Matching the effects of immersion to game mechanics in a cooperative game

Rúben de Carvalho Nogueira Vines

Abstract—In current times we grant questionnaires as the standardized way of measuring several experience-driven feelings. The same applies for the likings of immersion. However, taking a questionnaire is usually a tedious and leisurely task that test subjects usually show aversion or discontent in performing. Therefore we focused a study on finding if a complementary or alternative tool could be created in measuring immersion.

We hypothesized that the actions a player takes when playing a game can be inline with their reported immersive feeling. In case this revealed to be true, we could now have an in-game implementation that would have measuring capabilities comparable with a questionnaire, while also providing several other benefits. For instance, we could now withdraw the dull task of answering a questionnaire; give game developers detailed information on how their game mechanics were affecting the immersive experience; and lastly, open up run-time video game adaption possibilities.

A test was conducted to analyze the correlation between some pre-established game mechanics and actions - metrics captured by our implemented model - i.e. with immersion values reported via questionnaire. Pairs of player (n = 20) were asked to play the first level of the cooperative game Dark Things About. The most promising correlation value with immersion (W=-0.352, p-value=0.128) happened with the in-game distance covered, using a Spearman’s rho correlation test. This result is inconclusive when trying to establish a connection with immersion, but the used methodology suggests a new manner of analyzing gameplay data, opening a guideline in possible future approaches.

I. INTRODUCTION

Video games differ enough to suit several personalities or motivations (socializers, explorers, etc.), levels of skill or dexterity (point-and-click and controller-driven games), technological requirements towards realism (2D, 3D, VR), genre preferences and many other personal characteristics. With the computerized representation of games there was the arising of new manners of perceiving and feeling towards this media. For instances, in the turn of the millennium the concept of immersion came into play. Douglas and Haragdon - in The Pleasure Principle: Immersion, Engagement, Flow - introduced this affective experience into the gaming world. They highlight the interactivity of the interface and the “ability to take situated action and to savor the results [...]”. In their eyes “interactive games fulfill their promise as immersive when they offer us [...] interactions enabling us to enjoy virtual experiences either unattractively risky or denied to us in everyday life” [1].

Immersion became an important keyword both in the scientific community and the marketing side of the gaming industry. Its meaning and effect is not the same between authors and has evolved in the last two decades. However, even though unclear, it is seen as fundamental part of the experience by either developers or players.

Even through most developers would argue their intent is to create the most immersive experience possible, their pursue is probably empirical and unaligned with their intentions. So what holds a developer from purposely pursuing a more immersive game? Maybe the ambiguity of the term; the standardized static way of measuring it; or the lack of tools to adapt their game into a more immersive one.

The referred standardized way of measuring immersion is through questionnaires. Their usage comes usually after user-testing sessions which may be inconvenient for two main reasons: it extends the session’s duration, and it is performed after the highest levels of engagement between the player and the game are over. These processes are usually tedious and dull, making this very important tool for developers unappealing to participants.

With the increase in popularity of real-time video game adaptation, it is more and more common to retrieve players’ in-game inputs in order to dynamically change, place or generate content to better suit their needs. For that reason, a dynamic method of measuring effects like immersion could potentially be equated and used in such methodology. Adapting a player’s immersion to create the best possible experience. However, since the current method is purely static it obviously does not take into account for this potential.

Our hypothesis in approaching this problem is by trying to understand if the way we feel towards the game we are playing can be caused by the way we play the said game. Either by being more of less effective, more or less goal-aligned, having more or less skill, and many other performing qualities and metrics of a given player when playing a specific game.

Metrics are quantifiable actions that occur recurrently throughout the gameplay. For this particular project and game, some of these actions can be: successful and unsuccessful interactions, number of deaths, distance covered or distance between both players. With a cooperative two-player game at hand named Dark Things About we assessed if some of these metrics can correlate, and admittedly cause, the immersion values obtained by a standard questionnaire (particularly IEQ).
II. RELATED WORK

A. Social Play & Social Presence

The cooperative aspects of our testbed game made us study further the social interactions while playing games. For instances, social play is seen as playing in the company of someone. This type of play can happen by playing a game with or against someone, surrounded by a crowd or any other type of way that can simulate this interaction (e.g. via internet). To see if this simulated way of interaction produces a comparable effect to the real-life interaction, the levels of social presence are measured.

