Comparing Eye Tracker Input to Traditional Input for Measuring Player Performance, Immersion, and Engagement in an Emergent Narrative

Balázs Gyula Koltai
Aalborg University
Department of Architecture, Design, & Media Technology
Aalborg, Denmark
bkolta15@student.aau.dk

Mózes Adorján Mikó
Aalborg University
Department of Architecture, Design, & Media Technology
Aalborg, Denmark
mmiko15@student.aau.dk

Robert Wittmann
Aalborg University
Department of Architecture, Design, & Media Technology
Aalborg, Denmark
rwitt15@student.aau.dk

ABSTRACT

In this paper we compare eye tracking input (ETI) to traditional input (TI), and evaluate what effect it has on player immersion, engagement, and performance. There is little research done in this field, and to our knowledge, engagement has not been measured with eye tracking before. Based on our research, we developed a custom game with ETI in mind that has an emergent narrative and focuses on exploration and visual searching. We examine how ETI affects our dependent variables. Due to complications involving limited test equipment, we were unable to reach valid results in our testing. Even so, initial results seem promising and encourage further testing.

KEYWORDS

Eye tracking, Game, Immersion, Engagement, Performance, Emergent Narrative

1 INTRODUCTION

From open world sandbox games to action-packed shooters and laid back puzzle games, there are a large number of different genres today. Estimates show that the market will likely more than double its size to become a $300bn dollar industry by 2025 [1].

Gaming is traditionally receptive to novel input devices driving their commercial adoption and proliferation [36]. The most commonly used input methodologies include mouse, keyboard and gamepad. Nowadays eye trackers are also becoming a commercially available options to home users, with Tobii manufacturing the Tobii Eye 4C specifically for gaming. Tobii currently has 152 eye tracking enabled games across 16 genres listed on their website [2].

The area of eye tracking in video games in relation to immersion and engagement is not a well-researched field, as most eye tracking research is focused on applications relating to health, user-experience evaluation, marketing, and accessibility [36]. Eye tracking enabled games usually substitute or augment traditional input with eye tracking input, such as aiming at gaze and dynamic lighting based on gaze location. This means that games designed specifically for eye tracing interaction are scarce.

Previous studies covered eye tracker use across game genres and future possibilities with this input modality, and categorised the involved eye movements [13, 36]. Studies confirmed that the use of eye trackers can increase player immersion in video games [11, 29, 32]. In these studies, either commercially available titles were modified to take eye tracker input, replacing traditional input like keyboard and mouse, or custom-developed games were made to make use of eye tracker technology. Aside from immersion, player performance was also measured in these papers. Results indicate that eye tracker input has a universal effect of increasing player immersion across game genres, while its effect on player performance depends on the game genre and the tasks players have to do [11, 14, 17, 18, 26, 32].

Among the reviewed studies, we noticed a trend that very little focus was put on narratives and engagement, and among those that featured narratives, emergent narratives were absent.

We set out to develop a game with an emergent narrative specifically with eye tracker input in mind. Our goal was to measure how eye tracker input affects player immersion, engagement and performance in the game compared to traditional input. Our game required participants to perform a visual searching task while traversing the game world, which was altered based on player action.

Our findings show that average dwell time on certain objects was shorter in the eye tracker input condition than in the traditional input condition. While engagement scores had a strong correlation with immersion scores, we could not find correlation of these scores with our other measurements, and 4 out of 5 measurements were not statistically significant between the two conditions.

2 BACKGROUND

2.1 Eye Tracking

The term gaze is usually defined as the point in space where a person is looking at [36]. We will continue to use this definition of the term. Gaze tracking or eye tracking refers to methods of gathering data on the users' gaze. In human-computer interaction (HCI), eye tracking has been used in a number of experiments, ranging from increasing immersion [11, 39] to dynamic difficulty changes in games [4] and guiding the users' gaze [5].

As technology advanced, different eye tracking devices were in use. At first the devices had to be attached to the user's skin or eye. Chennamma et al. gathered the evolution of these devices [9]. The first methods of eye tracking were called Electro-Oculography, where electrodes were attached to the area around the eye, and by measuring the potential differences in the areas, the gaze direction could be estimated. Another, very accurate method called Scleral Search Coils required modified contact lenses inserted in the user's eyes, then the coils in the lenses could be used to find the orientations of the eyes in a magnetic field. This method, however, was not used in an HCI setting. Infrared Oculography measures the intensity of infrared light reflected back from the surface of the eye, and it can be more accurate than electro-oculography, but it is more sensitive to external light.

Video Oculography methods, like the screen-based Tobii eye tracker devices use a method called "pupil centre corneal reflection". In their implementation, IR illuminators shine a pattern into the user's eyes, and the cameras are recording the reflected patterns from the surface. These prukinje reflections [33] are then sent through the device's image processing algorithm, where their positions and intensities can be used to calculate gaze direction. As our research is focused on games, and we intend to measure immersion and user engagement, we have decided to use a commercially available device, namely the Tobii 4C, to conduct eye tracking.

The most common measures in eye tracking studies are saccades and fixations [33]. Fixations are periods where the eyes focus on an object and saccades are fast eye movements between fixations. Scanpaths are a cycle of a saccade-fixation-saccade. Combinations of the specific measures in these categories can be used to evaluate different uses of eye tracking. The third type of eye movement is called Smooth-Pursuits [36]. This movement happens when the person is tracking a moving object with their eyes, matching their relative velocity. Smooth-Pursuits cannot be faked, thus can be used to see if the user is actually following a target. However in our research this type of movement was not relevant.

2.2 Previous work

Use of eye tracking in video games was investigated with regards to immersion and performance [11, 29], and to see how well eye tracker input (ETI) performs in comparison to traditional input (TI), like keyboard and mouse [11, 14, 17, 18, 26, 32].

2.3 Immersion

Immersion is usually defined as

> The perception of being present in a non-physical world [11].

