Video games for visual interface (eye-tracking)

Wafaa EL HUSSEINI
MOSIG - Grenoble INP
Grenoble, France
wafaa.el-husseini@grenoble-inp.org

Supervised by: Didier SCHWAB.

I understand what plagiarism entails and I declare that this report is my own, original work.
Name, date and signature:
Wafaa EL HUSSEINI - June 12, 2019 WH

Abstract

People with disabilities face challenges and difficulties in interacting with their environment, on a daily basis. They tend to rely heavily on the help of others especially when it comes to dealing with the digital world. Gaze is considered one of the most natural and easy form of expression for these people, and that can be used as an important form of Augmentative and Alternative Communication (AAC) [1]. However, gaze trackers and applications that use these trackers are not widely available, especially not at affordable prices. Using Eye-trackers and the GazePlay[1] open-source platform, we try to improve the way these people interact with computers and their environment, while developing their cognitive skills.

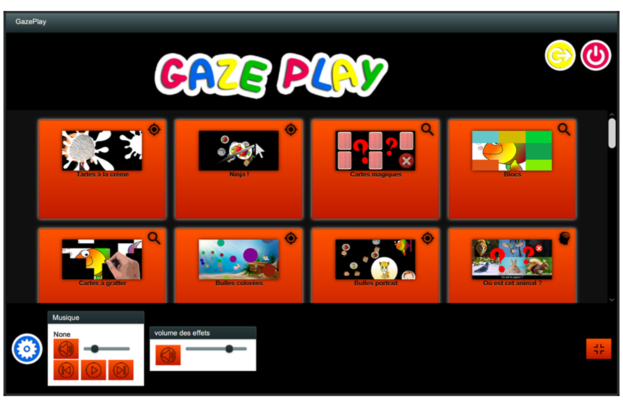


Figure 1: GazePlay Main Menu

1 Introduction

When it comes to motor-disabled individuals, the lack of autonomy is a big concern. Their inability to perform even basic movements causes them to be unable to make decisions on their own or to be in control of their own body and of the environment around them. It is usually the case however, that they are in control of their own gaze. Exploring their capacity by using their gaze can provide them with the sense of autonomy they need, while being a good use of their time.

The GazePlay project has both social and research related objectives. The social benefit of this project is to be able to cater for the entertainment and educational needs of a portion of the society that is often ignored, or not taken into consideration. The research aspect of this project lies in the study of the Gaze and it potential uses in several fields. Gaze-detection technology is still lacking in terms of accuracy today. However, disabled kids with slower reaction and movement time than regular people, can benefit greatly from the available services.

Motor disabled children being our main targeted users, this paper will detail the work done to develop the GazePlay platform through discussing the implemented games and their importance, then through explaining the extracted metrics based on these games and how they can be analyzed and used to improve the user experience, and more generally the gaze-tracking.

2 Context & Background

2.1 Eye-Tracking and Video Games

Eye-tracking, which dates back to the 1900s, is a way to detect the position of the person's gaze fixation on the computer screen and use that information to execute commands and perform tasks based on this fixation [3]. Eye-tracking has been finding its way into video gaming in recent years, with eye-trackers being used as tools to improve the Gameplay[2] experience. Researchers have been investigating ways to improve the gaze detection and compare it to traditional input (through mouse, keyboard, and gamepad, etc) in terms of accuracy and responsiveness [4]. Gaze detection can therefore be an additional dimension used to improve the Human Computer Interaction (HCI) experience, while gaming or even while performing simple computer tasks. We can imagine the gaze replacing the key-board controls when it comes to changing the camera view of a 3D scene, to mirror real life

[1] https://gazeplay.net [2]

[2] Gameplay refers to the way a player interacts with a video game.

gaze focus movement, similarly to virtual reality (VR) headsets.

2.2 Gaze-based Computer Controls

As opposed to the traditional ways to provide user input to the computer, with tools such as a mouse, a keyboard, a TrackPad, a joystick etc., eye-tracking has a limited range of provided inputs, as the eye motion is limited to: blinking[3], fixation[4], pursuit[5] and saccade[6] [1]. Applications implementing gaze-based functions have to take this restriction into consideration and be simple enough to only be controlled through these types of input.

In addition, not all of these gaze movement can be easily and accurately detected by the eye-tracker, blinking for example. Blinking is usually very fast that the tracker may miss recording this action. Blinking therefore might not be a good command to use for executing commands. Fixations and saccades are more easily detectable, on the other hand.