Biocca and Charms created a questionnaire - Networked Minds Measure of Social Presence [2] - for this exact purpose: understanding one’s sense of social presence when interacting with others, in this case, in a mediated way. In Guide to the Networked Minds Social Presence Inventory - by the same authors - social presence is seen as: “the degree to which users of a medium feel that mediated others are spatially co-present, psychologically accessible, and behaviorally interactive.”[3] Additionally, Cairns et al. define recognizes social presence to happen “when there are actions that have social meaning [...] or social actors who can affect your mood and your actions as a response to the social situation”.[4]

The previous quote is referenced from Who but not where: The effect of social play on immersion in digital games - Cairns et al.’s endeavor on understanding the influence of social presence on immersion (as a result of social play). This research was influenced by Gajadhar et al.’s earlier effort on understanding the same [5]. Results showed that higher social presence (by either playing online or in co-location) meant higher immersion. Not only so, but restating Ravaja’s [6] claims, playing with friends (as opposed to strangers) resolves in an overall better game experience.

B. Immersion, Flow & Presence

Gordon Calleja¹ states that regarding virtual environment habitation (being “in there”) people tend to define it deep into a spectrum of definitions that can go from “the deeply engaging to the hopelessly addictive, or any shades in-between”. Amongst that spectrum, immersion and presence usually come along, sometimes separately, many times interchangeably. For disambiguation “the game’s world either being games research, game’s industry or general gaming population [...] uses immersion as the preferred term”. In-Game: from immersion to incorporation [7] is Calleja’s contribution dedicated to the exploration of all headings of many ambiguous terms used for engagement by the gaming and scientific community. For instances, a strong argument against Slater’s view on presence and immersion being necessary to each other is held.

The look on presence as something that comes from high-fidelity mediation is pointed out as a flaw by Calleja. An example is given where someone who has never heard an orchestra or been to a theatre before wouldn’t feel present in the ambient created by the sound system because there is nothing familiar that the said person can be transported to. Therefore in his eyes high-fidelity systems are not the cause of presence but representing an ambient with fidelity may be a big step in feeling present. This way, the feeling of immersion comes from a continuous stream of stimuli and can’t be turned into an “on and off switch” as Slater seems to state.

Flow is probably the least misinterpreted term out of the ones associated with enjoyment. Mainly because Csikszentmihalyi’s early endeavours[8], [9] gave foundation to many other researches. Here, described as the process of optimal experience, flow, is the “state in which individuals are so involved in an activity that nothing else seems to matter”[9]. The eight dimensions of flow experience are described as the following: clear distinctive objectives or goals and immediate feedback; high opportunities for acting matched by one’s perceived ability to act; one-pointedness of mind; absolute concentration on task at hand; sense of potential control; loss of consciousness; altered sense of time; and finally, autotelic experience.

Brown & Cairns, in a user study, proposed immersion to be three-layered state where a person needs to break consecutively barriers of immersion to achieve higher layers or levels. Firstly comes engagement which is “the lowest level of involvement [...] and must occur before any other level. To lower the barriers to enter this level, the gamer needs to invest time, effort and attention”. Then it comes engrossment, where a player needs to “become further involved [...] and become engrossed”. Specifically, when a player feels engrossed its “emotions are directed directly affected by the game”. Lastly, and more controversially, comes total immersion which the authors describe as being the same as presence. In this state “participants described being cut off from reality and detachment to such an extent that the game was all that mattered”[10].

Jennett et al. in a study that hypothesizes several ways of measuring immersion, characterizes immersion as lack of awareness of time, loss of awareness of the real world, involvement and a sense of being in the task environment[11]. Cairns et al. takes this latter approach and extends immersion to “a cognitive experience wherein thoughts of the player are wholly absorbed in the action within the game [...]”. Once again, they make a correspondence between immersion and flow. This time, in their understanding “flow corresponds to the extreme end of immersion where a person is so immersed [...] they enter a flow state”[4].

Lastly, Lombard and Ditton separated the concepts of spatial and social presence as a way of trying to break down this complex and controversial term. Presence is a way to address the feeling people have when interacting with media. Being present in the medium means having the illusion of non-mediation when experiencing mediated environments (i.e. losing the perception of media interaction when that is happening in the actuality)[12].

C. Experience Evaluation

Biocca and Harms created Networked Minds Measure of Social Presence not as a questionnaire designed for games but instead as a more general way of assessing the user’s sense of social presence [2]. Social presence in this questionnaire was conceptualized as a combination of six different sub-dimensions: co-presence; attentional allocation; perceived message understanding; perceived affective understanding; perceived affective interdependence; and perceived behavioural interdependence.

Social Presence in Gaming Questionnaire (SPGQ) also regards social presence, but this time with several applications in experiments with video games. This questionnaire designed by Kort et al. consists of 17 items scored on a Likert scale (0 to 4) [13]. Those items are unevenly dispersed in three sub-scales:

1) Psychological Involvement - Empathy (6 items): sense of feeling to be in the same enjoyable situation as the other players;
2) Psychological Involvement - Negative feelings (5 items): extent to which players were competitive or aggressive towards other players;
3) Behavioural Engagement (6 items): degree of feeling that the other players were influencing the game.