According to findings by Vidal et al., Ejdemyr, Smith and Graham, Isokoski et al., and Jönsson, use of eye tracking as an input modality elevated players' sense of immersion while playing [11, 13, 18, 32, 39]. The results regarding immersion come from multiple game genres which might point towards gaze input having a universal effect in increasing immersion. Ejdemyr, Isokoski et al. and Vidal et al. used a custom-developed game while Smith and Graham and Jönsson used commercially available titles where input had been modified for the ETI condition. In cases where two input modalities were tested with a game, the game versions were identical except for which input was used. In Vidal et al.'s case, no input methodologies were compared, as the study examined the use of social gaze interactions as a game mechanic. Players reported elevated sense of immersion, as the use of this mechanic led to feelings of presence in the game world [39].

Tamborini and Bowman consider three categories of presence: *spatial presence, social presence* and *self presence.* Spatial presence refers to the sense of being physically located in a virtual environment. Social presence refers to how users in a virtual environment experience virtual social actors as if they were actual social actors. Finally, self presence refers to the users experiencing their virtual selves (avatars) as if they were their actual selves [35]. Vidal et al. do not specify what they meant by presence, but in the context of their game, all of the above three dimensions of presence apply.

With the exception of the Royal Corgi that featured a branching narrative, the games that were used in the above studies did not put emphasis on narratives [39]. To our knowledge, immersion has not been tested for in games with an emergent narrative. In our experiments, we tried to see how input methodology would affect immersion in our game, which features an emergent narrative.

2.4 Player performance

Player performance is defined by Ejdemyr as

> The measure of how well the player completes the objective of the game [11].

In the reviewed studies, player performance measurements included time to complete tasks [11, 17, 26, 32], score [11, 18, 32] or progression through the game levels [29].

Smith and Graham used 3 commercially available games for their studies. The 3 games were from 3 different genres: A First-Person Shooter (FPS), a Role-Playing Game (RPG), and an Arcade game. Each utilised input modalities in different ways. In the FPS and RPG games, completing the tasks in the ETI condition took longer than in the TI condition. In the Arcade game players achieved a higher score in the TI condition than in the ETI condition. However statistical difference was not found for the FPS and RPG games, only for the Arcade game [32].

Ejdemyr developed a custom First-Person Adventure (FPA) game in which the goal was to find 6 keys as fast as possible, each awarding a certain amount of score. The game also featured enemies, and making contact with them decreased the player's score. Average completion times in TI condition were about 27% faster that in the ETI condition, and 90% of the participants got a lower or equal score with playing the ETI version of the game. The study determined the differences to be significant [11].

Orlov and Apraksin compared gaze-contingent input to mouse-based control in a custom-developed strategy game, where users had to find the unit with the lowest health among 5, 10, 15, 20 and 25 units in five trials. Selecting the unit by moving the mouse pointer onto it (TI condition) or by looking directly at the unit (ETI condition) displayed the unit's health. Recognition times in the ETI condition were lower than in the TI condition. They conclued that in a visual searching task such as this, gaze-contingent interaction is 1.5 times more effective [26].

Isokoski and Martin's study did comparisons between mouse and keyboard input, gamepad input, and ETI, in an FPS game. While the study uses very limited data, their conclusion is that gamepad input is on par with ETI, but both of them fall short of the performance of the keyboard and mouse input [14].

Jönsson compared TI and ETI in a Shoot 'em Up (SEU) game and an FPS. In the SEU with TI, aiming was controlled with the mouse, and with ETI aiming was controlled with gaze. The majority of players thought ETI was more accurate, more fun, more natural and faster than TI, while TI was considered more difficult. With ETI, players were able to achieve a higher score.

In the FPS three different setups were used. One where the mouse controlled the camera and the aiming (FPS Mouse, or TI), one where the players' gaze controlled the camera and aiming (FPS Eye, or ETI), and one where the camera was controlled with the mouse but aiming was controlled with the eyes (FPS Combined). TI seemed to offer better performance than ETI. Comparing the three conditions showed that Combined was overall favoured over TI but not over ETI [18].

With regards to control naturalness, Ejdemyr, Isokoski and Martin, and Jönsson all mention that participants possibly favoured TI because they have more experience with TI than ETI [11, 14, 18]. Smith and Graham mention that ETI was favoured more in the RPG game because it reduced the amount of effort needed to complete interactions in the game. While in Jönsson's thesis players were able to achieve a higher score in the SEU game with ETI, it was because the game was specifically chosen to allow for this interaction as the targets just appeared on-screen and they did not move, making it easier to shoot them. In contrast to this, in Smith and Graham's study, the Arcade game featured moving targets which resulted in lower average score with ETI than TI [18, 32].

We can see that when talking about player performance, TI and ETI offer advantages in different kinds of games. FPS games in particular seem to benefit more from TI than ETI, while slower-paced games where pinpoint accuracy is not important work better with ETI. Smith and Graham's results show that the RPG game, where the camera remained fixed relative to the avatar, got the most favourable results for ETI, which is promising since our game's camera is similarly fixed. Orlov and Apraksin's results are of particular importance to us, as our game design is similar to what they used in their study, with our game also requiring players to perform a visual searching task. According to their study, participants' recognition times were faster with ETI than in TI. Based on the findings, we expect player performance, especially dwell times on certain objects to be shorter with ETI than TI in our game.

2.5 Not looking as a game mechanic

Smith and Graham mentions that the most obvious eye-based interaction is pointing tasks, but eye-based input is also commonly used as an auxiliary input channel to the hands [32]. Contrary to this, games can also make use of not looking as a game mechanic. This is quite unlike previous work mentioned above: while ETI is still used in these games, the player is essentially punished for looking at the game objectives, and these games rely on the participants' peripheral vision. Some games were made or adapted by Vidal et al., Gomez and Gellersen, Lankes et al., Velloso et al., Ekman et al. and Newn et al. to utilise this mechanic [12, 20, 21, 24, 29, 37–39].