Using the available technology, it is therefore important to be able to set standards for interpreting the eye movement we detect and apply or adapt them to computer commands. For example, we could have standards on the duration of the fixation on a certain screen location, that can be interpreted as a ”Key Press” command. Similarly for scrolling.

2.3 Benefits for people with disabilities

People with motor disabilities or cognitive impairment may not be able to control their hands or feet movement as they wish to, therefore using typical tools to manipulate a digital interface (computer, phone, tablet) would be extremely difficult.

A study made on individuals with the Rett[7] syndrome, a rare syndrome that can result in the inability of the person to use conventional communication methods or control their own movement, showed the positive impact of the Eye Gaze technology in developing cognitive skills for the people who constantly used it [5]. After constant use, the families of users reported an increase in the awareness and engagement of their children, as well as in their understanding of the languages.

Gaze-based activities are therefore important as a form of communication, education as well as entertainment. We explore there benefits through our platform.

2.4 GazePlay: An Overview

GazePlay is a platform containing several short and simple games that can be played with the gaze, using eye-trackers. Figure 1 shows the main menu of the platform with a snapshot of some of the included games. GazePlay aims to entertain the users all while teaching them three main skills [1]:

- **Action and Reaction:** by showing the child the consequence of his gaze location on the screen. For example, in shooting games, looking at a target removes it from the screen, the target is therefore shot and the goal achieved.
- **Selection, through fixation:** the child is encouraged to focus/fix his gaze on one area on the screen in order to receive feedback. We generally include a progress indicator in our games, where relevant, to let the user know for how long to fix their gaze before they can receive feedback. For example, in Maths games, three options are given: selecting a correct answer until the progress indicator disappears results in a positive feedback while selecting a wrong one results in a negative feedback.
- **Memorization:** the child is encouraged to improve his short-term memory through games such as the hidden pairs. Similarly the child is encouraged to memorize letters or words by repeatedly seeing or listening to the said word.

By training children to use Gazeplay, they are able to develop these skill and better master their eye muscle movement. Children will then be able to better communicate with their environment. Gazeplay therefore aims to provide a more universal and affordable means of communication for people with disabilities, as opposed to how currently solutions are implemented on a per-case basis which makes it very costly for the patients.

3 Games and Interface Improvements

The majority of the work focused on increasing the number of games offered by the platform as well their variety. The aim was to tackle various learning goals. In addition, it was also important to make this service available and accessible to a wide audience, namely the Arabic speaking community.

(a) Arabic - RTL (b) English - LTR

(c) Configuration Panel - AR (d) Configuration Panel - EN

Figure 2: HomeMenu Screen Layout Comparison

3.1 Arabic interface

Gazeplay is made available for several languages, including English, French, Arabic, Albanian, Chinese and many others.

[3]The act of shutting and opening the eyes very quickly.

[4]The act of fixing your gaze at a specific point for an extended duration.

[5]The act of following a certain target with the gaze.

[6]A brief movement from one area of interest to another.

[7]RTT is a rare neurodevelopmental disorder found predominantly in females, characterised by stages of stagnation, regression and (possible) recovery in skills following seemingly near-normal development in the early years of life [5].

Arabic, however, is one of the few languages used today that can only be written from right to left (RTL), which imposes several design constraints.

According to a study by UXBERT Labs on Arabic user interfaces, ”Users on Arabic language sites start from the top right corner of the page, scan across the top and then scan down the right hand side of the content” [6]. In the context of GazePlay, this means that the focus of the Gaze of the player will be on the right side before moving to the left.

In addition, the Arabic language poses additional font sizing and styling constraints, due to both the difference in the shape of its letters, as well as its wordiness. An adapted interface with all the text positioned to the right side, and the illustrations, and layouts modified according to the new text, was implemented for better readability of the interface. Figure 2 shows a side-by-side comparison of the two layout versions.

3.2 Games implementation

Based on the original implementation of Gazeplay which relies on the JavaFX API, and sets a solid foundation for the translation of the gaze data from eye-tracking devices to commands used within the games, we extended the games selection by developing several new simple games, to cater for the educational needs of the children.

The games should be simple and easy to understand for children who are less than 10 years old. The games should have a clear educational goal all while being catchy and entertaining for the kids.

We introduced new skills that the targeted children can benefit from, being literacy (or language processing) and logical thinking and problem solving through mathematical problems.

A Step Towards Literacy

The UNESCO defines literacy as ”the ability to identify, understand, interpret, create, communicate and compute, using printed and written materials associated with varying contexts” [7].

In GazePlay’s context, literacy is the knowledge and competence of the child in the area of reading, recognizing and using the intended language. Because of the potential difficulties in writing or speaking that these children may have, the minimum objective of our games is to help them recognize and understand the language.

