Eye-tracking for Sense of Immersion and Linguistic Complexity in the Skyrim Game: Issues and Perspectives

Alessandro Cierro
Université catholique de Louvain
Louvain-la-Neuve, Belgium
alessandro.cierro@uclouvain.be

Thibault Philippette
Université catholique de Louvain
Louvain-la-Neuve, Belgium
thibault.philippette@uclouvain.be

Thomas François
Université catholique de Louvain
Louvain-la-Neuve, Belgium
thomas.francois@uclouvain.be

Patrick Watrin
Université catholique de Louvain
Louvain-la-Neuve, Belgium

Sébastien Nahon
Université catholique de Louvain
Louvain-la-Neuve, Belgium

ABSTRACT

As part of an experimental study aimed at evaluating the linguistic and paralinguistic factors that can influence the sense of immersion in an open-world video game, we have partially opted for an eye-tracking data collection protocol. In doing so, various problems emerged in the course of the research and we therefore propose to report and analyze them in this article in order to provide useful feedback for further research. The first set of problems is of a technical nature and relates to the difficulty of collecting reliable eye tracking data in an open and complex game environment. Our second concern is about the difficulties that may appear depending on the morphological characteristics of the players. The third issue is about player's familiarity with the game and the experimental parameters. And lastly, we discuss some post-processing issues for the analysis. The reflections raised from these few difficulties allow us to discuss some challenges for future oculometric research in complex video game environments.

CCS CONCEPTS

• **Human-centered computing → User studies**.

KEYWORDS

open-world video games, immersion, user experience, language complexity

ACM Reference Format:

Alessandro Cierro, Thibault Philippette, Thomas François, Patrick Watrin, and Sébastien Nahon. 2020. Eye-tracking for Sense of Immersion and Linguistic Complexity in the Skyrim Game: Issues and Perspectives. In *Symposium on Eye Tracking Research and Applications (ETRA '20 Short Papers), June 2–5, 2020, Stuttgart, Germany.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3379156.3391836

ETRA '20 Short Papers, June 2–5, 2020, Stuttgart, Germany

ACM ISBN 978-1-4503-7134-6/20/06…$15.00
https://doi.org/10.1145/3379156.3391836

1 INTRODUCTION

The current contribution is based on the initial data and analyses conducted as part of a research program on the evaluation of the sense of immersion and presence in complex videogame and virtual reality environments. In this program, we first wanted to explore the possibility of using eye-tracking data to identify indicators to investigate this sense of immersion within a complex gaming environment, namely *The Elder Scrolls V: Skyrim* (Bethesda Game Studio, 2011). Immersion has been hardly investigated in such environments and previous studies using eye-tracking have mostly studied gaming aspects. The originality of our research project lies in the fact that we also want to take into account the linguistic aspects affecting this sense of presence and immersion, and their interaction with gaming characteristics. Such research objectives have led us to investigate the frontiers of what is currently technically possible with eye-tracking systems. We have been confronted with several technical constraints of eye-tracking tools and software in the context of open-world games research. This paper reports the lesson learned as well as an experiment in which we correlated various subjects characteristics with the confidence of the eye-tracker recording system. Based on these findings, we provide perspectives on material development and future analyses.

2 BACKGROUND

Research on the use of eye-tracking tools in the context of video gaming practices has focused on eye-tracking as a substitute for game controllers, or as a tool for collecting and analyzing eye-behaviour [Almeida et al. 2012; Ejdemyr 2016]. As part of an analytical approach using eye-tracking tools, working hypothesis generally concern three main aspects: interactions, attention and/or emotional states [Iqbal 2015].

