibration points list like the one shown below, note that all calibration points have coordinates in black.



Count: 5

xPu	yPu	xDeltaP...	yDeltaP...	xScene	yScene
331,42	216,19	32,75	-74,46	268,79	94,54
273,62	215,54	3,67	-73,79	375,78	81,66
332,60	237,33	34,18	-59,40	276,21	150,87
283,20	241,11	4,77	-57,70	369,65	154,02
307,14	219,98	20,10	-70,95	317,21	114,83

- 22: after this you can choose the location where you want to save your file in the save settings tab then you can start your recording by clicking the button that says “start recording”. When you’ve finished recording, make sure that you export the video by clicking button that says “create movie” or it will not be saved.



○ Save Movie Only ○ Save Text Only ● Save Both ☑ Warn on file overwrite

GB available: 133.72

Saving at: 0.00 f/s)

/Users/royhessels/Desktop/PSData/pp5/

○ 320x240 ○ 420x315 ○ 640x480 ● Native

max high lossless

○ MPEG4 ● H264 ○ JPEG ☑ Open Movie

compression status:

-