Statistical determination of area of interest scores in low quality eye-tracking data

Tim Van Wesemael

December, 2017

Abstract

This report examines methods to interpret low quality eye tracking data. The most important part is area of interest assignment, for which two methods that make use of statistical adjustments are introduced. These methods take the quality of the data into account, which makes them suited for data with much noise. One method eliminates certain gaze points that are situated too close to an AOI border. The other uses a normal distribution and quasi-Monte Carlo integration to give multiple AOIs, for every fixation point, a continuous score between 0 and 1. For all AOIs this estimates the chance, based on the data quality, of a fixation point lying in that respective AOI. Especially the latter proves itself to be very useful because no gaze data is lost.

Contents

1 Introduction

Eye-tracking is a promising method in contemporary behavioural research, especially for non-verbal subjects or as an implicit measurement for attention and interest. During such an experiment, an eye-tracker records point where the participant is looking at, while stimuli are shown. Many measurements make use of *areas of interest* (AOIs) [1]. These are certain regions on the surface, usually a screen, where the participant is looking at. The exact definition of these fields depend on the research and is largely a choice left to the experimenter. As a result, many similar studies use differently defined AOIs, and are thus difficult to compare [2]. Therefore it is important to define *areas of interest* which are set by objective rules. Once they are properly determined, a simple hit-or-miss algorithm determines the region the participant is looking at. In this paper other more complex methods will be examined. By taking the data quality into consideration, these methods will significantly improve the results.

An important parameter which will be used by the algorithms described in this paper is the data quality. This is determined by a process called *calibration validation*. Noise and data loss can be induced by many factors, e.g. inaccuracies in the eye-tracker, micro-trembles in the eye, a participant who loses focus [1]. Especially if the participant is young, the data quality is extra sensitive to the last factor. The methods proposed are specifically developed to deal with these inaccuracies. This is achieved by making algorithms adaptive to the data quality of the individual participants, so the results are based on the values calculated by the *calibration validation*. The data will be interpreted by two different methods, both introduced in this paper.

The key element of the first method is to define margins of low confidence between AOIs. Gaze points falling within these margins are remove to disambiguate the dataset. Thus, only data for which the AOI can be fairly certainly decided is taken in consideration. The results of the *calibration validation* determine the size of these margins. For very low quality data, the margins will span a large area, so many gaze points will be ignored. To solve this, a second method without data loss is introduced and examined too. The second algorithm does not abolish any data points, however it is more complex than a simple *hit-or-miss* function. For one gaze point the algorithm can give multiple AOIs a score that is calculated using a normal distribution. The score is dependent on the probability of a gaze point falling into the corresponding AOI, so the cumulative sum of all scores is one and the shape of the distribution is again influenced by the *calibration validation*. Hence the worse the quality of the data is, the wider the distribution of the scores is and vice versa. The main challenge of this method is the determination of the scores. This is solved by using *quasi-Monte Carlo* integration and *Halton sampling*.

This report will conclude with a comparison of the methods, using two experiments that are embedded in a broader research project. One experiment, *houses-faces*, assesses preferential looking for houses versus faces, compared between a group of typically developing boys (TD) and a group of boys diagnosed with autism spectrum disorder (ASD). The second experiment, *upper-lower* assesses preferential looking to the eyes versus the mouth of a face, compared for the same groups of children. Faces, and especially the eyes, are highly relevant stimuli for our social interactions. Hence, a social preference emerges very early and even within minutes after birth, infants are more attracted to the human face and social stimuli over nonsocial stimuli [3]. However, previous eye-tracking research suggests that children with ASD might not show this preference for social stimuli [4] and for the eye region [5]. For each experiment, a comparison of the methods will be shown.

2 Determining the data quality

During the eye-tracking testing many factors can influence the recording data. Natural causes such as micro tremors as well as machine inaccuracies can induce unexpected variations in the data. Because this project handles with a screen mounted eye-tracker, which allows the the test person to move, and with participants whose recorded gaze points are prone to inaccuracies, much noise and data loss is expected. Some precautions are needed when dealing with low-quality data. In this paper, a filter is applied first to the raw output of the eye-tracker. Second, through the use of a *calibration validation* process, parameters determined by the data quality are calculated.

2.1 Filter

A filter serves two important purposes in eye-tracking research. First it makes the data better suited for interpretation by deleting random noise and interpolating missing data. Besides it also identifies and labels gaze events, such as fixations, saccades or smooth pursuits [1]. Many different filter algorithms are published, one of the more popular being *identification by velocity threshold*. This filter is build into the Tobii eye-tracker that is used in this project [6]. However, results given by that filter are not used. Instead, the raw data is filtered by a *identification by 2-means clustering* or I2MC algorithm [7].

The I2MC filter works by dividing every gaze point in a time frame in two clusters after using Steffen interpolation to replace missing data. Every cluster is the collection of points that are situated closest to the mean coordinate pair of their cluster. Depending on the number of times the line that connects the gaze points in a chronological order crosses the border between the two clusters, it is decided if that time frame is a fixation candidate. Many border crossings imply random noise and a fixation, few imply a noticeable difference between the clusters and thus eye movement. The algorithm then checks which candidates should be merged and which should be abolished. The filter is robust to noise and is better suited for low quality data than other filters [7]. However it is only able to identify one gaze event, a fixation. Since no other events are used, this is not a problem for this study, which makes the I2MC filter a great option.