First of all concerning attention, the data collected by eye-trackers relate to latency time, duration and number of fixations, or more or less recurring gaze points [Chen and Tsai 2015; Ejdemyr 2016]. Most of the time, the software associated with these tools allows graphic rendering (e.g. eye paths, heat maps) in relation to areas of interest (AOIs) determined by research objectives [Chen and Tsai 2015; Iqbal 2015; Mangiron 2016]. These AOIs are extremely dependent on the nature of the game, which implies a strong orientation of the research towards either "Tetris-like" games with a fixed background and moving objects [Chen and Tsai 2015]; or basic games developed by the researchers or colleagues [Ejdemyr

2016; Mangiron 2016]; or in the case of more advanced commercial games, First Person Shooters (FPS) type games where the useful areas are relatively consistent — i.e. the weapon's tip or some control windows — despite the changes of scenery and actions [Almeida et al. 2012; Iqbal 2015]. A notable exception is a study of an adventure game by Renshaw et al. [Renshaw et al. 2009], but it only brings the potentially complex analysis back to simple emotional states (frustration or boredom).

This leads us to the second perspective, as regards emotional analysis. Using analytical tools, this perspective attempts to infer, from visual and attentional behaviours, certain types of emotion. The register of emotions offered by complementary tools —e.g. face readers [Iqbal 2015] — is often limited (6-7 main types) in addition to inference from low-variable facial and eye data inevitably leads to a lack of nuance and a "black box" feeling for non-experts in physiology and biometry. As Almeida et al. [Almeida et al. 2011] point out, inferring visual attention from fixation points is already reductive in view of the microsaccades that the eyes are capable of performing, which corneal reflection eye-trackers [Ejdemyr 2016] do not capture. Besides, if we are interested in complex psychological states such as immersion or presence, it is necessary to take into account multiple factors aside eye-behaviour data to go beyond the parameters of selective attention or engagement [Witmer and Singer 1998].

These considerations lead us to the third domain of application of eye-tracking in the context of data analysis, which is the most promising for us, namely the evaluation of interactions. If several innovative experiments try to evaluate the potential of eye-tracking in the interaction with(in) the game, especially in terms of control [Chen and Tsai 2015] or collaboration [Špakov et al. 2019], the use of eye-tracking tools, in addition to other methods such as "think aloud protocol", interviews or questionnaires [Iqbal 2015; Mangiron 2016] seems very promising to assess differences between players in the ways they interact and experience the game. Indeed, the use of eye-trackers allows to capture certain indications on visual factors facilitating or hindering the "natural" interaction, while acknowledging that the eye-tracking device itself being one of those factors.

As part of our research, like Mangiron [Mangiron 2016], we are particularly interested in the role of textual components in the proper progression and understanding of a game but in a fairly rich and more complex game environment. The role of the language comprehension processes in access to meaning has indeed been long demonstrated in various contexts (reading, learning of a second language, negotiation, etc.), but hardly investigated in game studies.

3 DESIGN OF THE EXPERIMENT

In our research program, we wish to evaluate the influence of several variables on immersion: being a player, being a native speaker and various playing settings (in particular VR/non-VR). This last criterion will be taken into account in a comparative manner in future research. However, it has influenced the choice of eye-tracking tool (see below).

The experiment on *The Elder Scrolls V: Skyrim* was conducted with 50 participants in a computer room of our university. These individuals were selected, using snowball sampling with six very different entry points, on the basis of various criteria relevant to our working hypotheses on the influence of language understanding and experience on immersion and presence in game. Thus, the first selection criterion was the fact of being a native French speaker or a learner of French as a foreign language (with different degrees of proficiency in French). The second criterion, crossed with the first one, was the fact of declaring oneself a non-player, a player of other video games or a Skyrim player. In addition, in order to correspond to the reality of the practices, we ensured a good gender balance among the participants (30 men, 19 women, 1 other). The experience took place for the participants facing a PC version of the Skyrim game, in French version (subtitles and audio activated), and a keyboard-mouse type configuration. In order to record the ocular pattern of the participants, we opted for a wearable eye-tracking solution: *Pupil Labs Core* 120Hz binocular glasses [Kassner et al. 2014].The advantages and limitations of this technical choice will be discussed later. In addition, an audio headset was placed over the participant's ears. As for the researcher, he was placed behind a master computer allowing him to observe the recording of the gaming session, and to possibly intervene in the event of a problem (calibration, avatar stuck in a zone, visual discomfort of the player, etc.).

