Video games for visual interface (eye-tracking)

Johana Marku
Grenoble INP
Grenoble, France
johana.marku@grenoble-inp.org

Supervised by: Didier SCHWAB.

I understand what plagiarism entails and I declare that this report is my own, original work.

Johana MARKU 12^{th} June 2019

Abstract

Adjusting to life when having a disability is a hard challenge, especially when the ones facing it are children. New technologies and scientific research have come up with ways these people can interact and gain some control over their surroundings. By exploiting eye-tracking and the cognitive benefits from video games we aim at teaching impaired children to develop skills and feel as independent and included as possible. Thus, professionals from various fields can contribute to improve these children's life through examination of the eye-trackers metrics and pursue further research. Scanpath is the eye-tracking metric we will focus on in this report. We will present it's importance and our approach to implement it into GazePlay. Moreover, we will discuss about the limitations faced when generating such a metric and propose ways to overcome them.

1 Introduction

With regard to developing and enhancing augmentative and alternative communication(AAC)[1], exploiting the gaze, which is one of the most natural and simple ways the human perceives and interacts with his surroundings, would give to impaired people the control and digital presence they are otherwise deprived from.

In the case when impairments occur to children and infants, who start their journey of learning about the world around them without previous experience, it is a hard challenge even to acquire basic knowledge. Such children not only have to learn to play and gain knowledge but also be conscious and realize the power of their eyes, the consequences of actions triggered by their gaze through AAC tools such as eye-trackers. To give a general idea of this tool, one may think of it as an infrared camera that provides raw coordinates from which it is possible to estimate fixation, saccades[2] and/or blinking.

Eye-trackers have passed through considerable evolution before reaching to the commercial types available nowadays at an accessible price. They were really expensive to the pockets of families having a disabled person, reaching a price tag of over €1000. Tobii, the most important firm to sell eye-trackers released their Tobii4C accompanied by their C++ library in 2016, at an affordable price. [Schwab et al., 2018]

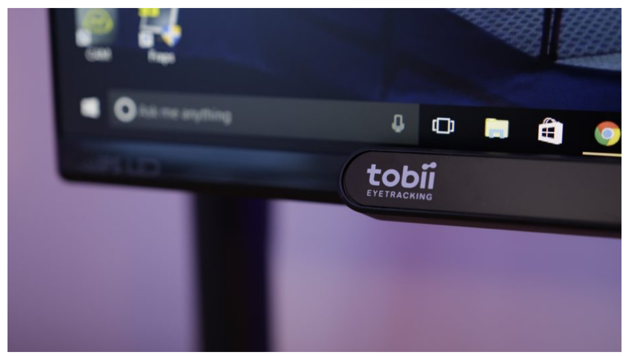

Figure 1: Tobii 4C eye-tracker

2 Learning through video games

In order to teach children and especially disabled ones, the best way is to make learning fun to them. In addition to that, children with multiple disabilities can't follow a traditional learning approach by frequenting public schools like their non-disabled peers. A good solution to this barrier is learning via video games specialized in developing certain skills for its player. These are the so called *serious games*[3].

The field of medicine has a history of embracing games as a means to engage patients behaviorally to improve their health outcomes. There are early reports of case studies using video games with patients experiencing diseases or physical disabilities [Kato, 2010]. To support the choice of video

[1] AAC is an umbrella term that encompasses the communication methods used to supplement or replace speech or writing for those with impairments in the production or comprehension of spoken or written language.

[2] Refer to definition in section 4.

[3] This term is used to describe video games that have been designed specifically for training and education.

games as a teaching medium it is crucial to mention their cognitive benefits. Preliminary research has demonstrated that cognitive advantages manifest in measurable changes in neural processing and efficiency. Moreover, scholars have also speculated that video games are an excellent means for developing problem-solving skills. In-game puzzles range in complexity from finding the quickest route from A to B, to discovering complex action sequences based on memorization and analytical skills. Game designers, often provide very little instruction on how to solve in-game problems, providing players with a nearly blank palette from which to explore a huge range of possible solutions based on past experience and intuitions. Another cognitive benefit, video games seem to be associated with, is enhanced creativity. [Granic et al., 2014] In addition, we must take into account that it is crucial to keep the child's motivation alive while doing a task. This is very well achievable through video games since they offer a way of rewarding the player after each task completion.