Game Experience Questionnaire (GEQ) the questionnaire used by Gajadhar et al.[5] and refuted by Cairns et al.[4], is a very extensive questionnaire that serves the purpose of measuring several aspects of the game experience. The questionnaire is divided in three modules that serve to address the gaming experience as a whole, each using a 5-point Likert scale[14]. The said modules are: core module - assesses the game experience as scores in immersion, flow, competence, positive and negative affect, tension and challenge; social presence module - investigates psychological and behavioural involvement of the player with other social entities; post-game module - assesses scores in positive experience, negative experience, tiredness and returning to reality after they stopped playing;

Finally, the Immersive Experience Questionnaire (IEQ), designed by Jennett et al., is a 31 question-based tool which utilizes a 5-point Likert scale to score immersion in a range from 31 to 155 points [11]. It is composed by five different components with an uneven amount of items:

1) Cognitive Involvement (9 items): experience of focusing on the game;
2) Emotional Involvement (6 items): strength of feelings experienced while playing;
3) Real world dissociation (7 items): sense of losing awareness of the world around and increasing awareness of the game;
4) Challenge (4 items): experience of being challenged by the game;
5) Control (5 items): extent to which player felt in control whilst playing.

D. Discussion

Cairns et al.’s work on immersion is seen by many as a mark in the literature. Thus, this research is going to take into account these authors views, terminology usage, approaches and further details henceforth. More specifically, using or adapting several points into this project: the same game experience measuring questionnaires (IEQ and SPGQ); ‘co-location’ scenario; and accounting for the nature of social connection (i.e. playing with friends).

Regarding terminology, and using the same reasoning of the prior paragraph, immersion shall be seen as the lack of awareness of time, loss of awareness of real world and a sense of being in the task environment (Jennett 2008 [11]); presence as a way to address the feeling people have when interacting with media (Lombard and Ditton 1997 [12]); social presence happens “when there are actions that have social meaning[…] or social actors who can affect your mood and your actions as a response to the social situation”; and finally, a bridge between two terms where flow is seen as “the extreme end of immersion where a person is so immersed […] they enter a flow state”.

III. CASE STUDY

A. Dark Things About

DTA is a survival horror two-player local cooperation game. Its development started in the beginning of 2017 in the Information Systems and Computer Engineering Master’s course with the partnership of Laboratório de Jogos do Instituto Superior Técnico, and has continued in parallel as a side project ever since.

Unreal Engine 4 is the supporting game engine for this game, aided by an audio workstation named Wwise and 3D-modeling tools such as Blender and 3DS Max. All this stated programs are highly compatible with the game engine, hence their choices. Apart from compatibility, this engine seems to be keeping up in the vanguard of engine technology comparing to its free-to-use competitors. Opening up the possibility of importing external plugins and tools with no great issue. The engine uses C++ as a programming language and an alternative visual scripting method called Blueprints.

Using this game as a basis for scenario testing brings added leverage: no need to adapt to another engine, study tools to mod an existing game, or start a testbed game from scratch; time and effort put into creating a better game experience, will also result in a better game experience for the future test subjects of this game; results from this project will be conveniently used in the game (either gameplay or knowledge wise), hopefully creating a better project and game overall.

The core mechanic of the game is a compromise that happens by one of the players being attached to the camera (game’s view perspective) while the other player holds a light source. The main idea is to cooperate in such a way that the camera player can keep the light player at focus, while the light player enlightens the path of his partner (represented in Figure 1).
These mechanics are interchangeable at any time, meaning that by being close to one another and willingly clicking a given button at the same time will make the camera player hold the light, and vice-versa. The lantern one of the players holds can also slow down or stun enemies. In the case of holding the camera, a player can also see through its character perspective. Which means if a character holding a camera is a medic then medicine related objects will be highlighted in its view; while if it is a priest he will see a more divine version of the world. Lastly, using a power up with the lantern can increase the intensity and radius of the light source. To be allowed to do it, consumable flasks need to be picked up along the way.

B. Game Scenario

For the existing game scenario, players are presented with the tutorial level, so that they can potentially learn and exercise the mechanics while also exploring the scenery. Prior gameplay testing has proven that the time to beat the level is around the 20 minute mark, which usually is enough to become engaging without becoming exhausting. The two following images (Figures 2 and 3) show an overview of the map played:

![Fig. 2. Village sky overview with interaction labels](image)

In the figure 2, on the reachable player ground (delimited with a green shading) we can see several possible interactions a player can go through while completing the level. When starting the game both players spawn next to each other on a village (route indicator 1) and are prompted, via a cutscene, with the immediate task of reaching a house (route indicator 2).