Gomez and Gellersen classified these games into five categories [28]. In the first category titled "Might not look", looking can create unintended outcomes, such as in The Royal Corgi, where characters can get upset at the player for looking at certain things during interaction [39]. In the second category called "Cannot look", players can only achieve the objective if they keep their eyes off of certain elements in the game, like in "Shynosaurs". Here players have to move creatures into an enclosure using the mouse while looking at predators to prevent them from taking those creatures [21, 38]. In these games looking is still permitted, but makes the game more challenging. The third category includes games that rely on players' awareness, meaning they "Should not look" because that will reveal their intentions to the other players. Examples include games where players' gaze location is visible to others on a shared screen. For example in a digitised version of Ticket To Ride, where players have to construct a railroad according to a road plan unique to each player. Gazing at the segments making up the road plan can indicate their intended route to other players, which may lead to the other players deliberately preventing them from constructing their desired path [20, 24]. The fourth category, "Must not look" includes games where the player is not permitted to look because it will lead to them losing the game. In SuperVision, looking at forbidden objects reduces the player's remaining attempts at the game [29, 37]. The final category, "Not looking" includes a game where some actions can only be completed if the game does not sense the players' pupil, like when they blink or keep their eyes closed [12].

Gomez and Gellersen comments that not looking can be implemented as an engaging and playful game mechanic [29]. This allows us to implement not looking as part of our game without it negatively affecting the overall game. This could be done by having a temporary drawback that prevents players from progressing for a short period of time, but not have any permanent consequences.

2.6 The "Midas Touch" problem

One common problem that we have to bear in mind when using ETI is what is referred to as the "Midas Touch" problem [11, 13, 15–17, 28, 32, 36]. This problem is defined as

> the trigger or selection of objects when we look at them without the intention to interact with them [28].

There are several ways of overcoming this problem: using buttons to initiate an action [15, 32], not using gaze for activation of game elements [11], remapping input area [16] or by implementing a dwell time, which means activation of object only occurs after looking at the object for a predefined amount of time [15, 32, 36]. However, this has the effect of slowing down interaction, which can be a problem depending on what type of game it is implemented in. Since the TI and ETI versions of our game will be identical outside of input methodology, the dwell time will be present in both versions. The game also does not require fast reaction times, and despite the total playtime being measured, no adverse effects occur if a player takes longer to complete the game. Thus, we do not expect it to be a problem in our case.

2.7 Narratives

In narratives there are two main categories: linear and non-linear – each of which are explained below. Linear narratives can be further divided into non-augmented and augmented linear narratives. For non-linear narratives there is a similar separation, between branching and emergent narratives. The difference between branching and emergent narratives may not only be found in different ways of interacting with the medium, instead we have chosen to look at the change being in ergodicity. Ergodicity here refers to the non-trivial way of interacting with a medium, so where linear narratives are straight forward, non-linear narratives require (to different degrees) the user to exert some effort to progress [22].

Linear narratives have a predetermined structure made by the author [30]. This means that independent of user interpretation all information regarding the narrative is provided in the same order each time the narrative is experienced. Though this is most often found in books, some video games also employ this structure – arcade games are an example of this. Different approaches have been made to increase the interactivity and ergodicity of narratives, where the former resulted in e.g. augmented narratives and the latter in non-linear narratives.

Augmented narratives pertains to the use of extra-textual information used to enhance the experience of the reader e.g. playing the sound of birds chirping when the setting is in a forest. To increase the interactivity of linear narratives, Kunze et al. measured eye tracking, blinking, and temperature to determine the engagement of the participant while reading non-engaging and engaging pieces of literature [19]. This could be used for augmenting the narrative if the level of user engagement reached a specific threshold. For a more varied look at how text-augmentation could be used Biedert et al. developed several prototypes to enhance the reading experience in different ways such as live word translation and footnote generation. Their prototype could also analyse the importance and frequency of words to increase focus on the important parts of the text through fading, allowing the reader to more easily skim through the text. [6]

Branching narratives allow the user to alter the course of the story. A common issue found in non-linear narratives is the inverse proportion between authorship and interactivity [30], commonly referred to as the *narrative paradox*. The paradox means that as the complexity of the branching narrative increases, the user comprehension lowers from the intended message by the designer-author. A different (but not necessarily self-consistent) view on this is by addressing it through the Author-Audience Distance (AAD) [7]. Hutchison et al. split the narrative into *narrative intelligibility* and *narrative closure*, where the former controls the AAD by moving the narrative along the abstract-didascalic spectrum as seen on Figure 1, the latter relates to the coherence and completeness of understanding the narrative regardless of the intended understanding by the designer-author. To successfully convey the intended narrative both aspects have to be taken into considerations though this does not mean either has to be included.

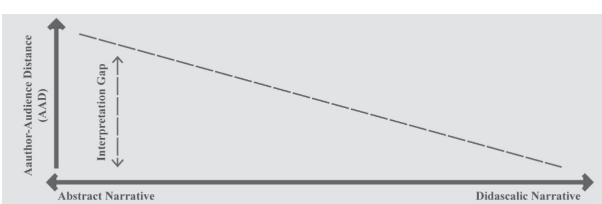

Figure 1: The y-axis shows the distance between intended designer-author message and user received message. The x-axis shows the abstract-didascalic spectrum of the narrative. [7]

As the abstraction (or complexity) increases the design reaches a point where the narrative intelligibility becomes so large that all the control of the narrative happens at run-time rather than being authorial [34]. When a narrative includes closure but disregards narrative intelligibility, the aim of the designer-author is often to allow the user to understand the narrative in the way the user sees fit. This is the case with emergent narratives where the narratives often disregard AAD.