Literacy is a very important skill for the kids that Gazeplay heavily addresses in the latest release of the game. Recall that one of the main goals of this platform is to help break the communication barrier between motor disabled children and their environment. The first step towards achieving literacy is to teach children the language of the people from their environment.

To achieve the literacy goal, several games based on letters recognition were implemented. These games shared common behaviors; the Find The Letter game goes as follows:
- Player is asked to select a certain letter.
- The pronunciation of the letter is played upon selection for the child to get used to the sound of the letter and be able to associate the shape to the sound.
- Similarly for basic words: colors, animals, numbers etc.

Through repetition the child will improve two skills: gaze-based selection and literacy.

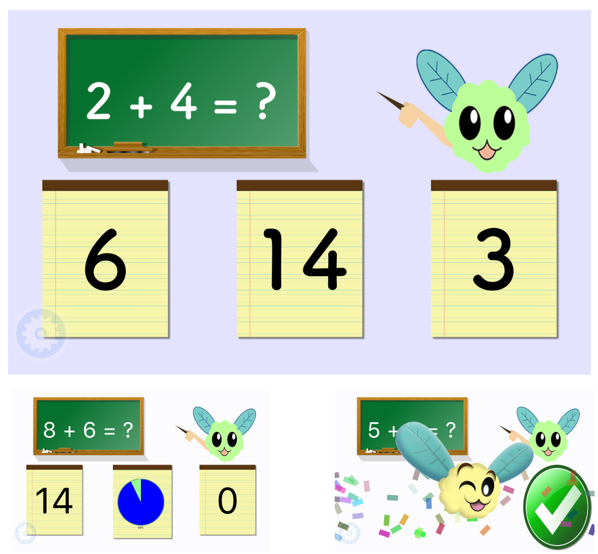
Currently, literacy games are limited to the English and French languages only.

Math and Logical Reasoning

Over the years, there have been many studies that linked Mathematical knowledge and skills at early ages to better performance in various subjects in their future educational years. Specifically, when it comes to ”pattern recognition, measurement, and advanced number,” results ”are most predictive of eighth-grade outcomes” [8].

Although in the case of motor disabled children, regular schooling might not be possible, we cannot ignore the importance of mathematics as part of their education. They should also have the opportunity of early exposure to mathematics.

To teach the children basic math, a simple question and answer math game was developed. The game features a random math operation question with two operands, and 3 sets of possible answers. Selecting a correct answer will result in positive feedback. As discussed earlier, the game is simple to understand and has a clearly defined goal. Figure 3 shows screenshots from the Math101 game. The layout of the screen is designed so that the child is not overwhelmed. The areas where the player should fix his gaze to play are clear, and accurate feedback is given based on their selection.



(a) Progress Indicator (b) Correct Answer Feedback

Figure 3: Screenshots from the Math101 game

When playing this game for a long time, the child will be able to memorize the result of these operations and answer them in a quicker way, and later advance to other levels. The game has several levels, each one focusing on a one of the basic math operation (101: addition, 102: subtraction, 103: multiplication, 104: Division and 201: all operations).

4 Gaze Metrics

We developed several metrics to analyze the user behavior when interacting with the platform so that we can extract relevant information. During gameplay, the recorded data is a timestamped sequence of points denoting the recorded screen location visited by the gaze.

The metrics we can extract out of these points are either based on the path taken by the gaze throughout the gameplay, or on the areas on the screen that are most visited, denoted by areas of interest [9].

4.1 Path-based Metrics

Figure 4 shows an example of metrics that are based on the gaze path, namely the heatmap and the fixation sequence. Heatmaps describe how "hot" a screen region is depending on how long it was visited. Fixation sequences, also known as scanpath show the path of the gaze in the order in which it moved and highlights interesting fixations that are of a bigger duration than a set threshold[10]. We can think of these fixations as a "select" command on a certain point.