Children develop beliefs about their intelligence and abilities. Those beliefs underlie specific motivational styles and directly affect achievement. Children who are praised for their effort (e.g., You worked so hard on that puzzle!) develop an *incremental theory of intelligence*; they believe intelligence is malleable, something that can be cultivated through effort and time.[Granic et al., 2014] Therefore, by using video games with a well thought achievement scheme it is possible to keep expanding their learning by maintaining them motivated.

3 GazePlay

GazePlay is a free and open-source software which gathers around 40 mini-games playable with an eye-tracker. It is compatible with all eye-trackers which are able to control the mouse cursor, with Tobii EyeX and Tobii 4C on Windows and the Eye Tribe Tracker on Windows or MacOs X.[Schwab et al., 2018] All the games implemented aim to develop a skill from this set : Action-Reaction, Memorization and Selection.

Figure 2: Screenshot of the Gazeplay software

While implementing the game its creator must keeps in mind the skill aiming to develop into the player and also the games' simplicity, taking into account the target group, which are multi-disabled children.

4 Metrics for Eye-tracking

Eye-tracking (or oculography) is a research method that uses an eye-tracker device to track the point of gaze or eye movement of a person during a task execution[Borys and Plechawska-Wójcik, 2017]. The emergence of interactive eye-tracking applications has made eye-tracking research grow considerably. These applications cover domains like Neuroscience, Psychology, Industrial Engineering and Human Factors, Marketing/Advertising and Computer Science. The big variety of the current eye-tracking applications can be categorized into *diagnostic* or *interactive*.

For diagnostic applications, the eye-tracker gives quantitative and objective evidence of the user's visual and attention processes. His eye movements are recorded to establish the user's patterns over stimuli. As for interactive applications, the eye-tracker serves as an interface modality. The interactive system must interact with or respond to the user based on the eye movements. [Duchowski, 2002]

Software engineering (SE) researchers use eye-tracking technology to study the cognitive processes and efforts involved in different SE tasks. Eye movements retrieved by the eye-tracker are crucial to cognitive processes because they focus the user's visual attention to a visual stimulus that is processed by the brain.[Sharafi et al., 2015] Before continuing to introduce the actual metrics used by these researchers it is necessary to know the terminology behind it.

> *Fixation*: The stabilization of the eye on a stimuli for a period of time (200-300 ms). Fixations and cognitive processes are considered as connected under the assumptions that as soon as a user sees the stimuli, he/she tries to interpret it; also the user fixates his/her attention on the stimuli until he/she comprehends it.
>
> *Saccade*: The quick and continuous eye movements from one fixation to another. Saccadic eye-movements are within 40-50 ms. They are usually voluntary, contrary to micro-saccades which are small erratic eye movements that are involuntary.
>
> *Pupil dilation*: The widening of the pupil, which allows more light to get into the eye in low light conditions. Also happens when a user's mood or attitude changes or during complex cognitive tasks.
>
> *Scanpath*: A series of fixations in chronological order that represents a user's pattern of eye movements. [Sharafi et al., 2015]

Before conducting an eye-tracking study, researchers must choose metrics that are relevant to the multiple tasks and their inherent cognitive activities for each case study individually[J. K. Jacob and Karn, 2002]. [Sharafi et al., 2015] divide metrics into 4 main groups: metrics based on fixations, metrics based on saccades, metrics based on scanpaths and metrics based on pupil size and blink rate. The classification is made by taking into account the eye-movement used to achieve the metrics' result. The reader who wants a deeper understanding of these metrics and the way they are calculated can consult [Sharafi et al., 2015] fourth section. Whilst, we will focus on metrics based on scanpaths in the following section.