As stated by Hutchinson et al. more focus should be placed on narrative closure rather than intelligibility to attain/retain user engagement [7]. As the focus of this study does not require a low AAD, emergent narratives lends themselves well as the narrative intelligibility can be omitted in the design. With the purpose of developing a prototype which would focus on interactions through gazing, exploration became a key concept. By developing an emergent narrative, the prototype could facilitate the exploration of the user in the setting, while still allowing the user to have some form of narrative closure.

3 PROTOTYPE DESCRIPTION

A common thread in studies reviewed above was comparing traditional input to eye input. These studies had the participants carry out a task that would be typical to a given game genre to see how effective eye input can be. To the best of our knowledge, no study exists that investigates the new modality's effect on performance, immersion, and engagement in an emergent narrative game or experience. Seeing this as an apparent gap in research, we decided to implement a system designed to test player performance, immersion, and engagement in relation to an emergent narrative experience. The games from the reviewed studies had a goal for the participants to achieve, so we added an achievable objective to our game as well.

Additionally, our game was designed primarily with gaze input in mind which is substituted with mouse input in the TI condition, while some studies did this the opposite way. Our experiments will show whether this has any bearing on the results compared to the results from the reviewed studies.

3.1 Design

To aid us in designing the game, we looked at the design of the games used in the previous studies. When designing game features, we considered whether they fulfill the following guiding principles:

- the game should inspire looking around in the game world
- the game should inspire exploration
- the player's actions should have a lasting effect on the world.

The first item was inspired by Orlov and Apraskin's study, where players had to find a specific unit among many others [26]. We based exploration on the game developed by Ejdemyr, where players had to navigate in the world trying to find the six keys that enabled them to complete the game [11]. Finally, the last item was inspired by The Royal Corgi by Vidal et. al, in which players' choices had permanent effects on the outcome [39].

In our game, players start on a map partially populated with objects. The objects can be interacted with using their eyes (ETI condition) or the mouse (TI condition). As the player traverses the map, additional objects procedurally spawn, further populating the game. A hidden goal is to find 4 golden objects. Though finding all the golden objects prompts a message stating that the game is "finished", the player can also end the session without achieving this.

3.2 Mechanics

The game world and the objects therein are completely desaturated to begin with, except for the dark and golden objects. Objects and terrain falling within the above mentioned cone of vision will begin to saturate gradually as shown in Figure 2. If the player fixates their gaze or hovers their mouse pointer over an object for an extended amount of time, those objects then become permanently saturated – also referred to as "activated" – unlike the rest of the objects, which gradually desaturate once they are out of the cone.

Activating an object in such a manner has another in-game effect as well. The game has several biomes to speak of. The individual biomes have objects specific to those biomes spawn, such as cacti in the desert, or a windmill in the farm biome. Activating an object enables the player to change what surroundings will appear around them as they navigate the game world, e.g. activating a cactus will cause more desert biome to spawn.

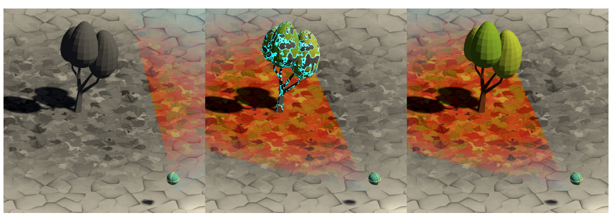

Figure 2: The transitional effect on the saturation of a tree object in the forest biome, as it leaves/enters the cone of vision.

In order to create a goal that gives the player additional incentive to keep playing, "golden" objects will appear that the players may find throughout a session. An example of these can be seen on *B)* in Figure 3. These objects appear only if the biomes they represent are being spawned. Upon finding a golden object, text showing the amount found and total amount appears above the object. Other than that, there will be no UI elements visible to keep players from being distracted.

To provide some challenge that has negative consequences, "dark" objects appear randomly in all biomes. An example of these can be seen on *C)* in Figure 3. Activating such an object will temporarily disable the players' cone of vision as well as object spawning. As seen in subsection 2.5, it is possible to include "not looking", or more specifically the "Should not look" category, as an engaging and playful game mechanic. Though it is implemented in the game, we are not investigating the effects not looking may have on player immersion and engagement.

The game uses orthographic projection. The players can navigate in the game environment by controlling an avatar using the WASD keys on their keyboard. They also control a cone of vision using their mouse in the TI condition, or by using their gaze via a Tobii 4C eye tracker in the ETI condition. Other than the input method, the two versions are identical.

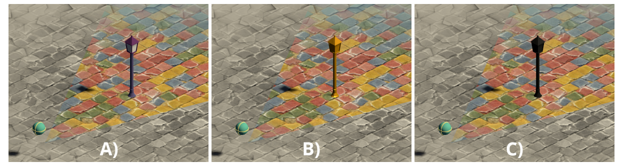

Figure 3: A) A normal lamppost in the village biome. B) A golden lamppost in the village biome. C) A dark lamppost in the village biome.

3.3 Hardware

In the studies we have reviewed, Tobii eye trackers were the most often used devices [10, 11, 14, 16, 17, 25, 29, 32, 39]. The specific model used in the study varies from paper to paper. Other studies used SensoMotoric Instruments eye trackers [15, 23, 26, 31]. Based on this, we chose to use a Tobii eye tracker for our study. The specific type is Tobii 4C, which is also the eye tracker Tobii manufactures for gaming.

We had users use their own computers and laptops while we provided them with the Tobii 4C eye tracker. This means we did not control for the hardware used for each test. Studies used anything between 15" to 27" displays with various resolutions. For TI input, standard USB keyboards and mice have been used, and for ETI input various versions of Tobii eye trackers were the most commonly used, thus we do not require participants to use any other hardware for display and input than what they already have.