Tutorial points 1, 3 and 4 are connected with the previously mentioned fuel flasks. The first, teaching the players how to collect them; the second demonstrating their use - increasing the intensity of the light; and the last giving a purpose for its use - scaring an enemy that is present in danger zone 1. Tutorial point 2 is a small challenge that helps the players settle in the unusual setting of only one of the players controlling the game’s perspective.

Clickable items 1, 2 and 3 are objects that can be interacted while nearby, serving the purpose of guiding the player to the pretended destination (route indicator 2). Clickable items 4 and 5 (figure 3) can be interacted in the same fashion, but in this case they are levers that integrate two puzzles that are to be cooperatively beaten. In here, danger zone 5 concludes the possible places a player can die in this level - in this case, by pressing an incorrect lever.

Most consumable items represented around the map are mostly fuel flasks that can be collected and used. The main exception is consumable item 1 (figure 3) that is a key to the destination door (route indicator 3). Lastly, the remaining type of interaction existent in this world is the possibility of opening a letter. Doing so happens with the action of one of the players, but can only end when both accept to close the letter.

C. Data Gathering

Since our hypothesis means to match in-game actions with the values of player immersion, we can’t be sure at first of which actions can be responsible for that. Therefore, we needed to implement a data retrieval system that could gather all the necessary information to create our quantitative measures, while also opening up the possibility for creating future ones with the data extracted. For this reason we divided the capture and transformation processes of the data in two separated steps.
Data gathering process was implemented within the game’s code using \textit{Unreal Engine 4} blueprint scripting. The main goal was to capture every action of both players during a session, making it so it would be theoretically possible to reconstruct the entire session after it was over. Following the world’s description, given in the latter sub-chapter, we can divide the possible outcomes of data at a given time in three key types: discrete and without gameplay value (e.g. player pressed a button on a controller); discrete and with gameplay value (e.g. player pressed a button and interacted with a game object - changing the world’s state); and lastly, continuous (e.g. the players’ positions at all times). The following activity diagram (Figure 4) envisions the flow of this process:

![Activity diagram for the run-time data gathering process](image)

From the diagram above we can perceive the ongoing activity that is executed after one of the players presses the start button on the main menu of the game. When doing so, the game fetches the current file number from a given local directory, stores it and immediately creates a file that will save both player positions at every possible timestamp. This file’s naming convention is “Distance\_FILEID”, where “FILEID” is an integer correspondent to the amounts of games already existent in that directory. The reasons behind separating this continuous data gather from the remaining has to do with the large amounts of information comparatively generated. On the discrete side every variable is updated in run-time and dispatched to a file when the session ends. As for the continuous (i.e. player positions), the respective variable is printed to its file in accordance with a timer.

“Metric\_Report\_FILEID” is the file’s naming convention for every other discrete piece of information. Every time an action occurs either in a player controller or in the game world, a correspondent log is added to a \textbf{preliminary metric value}. These values are grouped among homologous actions, where every line for a given group means a new action of the same kind.

When an action is logged, it is included with other existing ones that are from the same category. Along with that action other four key points are captured: a timestamp of when was it performed, by whom was it performed, and both players positions at that moment. Retrieving these additional details opens the possibility of creating metrics derived from the primary act.

The data capturing process resolves when the last action on the level is accomplished (i.e. opening up a door). Doing so will result in the two log text files above mentioned. Lastly, a preventive way of printing the files without reaching the end of the level was also implemented with a keyboard input.

\textbf{D. Data Treatment}

Upon finishing a gameplay session two text files would be created containing thousands of line entries. But as previously briefly described, these were not more than preliminary metrics. Most do not constitute what we find to be worthy of representing the gameplay experience on their own.

We conjecture this line of thought on the literature, and argue that since immersion is achieved in the presence of several factors that only make seen for an overall experience, the metrics that compose the experience should also describe it as a whole. We tried to transpose to metrics the actions that could potential either bring cohesion or disruption to a player’s satisfaction. Doing so was based on three factors: whether the measure is based on an important gameplay mechanic; if an action happens usually either constantly or very frequently while playing; and finally, if an action is something players are familiar with upfront - by either being perceptibly intuitive or commonly seen in other games. With this in mind and from the existing data, we created the following quantitative in-game metrics:

- **Deaths**: Number of times a given player died [and respawned];
- **Successful interactions**: Number of times a player pressed the correct input near an object that required that same input;
- **Failed interactions**: Amount of non-fruitful or unresponsive input actions - i.e. when a player pressed a button and a successful interaction didn’t occur;
- **Time as Camera\(\text{(ms)}\)**: Time spent while holding the game’s perspective (one of the main mechanics in the game);  
- **Time away from focus\(\text{(ms)}\)**: Time spent while holding the lantern and not appearing on the viewport of the game;  
- **Distance covered\(\text{(in-game\ meters)}\)**: Amount of ground that was stepped on;  
- **Time stopped\(\text{(ms)}\)**: Time spent while not changing position;  
- **Average distance to other\(\text{(m)}\)**: Average distance between players during the session.