Eye-tracking has demonstrated to be an important tool in medicine. It has been used for medical diagnosis, progression or recurrence of disease over time and even as an objective measure of treatment's effectiveness. Eye-tracking metrics might be applied to different aspects of oculomotor behaviour, depending on the type of analysis. The most common metrics are based on fixations and/or saccades. These types of metrics might be used to define observer's engagement. Long fixation duration is correlated with high cognitive workload and higher cognitive effort. Metrics based on saccades are related to searching sequence of particular areas of interest. As opposed to fixations, during saccades visual information is not processed. The saccadic latency metrics allows to detect Parkinson disease, given the fact that an increased saccadic latency might be related to this illness. Through saccadic based metrics it is also possible to diagnose neural disorders affecting frontal cortex or the basal ganglia by conducting the *anti-saccadic* task. This task consists in the patient looking to the opposite direction of a displayed point. Another phenomenon detectable by eye-tracking is smooth pursuit. It consists in an observer following a movement is a presented stimulus. Problems with maintaining smooth pursuit eye movement (eyes move faster/slower than the object in motion) might indicate schizophrenia, autism, Alzheimer disease or Parkinson disease[Borys and Plechawska-Wójcik, 2017].

Despite delivering great results, eye-tracking comes with it's difficulties. The need to constrain the physical relationship between the eye-tracking system and the participant is one of the biggest barriers to incorporation of eye-tracking in more studies. At present the researcher has the choice of a system mounted remotely that restricts the participant's movement or of a firm and uncomfortable headgear. Moreover, eye-trackers produce data that can reach large quantities depending on the sampling rate of the eye-tracker and the duration of a session. Even if there exist software that allows the quick extraction of fixations and saccades, there is no standard technique to do this. Thus, the results can change drastically even for minor changes in the parameters definition a fixation and/or saccade. On top of that a scene that changes dynamicaly, caused by head/body movement provides additional challenges to automating eye-tracking data extraction. Even if the two previous problems were not considered there still remains another challenge, that of data interpretation. Eye-tracking data analysis can proceed either *top-down* : based on cognitive theory or design hypotheses, or *bottom-up*: based on observation of the data without predefined theories relating eye movements to cognitive activity. Since there is yet no standard in eye-tracking metrics and even for the fundamental terminology researchers have a hard time conducting their experiments.[J. K. Jacob and Karn, 2002]

5 Scanpaths

Scanpath is an interesting metric that was recently included in the GazePlay software. Before continuing with some insight on it's actual implementation we will give some information about this metric.

We will define the *scanpath* as eye movement data collected by a gaze-tracking device, where information about the trajectories(paths) of the eyes is preserved. This data is recorded while a participant scans the visual field, views and analyses any kind of visual information. Such data consists of gaze-direction, fixation position and duration, and saccade duration[Babcock et al., 2004].

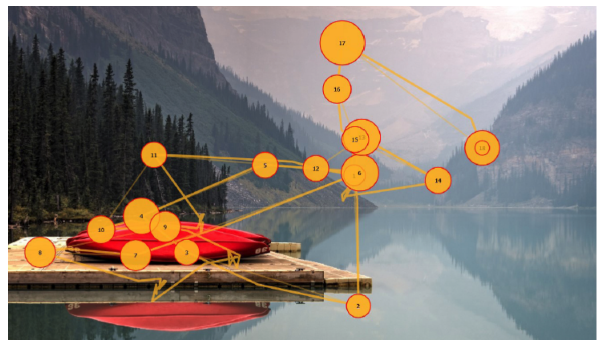

Figure 3: Example of scanpath of a static image

Eye-tracking has been very useful for detecting salient regions in interfaces that naturally attract the visual attention of users. Scanpaths are achieved through eye moves between different salient regions. There exists a relationship between scanpath direction and cognitive load during the navigation of an interface, as shown by research. Thus by analysing scanpaths it is possible to give insight into the cognitive load required to navigate an interface for specific users[Chakraborty and P. McGuire, 2016].

Although being an important metric, scanpath comes with it's challenges. There are several issues related to the implementation of this metric that we will discuss below.
Given the fact that this is a psycho-physiological measure, eye movements may contain gaze data that are unintentional and unconscious. Distracting events, either in the environment or visual stimulus may impact the eye movements and lead to noise in the collected data. As a consequence, there could be incorrect interpretations of data[Sharafi et al., 2015].
Moreover, the precision of data can be affected by a variety of factors. Such as individual factors that differ amidst participants like different eye physiologies or the relative position of the participant to the eye cameras of the tracker. Also environmental factors like lighting conditions, and effects coming from the eye-tracking hardware and software add up[Saez de Urabain et al., 2015].