2.2 Calibration validation

The *calibration validation* is a process which takes place after the eye-tracking recorder is calibrated. The participant is then shown a few fixation crosses, where he ought to stare at for four seconds. During the test the researcher can check on an other screen if the participant is looking and if the eye-tracker is calibrated with sufficient accuracy [1]. In this project a script evaluates the filtered *calibration validation* data and returns three values, which in this project, are referred to as *hit rate*, *accuracy* and *root mean square*.

These parameters all relate to the *visual angle*, the angle between the vectors to the expected and real gaze point, relative to the eye. It is important to notice that the *visual angle* does not indicate a single point on the screen (except if it is 0). Rather it defines a certain ellipse of which the border consists of possible gaze points. If $\vec{r}_{eye}$, $\vec{r}_c$ and $\vec{r}_g$ are respectively the position vector of the eye, the expected gaze point i.e. the fixation cross and the real gaze point in a three dimensional coordinate system, then the visual angle, θ is determined by this formula [8]:

$$\theta = \arccos\left(\frac{(\vec{r}_c - \vec{r}_{eye}) \cdot (\vec{r}_g - \vec{r}_{eye})}{||\vec{r}_c - \vec{r}_{eye}||\ ||\vec{r}_g - \vec{r}_{eye}||}\right) \tag{1}$$

A certain upper threshold for the *visual angle*, θ_{max} is defined. Now, for every fixation cross shown to the participant c_i it is expected that multiple fixations g_{ij} are detected. Let θ_{ij} be the

visual angle between c_i and g_{ij} and let t_{ij} be the duration of that fixation. Besides, let n_c be the number of fixation crosses and $n_{g,i}$ be number of recorded gaze points during the time that cross c_i was shown. Then the set of visual angles which will be used in the calibration validation, Θ, is:

$$\Theta = \{\theta_{ij},\ i = 1, 2, ..., n_c \text{ and } j = k_i, k_i + 1, ..., n_{g,i}\} \tag{2}$$

with

$$k_i = min(\{j \mid \theta_{ij} \leq \theta_{max}, j = 1, 2, ..., n_{g,i}\}),\ i = 1, 2, ..., n_c.$$

In words this means that all fixation points recorded before the gaze point of which the *visual angle* with the corresponding fixation cross is smaller than the threshold, are not taken into account. As an effect, data recorded during the period when the participant was switching between looking from one gaze point to another, is discarded. With these angles, the *hit rate* a, *accuracy* θ_{err} and *root mean square*, θ_{RMS} can be defined as follows, with the help of a hit-or-miss function h:

$$h(\theta) = \begin{cases} 1 & \text{if } \theta \leq \theta_{max} \\ 0 & \text{else} \end{cases} \tag{3}$$

$$a = \frac{\sum(h(\theta_{ij}) \cdot d_{ij})}{\sum(\theta_{ij} \cdot d_{ij})} \qquad \forall \theta_{ij} \in \Theta \tag{4}$$

$$\theta_{err} = \frac{\sum(\theta_{ij} \cdot d_{ij})}{\sum d_{ij}} \qquad \forall \theta_{ij} \in \Theta \tag{5}$$

$$\theta_{RMS} = \sqrt{\frac{\sum(\theta_{ij} \cdot d_{ij})^2}{\sum d_{ij}}} \qquad \forall \theta_{ij} \in \Theta \tag{6}$$

The *hit rate* is thus the fraction of the time that the visual angle is smaller than the predefined threshold. The lower the value of this parameter is, the lower the quality of the data. This can be used as a warning for datasets of abysmal quality. The other two parameters are the mean of the *visual angle* and root of the mean of squares of the *visual angles* respectively. The latter two values are important during the attribution of the *areas of interest*.

3 Defining the areas of interest

The *areas of interest* are defined according to the needs of the researcher. In simple paradigms like *houses-faces* two rectangle AOIs that span the two zones where the pictures are shown would suffice. Two rectangle AOIs can as well be defined in the *upper-lower* task: one for the upper part and one for the lower part of the face. However, if one is interested in the distribution of fixation points to the eye vs mouth region of a face, the simple rectangle AOIs are not optimal, as they contain additional information beyond the eyes and the mouth (e.g. nose area). A proposed standard for the division of faces into AOIs is by using a *Voronoi diagram* [2].

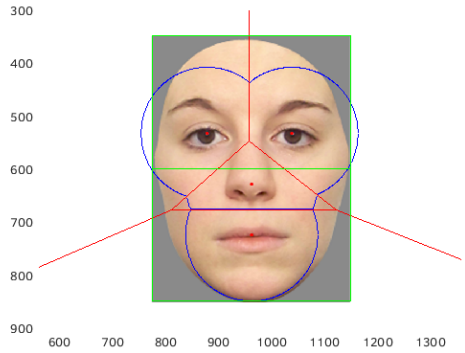
A *Voronoi diagram* is a subdivision of a surface, S, in multiple polygons, R_i, defined by a collection of preset points or sites on that surface, $\vec{P_i}$, such that every point in a polygon, R_i, is closer to the site of that polygon, P_i, than to any other site. Or in mathematical notation [9]:

$$R_i = \{\vec{x} \in S \mid ||\vec{x} - \vec{p_i}|| \leq ||\vec{x} - \vec{p_j}||\ \forall j \neq i\}. \tag{7}$$