The research protocol can be summarized as follows: 1) The participant is welcomed by the researcher who explains the protocol and the research objectives; 2) The participant answers a first questionnaire allowing the collection of socio-demographic information; 3) The researcher starts the calibration of the system and, as soon as a sufficient degree of confidence is reached, he launches the Skyrim game; 4) The player must then follow the whole "tutorial" part of the game, which lasts approximately 45 minutes to one hour (until his/her avatar leaves the cave and separates from the NPC[1]); 5) Once the eye-tracking and headset equipments are removed, the participant completes a second questionnaire about his/her feeling of immersion and presence in the game, information about his/her morphological characteristics, and items designed to measure understanding of certain elements of the narration; and 6) The participant signs a consent form for the use of the data and receives a voucher as a thank you. Three participants agreed to repeat the experiment a second time so that we can assess the consistency of the data collected by the eye-tracking system.

4 EVALUATING EYE-GAZE ISSUES AND SOLUTIONS

The above experiment was designed to investigate the relation between immersion and various game characteristics (language, interactions, etc.). However, in this paper, based on our field experience, we specifically discuss the opportunities and shortcomings of using eye-tracking systems for game studies for complex game environments, focusing on four issues influencing the reliability and usability of the data collected: (1) technological issues related to eye-tracking systems; (2) the impact of morphological features of the players; (3) the impact of game habituation; and (4) post-processing software issues.

[1] Non-Player Character.

4.1 Issue 1- Technological issue: wearable eye-tracking system

The first type of problems encountered with regard to the initial objectives of this research concern the technical choices that were made and which hindered the smooth running of the data collection protocol. As a matter of fact, we opted for a wearable eye-tracking system that offers the advantage of great versatility, namely the *Pupil Labs Core* binocular [Kassner et al. 2014]. A priori, it is not to be preferred compared to a screen-mounted solution [Lolatto et al. 2020] as long as you are only interested in what is happening on the screen. However, this solution offered advantages: it can be reused in other research settings and it allows to capture disruptive aspects of the screen experience, such as player looking at one's hands. In addition, *Pupil Labs* offers open-source software where most of the others are proprietary, the accuracy of their core system has been stressed by previous research [MacInnes et al. 2018] and they have virtual reality eye-tracking solutions for the future research. This choice has nevertheless led to some complications in our study. The possibility of moving the head created the need to regularly recalibrate the Area Of Interest (sometimes every 7 minutes), especially for the less experienced players who showed a much more erratic focus of their attention and who regularly left the screen to look at their hands. This need for recalibration was amplified by working on a game with many dynamic scenes [Renshaw et al. 2009] and the need to identify reading behaviours [Mangiron 2016] in various screen areas including subtitles. In order to compensate for this first limitation, we opted for a large number of participants compared to most research, which makes it easier for us to accept a certain loss rate in the data but has created an additional workload.

For the future of the research, it would be useful to reproduce the protocol with a screen-mounted system to evaluate the influence of eye-tracking settings on the immersive experience. The question will then be whether a more rigid posture required by a screen-mounted system is less or more constraining than a head-mounted system that the player would not be able to completely disregard.

4.2 Issue 2- Morphological issue: vision troubles and eyeballs size

A second problem concerns the morphological characteristics of persons, which also has an influence on the quality of the recorded data. Indeed, having been able to test our protocol on a large number of participants, we identify several aspects that could pose a problem. The first concerns visual disorders and the wearing of glasses. When possible, subjects with eye disorders used lens. For those wearing glasses, problems of comfort and reflections quickly arose, and they generally took the test without glasses. As a results, some specific vision disorders (notably astigmatism and hyperopia) seem to raise problems of reliability, especially when the gaze was at the limit of the defined AOI. Due to the fact that the eye-tracking solution was wearable, some asymmetries of the face or eyes required a longer adjustment time and despite this, two recorded files had to be discarded due to a lack of consistency (remains 48 subjects). In order to investigate these issues in a more systematic fashion, we conducted a correlational analysis between the confidence scores of the eye-tracking system and various morphological characteristics of our subjects. Confidence scores are automatically produced by the Pupil detection algorithm for each pupil datum; they represent the quality of the detection result and low values may correspond to a blink. Three variables of interest were collected: confidence score for the right eye, confidence for the left eye, and confidence for both. Explanatory variables were gender; age; a self-declared assessment of the discomfort caused by the system (1 to 5); eye shade (dark/light); eye color; five binary variables tracking the presence of eye disorder (presence of any disorder, nearsighted, astigmatic, farsighted, and presbyopic); and diopter of each eye. In-house Python scripts using Pandas [McKinney 2010] were used to process all variables and compute Spearman rank correlations. Spearman correlation was selected as we had a combination of continuous, ordinal and binary variables with missing data.