The task at hand was now to parse the existing information from the generated text files to usable metrics. For that, a \textit{Python} script was created using \textit{Thonny}², a Python’s IDE. When executed, it went through every existing log with the predefined names, and recovered information from both files for each pair. Dealing with each category of information separately, sequenced by the order coming in from the metrics file.

Apart from capturing (and counting) the amounts of times an event occurred, it also dealt with converting the formats of timestamps and player positions into more convenient

²[https://thonny.org/](https://thonny.org/) - Free Python IDE
units. Some measures were directly obtained by counting the amounts of occurrences (e.g. deaths); other made use of the stated format conversion to get the pretended sums (e.g. distance covered and time stopped); and a few were derived from other simpler metrics (e.g. time as camera was comprised from knowing which player started with the lantern, and the timestamps where players swapped the camera with the light).

The exporting process of these values to a statistical program only made sense, however, after obtaining the immersion values for each session. That process was achieved via a questionnaire after the conducted test scenarios - described in the next chapter. After doing so, these immersion values would be saved in another text file in the same directory as the others. Finally, a csv file would be created containing the above mentioned metrics in addition to the immersion value. Having a line entry for each individual participant. This way, importing the values into IBM’s SPSS Statistics 25 was a simple procedure.

IV. EVALUATION

This chapter is composed of by the testing phase of our project. Firstly with two preliminary tests that were carried out before proceeding into our finalized test. In the first one, we knew immersion was aligned with our intents for this project, but at this stage we conjectured if other experiences alike could also be studied. So we focused on the values of Social Presence while playing DTA, since the testbed game has a purely cooperative component associated with it and we wanted to understand if that component revealed itself through questionnaires after a gameplay session.

A. Preliminary Tests

We assembled this first test during Lisboa Games Week - a games' convention that happens every year in Portugal - where the game "Dark Things About" was present as a finalist for the PlayStation Talent Awards. This event occurred between the 15th and 18th of November 2018 in a major exhibition room in Lisbon, where there would be a large affluence of people with high interest in games.

During the four days of the event, 22 participants agreed to participate in the experiment and 21 valid answers were taken from the test (1 user didn’t properly submit his form). Out of these 21, 14 were male and the remaining female. Their ages were between 16 and 32 years-old (M = 20.48, SD = 4.43).

Regarding their professional relation in the video game industry 9 people answered with none, other 9 said they were students in that area, and only 3 users were professionally related with video games. Inquired for the frequency of which each person played video games: 14 said that they scheduled their time in order to play, while the rest only played occasionally and when given the opportunity. Then we questioned about multiplayer games, and learned that: 1 person played rarely them; another said that every game he played was multiplayer; nine users reported that the majority of games they played were multiplayer, and the remaining 10 said only some of the games they played were of that type. Concerning the 3rd person adventure genre (a broad definition in which Dark Things About fits), results showed that 2 players don’t play this kind of games; 9 had played at least one of these games in the past while the remaining sample said it was one of their most preferred genres. Eight people had played a Dark Things About demo before that day.

A two-section questionnaire was created in order to inquire the participants. Firstly, the already reported section that was meant to characterize the sample’s demographics. This section was reused for all the other evaluation questionnaires of this project, having only minor corrections and adjustments until the finalized version. And then, a Social Presence related section used to determine the involvement of player with their gaming partner.

SPGQ was used as the the core for this measurement. This questionnaire is based on the GEQ’s Social Presence Module having however a more extensive analysis in the subject by adding 4 additional items to the original questionnaire. This survey is composed by three sub-scales: psychological involvement - negative feelings, psychological involvement - empathy and behavioural engagement. All the questions had their order randomly assigned and no association to which sub-scale they belonged to.

To make up for the uncommon conditions the test was conducted, a pair would only be invited to partake in our study after reaching the final end of the demo. Note that this would be independent of the enjoyment and difficulty shown for the duration of the playthrough. Only then, and with their acceptance, we would administer the previously described questionnaire.