A variant on the *Voronoi diagram* is the *limited-radius Voronoi-tessellation* (LRVT) [2]. This is the standard definition with an extra restriction on maximum distance, d_{max} to the nearest site:

$$R_i = \{\vec{x} \in S \mid ||\vec{x} - \vec{p_i}|| \leq ||\vec{x} - \vec{p_j}|| \wedge ||\vec{x} - \vec{p_i}|| \leq d_{max}\ \forall j \neq i\}. \tag{8}$$

(a) houses-faces: rectangles (red)



(b) upper-lower: rectangles: green and LRVT: blue and red

Figure 1: AOIs during the different paradigms

As a result not every point in S will belong to one of the polygons R_i. In eye-tracking research this is useful if the stimuli are only a small portion of the viewing surface, as is the case in the *upper-lower* task were the face is not scaled to fit the complete screen.

In conclusion, as visible in figure 1, for the *houses-faces* experiment for every picture a rectangle AOI that shares it borders with the stimulus, is defined. For the *upper-lower* task two types of *areas of interest* are defined. Two rectangle AOIs divide the face in an upper and lower part. In addition a *limited-radius Voronoi-tessalation* is created over the stimulus with sites located at both of the eyes, the nose and the mouth. The two types of AOI do not interfere with each other.

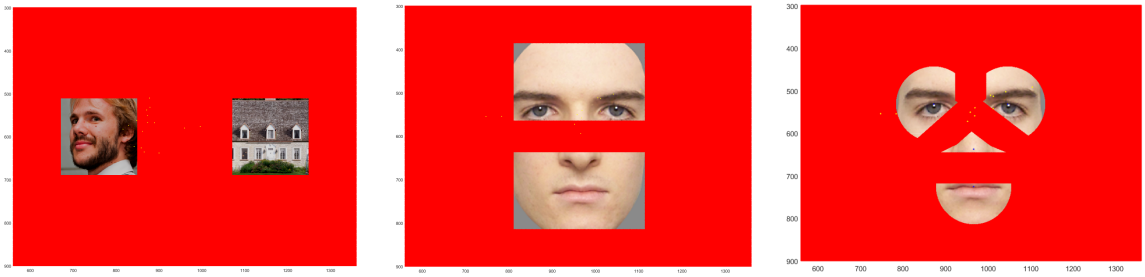
4 Broad borders approach

The main problem this paper addresses, is the determination of *areas of interest* in eye-tracking data of low quality. Because of the noise, there is an uncertainty about the true gaze point for every recorded data point. There can thus not be an AOI attributed to a gaze point near to a border with absolute certainty. The first method solves this problem by ignoring every gaze point of which the AOI cannot be decided with enough confidence. The factor that represents this grade of certainty is the *mean error angle* returned by the *calibration validation*. This implies that the worse the data quality is, the smaller the AOIs will effectively be. This is visible in figure 2. An obvious downside to this is that much data could get lost, especially during tasks with *areas of interest* bordering each other, e.g. the *upper - lower* paradigm. However, the gaze points that are interpreted can be assigned unambiguous to an AOI.

To determine the certainty, the smallest *visual angle* between the recorded gaze point and any point on an AOI border is calculated. If this value is smaller than the *mean error angle* calculated during the *calibration validation*, the point will be abolished. To determine the smallest angle, the point on an AOI border closest to the gaze point $\vec{r}_g$ must be known. Let this point be $\vec{r}_g\,'$ and let it substitute $\vec{r}_c$ in equation (1). The smallest *visual angle* is now known.

4.1 Rectangular AOI

For a rectangular shaped AOI it is not hard to find $\vec{r}_g\,'$. If the horizontal borders are denoted by $y = y_1$ and $y = y_2$, the vertical ones by $x = x_1$ and $x = x_2$ and $\vec{r}_g = (x_0, y_0)$, then there are



(a) houses-faces (b) upper-lower, rectangular (c) upper-lower, Voronoi

Figure 2: The broad borders approach for both paradigms.

four candidates $\vec{r}_{g,i}\,'$ for $\vec{r}_g\,'$:

$$\begin{aligned}\vec{r}_{g,1}\,' &= (x_1, y_0)\\ \vec{r}_{g,2}\,' &= (x_2, y_0)\\ \vec{r}_{g,3}\,' &= (x_0, y_1)\\ \vec{r}_{g,4}\,' &= (x_0, y_2).\end{aligned} \tag{9}$$

This is because $\vec{r}_{g,i}\,'$ must be an orthogonal projection of $\vec{r}_g\,'$ on one of the AOI borders. Hence

$$\vec{r}_g\,' = \vec{r}_{g,i}\,' \tag{10}$$

such that

$$||\vec{r}_{g,i}\,' - \vec{r}_g|| < ||\vec{r}_{g,j}\,' - \vec{r}_g|| \;\forall j \neq i. \tag{11}$$

The smallest *visual angle* is thus the one between the gaze point and the closest border of the AOI.