3.4 Implementation

We set out to create a game with an emergent narrative, implemented in Unity using C# for scripting. Since the game is procedurally generated, instead of hand-building the game environment, we went with creating tools to easily modify and author the game, which consisted of two major parts. The first one is an asset database, where we can easily import and modify the prefabs into the game, without having to go through the individual objects. Second, we have implemented a custom event system where we can use scriptable objects to raise events across the scene without needing hard references in game objects, allowing for more encapsulation and easier debugging during development. The eye tracking portion of the game was implemented using the Tobii API, which provided out of the box classes and methods for gaze tracking and object focus checking.

4 METHOD

To test the above described prototype, we created two versions of it: one using ETI and one using TI. The two versions of the game were identical aside from the input modality. We conducted tests to determine the effects of ETI on player performance, immersion, and engagement in an emergent narrative game.

4.1 Measurements

Immersion: We measured participants' immersion using the Immersion Questionnaire by Jennet et al. The questionnaire consists of 31 questions, and answers were given on a five-point scale where (1) is "Not at all" and (5) is "A lot" [3].

Engagement was measured using User Engagement Scale Short Form (UES-SF). This questionnaire consists of 16 questions that are rated on a five-point scale where (1) is "Strongly Disagree", (2) is "Disagree", (3) is "Neither Agree nor Disagree", (4) is "Agree" and (5) is "Strongly Agree" [27] These questionnaires are filled out after the session has concluded.

Player performance was measured by logging the time taken and distance traveled to complete the task of finding all the golden objects in the game. Furthermore dwell times were also logged so we could compare recognition times for dark objects between TI and ETI conditions.

Player performance is used loosely in this study, due to the hidden objective. This means that the definition in 2.4 does not apply completely to our study. Dwell times are the strongest measurement for this in our case, as that was affected most by the input modalities.

4.2 Participants

A total of 20 participants partook in our study. We recruited the 10 participants for the TI condition through the university's social media. For most of the ETI condition, we distributed our eye tracker to friends and asked them to pass it along. One person had their own eye tracker, whom we contacted through the university. 15 males and 5 females participated, with ages ranging from 23 to 31 (mean age 25.85). All participants had experience playing video games, but only 6 out of 10 had used ETI before.

4.3 Experimental Design

For this study, we used a between-subjects design. Participants were assigned to one of two groups of 10 (ETI m = 5 f = 5, TI m = 10). One group tested with TI and the other tested with ETI. The independent variable was the input modality and the dependent variables were player immersion, player engagement, and player performance.

4.4 Procedure

Built versions of the game were distributed to the participants online. Those doing the ETI condition received the Tobii eye tracker as well. A set of instructions were also provided that they had to follow to properly set up the environment. The document also gave the participants instructions to explore the world in the game. They were each given an ID that they had to input in the main menu before starting the game. Those doing the TI condition just had to launch the game, input the ID and play until completion. Those doing the ETI version had to plug in the Tobii eye tracker provided to them, then install the necessary software and calibrate it. Following successful calibration, the participants input their ID and proceeded to play the game. In both conditions, upon completion, the game generated a .json file with the necessary data, such as ID, seed value, dwell times, time elapsed, etc., and a screenshot of their game map. Following this, they were asked to fill out the two questionnaires for immersion and engagement. After completing the questionnaires, participants were asked to send the .json file and the screenshot to us for analysis.

5 RESULTS

Our study's results can be categorised into three larger groups with the last group consisting of three different measurements:

- **Immersion** was measured using the Immersion Questionnaire. Scores from the Likert scale (1-5) given to the 31 questions were summed up. Questions 6, 8-10, 18 and 20 were scored inversely, so a response of "4" was logged as a "2".
- **Engagement** was measured using Engagement Scale Short Form. Scores (1-5) were summed up and divided by 12. Questions 4-6 were scored inversely.
- **Player performance**
  - **Dwell time on Dark objects** was measured by taking the average of dwell times on all 8 dark objects.
  - **Distance traveled in game** was measured in meters.
  - **Time spent playing** was measured in seconds but was converted to minutes.

After testing for normal distribution and homogeneity of variance, an independent t-test was performed on both Immersion and Engagement to see if differences were significant.

Users reported slightly higher immersion scores with ETI (m = 85.3, SD = 16.50) than TI (m = 83.4, SD = 20.13). The independent t-test determined that the difference was not significant ($t(17.3) = 0.23$, $p = 0.82$).

Slightly higher engagement scores were reported with TI (m = 2.76, SD = 0.50) than with ETI (m = 2.9, SD = 0.52) The independent t-test determined that the difference was not significant ($t(17.9) = 0.34$, $p = 0.57$).

For player performance measures, a Mann-Whitney test was performed. Players' dwell time on dark objects was lower in ETI (m = 9.81, SD = 6.13) than in TI (m = 11.69, SD = 6.85), however the test determined that the differences were not significant ($W = 41$, $p = 0.52$).

On average, players spent more time playing with TI (m = 31, SD = 26,66) than in ETI (m = 16, SD = 12,66), however the difference was not significant ($W = 29$, $p = 0.13$).

In the TI condition, average travel distance was higher (m = 7658, SD = 6377) than in the ETI condition (m = 2833, SD = 2439), and testing revealed that the difference was significant $W = 20$, $p = 0.02$.

Even though we recorded a high value with one participant for distance traveled (21080 meters), calculations showed that this was not an outlier. This distance correlated to a high play time as well (89 minutes). We asked the participant about why they played for so long, and they responded that they tried to figure out how the mechanics of the game worked, but based on data, we assumed that they had softlocked the game and could only spawn one biome. Despite this, they still continued to play the game for an hour and a half. From another participant we recorded an immersion score of 127, but analysis showed it was not an outlier.

There was a strong correlation between immersion and engagement scores, but these scores did not show correlation to time played or distance traveled.