None of the explanatory variables were found significant for the right eye, whereas, the only significant variable for the left eye is the discomfort level ($\rho = -0.39$, p = 0.005). This variable is also significant for both eyes ($\rho = -0.31$, p = 0.03). The negative correlations mean that greater levels of discomfort indeed lead to less reliable data for the subject. Interestingly, subject self-assessment is the best predictor of poor data capture. Although variables encoding vision disorders did not reach significance, it is worth mentioning that most of them were very close to significance for measures related to both eyes —presence of disorder ($\rho = 0.27$, p = 0.053); presbyopic ($\rho = 0.26$, p = 0.063); farsighted ($\rho = 0.25, p = 0.074$)—, but were far from significance when considering a single eye. Such finding would require further investigation to verify whether a greater sample would allow us to detect a significant effect for eye disorder on the reliability of data as well as to explain why this effect can be observed only when the two measures from each eye are put together by the system.

4.3 Issue 3- Behavioural issue: expert vs. novice, native vs. non-native

In our preliminary experiments, being a seasoned or a novice player seems to lead to different gaze patterns. An expert player concentrates his attention and the variations are quite fluid and linked to the anticipation of action. On the opposite, a novice player is quickly overwhelmed by all the stimuli offered by the game, which sometimes makes him visually "groping around" useful information. This is especially true for textual stimuli that are largely ignored by novice during phases of intense action. Concerning the differences between native or non-native players, the influence is noticeable in particular in the identification of NPCs that address the player during the dialogue phases, a native player doing it more directly.

Beyond these first observations on our data, we also wonder to which extent these differences in gaze patterns may have an impact on the reliability of the collected data. Our assumptions were that an expert and native player would produce more reliable data than a non-expert and a non-native player. We also postulated that expert player would exhibit different fixation patterns than non expert, possibly leading to a better capture of data. To further investigate these hypotheses, we performed correlational analyses between the confidence scores of the eye-tracking system, three fixation-based variables and various behavioural characteristics of our subjects. The confidence variables were the same as above (confidence score for the right eye, confidence for the left eye, confidence for both). Behavioural features were gender; age; three Likert scaled-variables measuring the subject habituation level to video games (1 to 5), and the subject habituation level to Skyrim (1 to 4). For the fixation variables, we first had to define what can be considered as a fixation. To this aim, we used a dispersion-based algorithm, as described in [Salvucci and Goldberg 2000], that is available in the Pupil Core fixation detector. It is based on the idea that fixation points tend to cluster together, so we can detect a fixation if consecutive points within a given moving window does not exceed a predefined dispersion threshold. Typical values for the minimum duration threshold spans from 100 to 200 ms, but we finally selected a value of 250 ms[2]. The degree of dispersion was set at a visual angle of 1.5°, which is the standard setting in the Pupil Player fixation detector [Kassner et al. 2014]. Once we had identified fixations, three variables were defined: the proportion of fixation located in the central zone of the screen (from 40% to 60% of the screen width and length); the average fixation duration; and the average Euclidean distance between a given fixation and the previous one. The latter variable aims at detecting players adopting more saccadic eye movements, which may lead to poorer data capture.