4.2 Voronoi AOI

The determination of the smallest *visual angle* in a *voronoi diagram* relies on the same principle of orthogonal projection. However, the borders are often not parallel to one of the coordinate system axes. So the determination of $\vec{r}_{g,i}\,'$ in this case asks for a more general approach: the projection $\vec{p}\,' = (x_0', y_0')$ of a point on any straight line

$$q \longleftrightarrow ax + by + c = 0 \tag{12}$$

is given by [10]

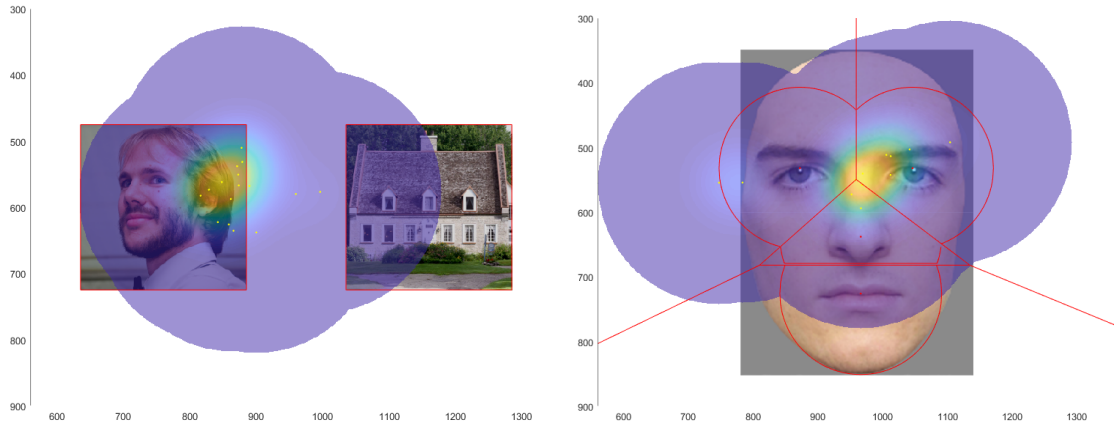
$$\begin{aligned}x_0' &= \frac{b(bx_0 - ay_0) - ac}{a^2 + b^2}\\ y_0' &= \frac{a(ay_0 - bx_0) - bc}{a^2 + b^2}.\end{aligned} \tag{13}$$

Due to the definition of a *Voronoi tessellation* the border q_{ij} between two AOIs, R_i and R_j, defined by the sites, $\vec{p}_i = (x_i, y_i)$ and $\vec{p}_j = (x_j, y_j)$ is an equidistant line between those two points. So, any point $\vec{p} = (x, y)$ on that border satisfies the equation:

$$||\vec{p} - \vec{p}_i|| = ||\vec{p} - \vec{p}_j|| \tag{14}$$

$$\Leftrightarrow \sqrt{(x - x_i)^2 + (y - y_i)^2} = \sqrt{(x - x_j)^2 + (y - y_j)^2} \tag{15}$$

$$\Leftrightarrow 2(x_j - x_i)x + 2(y_j - y_i)y + x_i^2 - x_j^2 + y_i^2 - y_j^2 = 0. \tag{16}$$



(a) houses-faces (b) upper-lower

Figure 3: The normal distribution approach for both paradigms.

It thus fits the equation for a straight line (12) with the following parameters:

$$\begin{aligned} a &= 2(x_j - x_i) \\ b &= 2(y_j - y_i) \\ c &= x_i^2 - x_j^2 + y_i^2 - y_j^2. \end{aligned} \tag{17}$$

So for every gaze point $\vec{r}_i$ the closest point on an AOI border is determined by equation (17) and (13), if the two closest sites $\vec{p}_i$ and $\vec{p}_j$ are known. The visual angle between these two points indicates then if the gaze point is situated too close to a AOI border.

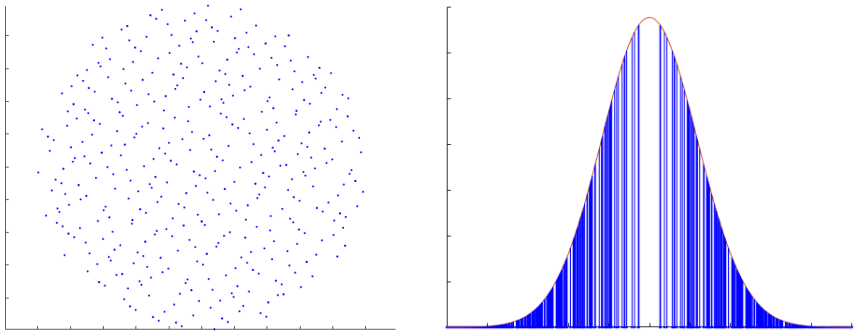
The method described above only works for borders between Voronoi cells. If the *limited radius Voronoi tessalation* defines the AOIs, a new method needs to be established for the borders formed by the maximum radius. The approach for this problem is to define a new $d_{max,i}$ for every fixation point $\vec{r}_i$:

$$d_{max,i} = d_{max} - ||\vec{r}_i - \vec{r}_{eye}|| \tan(\theta_{err}) \tag{18}$$

with $\vec{r}_i$ the gaze point and $\vec{r}_{eye}$ the position of the eye in the three dimensional space. Assuming the participant sits right in front of the stimulus and the deflection of the gaze point from the orthogonal projection of the eye position on the screen is small relative to the distance between the participant and the screen, this method is equivalent to checking the smallest angle between a gaze point and a border. This is because for every *visual angle* the collection of every possible gaze point is the border of an ellipse. However if the mentioned conditions are valid, this ellipse becomes very similar to an circle, so the second term in equation (18) represents the minimum distance such that the current visual angle is bigger than the threshold, independently of the direction.