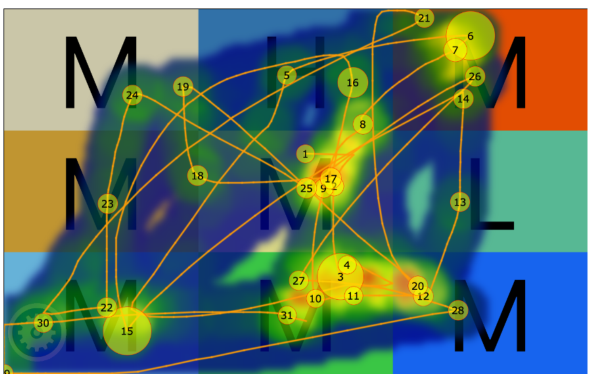


Figure 4: HeatMap & Fixation Sequence on 'Find the Letter'

4.2 The Area Of Interest[8] Concept

Tobii defines an Area of interest as a concept used "to calculate quantitative eye movement measures"[11]. As the name suggests, it is an area on the screen that is of interest and can be studied. "While not strictly a metric by itself, it defines the area by which other metrics are calculated" [9]. Once we are able to define these areas, we are able to make computations per area. For example, we can count the number of fixation on one area as a factor of the total number of fixations and determine the most visited area (Ratio).

4.3 Defining the Areas of Interest

There are several methods for defining AOIs that have been studied in recent years.

An important consideration to remember when defining AOIs is to know the nature of the game we are analyzing. As such, we distinguish two types of games in GazePlay:

[8] Also referred to as AOI. This paper uses the two terms interchangeably.

Static scene games

- The layout of the game rarely changes.
- Each part of the screen is exclusively reserved for one predefined purpose.
- The scenes of these games do not evolve over time.

Dynamic scenes games

- The layout of the game, and the location of its artifacts are randomly generated.
- There is no part of the screen that is reserved for a predefined purpose.
- The scenes of these games evolve over time.

For the static scene games, the AOIs can be predefined prior to playing the games, along with their function. In this case, the AOIs are assumed to be applied to a picture, and the calculated metrics will also be performed on a picture. The analysis of the static AOIs is said to be made *a priori*, as it requires analyst intervention before gathering data.

For the dynamic scenes games, the AOIs need to be computed and updated over time, and we cannot expect their function until after the calculations are done. In this case, the AOI are computed based on a video, and the calculated metrics will also result in a video showing the details. The video will have to be analyzed to define the AOI functions [10]. The analysis of the dynamic AOIs is said to be made *a posteriori*, as it requires analyst after the collection of data.

In this paper we discuss the static areas of interest.

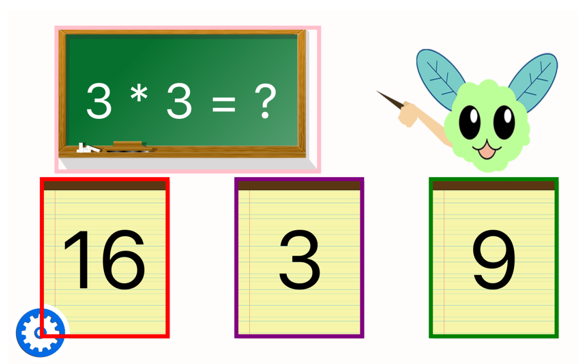


Figure 5: Static AOI boxes for Math games

4.4 The Static AOI

All of the previously developed games (math and letters) fall under the category of static scene games. To be able to analyze these games, we implemented the static AOI and its corresponding metrics.

Recall that the static AOI is based on a static screen view, namely a picture. The AOIs are defined beforehand and then user response is analyzed.

If we take the Math101 game from Figure 5 as an example, we identify four main static AOIs: The question area and the three answers areas. The question area can provide interesting insight on the difficulty of the game as we can study the average time spent looking at the question per round, and compare it to the time spent attempting to answer. We can also measure how many times a player visits the question per round which should give us an insight on how easy to understand the instruction is. The quit button can be considered an additional AOI[9].

The static AOI does not take into consideration the time factor. It assumes that the areas on the screen have a fixed function over time and computes area based metrics accordingly. This limitation translates into the AOI definition not being accurate across rounds, but being accurate per round. To further explain this point we take the Math101 game as an example. It is true that the game has three answer areas, however only one of these areas contains the correct answer at each round. Also, at each round the correct answer location is randomly assigned. This means that the function of the answer area is changing over time for new rounds, although their location is still the same.

To address this issue, we made sure to calculate the metrics and store them on a round by round basis. In this way we can keep track of the accurate function of the area, and then calculate the average of our metrics based on the accurate function. For example, we can calculate the number of times the player gave a correct answer from the first try, as a function of the number of rounds played.

4.5 Static AOI based Metrics

One important advantage the static AOI method has over the dynamic AOI method, when it comes to static games, is the fact that you can automate the analysis process of the extracted data, meaning that you can directly have meaningful analytical sentences covering all the AOIs. For example a possible sentence to be extracted is: "You spent X time looking at the question, and Y time attempting to answer, only Z% of which were correct answers." Such analysis is not possible for dynamic AOIs due to undefined functions.

