Adapting presence measuring methods for game development

Korsgaard, Dannie; Nordahl, Rolf; Stenholt, Rasmus

Published in:
Proceedings of CHI 2008

Publication date:
2008

Document Version
Publisher's PDF, also known as Version of record

Adapting presence measuring methods for game development

Dannie Korsgaard
dkor05@media.aau.dk

Rolf Nordahl
rn@media.aau.dk

Rasmus Stenholt
rs@media.aau.dk

Medialogy, Aalborg University Copenhagen, Lautrupvang 15, DK-2750 Ballerup

ABSTRACT
As the game industry expresses a growing demand for effective evaluation methods, it is worth investigating if the commonly used questionnaires, with advantage, can be replaced by alternative ways of measuring user experience in interactive environments. This paper describes an experiment where an existing presence measurement method is modified for use in computer game development. 39 subjects were part of the experiment, which was designed to test applicability of the adapted presence measuring method. Besides playing a game prototype, test participants were asked to press a button when a visual signal, triggered by an in-game event, would appear on the screen in the periphery of sight. Noting how strong the signal was is assumed to infer how strong the stimuli had to be in order to break the immersive presence. The results indicated that the adapted method with observations from the test is more useful, than questionnaires, in determining immersive presence at certain events.

Author Keywords
Computer games, presence, immersion, measurement, measurement method, game development, user experience.

ACM Classification Keywords
D2.5 Software Engineering: Usability testing.

INTRODUCTION
Game development budgets have over the years reached scales comparable with those of Hollywood film productions [1]. This entails that the development of a game is of high financial risk. In the corporate world, test and evaluation methods can be valuable tools to provide the foundation for informed predictions and decisions. Applying methods, which indicate how well a game translates into a positive user experience, might help the developers make the right design decisions, in order to create a commercially successful game.

For the aforementioned reasons, test and evaluation methods concerning the user experience in games have received an increasing amount of interest nowadays. As a contribution to this discussion the authors (students at the 6th semester on the Medialogy education at AAU Copenhagen) wished to investigate whether measuring methods customized from virtual reality (VR) research can be applied in this matter.

One of the significant fields of study within VR research is to investigate how this technology makes the user feel as if he/she is a part of the presented environment, i.e. is present [4,5]. The sensation of being present is generally a very intangible umbrella concept gathering more psychological constructs within the same notation. It has therefore been hard to measure and document [2, p.1]. To overcome this problem, presence measurement methods have emerged [2, p.1], which apply different approaches in providing an indication of the amount of presence experienced. As an example studies, which use psychophysiological measurement (measures of heartbeat, muscular responses etc.) as an indicator of presence, justify this assumption by the theory that if one is present in a virtual environment the body of the person will react on stimulus in the virtual environment in the same way as in the real world [2, p.44]. Our experiment attempts to adapt such existing methods to give game developers a tool to handle the intangibility of evaluating immersive capabilities of a game experience, according to the definition of presence as immersion in [3].

MEASURING PRESENCE
Questionnaires are one of the generally accepted and common ways to measure the effects of presence [2, p.4]. They are cheap and easy to produce and work with. However the participant answers questions retrospectively, which can be a negative aspect as it relies on the memories of the participant, which may be incorrect. Even though questionnaires are easy to adapt, another issue rises when looking at the nature of computer games. They are not usually linear constructs, in the sense that the content is interactive so the experience may differ from one user to another. Therefore it can be difficult to match the questionnaires with each individual player's game experience. This is particularly problematic if the game

developers wish to evaluate specific situations, or events, that occur in their game. The last and most significant shortcoming of questionnaires is that they rely on a subjective judgment from the user. Feelings, prejudices, personal impressions etc. can make an impact on this judgment, which may be undesirable.