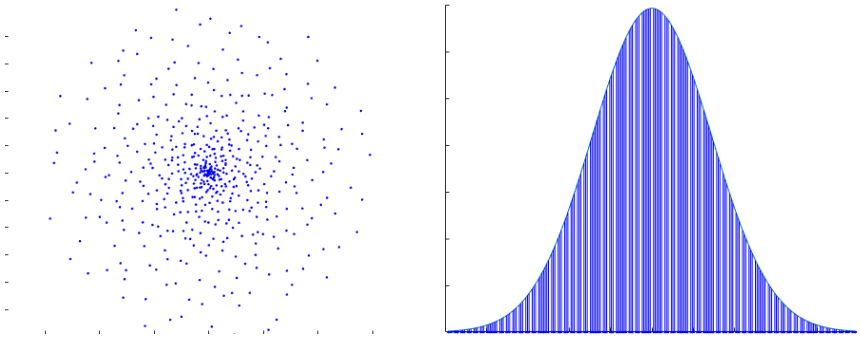
5 Quasi-Monte Carlo normal distribution approach

A major downside to the method described above is the great amount of data loss. Especially if the AOIs border each other, as is the case in the *upper - lower* paradigm, many gaze points are expected to fall into the *broad borders* and thus fail to be attributed to one or the other AOI. The second method proposed in this paper solves this problem by interpreting every recorded gaze point. However it does not attribute every fixation point to one AOI. Instead, for every gaze point, a continuous score between 0 and 1 is calculated for every AOI in a way that the



(a) Uniform on the screen

(b) Angles concentrated around the edges



(c) Concentrated around gaze point

(d) Uniform angles

Figure 4: Difference between samples generated uniformly on the screen and with a uniform visual angle.

cumulative sum of these scores equals 1 (including the score for a region outside of any AOI). The higher the score, the larger the probability is that the corresponding AOI contains the recorded gaze point. Again, this method takes a parameter of the *calibration validation* into account, the *root mean square* θ_{RMS}. As a result of the scoring algorithm, there will not be an unambiguously determined AOI for every gaze point. As an effect, this method is not suitable for all AOI measurements. Similar methods have been used in a few rare cases, but are not documented well [1, 11].

The general outline for the algorithm will be described next. First random sample points around the gaze point will be generated. For each sample point the score will be decided based on an a normal distribution around the gaze point. Finally, the algorithm determines which AOI the sample belongs to and assigns the score of the sample to the AOI. As a result the closer an AOI is to the recorded fixation point and the larger the region of the AOI near that point is, the higher the score of that AOI will be. This can be seen in figure 3.

5.1 Normal distribution

The score for every sample point is based on a normal distribution of visual angles around the recorded fixation point. Because the visual angle only indicates the distance between two points, it does not denote a single point on the screen. Instead, as said above, a single visual angle can indicate any point on an ellipse. So to select a single point on the screen, a direction needs to be defined too. Assuming the probability of this direction is uniformly distributed, the expected value of the visual angle is zero. This is because the occurrence of a certain visual angle in one direction has the same probability of the exact same visual angle in the opposite direction. Since

the expected value is assumed 0, the *root mean square* of the *error angle* is a good indicator for the standard deviation. So, for a gaze point $\vec{r}_i$ let $\vec{s}_{ij}$ be a sample point around it and $\theta_{s,ij}$ the angle between $\vec{r}_i$ and $\vec{s}_{ij}$, calculated using equation (1). The score belonging to $\vec{s}_{ij}$, S_{ij} is thus calculated by:

$$S_{ij} = N(\theta_{s,ij}, 0, \theta_{RMS}) \tag{19}$$

with

$$N(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

being the normal distribution [12].

The normal distribution is only sampled and not analytically integrated. The latter being generally used to denote a probability. However *Monte Carlo* integration allows to approach the integral using only samples [13]. If I is the exact value of the integral $\int_a^b f(x)$ and n is the number of samples taken, than

$$I \approx \frac{b-a}{n} \sum_{i=1}^{n} f(x_i). \tag{20}$$

5.2 Sample generation

The spawning of sample points has a great influence on the final result. Variations in the distribution of sample points might cause great fluctuations in the outcome. Also, if the number of taken samples is too small, the calculated score might misrepresent the correct one. In contrast, if the number is too high, valuable computing time is lost. A solution to this problem is to use formula (20) to approach the integral. Since the integral of the normal distribution over a certain interval equals the probability that the interval contains the variable and the sample interval ought to cover every significant point of the distribution, the integral must equal one. So the difference between (20) and one is a good indicator for the quality and quantity of the sample points.

Because *visual angles* are evaluated, it is important that they are uniformly distributed over the interval. Otherwise the *Monte Carlo* method (20) will not converge to one. For example if the sample points are generated uniformly on the screen, the distribution of *visual angels* will be concentrated around the outskirts of the interval, as visualised in figure 4. So to have a correct estimation, first the *visual angles* are generated for which the score is immediately calculated. Afterwards a random direction will denote the sample point on the screen. To assure the samples are uniformly distributed, a *quasi-random* algorithm is used [14]. Such an algorithm produces points in a deterministic way, however, they keep the output sequences' *discrepancy* as low as possible. This means that the points are distanced uniformly [15]. These characteristics are inherent to the *Halton sequence*, if small bases are used (e.g. 2 and 3) [16]. Another advantage of a *quasi-random* method is that every score needs to be computed only once. Since the the algorithm outputs the same deterministic values every time, the visual angels, directions and scores only have to be computed and saved one time. For every subsequent gaze point, the saved *visual angle* and direction decide the sample point. The *area of interest* which contains this sample point receives the score saved along the first two parameters.