Before adding the scanpath as one of the new features to the *Statistics* interface of the GazePlay software, heatmaps were available. Heatmaps represent visual attention by indicating not individual fixation points but clusters of fixation points focusing on an area of visual interest [Borys and Plechawska-Wójcik, 2017]. This metric is similar to the scanpath in terms of spatial information mostly and temporal as well (depicted by the color scheme of the heatmap). Although, it does not provide the sequence of eye movements followed by the participant during playtime.

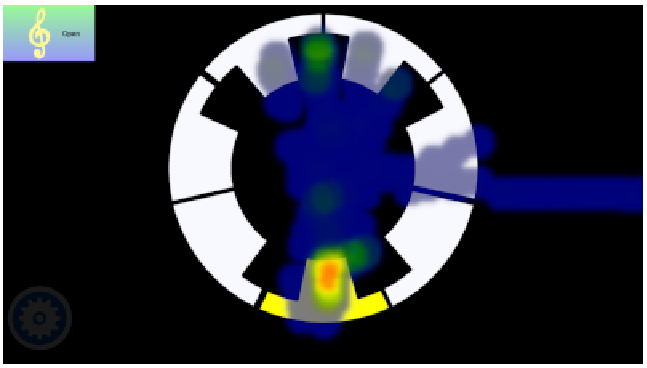

Figure 4: Heatmap output of Piano game from GazePlay

Scanpaths are based on both spatial and temporal information; when and where a participant looked. Thus, this is the core of the implementation for this metric. We defined a data structure containing the *x* and *y* coordinates of the fixation point, the *first gaze time* on this point and the *duration of the gaze* on this point. This will be used to build a picture of what is prioritized by a participant when they see a visual scene.

The first step took to build the scanpath was to collect all the data for each point of the visual stimuli where the participant gazed on. It is important to take into account a technical detail of the eye-trackers in this step. The eye-trackers sample the data according to their frequency, thus it is necessary to avoid collecting data for the same point coordinates(not to confuse for points that are revisited).

After having collected the data the next step is to build the image representing the scanpath. We went through three approaches before reaching the current result of the scanpath available in the GazePlay software.

The first approach was related to the geometric properties of the scanpath. The goal was to find a way to represent the fixation points of the scanpath's polyline. One way of representing them is to set a fixation point on the breaking points of the polyline. To do so we calculate the angle formed by 2 consecutive line pieces of the polyline and represent a fixation point only if the angle between these lines is a value of the angle threshold interval. The illustration below shows the idea behind this approach. The saccades would be represented as the lines connecting the fixation points.

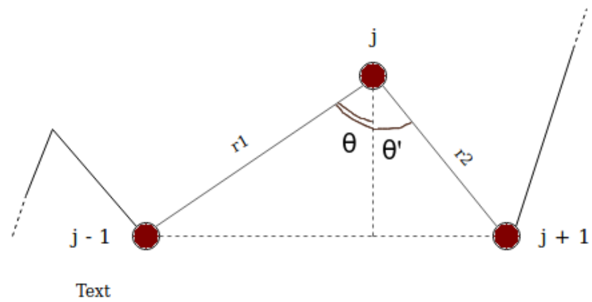

Figure 5: Geometry of the scanpath's polyline

The second approach is based on the fixation duration. This is a more intuitive and simple way of achieving the scanpath. A fixation point would be included in the scanpath image only if it's duration is bigger than a given threshold. For this threshold we used the fixation's average duration.

The third approach includes Vertex Reduction. After testing the first two approaches one can observe that the output is cumbersome and misleading. In the corners of the scanpath's polyline there are way too many fixation points represented. This is due to the fact that the eye cannot focus on an exact point of the screen but rather on a neighbourhood of points, when gazing at stimuli. In vertex cluster reduction, successive vertices that are clustered too closely are reduced to a single vertex. It is in our interest to reduce this cluster of points to a single fixation point to represent in a corner of the scanpath polyline.

Vertex cluster reduction is a brute-force algorithm for polyline simplification. Following this algorithm, a polyline vertex is discarded when its distance from a prior initial vertex is less than some minimum tolerance $\delta > 0$. After fixing an initial vertex V_0 , successive vertices are tested and rejected if their distance to V_0 is smaller than δ . In the case of a vertex being further than δ, it is accepted and becomes part of the simplified polyline[Sunday, 2012]. Below is a visualization of the procedure.