Awarded in a 5-point Likert scale, each question can be valued between 0 and 4 - respectively from lowest to highest agreement for a given item. Calculating the mean and standard deviation of every subject’s answers for each SPGQ sub-scale, we got the results given in Table I.

<table>
<thead>
<tr>
<th>PI – Empathy</th>
<th>PI – Negative feelings</th>
<th>Behavioural Engagement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>3.05</td>
<td>0.60</td>
</tr>
<tr>
<td>Std Dev</td>
<td>0.37</td>
<td>0.21</td>
</tr>
</tbody>
</table>

TABLE I
SOCIAL PRESENCE REPORTED BY TEST SUBJECTS

Psychological Involvement - Empathy - i.e. feeling to be in a situation as enjoyable as the other player - and Behavioural Engagement - impression on whether the other player influences the game - came recorded both very high scores. Psychological Involvement - Negative Feelings - i.e. competitiveness and aggressiveness shown towards others - produced a convincingly low result. All these results neared advisable maximums and minimum, without ample variation. Aligned with the most desirable results for the present type of game. Thus, suggesting that our hypothesis of finding a correlation using values with such characteristics would

theoretically be very difficult. So, we discarded the study of social presence in addition to immersion, and decided to narrow our focus solely to the latter.

On a second experience, and in order to validate our process and better execute the last evaluation phase of the project, we conducted a pilot test. This experiment was carried in parallel with the development of the data retrieval and transformation processes.

With the participation of 4 MSc and PhD students (2 of each) with high interest and academic relation with games, we solidified our process. In this test, we substituted the second section of the latter mentioned form by the intended immersion questionnaire (IEQ). We ran and assessed the entirety of the process: briefing, gameplay session, questionnaire delivery and debriefing. The obtained results helped us in solidifying our process, correcting minor issues in the questionnaire or in the design of the game scenario, as well as, dealing with problems found in the data gathering implementation.

B. Final Test

1) Procedure: After having a functioning implementation of the player’s in-game metrics gathering and treatment processes, it was possible to address the main question between a player’s performance and their feeling of immersion. The carried out experiment counted with twenty (20) participants - 5 females and 15 males - that played the first level of Dark Things About. This sample was composed of 10 pairs of Instituto Superior Técnico’s students (4 female/male pairs, and 6 only male pairs), registering an average of 22.3 years (SD = 0.19). The already mentioned demographic’s section of the questionnaire was once again used, characterizing the subject’s background and desirability to play video games.

On the topic of the testers’ professional relation with games 11 users responded with having no relation, 7 having a purely academic relation and 2 being researchers in this field of study. When asked about how often they played video games, 1 person admitted to never play video games, 9 showed that they played when the opportunity presented to themselves, and the remaining 50% of the population revealed to purposely reserve time on their schedule to play video games. Regarding the question on the frequency of playing multiplayer games: 6 reported that they have played only a few games while 12 play mostly this type of game; the remaining 2 played exclusively multiplayer games. As for the interest in the 3rd person adventure genre: 45% of the sample was familiar with it and had played at least one game of this genre, 5 people said it was their favorite and other 5 said it wasn’t part of their preferences; 1 person restated not to be interested in video games. Lastly, 15 out of the 20 players had never played any demo of the testbed game.

In order to try to test our hypothesis and correlate the in-game player metrics and their own immersion levels, information was gathered by two means: in-game data - captured by the game while the players went through the test level; and immersion level - collected via questionnaire presented after the gameplay session. The used questionnaire is IEQ, scored in a 5-point Likert scale from 1 to 5: where 1 usually the highest disagreement depending on the formulation of the question; opposed to 5 which means the symmetric value in this said spectrum. The resulting sum of the items varies between 31 and 155. A lesser number means a lesser degree of immersion, and a higher number means a higher degree.

The gathered in-game data is preserved as a single file for both persons on that test. Processing the retrieved information is the coming step in transforming the large amounts of data into usable metrics for each single player. As described in the previous chapter, the following metrics were considered: Deaths, Successful interactions, Failed interactions, Time as Camera, Time away from focus, Distance covered, Time stopped and Average distance to other.

The same process achieved in the pilot test was now reused. We asked the pair of players to play our game scenario with minor spoken interventions in case of game design discussion or to resolve eventual in-game setbacks. When the pair reached the end of the demo, a log containing information for that session would be saved automatically. The pair would then be thanked for playing the game, and requested to fill the experience questionnaire (prepared from the last preliminary test). Then, the questionnaire would be handed out.

Following the user tests, every file recovered from the gameplay sessions would be parsed and agglomerated turning its raw information into the previously described metrics. This process was achieved running a Python script for all existing files. Then, a single csv file would be generated containing the metrics’ information for every individual.