6 DISCUSSION

Our results show no significant differences between immersion and engagement scores for the two conditions. This could be due to a number of different factors. When talking to some of our test subjects after testing had concluded, it seemed that they were unsure as to what the goal of the game was or what they had to achieve. The golden objects were never disclosed as a part of the game, and were only discovered as a mechanic during testing. While the golden objects indicated how many were found out of the total four, no test subject managed to find all four of them. Perhaps if finding the golden objects was made an explicit goal to the players before testing commenced, they would have spent more time trying to find them. Likewise, understanding how the golden object spawning worked, might have prevented participants searching for over 30 minutes without finding them all. This issue can be related to the players not knowing exactly how object spawning worked in general. We did not include instructions regarding the golden objects in the technical documentation because we felt that not knowing this mechanic might encourage further exploration by the players. Likewise, how the dark objects worked was not explained for the same reason.

Results indicate that players were more immersed in the ETI condition than in the TI condition, despite play time being shorter in ETI than in TI on average. While engagement scores were higher in TI, this could be the result of the simplicity of the TI setup: players could download and play the game at their leisure, but with ETI, they had to set up and calibrate the Tobii eye tracker themselves and recalibrate it if multiple participants tested from the same computer. Calibration of the eye tracker also took time and if participants considered it a part of the test, that may have impacted time spent playing the game. With ETI, time outside the setup was also a factor as they had to pass the sensor along to other participants after they were done the testing. Though a possible solution could have been to ask the participants to play a specific amount of time (e.g. 10 minutes), this could have affected the results as there could have been an imposed stress of "completing" the test before the time ran out. On the other hand, having an unknown time limit could have ended the session too abruptly, negating the narrative closure, and thereby also negatively influencing the engagement of the participant.

Regarding time spent playing the game and distance travelled, we see a correlation between the two measures. This shows that participants kept moving about the game world and they were not spending much time being idle, but we can also see from the individual results that the rate of movement was not the same for all of the participants. Additionally, since nobody collected all four golden objects, time and distance cannot be used for reliably measuring player performance.

Some participants had exceedingly high gameplay times. This could be explained by the fact that players were not introduced to the game's mechanics. This could have been prevented by preliminary testing, but we had to prioritise the final test because of the availability of possible participants.

Results also show that there was a 20% higher average dwell time on dark objects in the TI condition. This lines up with our expectation based on our research, but the difference was not statistically significant. We might have gotten a different result if the dark objects spawned during gameplay. A learning effect might have taken place where the participants remembered where the dark objects were and started avoiding them entirely. By having dark objects spawn during the session a bigger difference might have occured between dwell-times in the two conditions.

A factor that was not taken into consideration for the results was the narrative aspects. Narrative closure is an important part when considering engagement, but there was no assessment to whether this was achieved. Since the intelligibility is removed from the narrative, it has to be achieved at the system level instead, and as this might have been an issue, with participants not knowing how the game worked, it could have had a negative effect on the results.

6.1 Limitation

Our testing equipment was limited, as we only had access to one Tobii eye tracker that the participants were asked to pass along to the next person at their home address. Distances between participants varied, but this coupled with the setup required for the use of the eye tracker may have impacted the amount of time spent testing the ETI version.

Due to the uncontrolled nature of the testing, we are unsure how participants approached the testing procedure and what their environment was like. We think that in a more controlled environment with better guidance to the participants might lead to different results, which mean that this causes the results to have lower reliability.

Even though we reached out to some participants after looking at their results to probe them further about their experience, this may not have been as effective than if we could ask follow-up questions right after they concluded their sessions while the experience was fresh. Since the test was administered by the participants at a time convenient to them, some of them carried it out at late hours and we could only look at their data the following day.

Though the gender distribution was 50/50 in the ETI condition, the TI condition only consisted of male participants. We do not believe this has had an effect on the final results.

7 CONCLUSION

We conducted a study on how eye tracker input affects immersion, engagement and player performance in a game with an emergent narrative. Eye tracking has been studied for its effectiveness compared to traditional input, and we took the previous findings into account when making design decisions.

We created a game with gaze interaction in mind and tested it with eye tracker input and traditional input, where subjects played only one version of the game. The game had an orthographic view and featured a visual scanning task because research showed that these design elements compliment ETI the most.

Our results show no significant difference in our measurement outside of distance travelled in-game, but it may be due to the nature of conducting the test remotely. Further testing in a controlled environment may lead to different, but possibly more valid data.

8 COVID-19 DISCLAIMER

At the time of conducting our study, lockdowns and social distancing rules were in effect due to the COVID-19 pandemic. This meant that we could not conduct the test personally, and we had to distribute the game to participants who self-administered the test based on the guidelines we supported them with. This also means that we were limited in what we could measure as part of our experiments. Furthermore, we were unable to conduct the tests on the same hardware setup, so participants had to use their own hardware. Had we been able to do this in person, there would have been opportunity of further measurements.

Cairns et al. describe their approach to objectively measure immersion by requiring participants to complete a control task -which was chosen to be a tangram puzzle- before and after playing a game for a set period of time, which is interrupted once to fill out a questionnaire [8]. We felt that doing this remotely would be difficult to control and facilitate, as participants would have had to set up a timer and measure control task completion times, and without knowing the correct solution, they might not have completed it successfully. Furthermore the participants having had to consult an in-depth guide on how to carry out this multi-step experiment may have affected the result to an unknown degree.

As the testing was self-administered several limitations arose in terms of data gathering. For example, if we used a setup dedicated to testing, we could have recorded the entire gameplay as well as body language for analysis. We did not pursue this as additional technical setup could cause further inconsistencies in the final test setup. A more technical setup would also allow for additional eye-tracking data, such as temporal data showing the participants' gaze over time, which could then be compared between conditions. For engagement, we could have employed sensors to measure biometric data, such as skin conductance or heart rate, but since these require skin contact and further setup by the user, we abandoned these measurements.