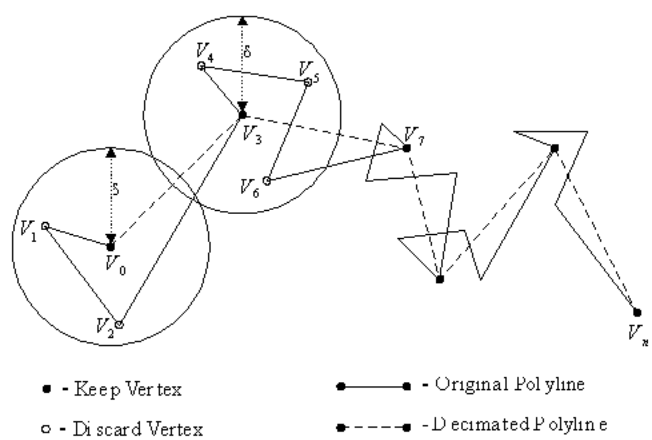

Figure 6: Vertex Reduction

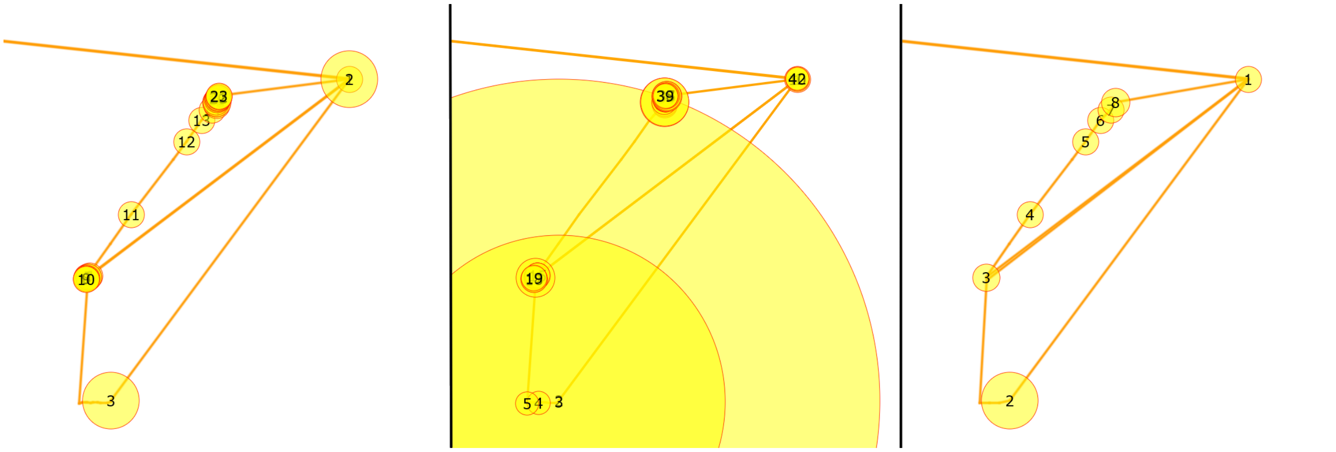

Figure 7: Scanpath output of the *Where is the animal?* game using various methods. The left scanpath is achieved by implementing the geometric approach, while the one in the middle using fixation length threshold and finally the one on the right by combining fixation length threshold and Vertex Reduction. We got these outputs after a playtime of 10.9s using a Tobii 4C eye-tracker.