5.3 Motivation

Because the method described above works with probabilities, it might return some unexpected results in particular conditions. For example, when a fixation point is situated near to a border

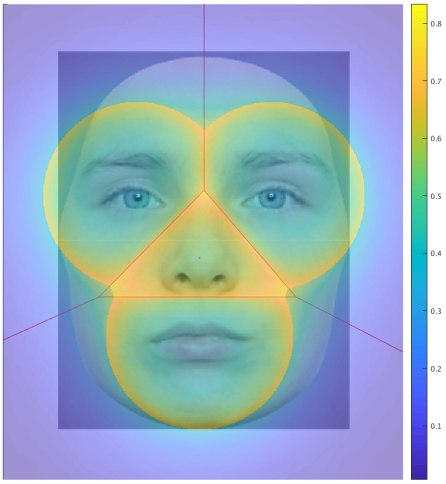


Figure 5: A heatmap of the probability that a gaze point gets wrongly attributed to a LRVT AOI during the *upper-lower* paradigm when the standard deviation of the angle deflection is 0.015 radians and the participant sits at a distance of 60 *cm* from the screen. The brightest yellow areas do indicate a probability of a little over 80% for a gaze point to be misattributed. The green areas around 50% and the darkest blue ones denote a very low probability of 0 to 10%.

of an AOI that is convex around the fixated point (so the border curves towards that point), the neighbouring AOI might receive a higher score than the AOI in which the recorded gaze point is situated. This is because of the uncertainty in the location of the actual gaze point and the fact that the other AOI spans a large area near the recorded gaze point, so there is a relative large probability that the actual gaze point is situated in that AOI. In another case it is possible that, when the RMS is large, so the distribution spans a large area, an AOI might receive a score, even though there is a great distance between the recorded gaze point and the AOI. However in this situation the calibration validation has not returned favourable values and as an effect the recorded data might not be reliable at all. This method may not always return the best results, depending on the data quality and the size and shape of the AOI. However to illustrate that the probabilities described above have a clear benefit in practical circumstances, the case of the LRVT is investigated.

To get a notion about the risky regions during the *upper-lower* experiment, a heatmap that displays the sensitivity to misattribution of ever pixel of the stimulus (see figure 5). In more detail, for every pixel the probability is calculated that a gaze point found at that position will be situated in an other AOI after Gaussian noise with a standard deviation of 0.015 radians (a typical value during the experiments described in section 6) in a random direction is inflicted on the visual angle from an eye 60 *cm* from the screen. The danger spots are as expected near to AOI borders. Especially the sharp corners of the triangular nose AOI are vulnerable to misatrribution, having a probability of around 80%. More central regions of the AOIs have a lower, but still significant misattribution probability of around 30% to 50%. The experiments described in section 6 and more specifically figure 6 show that many fixated regions overlap with risk zones. In this case the AOI score attribution based on probabilities is beneficial because it does not assume a fixation point that is situated near a border is certain to belong to the AOI it is in, but it credits other probable AOIs as well.

(a) Standard

	Relative to	Houses	Faces
ASD	all points	0.2398	0.3161
	points in AOI	0.4015	0.5985
TD	all points	0.1884	0.3632
	points in AOI	0.3415	0.6584

(b) Broad borders

	Relative to	Houses	Faces
ASD	all points	0.1947	0.2653
	points in AOI	0.4233	0.5767
TD	all points	0.1577	0.3172
	points in AOI	0.3321	0.6679

(c) Normal distribution

	Relative to	Houses	Faces
ASD	all points	0.2309	0.2998
	points in AOI	0.4108	0.5891
TD	all points	0.1800	0.3398
	points in AOI	0.3887	0.6113

Table 1: Comparison between relative AOI time distributions during the houses-faces paradigm for the classical method and the two methods introduced in this paper.

(a) Standard

	Relative to	Eyes	Mouth	Nose	Upper	Lower
ASD	all points	0.4738	0.0495	0.1522	0.5362	0.1706
	points in AOI	0.7014	0.0733	0.2253	0.7586	0.2414
TD	all points	0.4172	0.1152	0.3422	0.4880	0.4034
	points in AOI	0.4770	0.1317	0.3913	0.5026	0.4974