Interviewing the participants would also have given further insight, but since they did the test at their convenience, sometimes in late hours, this was not always possible. Furthermore, due to personal scheduling, interviews would have taken place almost an entire day after testing was concluded. This might have given different results than if the interviews were conducted immediately following testing or in person. It was also hard to find participants willing to do a longer interview after having completed a long play session followed by a lengthy survey. This is especially true in cases where participants had to set up the eye tracker as well, as that added about 10 minutes to the testing session.

ACKNOWLEDGMENTS

REFERENCES

[1] [n. d.]. Global Video Games Thematic Research Report 2019-2025. https://advance-lexis-com.zorac.aub.aau.dk/document/?pdmfid=1516831&crid=b6411a0c-cd4d-4e69-9d7a-230dca4d299d&pddocfullpath=%2Fshared%2Fdocument%2Fnews%2Furn%3AcontentItem%3A5XD1-HB41-DY2R-450T-00000-00&pdcontentcomponentid=222091&pdteaserkey=sr0&pditab=allpods&ecomp=kb63k&earg=sr0&prid=049ea820-95e7-4fc1-9cbc-415b05ea4629

[2] [n. d.]. Tobii Gaming | PC Games with Eye Tracking gameplay. https://gaming.tobii.com/games/ Library Catalog: gaming.tobii.com.

[3] 2008. Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies* 66, 9 (Sept. 2008), 641–661. https://doi.org/10.1016/j.ijhcs.2008.04.004 Publisher: Academic Press.

[4] 2017. Hint Function in Room Escape Games by using Eye-tracking. https://doi.org/10.15002/00021512

[5] Reynold Bailey, Ann McNamara, Nisha Sudarsanam, and Cindy Grimm. 2009. Subtle Gaze Direction. *ACM Trans. Graph.* 28 (2009). https://doi.org/10.1145/1559755.1559757

[6] Ralf Biedert, Georg Buscher, Sven Schwarz, Jörn Hees, and Andreas Dengel. 2010. Text 2.0. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems - CHI EA '10*. ACM Press, Atlanta, Georgia, USA, 4003. https://doi.org/10.1145/1753846.1754093

[7] Luis Emilio Bruni and Sarune Baceviciute. 2013. Narrative Intelligibility and Closure in Interactive Systems. In *Interactive Storytelling*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Hartmut Koenitz, Tonguc Ibrahim Sezen, Gabriele Ferri, Mads Haahr, Digdem Sezen, and Güven Catak (Eds.). Vol. 8230. Springer International Publishing, Cham, 13–24. https://doi.org/10.1007/978-3-319-02756-2_2 Series Title: Lecture Notes in Computer Science.

[8] Paul Cairns, Anna Cox, Nadia Berthouze, Samira Dhoparee, and Charlene Jennett. [n. d.]. Quantifying the experience of immersion in games. ([n. d.]), 7.

[9] Yuping Chen and Meng-Jung Tsai. 2015. Eye-hand coordination strategies during active video game playing: An eye-tracking study. *Computers in Human Behavior* 51 (Oct. 2015), 8–14. https://doi.org/10.1016/j.chb.2015.04.045

[10] Soussan Djamasbi, Siavash Mortazavi, and Mina Shojaeizadeh. 2015. Baby Boomers and Gaze Enabled Gaming. In *Human Aspects of IT for the Aged Population. Design for Everyday Life (Lecture Notes in Computer Science)*, Jia Zhou and Gavriel Salvendy (Eds.). Springer International Publishing, Cham, 479–487. https://doi.org/10.1007/978-3-319-20913-5_44

[11] Niclas Ejdemyr. 2016. *Eye Tracking as an Additional Input Method in Video Games : Using Player Gaze to Improve Player Immersion and Performance.* http://urn.kb.se/resolve?urn=urn:nbn:se:bth-11824

[12] Inger M. Ekman, Antti W. Poikola, and Meeri K. Mäkäräinen. 2008. Invisible eni: using gaze and pupil size to control a game. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems (CHI EA '08)*. Association for Computing Machinery, Florence, Italy, 3135–3140. https://doi.org/10.1145/1358628.1358820

[13] Poika Isokoski, Markus Joos, Oleg Špakov, and Benoît Martin. 2009. Gaze controlled games. *Universal Access in the Information Society* 8 (2009), 323–337. https://doi.org/10.1007/s10209-009-0146-3

[14] Poika Isokoski and Benoît Martin. 2006. Eye Tracker Input in First Person Shooter Games.

[15] Howell Istance, Richard Bates, Aulikki Hyrskykari, and Stephen Vickers. 2008. Snap clutch, a moded approach to solving the Midas touch problem. In *Proceedings of the 2008 symposium on Eye tracking research & applications (ETRA '08)*. Association for Computing Machinery, Savannah, Georgia, 221–228. https://doi.org/10.1145/1344471.1344523

[16] Howell Istance, Aulikki Hyrskykari, Lauri Immonen, Santtu Mansikkamaa, and Stephen Vickers. 2010. Designing gaze gestures for gaming: an investigation of performance. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications (ETRA '10)*. Association for Computing Machinery, Austin, Texas, 323–330. https://doi.org/10.1145/1743666.1743740

[17] Howell Istance, Aulikki Hyrskykari, Stephen Vickers, and Thiago Chaves. 2009. For Your Eyes Only: Controlling 3D Online Games by Eye-Gaze, Vol. 5726. 314–327. https://doi.org/10.1007/978-3-642-03655-2_36

[18] Erika Jönsson and Erika Jönsson. 2005. *If looks could kill - An evaluation of eye tracking in computer games.* Technical Report.