Adapting an alternative

In [2], a collection of several lesser known presence measurement methods is collected. These methods are not directly designed to measure presence in games, but act as a good starting point for the making of such a method, that might also cope with the disadvantages of questionnaires. One of these alternatives is the selective attention measuring method [2, p.40], which exposes users to two tasks at the same time, and then attempts to measure presence in terms of, which task is given the most attention by the user. In [4], this method was used but relied on the participants' memory to identify how attention was distributed between the two tasks, which once again is an encounter with the retrospective problem.

The interactive possibilities of computer games can be used to make one of the presented tasks into a measuring method that enables the participants to give feedback on-the-fly and thereby eliminating the occurrence of the described retrospective problem. By gradually exposing participants to a stronger stimulus, a recording of when each participant reacts on this (e.g. by pressing a physical button) could be regarded as a secondary task, which might be useful as documentation for the amount of experienced presence. The shift of attention from the game to the stimuli could be described as a break in presence [5]. The intrusive stimulus is supposed to be activated at certain events during play. These events are to be chosen by the game developers and are likely to be some of the unique selling points that the developers wish to test.

THE METHOD

The stimuli used to provoke a break in presence could be visual-, sound- or even tactile signals. The important factor is that the signal can vary in strength, from undetectable to unavoidable. The stimuli signal which is chosen for this paper is a visual signal.

The assumption is that the player during a play session will treat the area around the center of the screen as the main point of visual attention, while the outermost areas of the screen will stay in the periphery of sight. As described in [3] the conceptualization of presence as immersion should entail that the test participant would shut out the world including perception of what is in the periphery of sight. If the intrusive stimulus is presented in the outermost areas one will be able to tell when a break in presence occurs.

Practical implementation

Placing a black border (a neutral zone) around the game visuals provided by the game makes up an edge, which separates two regions (Figure 1). The gradually changing visual simulation is made by altering the color from black to gray of a variable amount of pixels, formed as a circle, in a random position within the black border on the screen. The appearance of the circle or dot of a certain size should then either go unnoticed by or work intrusively and break the sense of presence and make the participant react by pressing a particular button on the mouse. As a participant can only either press the button or not, an extraordinary procedure is needed in order to make the method sensitive to different levels of presence.

Figure 1. The screen area divided in two: The actual game area and the neutral zone, in which the dot appears. Screen shot is taken from [6].

The binary search algorithm

When tracking improvements in game development, it is not enough only to know whether the player is immersed or not. To cope with this, a search pattern was implemented, which made it possible to test for different levels of presence. The algorithm works as the following. A maximum diameter is set for the dot (typically the width of the black border) such that the dot size can vary from 0 to this maximum. Half of this value is set as the initial dot size for every event, which should be tested. Now an odd number (n) of test persons will play the game and when an event occurs a dot, of the initial size, will appear a random place in the neutral zone with the participant either reacting by pressing the key or not. The gathered binary results will now be evaluated, for each individual event, by the computer. If the majority (over 50%) reacted upon the dot of the initial size, the dot would increase, for the next test round, with half of the amount of pixels from the current size to the maximum dot size. Likewise, if the majority did not react upon the dot, the size would decrease with half of the pixels from the current size to zero. After the evaluation of the first test round, a new dot size for every event has been computed and a new round of n test persons can start playing the game with the appearance of the adapted dots. By always adjusting the dot sizes with either the upper or the lower half of the available range of untested sizes, eventually one size will be left when enough test rounds

have been evaluated. This size is assumed be equal to the general amount of immersive presence experienced at a particular event. The accuracy of this measure will improve as n tends towards infinity.

EXPERIMENT

Before the experiment it was expected that the customized method would be better suited to measure immersive capabilities of specific events than questionnaires.

Procedure

To do a proper test of the adapted measuring method a prototype of a first-person shooter zombie-survival game was developed. With the prototype as a basis, tests were conducted on 28th and 29th of May 2007 in the facilities of Aalborg University Copenhagen. A total of 39 people participated, 34 male and five female, between ages 20-30 and all full-time students. Participants were screened: prior to the test they were asked, if they enjoyed playing first-person shooters and whether they disliked violence in games. A negative answer to these questions would mean that they can not be regarded as part of the game's target group and could therefore not be included in the experiment.