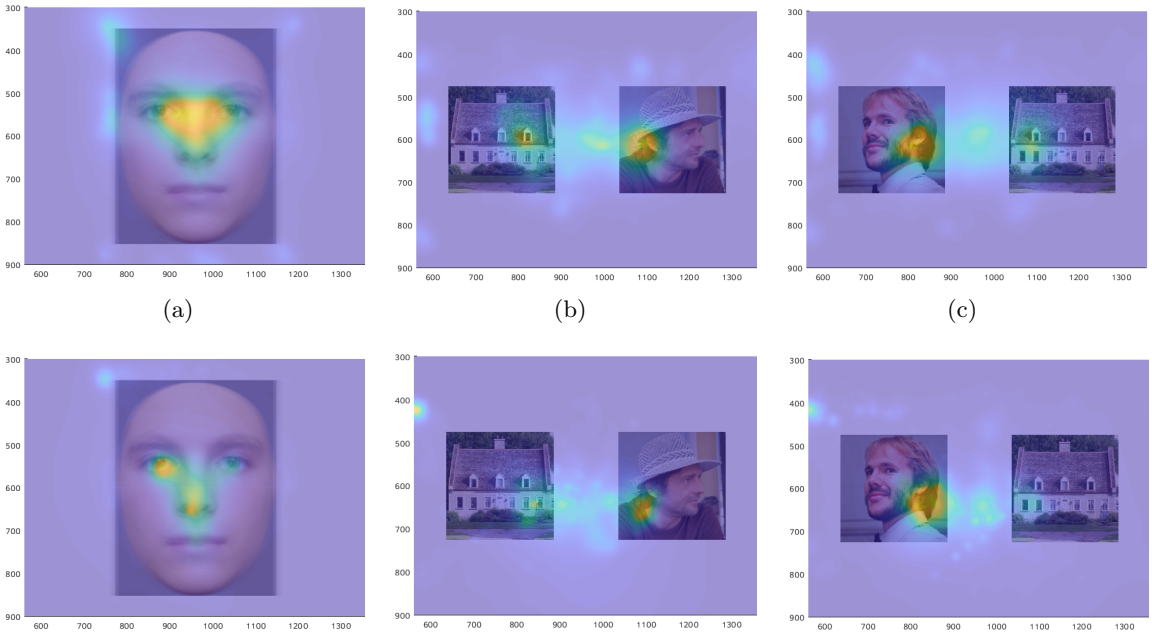
(b) Broad borders

	Relative to	Eyes	Mouth	Nose	Upper	Lower
ASD	all points	0.2937	0.0327	0.0719	0.4218	0.0975
	points in AOI	0.7374	0.0821	0.1805	0.8122	0.1878
TD	all points	0.3203	0.0851	0.2414	0.4018	0.3232
	points in AOI	0.4952	0.1316	0.3732	0.5542	0.4458

(c) Normal distribution

	Relative to	Eyes	Mouth	Nose	Upper	Lower
ASD	all points	0.4596	0.0527	0.1521	0.5232	0.1766
	points in AOI	0.6918	0.0793	0.2290	0.6976	0.3024
TD	all points	0.4018	0.1191	0.3158	0.4662	0.3928
	points in AOI	0.4802	0.1423	0.3774	0.5692	0.4308

Table 2: Comparison between relative AOI time distributions during the upper-lower paradigm for the classical method and the two methods introduced in this paper.



(a) (b) (c)

(d) Upper - lower (e) Houses left, faces right (f) Faces left, houses right

Figure 6: Visual representation of the gaze data for both the ASD group (above) and the TD group (below), for the upper - lower and the houses - faces paradigm interpreted by the normal distribution approach.

6 Comparison

The algorithms described above as well as a method without statistical correction interpreted the eye tracking data of group of 18 participants diagnosed with ASD as well as of a typically developing control group (TD) of the same size. All the participants are boys around the age of ten years. During the *houses - faces* paradigm faces are shown next to houses at a certain frequency. Each participant watches this sequence of thirty seconds four times. Two times the faces are shown left, the other two times, the houses appear on the left side. During the *upper - lower* paradigm the upper and lower halves of a faces are shown consequently to the participant. This happens in four trails of one minute. During each trail faces of the same gender are visible between 8 and 12 seconds. There are two female and two male trails. During the whole experiment a Tobii Pro X3-120 eye-tracker records the gaze data of the participants.

Table 1 contains the average results for each group for the *houses - faces* paradigm and table 2 for the *upper - lower* paradigm. These tables include the relative amount of time the participants looked at each AOI, both relative to all measurements and relative to gaze points inside of an AOI. In table 2 the *eyes, mouth* and *nose* columns contain the results when using the *limited radius Voronoi tessellation area of interest* definition and the *upper* and *lower* columns contain the results for simple rectangle AOIs. The calculation of these two categories happens completely independently. Also in that table, the *eyes* column refers the the sum of the left and right eye. The overall results for both groups when interpreted by the normal distribution approach are visualised in figure 6. In this plot there is a distinction between the experiments where the houses are shown left and the experiments where the faces are shown left. This distinction does not exist in tables 1 and 2. The results of the *houses - faces* paradigm enforce the believe that the ASD group has a lower preference for a social stimulus [4], i.e. the face, than the TD group, even though the preference still is noticeable. The results from the *upper - lower* paradigm show that the ASD group has an increased amount of attention for the eye regions,

relative to the TD group. This contradicts the hypothesis that the eyes are a social region which will thus be avoided by the ASD group [5].

Every value relative to all points calculated by the *broad borders* algorithm is significantly lower than the corresponding score given by the classical method. This is explained by the smaller size of each AOI due the the abolishing of uncertain fixation points. Especially for the *upper-lower* task, that contains bordering AOIs, the difference can grow big, because the participant does not try to look as clearly as possible at a certain AOI, but will gaze between them much of the time. Most values returned by the *normal distribution* algorithm are slightly lower than classical score too, although higher than the ones calculated by the *broad borders* approach. This is again due to gaze points too near to an AOI border. However, this time those points are not completely ignored, because only a fraction of the sample points fall outside of an AOI, so the fixation points still receive a score, albeit smaller than one. The cases where some values (e.g. the mouth) are higher when using *normal distribution* scoring is explained by the fact that AOIs can receive a score even though they do not contain a fixation point. So this will especially occur if the number of gaze points just outside of an AOI is significantly higher than the number of gaze points inside of the AOI. The results allow to conclude as well that the LVRT method is better than rectangular AOIs. The nose AOI is often more looked at than the mouth AOI. This result could not be concluded when solely using the rectangular AOIs.