2) Results: Three different approaches were made in order to thoroughly analyze the obtained data. Firstly, while tackling our hypothesis, we needed to understand if there were any indications on whether a player’s feeling of immersion can be identified by quantitative actions performed while playing.

Having every metric value and immersion level discriminated for each individual was the first step in this challenge. To understand whether or not gameplay and immersion correlate, we needed to perform statistical operations to get said correlation values. We did them using SPSS 25.

Due to the reduced sample size we proceeded with a non-parametric approach. Consequently, a non-parametric correlation test was also performed: Spearman’s rho. It appeared to exist a slight indication of correlation. Spearman’s rho ranked the correlation between "Immersion" and "Distance Covered" on a value of -0.352 for a p-value = 0.128 (n = 20). This value couldn’t be considered a high correlation, but it was a first step in considering new other approaches.

Our second approach was an attempt of further exploring the obtained results. Since the most promising metric was related with the amount of ground that was covered by the players we looked into the level design of our game in order
to assess if there was something that may be responsible for that result. Therefore, we made a separation of the level in two parts that we felt were considerably different in terms of producing ground coverage values (a forest with a wide area, and a house with a constrict area). However, repeating the same statistical analyses as before dismantled this line of thought. By neither achieving much different values of covered ground in both scenarios, nor getting any improvement on the correlations for both variations of the metrics (“in forest” and “in house”).

Lastly, our final approach to scrutinize the metrics was by exploring the correlation not with immersion, but within the detailed metrics themselves. When reevaluating the already performed non-parametric test (Spearman’s Rho), several correlations were revealed. The more relevant ones are demonstrated in the table below (and the dispersion of data in the later presented Figure 12):

<table>
<thead>
<tr>
<th></th>
<th>Failed Interactions</th>
<th>Time Away from Focus</th>
<th>Time Stopped</th>
</tr>
</thead>
<tbody>
<tr>
<td>Successful Interactions</td>
<td>0.452 (p = 0.045)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Time in Camera</td>
<td>-0.677 (p = 0.001)</td>
<td>0.430 (p = 0.012)</td>
<td>-</td>
</tr>
<tr>
<td>Average distance to other</td>
<td>0.506 (p = 0.022)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Distance covered</td>
<td>0.379 (p = 0.099)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

TABLE II
SPEARMAN’S RHO CORRELATION BETWEEN IN-GAME METRICS

3) Discussion: When comprehending the alignment of immersion with gameplay mechanics we were trying to canvass the possibility of this feeling being more than purely subjective. We were thriving for a more theoretical iteration on the immersive experience that is commonly accepted in today’s industrial and academical video game studies. Doing so could potentially open doors to run-time video adaptation possibilities. For that, immersion needs to surface as being reproduced (or correlated) by what a player objectively does while playing, therefore, adjusting these actions could ultimately also tune the experience.

Results obtained for this correlation were not conclusive but still incited future solutions to explore. In the first approach, when trying to understand if the mentioned correlation existed, the metric of “Distance covered” seemed to slightly align with our assumptions. By scoring -0.352 in Spearman’s Rho test, player’s appeared to be less immersed the more ground they covered. There are potentially several reasons for this to happen: the game may have had a lot of unnecessary, long or confusing paths - giving a sense of dissatisfaction to the ones who ended up walking the most. Splitting the information into two zones (“house” and “forest”) did not show to be enough to identify the issue. This line of thought ended up contradicted by the values of ground covered by the players in and outside the house - showing larger values for indoors (house: M = 90.7m, SD = 30.5; forest: M = 81.8; SD = 32.9). This may be because of the existing puzzles inside, where players may have travelled much in order to assess which were the correct levers to press to beat that challenge.

Immersion continues to been seen as an experience that is fairly more recognizable by its effects on a person using a medium, than by the ability of using that medium. However, there are still reflections to take from this study. Both in terms of results and approach. The intended contribution from this study was to provide game developers a strict and theoretical approach to their methodology of producing more engaging and immersive games. Or at least explore the feasibility on achieving that approach. The most impactful results of this study that could potentially serve the game development contribution as a tool is the correlations of metrics between themselves.

Looking at some of the strongest correlations inbetween metrics (latter presented in Table II) we can try to understand their meanings. Some will seem very obvious: correlation of “Successful interactions” with “Time stopped”, while others can be considered counter-intuitive: “Successful interactions” with “Failed Interactions”, respectively. Scrutinizing and understanding the reasons behind them can become a viable way to tune the experience of the players.