Results of this analysis first reveals that habituation levels to video games or to Skyrim in particular did not influence the reliability of the data capture in our experimental setting, as it was not significantly correlated with either confidence scores for eyes or fixation variables. In other words, being an expert player is neither correlated with different gaze patterns than those of non expert nor lead to more confident data capture. Our second assumption was that non native players would exhibit different gaze patterns from native players. In this case, correlation between being a native speaker and the proportion of central fixations is very close to significance (ρ = 0.28, p = 0.06). Non native players tends to spend less time in the center of the screen (mean = 19.7%) versus native players (31.8%). However, being a native or not does not lead to any difference as regards model confidence while analyzing the eyes. Other personal variables also seems to affect fixation patterns, such as age over centrality of fixations ($\rho = -0.30$, p = 0.04) and gender on the average distance between two consecutive fixations ($\rho = -0.35$, p = 0.02). In the latter case, women appear to move further from one point to another compared to men.

Our results must be taken with some cautious, not only because of the limited size of the sample, but also because of the various calibration issues described above, our computation of the fixation position is subject to approximation. In addition, if there are no statistically difference between fixation time or centrality within the areas of interest we considered, our qualitative data invites us to explore other ways to investigate potential eye patterns differences between players' profiles.

[2]Such value is standard in the field [Holmqvist et al. 2011]. Moreover, preliminary analyses revealed that a lower threshold tends to produce false positives for fixation points. As we were also interested in reading fixation points, it is generally acknowledged that minimum fixations of 250 ms are required to allow cognitive processing [Rayner et al. 2006].

4.4 Issue 4- Data visualization issue: analyze constant motion scenes

We reviewed default visualization that could allow us to identify variation between players' patterns, and consequently the statistical analyses to be conducted. The visualizations allowed by the post-hoc processing of eye trackers are mostly designed for 2D non-interactive content. These visualization can be either fixed representations (circle, cross or light), plots (polyline) or surface tracking (heat maps, fixations). These are either designed for temporary use (no tracking needed), or designed as synthetic forms of attention on relatively stable background elements (i.e. surface), and that despite the "video" format of the exported files. These visualizations tools are not suited for the analysis of data collected from interactive 3D content, such as video games or VR games.

The development of visualization techniques that accommodates the wide application range of eye tracking is an interdisciplinary challenge [Blascheck et al. 2014]. Several papers refer to fixation representation of 3D scan path [Stellmach et al. 2010], advanced attention maps for 3D virtual environments [Stellmach et al. 2010] or motion-compensated attention maps [Kurzhals and Weiskopf 2013]. However, current visualization tools remains insufficient, if the aim is to show longitudinal and comparative elements allowing advanced quantitative analyses (which seems particularly relevant in the case of complex video games). Some beta versions of software are developed to compensate for some current limitations (e.g. Blickshift Analytics). But there is clearly a need for more default progressive visualizations (dynamic heat maps, space-time frequencies, scan path), for they require a lot of expert manipulation of video files[3] at the moment.

5 CONCLUSION

In this paper, we wanted to highlight the difficulties and challenges of using eye-tracking analyses to evaluate players' feeling of immersion and the role of textual and linguistic components in the interaction with a complex open-world video game environment such as *The Elder Scrolls V: Skyrim*. Our initial analyses highlight the fact that the well-being of players with respect to the data collection protocol significantly influences not only the reliability of the data collected, but most likely the possibility of immersing oneself in the game. Beyond the attentional aspects captured, this indicates the need for eye-tracking data collection protocols that make the gaming experience as natural as possible. The consideration of eye disorders is another important aspect that we have investigated: due to the limited sample size, we were not able to detect significant impacts on data quality, but it is no surprise that some newer eye-tracking solutions, such as the *Pupil Invisible*, take eye disorders more into account.

Regarding to the evaluation of the linguistic elements, our analyses do not yet allow us to draw conclusions on the substance. At this stage, we can only draw very gross intermediate conclusions, based on our fixation analysis, that non native players tends to fixate less the central part of the screen than native players, possibly because they spend more time on subtitles. We plan to carry out future analyses of certain moments of play in order to assess more

[3]See for example visual-computing.org.

precisely potential differences between players and non-players and between French natives and non-natives. However, we can already point to the need for technical improvements both in the collection of data in multiple and variable areas of interest, and in the dynamic visualization of the information to allow for more refined interpretations of the evolution of ocular behaviors over time and in a comparative manner. It is clear that video game research involving eye-tracking analysis is still in its infancy, but that the issues it raises already lead to think about the future needs.