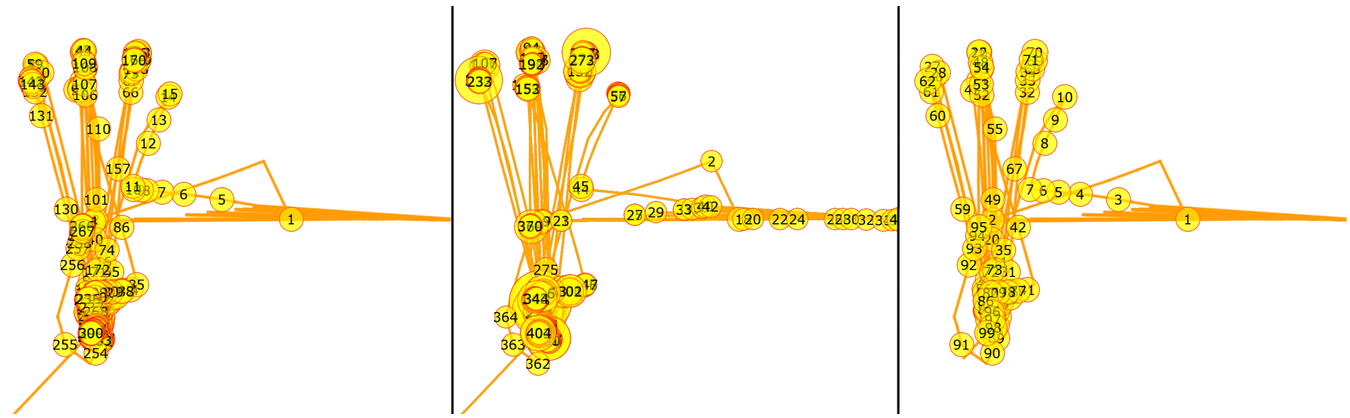

Figure 8: Scanpath output of the *Piano* game using various methods. The left scanpath is achieved by implementing the geometric approach, while the one in the middle using fixation length threshold and finally the one on the right by combining fixation length threshold and Vertex Reduction. We got these outputs after a playtime of 11s using a Tobii 4C eye-tracker.

After giving the methods used for implementing the scanpaths, we will make a comparison to see which one represents in the best way possible the visual scan of a participant. In the following images are shown the scanpath outputs for each of the methods. The lines represent saccadic movements of the participant's eyes while the circles represent the fixation points. The enumeration of the fixation circles represents the path followed by the participant during playtime. The radius is proportional to the time period of the fixation. Thus, a bigger radius for a fixation point means more time spent by the participant gazing there.

In this comparison we will include two types of games, a static and a dynamic one. By static game we infer that the visual stimuli of this game is not moving/or it's motion comes to an end rather fast, and by dynamic game we infer the fact that the visual stimuli is in constant motion. The scanpaths are outputs of a similar playtime for both of the games. Moreover the eye-tracker used to play the games is the same Tobii 4C eye-tracker.

First we will start by comparing the different scanpath outputs of the static game. As depicted in Fig.7 the path is rather simple, this due to the static nature of the game. The player does not have a lot of stimuli as he/she would have had in a dynamic game. We can see that the number of fixation points illustrated in the middle scanpath is the highest (39), compared to 23 for the left scanpath and only 8 for the right one. This is expected since for the left scanpath we used a Vertex Reduction algorithm to avoid redundancy in visualisation. We can observe that the left and right scanpaths are rather similar, but we agree that the left scanpath (achieved by geometric approach) introduces some redundancy which is avoided. It is worth mentioning that the middle scanpath shows some exaggerated radius for some fixation points, this is due to an accumulative approach for the fixation points used here which was discarded for the other two methods. After several tests of these methods in different static games we concluded that the best approach, for generating scanpaths, would be to combine Vertex Reduction for polyline decimation and exploitation of the natural fixation length threshold (like the rightmost scanpath).

We will move on to comparing the different scanpaths of dynamic games. The first thing we notice in this category of scanpaths in Fig.8 is the high number of fixation points present. This is due to the dynamic nature of the game, where the stimuli are higher and thus the visual reaction too. Like in the previous scanpaths, the highest number of fixation points is present in the middle one, counting 404 points. While the left scanpath contains a total of 300 and the rightmost 99. In addition, we can notice a simplification of the polyline in right scanpath, although very subtle. We may make the incorrect assumption that the fixation point illustrated in the middle scanpath is missing on the outermost scanpaths, but this is false because that corner in the polyline represents a saccade. For this reason in the other scanpaths this point is omitted via the fixation length threshold. We can conclude that even for this kind of games the best representation of scanpath is via the third, rightmost approach.