Test participants were divided into two groups to test the adapted method versus questionnaires. 26 participants, 25 male and 1 female were assigned to test the game using the secondary task measuring method (adapted group), while 13 participants, 9 male and 4 female would test just using questionnaires (regular group). Before playing the game, all participants were told about the objectives in the game and what the controls were. The adapted group was given additional instructions to press a mouse button when the dot would appear. Then the participants were asked to play the game at intervals of five persons per test round (n=5), after which they were asked to fill out the questionnaire.

The questionnaires contained seven questions, formulated according to the guidelines presented by the International Society for Presence Research [7] and meant to give an indication of the over-all immersive presence during the test session. In Addition there were three questions regarding the immersion level at specific events. All of the questions could be answered using a seven-optioned Likert scale [8]. As the game prototype evolved around the concept of survival horror, the gameplay was largely about stressing the player by exposing him/her to lots of attacking enemies. The main task was to survive for five minutes while fighting off an increasing number of zombies. A set of situations which could arise in connection with this task could be for example, events with many enemies near the player combined with little ammunition, which are assumed to create a more intense experience than events with just many enemies because of the limited means (ammunition) available to deal with the enemies. With that in mind, the questions concerning events were formulated as:

"How intense was the game when you had low/medium/high ammunition and were attacked by many/moderate amount of/few zombies?"

Since immersive presence [3] is a multi-dimensional concept, a problem was encountered at this point. Obviously it would be desirable to ask about as many of these psychological dimensions as possible, but doing so for each event would result in an impractically large questionnaire, so it was decided to focus on "intensity". This is problematic since the questionnaires will not measure the full effect of immersion, but rather just a small part of it.

It was expected that the level of immersion experienced would be ascending with the amount of zombies and descending with the ammunition available. For example, moments with low ammunition and many enemies nearby was assumed to be more intense, and thereby more immersive, than moments with few zombies and lots of ammunition. During the play sessions the three events were tested by the adapted group, where dots would appear when the criteria for the three events were met. Each dot could vary from 0 to 128 pixels in diameter and to avoid confusion, a dot would only appear the first time an event occurred for each player.

RESULTS

After the test sessions, a mean value of immersive presence could be calculated from the seven over-all immersion-related questions in the questionnaires. This immersion mean and the standard deviation is displayed in the top of (Table 1) for each group. The calculated means and standard deviations from the event-based questionnaire questions are listed in the rows below the immersion mean. All the means are based on a scale where 1 is low immersion/intensity and 7 is high immersion/intensity.

Parameters	Regular	Adapted
Immersion mean	4.49 ± 0.59	4.24 ± 0.97
Intensity, low ammo and many zombies	4.92 ± 1.62	5.22 ± 1.31
Intensity, medium ammo and moderate zombies	4.58 ± 1.38	3.70 ± 0.93
Intensity, high ammo and few zombies	3.77 ± 2.05	2.43 ± 0.95

Table 1. Displaying key values gathered from the questionnaires of the regular and adapted group.

An unpaired t-test assuming equal variances between the immersion means from both groups returned $p<0.4$. The results from the customized selective attention method are shown in (Table 2) with respect to each event.

Event	Final dot size	Times evaluated	Remaining uncertainty
1. High ammo, few zombies	62 px	6	± 0 px
2. Medium ammo, moderate zombies	92 px	5	± 1 px
3. Low ammo, many zombies	24 px	4	± 3 px

Table 2. Results from the adapted measurement method.