This advantage also translates into so metrics being more relevant for static AOI than dynamic AOI. We outline all the implemented metrics below:

Time To First Fixation

The exact time at which the player first fixated their gaze on the said AOI.

Average First Fixation Time

The average duration spent at an AOI per fixation.

Time Spent & Fixations count

Both of these measure measure the amount of time spent fixating on an AOI and can be used to compare the AOI to the other AOIs in the game.

Dwell Time

"Dwell time is the time that a gaze remains in a particular AOI, from entry to exit" [12]. Dwell time is therefore a more accurate measure of the total time spent on a certain AOI, as it does not ignore saccades.

[9] The quit button can be considered an AOI for all games. This AOI measures the time it takes for the kid to give up on a game and exit it. Therefore this AOI is a measure of how successful a game is in capturing the attention of the child.

Revisits

"The number of revisits provides information about how many times a participant returned their gaze to a particular spot"[9]. Revisiting a spot is only relevant when this spot is fixed. Therefore we count the AOI revisits per round.

Example: Average time per round the player revisited the question area.

Ratio

The ratio is a versatile metric that can indicate several things: as it is a measure of a part that is compared to a whole. In our implementation, we calculate the ratio as an overall metric and not per round. In other words, we calculate the number of rounds for which the user visited an AOI regardless of the AOI function, and divide it by the total number of rounds.

5 Conclusion

Gaze behavior is a skill that can be acquired by children at a young age. For motor disabled children, this means that they can use this skill to fight the communication barrier imposed by their disability. Our work on GazePlay has further improved the platform and contributed to making it more accessible and available to a wider range of audience. In addition, the continuous addition of new games of various types made it so the platform is more enriched so that it can contribute to the children's education and become part of their daily used tools. Furthermore, the newly developed static AOI metrics made it so that both care-taker of the kids and the analysts can analyze the performance of the children. Despite their limitations, static AOIs are very well adapted for most games on GazePlay. However, a good approach would be to try and test other algorithms that may bring down the gap between the complexity of the dynamic AOI and the straightforwardness of the static AOI.

References

[1] D. Schwab, A. Fejza, L. Vial, and Y. Robert, "The GazePlay Project: Open and Free Eye-trackers Games and a Community for People with Multiple Disabilities," in *ICCHP 2018 - 16th International Conference on Computers Helping People with Special Needs*, vol. 10896 of *LNCS*, (Linz, Austria), pp. 254–261, Springer, July 2018.

[2] "Open Source Gaze-Based Game Platform." https://gazeplay.net. [open source software].

[3] A. Kar and P. Corcoran, "A review and analysis of eye-gaze estimation systems, algorithms and performance evaluation methods in consumer platforms," *CoRR*, vol. abs/1708.01817, 2017.

[4] J. Antunes and P. Santana, "A Study on the Use of Eye Tracking to Adapt Gameplay and Procedural Content Generation in First-Person Shooter Games," *arXiv e-prints*, p. arXiv:1801.01565, Jan 2018.

[5] G. Townend, P. Marschik, E. Smeets, R. van de Berg, M. van den Berg, and L. Curfs, "Eye gaze technology as a form of augmentative and alternative communication for individuals with rett syndrome: Experiences of

families in the netherlands," *Journal of Developmental and Physical Disabilities*, vol. 28, pp. 101–112, 2016.

[6] UXBERT-Labs, "Designing an arabic user experience." http://uxbert.com/wp-content/uploads/2017/05/Designing-An-Arabic-User-Experience-Ebook-by-UXBERT-Labs-min.pdf, 2017.

[7] S. Montoya, "Defining literacy: Unesco." http://gaml.uis.unesco.org/wp-content/uploads/sites/2/2018/12/4.6.$1_0 7_4.6 - defining - literacy.pdf$, 2018.

[8] A. Claessens and M. Engel, "How important is where you start? early mathematics knowledge and later school success," *Teachers College Record*, vol. 115, no. 6, pp. 1–29, 2013.

[9] B. Farnsworth, "10 most used eye tracking metrics and terms." https://imotions.com/blog/7-terms-metrics-eye-tracking/, 2018.

[10] G. Drusch, J. M. C. Bastien, and S. Paris, "Analysing eye-tracking data: From scanpaths and heatmaps to the dynamic visualisation of areas of interest," in *International Conference on Applied Human Factors and Ergonomics*, (Krakow, Poland), 2014.

[11] TobiiPro, "Digging into areas of interest." https://www.tobiipro.com/learn-and-support/learn/steps-in-an-eye-tracking-study/data/digging-into-areas-of-interest/.

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