For instances, it is very understandable that the metric of making fruitful actions (“Successful interactions”) has a parallel with the time players spends stationary (“Time stopped”). Since every time an interaction of this sort occurs, the in-game character displays an animation where it locks the player briefly to a given position. In the optimal case these metrics would also align with immersion. If so, we
could adjust the experience by either inciting the player to either perform more interactions of this sort (if the correlation was positive); or otherwise, by locking the player’s position for lesser span of time.

“Successful interactions” to “Failed Interactions” correlation seems to match positively. Meaning that the more times a player performs a successful interaction, he also performs more incorrect ones. This result may seem abnormal or at the least counter-intuitive. A possible reason for this to happen is due to what we consider “successful” and “failed”. Successful is when a player presses a given button, near enough an interactable object that requires that input - ultimately interacting with it. Failed means not doing so. It may be that players that strive more to interact with every object in the game may not be receiving a clear enough feedback, therefore, pressing the same button over and over until achieving the desirable outcome - increasing the overall number of failed interactions. These are mere conjectures. They serve but to show the direction in which this project would go in case there was alignment with them and immersion.

V. CONCLUSIONS

Ultimately, our approach cannot provide a concrete answer on considering an alternative approach to questionnaires on measuring immersion. Since exploring our main hypotheses of finding a correlation between immersion and game metrics values deemed inconclusive. We believe, however, that our process of metric selection was substantiated, and that not every aspect of the game can be responsible for affecting or even inducing the state of immersion. Therefore, several points from our methodology can be taken to further explore this work’s basis.

The obtained most valuable converging point between the relation of immersion with gameplay was the Spearman’s rho correlation \((W=-0.352, \ p-value=0.128)\) between said immersion and distance travelled by players. The value was not strong enough to proceed with more conclusive answers on the influence of this metric in immersion. However, as a developer, making efforts in understanding the real impact of this value in the experience, may well be a enough in producing a better one. All in all, this project raised the degree of knowledge of the relation with DTA and the cooperative aspects of players who explore this game. And here, a parallelism with other developers and games can be drawn.

A. Future Work

Nevertheless, the real potential from this project may be achieved if a correlation with immersion can be strongly asserted. Therefore we ponder some possible variations to this project’s approach that may see different results from what were here obtained:
- **Single-player game.** We were supported by the literature’s positive endorsement on playing games in the presence of someone. However, considering a single-player game might reduce the disruption of a player’s data from outside factors (in this case another player);
- **Account for player preference.** Some players might enjoy playing the game as they want, and for as long as they want. Having a predefined locked goal to achieve (e.g., we asked people to play until reaching the end of the level) may be unaligned with what some players expect from the experience. Some might find it exhausting while others would have liked to proceed further. Giving the player the freedom to play as much as they want (in a game where doing so would be acceptable) might be enough to obviate this situation. In the case of reaching further in exploratory work, an effort could be made in understanding if personal preferences have in importance in identifying positive aspects of the game.
- **Larger user sample.** Large amounts of test subjects are usually crucial in statistical-based experiments. We inferred that our obtained data followed non-parametric distributions for this reason. Some eventual approaches (below presented) might even compel in having much wider sample. However, finding test subjects willing to participate in such experiences can prove to be quite a challenge, specially when using local-multiplayer games.
- **Simplified game mechanics.** Using a game with less complexity in mechanics or stricter gameplay branching possibilities can help narrow down the overall actions that can happen during a session; and so, giving more importance and simplifying the definition of what we can consider crucial (and therefore a metric) for a given game.

A conceptual error we might have incurred was with the linear-type correlation we were hoping for some metrics. For instances, when we studied the relation of the distance between the two players and immersion, we were conjecturing that a higher average distance value would result in a lower immersive one. What we didn’t account for was that this linearity may not exist. There may be a perfect value which may neither be too apart or too adjoining.

Creating a model to study such relations would need a more fitting approach. A Machine Learning model could be equated in relating large amounts of gathered data with reported immersion values. Doing so could increase the chances of finding these eventual patterns, but would be dependent on having a greater sample size.

Lastly, if the correlation between immersion and game metrics proved to be existent, achievable and strong, real tools could emerge from this knowledge. As an example, if a given metric proved to be aligned with immersion, a game developer could now take that into account in order to adjust a mechanic; and doing so, would consequently adjust a game’s perceived immersion. Another instance for a possible usage of this correlation would be with runtime adaptation. Where a game could monitor a player’s performance, extrapolating and adjusting (if necessary) that player’s immersion.

**ACKNOWLEDGMENT**

A huge thankful note to Prof. Carlos Martinho for supporting me, this project and Dark Things About. The availability and encouragement shown were outstanding. A special thanks to my family, friends and girlfriend. And finally, each and everyone that somehow contributed for this work.

**REFERENCES**