Assuming the adapted method is valid as a measurement of immersive presence, the size of the dot should correspond to the level of immersion experienced during the given event. "Times evaluated" differs between events because some events occurred less often than others. As a limited amount of test participants were at disposal, the binary search algorithm did not manage to reach a single dot size value for all of the events. Therefore another parameter must be listed with the results. "Pixel uncertainty" refers to the maximum amount of pixels (px), which the dot size can vary at the arrived point in the search algorithm.

DISCUSSION

By first comparing the two immersion means of each group, it is surprising that they come this close to each other due to the intrusive nature of the adapted method. Those who were exposed to the adapted measuring method experienced a slightly lower immersion level. But, a t-test ($p<0.4$) indicates that the difference is insignificant and that it could in fact be a result of the standard error.

Results from the event-questions in the questionnaire (Table 1), shows that the participants from both groups answered as expected. However it can be questioned if these answers are the outcome of the same logic thinking which made up the expectation in the first place. Meaning that the questions might be formulated in a way, which induces the participants' to rely on logical thinking rather than what actually happened during the test. Results from the adapted method (Table 2) show a different pattern. Event three was expected to have the highest value, but is here assigned the lowest. At first glance these results might seem very irrational. However observations, during the test sessions, revealed that event three often got triggered when the player had no ammunition and was in a process of searching for new supplies. When the player is literally browsing the screen in order to find ammunition it could be expected that dots appearing at this point may be easier to detect. This suggests that the processes behind the results from the adapted method are more complex than first expected and that they might be giving a better indication than the questionnaires regarding the events.

As well as the adapted method is assumed to be able to measure presence at specific events, it might also be useful for getting an impression of the over-all amount of immersive presence. Calculating the average from the final dot sizes will results in a mean dot size. In this case it can be said that the general level of presence during the play sessions was at a point where it required a gray circle of 59 pixels in diameter to break it. This over-all value can serve as a mean for comparison between different stages of development or even between different computer game productions.

CONCLUSION

The questions regarding the intensity level of each event, demonstrated how questionnaires can be hard to form in a way, which gives a valid answer. It also turned out that the quantitative results from the adapted method accompanied by observations might be able to give some insights of the complex causes and effects contained within a prototype of a game and thereby assist in the improvement of the final computer game.

Based on these results it is suggested that further research and testing is conducted in determining the reliability of this and similar methods, which attempt to objectively measure the immersive capabilities of computer games.

ACKNOWLEDGMENTS

The research behind this paper was conducted in a bachelor thesis by group 602 with supervisors R. Nordahl and R. Stenholt in the Medialogy education at Aalborg University Copenhagen. Furthermore the authors would like to thank ESN and the C. W. Obel foundation for financial support.

REFERENCES

1. Rosmarin, R. An article at Forbes.com analyzing the costs in getting Gears of Wars to the store selves. http://www.forbes.com/technology/2006/12/19/ps3-xbox360-costs-tech-cx_rr_game06_1219expensivegames.html
2. Baren, J. and Ijsselsteijn, W. "Measuring Presence: A Guide to Current Measurement Approaches". (2004).
3. Lombard, M. and Ditton, T. "At the Heart of It All: The Concept of Presence". Department of Broadcasting, Telecommunications, & Mass Media Temple University. (1997).
4. Darken, Bernatovich, Lawson and Peterson. "Quantitative Measures of Presence in Virtual Environments: The Roles of Attention and Spatial Comprehension". Cyberpsychology and Behavior. (1990). Vol. 2(4) pp. 337-347.
5. Slater, M. and Steed, A. "A virtual presence counter". Presence: Teleoperators and Virtual Environments. (2000). Vol. 9(5) pp. 413-434.
6. Warsow, a free first person shooter for Windows and Linux. http://www.warsow.net
7. International Society for Presence Research (ISPR). Online document with suggested ways of measuring presence through a questionnaire. http://www.temple.edu/ispr/frame_measure_t.htm
8. Likert, R. "A Technique for the Measurement of Attitudes". Archives of Psychology. (1932). Vol. 140