[19] Kai Kunze, Susana Sanchez, Tilman Dingler, Olivier Augereau, Koichi Kise, Masahiko Inami, and Terada Tsutomu. 2015. The augmented narrative: toward estimating reader engagement *(AH '15)*, Vol. 11. ACM, 163–164. https://doi.org/10.1145/2735711.2735814

[20] Michael Lankes, Bernhard Maurer, and Barbara Stiglbauer. 2016. An Eye for an Eye: Gaze Input in Competitive Online Games and its Effects on Social Presence. In *Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology (ACE '16)*. Association for Computing Machinery, Osaka, Japan, 1–9. https://doi.org/10.1145/3001773.3001774

[21] Michael Lankes, Thomas Mirlacher, Stefan Wagner, and Wolfgang Hochleitner. 2014. Whom are you looking for? the effects of different player representation relations on the presence in gaze-based games. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play (CHI PLAY '14)*. Association for Computing Machinery, Toronto, Ontario, Canada, 171–179. https://doi.org/10.1145/2658537.2658698

[22] Judy Malloy and Espen J. Aarseth. 1998. Cybertext, Perspectives on Ergodic Literature. *Leonardo Music Journal* 8 (1998), 77. https://doi.org/10.2307/1513408

[23] Ann McNamara, Thomas Booth, and Srinivas Sridharan. 2012. Directing gaze in narrative art. In *Proceedings of the ACM Symposium on Applied Perception - SAP '12*. ACM Press, Los Angeles, California, 63. https://doi.org/10.1145/2338676.2338689

[24] Joshua Newn, Fraser Allison, Eduardo Velloso, and Frank Vetere. 2018. Looks Can Be Deceiving: Using Gaze Visualisation to Predict and Mislead Opponents in Strategic Gameplay. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, Montreal QC, Canada, 1–12. https://doi.org/10.1145/3173574.3173835

[25] Mica Olsson. 2017. *Visual composition in video games : Visual analyzation using eye-tracking.* http://urn.kb.se/resolve?urn=urn:nbn:se:ltu:diva-64489

[26] Paul A. Orlov and Nikolay Apraksin. 2015. The Effectiveness of Gaze-Contingent Control in Computer Games. *Perception* 44, 8-9 (2015), 1136–1145. https://doi.org/10.1177/0301006615594910

[27] Heather L. O'Brien, Paul Cairns, and Mark Hall. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies* 112 (April 2018), 28–39. https://doi.org/10.1016/j.ijhcs.2018.01.004

[28] Argenis Ramirez Gomez. 2019. Exploring Gaze Interaction Design in Games: Playing with Vision. In *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts (CHI PLAY '19 Extended Abstracts)*. Association for Computing Machinery, Barcelona, Spain, 55–61. https://doi.org/10.1145/3341215.3356333

[29] Argenis Ramirez Gomez and Hans Gellersen. 2019. SuperVision: Playing with Gaze Aversion and Peripheral Vision. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, Glasgow, Scotland Uk, 1–12. https://doi.org/10.1145/3290605.3300703

[30] Henrik Schoenau-Fog, Luis Emilio Bruni, Faysal Fuad Khalil, and Jawid Faizi. 2013. Authoring for Engagement in Plot-Based Interactive Dramatic Experiences for Learning. In *Transactions on Edutainment X*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Zhigeng Pan, Adrian David Cheok, Wolfgang Müller, Ido Iurgel, Paolo Petta, and Bodo Urban (Eds.). Vol. 7775. Springer Berlin Heidelberg, Berlin, Heidelberg, 1–19. https://doi.org/10.1007/978-3-642-37919-2_1 Series Title: Lecture Notes in Computer Science.

[31] Baris Serim and Giulio Jacucci. 2016. Pointing while Looking Elsewhere: Designing for Varying Degrees of Visual Guidance during Manual Input. 5789–5800. https://doi.org/10.1145/2858036.2858480

[32] J. Smith and T.C. Graham. 2006. Use of eye movements for video game control. 20. https://doi.org/10.1145/1178823.1178847

[33] Jose L. Soler-Dominguez, Jorge D. Camba, Manuel Contero, and Mariano Alcañiz. 2017. A Proposal for the Selection of Eye-Tracking Metrics for the Implementation of Adaptive Gameplay in Virtual Reality Based Games. In *Virtual, Augmented and Mixed Reality (Lecture Notes in Computer Science)*, Stephanie Lackey and Jessie Chen (Eds.). Springer International Publishing, Cham, 369–380. https://doi.org/10.1007/978-3-319-57987-0_30

[34] Neil Suttie, Sandy Louchart, Ruth Aylett, and Theodore Lim. 2013. Theoretical Considerations towards Authoring Emergent Narrative. In *Interactive Storytelling*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Hartmut Koenitz, Tonguc Ibrahim Sezen, Gabriele Ferri, Mads Haahr, Digdem Sezen, and Güven Catak (Eds.). Vol. 8230. Springer International Publishing, Cham, 205–216. https://doi.org/10.1007/978-3-319-02756-2_25

[35] Ron Tamborini and Nicholas Bowman. 2010. Presence in video games. *Immersed in Media: Telepresence in Everyday Life* (2010), 87–109.

[36] Eduardo Velloso and Marcus Carter. 2016. The Emergence of EyePlay: A Survey of Eye Interaction in Games. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '16)*. Association for Computing Machinery, Austin, Texas, USA, 171–185. https://doi.org/10.1145/2967934.2968084

[37] Eduardo Velloso, Carl Oechsner, Katharina Sachmann, Markus Wirth, and Hans Gellersen. 2015. Arcade+: A Platform for Public Deployment and Evaluation of Multi-Modal Games. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '15)*. Association for Computing Machinery, London, United Kingdom, 271–275. https://doi.org/10.1145/2793107.2793145

[38] Melodie Vidal. 2014. Shynosaurs: a game of attention dilemma. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play (CHI PLAY '14)*. Association for Computing Machinery, Toronto, Ontario, Canada, 391–394. https://doi.org/10.1145/2658537.2662979

[39] Melodie Vidal, Remi Bismuth, Andreas Bulling, and Hans Gellersen. 2015. The Royal Corgi: Exploring Social Gaze Interaction for Immersive Gameplay. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, Seoul, Republic of Korea, 115–124. https://doi.org/10.1145/2702123.2702163