REFERENCES

Samuel Almeida, Ana Veloso, Licínio Roque, and Óscar Mealhala. 2011. The eyes and games: A survey of visual attention and eye tracking input in video games. In *Proceedings of SBGames 2011 (Arts & Design Tracks)*. 1–10.

Samuel Almeida, Ana Veloso, Licínio Roque, and Óscar Mealhala. 2012. Video games and eye tracking: Does the video game industry know eye tracking?. In *Play, Game, and Society, Proc. of Videojogos 2012, 5th Annual Conf. in the Science and Art of Video Games*. 269–281.

Tanja Blascheck, Kuno Kurzhals, Michael Raschke, Michael Burch, Daniel Weiskopf, and Thomas Ertl. 2014. State-of-the-Art of Visualization for Eye Tracking Data.. In *EuroVis (STARs)*.

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 (2015), 8–14.

Niclas Ejdemyr. 2016. *Eye Tracking as an Additional Input Method in Video Games: Using Player Gaze to Improve Player Immersion and Performance.* Master's thesis. Blekinge Institute of Technology. Bachelor of Science in Digital Game Development.

Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer. 2011. *Eye Tracking: A Comprehensive Guide to Methods and Measures(1st edition).* Oxford University Press, New York.

Faisal Iqbal. 2015. *Understanding User Interaction in a Video Game by using Eye Tracking and Facial Expressions Analysis.* Master's thesis. University of Tampere. Master of science in Human Technology Interaction.

Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction. In *Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Seattle, Washington) *(UbiComp '14 Adjunct)*. ACM, New York, NY, USA, 1151–1160. https://doi.org/10.1145/2638728.2641695

Kuno Kurzhals and Daniel Weiskopf. 2013. Space-time visual analytics of eye-tracking data for dynamic stimuli. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2129–2138.

Riccardo Lolatto, Giulia Rocco, Riccardo Mustoni, Chiara Maninetti, Riccardo Pastura, Andrea Pigazzini, and Riccardo Barbieri. 2020. Characterization of Eye Gaze and Pupil Diameter Measurements from Remote and Mobile Eye-Tracking Devices. In *XV Mediterranean Conference on Medical and Biological Engineering and Computing – MEDICON 2019*, Jorge Henriques, Nuno Neves, and Paulo de Carvalho (Eds.). Springer International Publishing, Cham, 201–208.

Jeff J MacInnes, Shariq Iqbal, John Pearson, and Elizabeth N Johnson. 2018. Wearable Eye-tracking for Research: Automated dynamic gaze mapping and accuracy/precision comparisons across devices. *bioRxiv* (2018).

Carme Mangiron. 2016. Reception of game subtitles: an empirical study. *The Translator* 22, 1 (2016), 72–93.

Wes McKinney. 2010. Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference*, Stéfan van der Walt and Jarrod Millman (Eds.). 51 – 56.

Keith Rayner, Simon P Liversedge, and Sarah J White. 2006. Eye movements when reading disappearing text: The importance of the word to the right of fixation. *Vision research* 46, 3 (2006), 310–323.

Tony Renshaw, Richard Stevens, and Paul D Denton. 2009. Towards understanding engagement in games: an eye-tracking study. *On the Horizon* 17, 4 (2009), 408–420.

Dario D Salvucci and Joseph H Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*. 71–78.

Oleg Špakov, Howell Istance, Kari-Jouko Räihä, Tiia Viitanen, and Harri Siirtola. 2019. Eye gaze and head gaze in collaborative games. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. 1–9.

Sophie Stellmach, Lennart Nacke, and Raimund Dachselt. 2010. Advanced gaze visualizations for three-dimensional virtual environments. In *Proceedings of the 2010 symposium on eye-tracking research & Applications*. 109–112.

Bob G. Witmer and Michael J. Singer. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments* 7, 3 (1998), 225–240.