7 Conclusion

This paper proposed two new methods to attribute a gaze point to an *area of interest*. Methods to define the AOIs and to filter the raw data have been largely adopted from other studies, but show themselves to be effective. The *calibration validation* process has been tweaked to return parameters that are useful for AOI determination algorithms. This enables the possibility to implement AOI attribution algorithms that take the data quality into account. As a result of this, the two methods are especially beneficial when the data is expected to contain much noise and loss. This is the case for example younger children tested by an screen mounted eye tracker. The methods are used on data from two groups of school aged children. One diagnosed with ASD and one control group.

The first method, *broad borders*, makes the predefined AOIs effectively smaller by ignoring gaze points that are situated too near to AOI borders. In this way, only the gaze points of which the AOI they belong to is clear, are interpreted. The *calibration validation* determines the exact distance threshold for fixations to count as valid. As an effect the data quality has an influence on the way that AOIs are attributed. A downside of this is that the worse the data quality is, the smaller the AOIs will become. This is especially a problem for paradigms where many, smaller AOIs border each other, because much of the gaze points are concentrated near the borders and thus fall into the uncertainty margins.

The second method takes every gaze point into account, but does not unambiguously attribute them to an AOI. Instead an algorithm calculates a score between 0 and 1 for every nearby AOI. AOIs which are close to or contain the fixation points receive a higher score, as do AOIs which span a great amount of surface around the gaze point. Again, the *calibration validation* influences the way the score are assigned. So, the worse the data quality is, the more the scores will be spread out. In contrast, the better it is, the more concentrated the scores will be around the gaze point. Since the algorithm takes every gaze point into account, as well as the data quality, it proves itself to be a very reliable method. However, it is not suitable for all experiments, since the AOIs are not unambiguously designated.

During the *calibration validation* only *visual angles* are included for evaluation. However, these angles do have a direction as well which is completely ignored. The methods could be improved by taken the direction into account as well. The methods are only tested for relative time distribution in this research. The comparative study between the different algorithms could be extended to more tests, to find more strengths and weaknesses of each algorithm.

References

[1] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. van de Weijer, *Eye Tracking: A comprehensive guide to methods and measures*. OUP Oxford, 2011.

[2] R. S. Hessels, C. Kemner, C. van den Boomen, and I. T. C. Hooge, "The area-of-interest problem in eyetracking research: A noise-robust solution for face and sparse stimuli," *Behavior Research Methods*, vol. 48, pp. 1694–1712, Dec 2016.

[3] C. C. Goren, M. Sarty, and P. Y. Wu, "Visual following and pattern discrimination of face-like stimuli by newborn infants," *Pediatrics*, vol. 56, no. 4, pp. 544–549, 1975.

[4] K. Pierce, S. Marinero, R. Hazin, B. McKenna, C. C. Barnes, and A. Malige, "Eye tracking reveals abnormal visual preference for geometric images as an early biomarker of an autism spectrum disorder subtype associated with increased symptom severity," *Biological psychiatry*, vol. 79, no. 8, pp. 657–666, 2016.

[5] W. Jones and A. Klin, "Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism," *Nature*, vol. 504, no. 7480, pp. 427–431, 2013.

[6] A. Olsen, "The tobii I-VT fixation filter," *Tobii references*, 2012.

[7] R. S. Hessels, D. C. Niehorster, C. Kemner, and I. T. C. Hooge, "Noise-robust fixation detection in eye movement data: Identification by two-means clustering (I2MC)," *Behavior Research Methods*, vol. 49, pp. 1802–1823, Oct 2017.

[8] T. De Laet, D. Vandepitte, and J. Vander Sloten, *Toegepaste Mechanica: Een vectorgebaseerde benadering*. Acco, 2016.

[9] A. Dobrin, "A review of properties and variations of Voronoi diagrams," *Whitman College*, 2005.

[10] R. Larson, *Precalculus: A Concise Course*. Cengage Learning, 2010.

[11] G. Buscher, E. Cutrell, and M. R. Morris, "What do you see when you're surfing?: Using eye tracking to predict salient regions of web pages," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, (New York, NY, USA), pp. 21–30, ACM, 2009.

[12] R. A. Fisher *et al.*, "012: A mathematical examination of the methods of determining the accuracy of an observation by the mean error, and by the mean square error.," *Stellar movements*, 1920.

[13] C. Lemieux, *Monte carlo and quasi-monte carlo sampling*. Springer series in statistics, New York: Springer, 1. ed., 2009.

[14] H. Niederreiter, "Quasi-monte carlo methods and pseudo-random numbers," *Bulletin of the American Mathematical Society*, vol. 84, no. 6, pp. 957–1041, 1978.

[15] W. J. Morokoff and R. E. Caflisch, "Quasi-monte carlo integration," *Journal of Computational Physics*, vol. 122, no. 2, pp. 218 – 230, 1995.

[16] A. B. Owen, "Quasi-monte carlo sampling," *Monte Carlo Ray Tracing: Siggraph*, vol. 1, pp. 69–88, 2003.