Even if we concluded on the most adequate approach for the generation of the scanpath, there is still room for improvement. Scanpath generation breaks down to fixation and saccade identification - that is, translation from raw eye-movement data points to fixation locations and implicitly saccades between them, on the visual display. Poorly identified fixations produce wrong metrics and thus biasing interpretation [Salvucci and Goldberg, 2000]. As mentioned earlier in this section there are some challenges to eye-metric calculation/generation. For our particular case we encountered problems with the device itself. The positioning of the eye-tracker in relation to the player induced inaccuracies, thus a well placed static behaviour is needed. This is an issue taking into account the nature of the software and its target group being children with impairments who can manifest continuous movements. In addition, since there is no standard method for fixation and saccadic identification we are not totally sure on the validity of our results and comparison to other researchers work is difficult to make because of this limitation. Our threshold values and polyline simplification methods may cause loss of data or would be negligible if we had a reliable source to compare.

6 Conclusions

AAC through eye-tracking is an excellent way of enriching the possibilities of interaction with their surrounding for children with multiple disabilities. Combining the power of cognitive learning through video-games and eye-tracking gives these children the possibility to learn in the most similar way possible to their non-disabled comrades. Eye-tracking metrics are crucial in this mission for the fact that they help the professionals behind the scenes prepare such games and also diagnose their medical status.

Scanpath is an important metric we chose to integrate into Gazeplay for the numerous reasons explained above. Although, like many other eye-tracking metrics it comes with various limitations. Apart from the physical limits, which can be overcome by redesigning the eye-tracking hardware to make it lightweight so it follows the head movement and avoids calibrating issues, there exist data extraction related issues.

These issues are related to the ambiguity of identifying fixations and saccades, which are the pillars of eye-tracking. So a step forward would be to finally arrive to a standardization of this identifying process. Apart from this, in our scanpath generation approach there is room for improvement concerning the fixation time length represented via the radius of a fixation point. Our approach is rather evasive and not well adapted to the Vertex Reduction algorithm used for the scanpath generation. We used a cumulative approach for the fixation duration of a point, meaning that the accepted vertices of the polyline have a duration equal to the summation of the discarted vertices. Perhaps a better approach would be to interpolate between these values and visualise the most representative one.

Furthermore, in other versions on GazePlay it would be of great interest to propose comparative algorithms for the generated scanpaths in order to understand the gazing pattern of the players.

References

[Babcock et al., 2004] Babcock, J., Cattaneo, Z., Dixon, T., Maringelli, F., Land, M., Nikolov, S., and Pelz, J. (2004). Online archive of scanpath data — *http://www.scanpaths.org (2019-06-12)*.

[Borys and Plechawska-Wójcik, 2017] Borys, M. and Plechawska-Wójcik, M. (2017). Eye-tracking metrics in perception and visual attention research. *EJMT*, 3:11–23.

[Chakraborty and P. McGuire, 2016] Chakraborty, J. and P. McGuire, M. (2016). Directional scan path characterization of eye tracking sequences: A multi-scale approach. pages 51–61.

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

[Granic et al., 2014] Granic, I., Lobel, A., and Engels, R. C. M. E. (2014). The benefits of playing video games. *The American psychologist*, 69 1:66–78.

[J. K. Jacob and Karn, 2002] J. K. Jacob, R. and Karn, K. (2002). Commentary on section 4. eye tracking in human-computer interaction and usability research: Ready to deliver the promises.

[Kato, 2010] Kato, P. M. (2010). Video games in health care : Closing the gap.

[Saez de Urabain et al., 2015] Saez de Urabain, I. R., Johnson, M. H., and Smith, T. J. (2015). Grafix: a semiautomatic approach for parsing low- and high-quality eye-tracking data. *Behavior research methods*, 47:53–72.

[Salvucci and Goldberg, 2000] Salvucci, D. D. and Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 71–78. ACM.

[Schwab et al., 2018] Schwab, D., Fejza, A., Vial, L., and Robert, Y. (2018). *The GazePlay Project: Open and Free Eye- trackers Games and a Community for People with Multiple Disabilities. ICCHP 2018 - 16th In- ternational Conference on Computers Helping People with Special Needs, Jul 2018, Linz, Austria.*

[Sharafi et al., 2015] Sharafi, Z., Shaffer, T., Sharif, B., and Guéhéneuc, Y.-G. (2015). Eye-tracking metrics in software engineering.

[Sunday, 2012] Sunday, D. (2012). Polyline decimation — *http://geomalgorithms.com/a16-_decimate-1.html (2019-06-11)*.