Agents and Avatars: Event based analysis of competitive differences

Mikael Fodor

University of Sussex
Brighton, BN19RH, UK
mikaelfodor@yahoo.co.uk

Pejman Mirza-Babaei

UOIT
Oshawa, ON, L1H 7K4, Canada
Pejman.m@acm.org

Judith Good

University of Sussex
Brighton, BN19RH, UK
j.good@sussex.ac.uk

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

Copyright is held by the owner/author(s).

CHI PLAY 2015, October 03-07, 2015, London, United Kingdom

ACM 978-1-4503-3466-2/15/10.

<http://dx.doi.org/10.1145/2793107.2810310>

Abstract

Investigations into playing against computer and human controlled opponents have shown higher levels of arousal against human opponents. Most experiments however measure this by using post play surveys. In contrast, the study reported in this paper used physiological measurements to allow for different events during competitive play to be analysed. Twenty participants played two death matches in Call of Duty: Black Ops 2. Although participants played against an agent in both matches, participants were told that they were playing against an agent in one match, and an avatar in the other. Significant differences in participants' arousal level were found when the enemy was first encountered, though measures became more mixed as the interactions played out. The believability of the avatar may have influenced the results, as not all participants seemed convinced of the humanity of the AI. Overall, the experimental results can form a basis for further investigation into this area.

Author Keywords

Video Games; Physiological Measurements;
Competition; Agents; Avatars

ACM Classification Keywords

H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces.

Introduction

Many games are full of character representations which the player can interact with in some way or another. These games can ask the player to engage with these characters in many different ways, and these interaction methods may ask the player to compete against or cooperate with them. In-game character representations controlled by computer AI are known as Agents. Agents can be seen all over game worlds, they are easy to replicate and are given rules for behaviour, perhaps some AI or a back story. Many games also allow players to engage with characters which are controlled by humans, known as Avatars. Avatars therefore have many different characteristics to agents, and the way that players interact with them also differs [1].

A 2010 study by Lim & Reeves [2] investigated player responses when interacting with in-game agents and avatars in World of Warcraft (WoW). They looked at two conditions: competitive and co-operative. The competitive condition asked the player to fight against an opponent who they were told was either an agent or an avatar. The co-operative scenario had the player trade with an agent or an avatar. The outcome showed that players responded to avatars with stronger reactions than they do with agents. They also found that competition resulted in higher physiological arousal than co-operation with the agent or avatar. This outcome was perhaps expected, as when interacting with other humans, people rely on heuristics and biases in order to attribute explanations to people's personalities and behaviours [3]. There is an idea of people as a motivated tactician, [4] which describes how cognitive short cuts are used to make evaluations and judgements to facilitate the achievement of goals.

The sort of heuristics and biases which are involved in everyday social behaviour seem to become active when gaming becomes sociable, whereupon the gamer looks to understand the behaviour of the opponent by calling upon the number of biases and heuristics available.

In competition, people look to understand the mind of their opponent in order to understand their behaviour and defeat them, whereas with a computer opponent the aim is to understand the behaviour of the program in order to defeat it, which seems to not trigger the same responses [1]. Humans have a sophisticated mechanism for making social judgements which rely on naïve theories, and a computer opponent may not elicit the same kind of response as a human opponent [5]. Players do not attribute the same emotions and motivations to agents as they do avatars, and the difference in mental processing methods seems to be demonstrated by the arousal levels of the players.

Physiological measures allow for the researcher to look at real time data about what the player is experiencing while playing the game [6]. This means that event based analysis is available to the researcher [7], rather than just post play evaluations [8]. This would enable us to look at individual interactions with agents and avatars within a game to see how they differ in the different conditions. Galvanic skin response (GSR) measures the level of sweat on a person's skin. Sweat glands are directly connected with a person's nervous system and when a person gets aroused in any way it is reflected in an increase of sweat [9]. Arousal also triggers an increase in heart rate, so measuring heart rate can also give an indication of arousal level. However, other stimuli can also raise and lower heart rate, such as exercise or even coffee. The combination

Figure 1: Sample frame from a video recording taken during the experiment. Top left graph: GSR
Bottom Left graph: Heart Rate

of using GSR and Heart Rate though may provide a more accurate measure of a person's arousal level.

The study mentioned previously [2] which looked at agents and avatars used the WoW game. WoW is a Massively Multiplayer Online Role Playing Game, which means that players typically spend a long time playing this game, much more than other types of games. The study also looked at competitive and cooperative tasks but used completely different tasks. This means that any comparison may have been unreliable. The competitive task was a fight and the cooperative task was a trading objective. We therefore wanted to use a game which could be used for analysing different conditions without changing the task drastically. A first person shooter (FPS) game seemed like the best option available for this. By using a FPS multiplayer mode both one on one deathmatch against agents and avatars was possible and the option to make this a cooperative task was also available.

For the first experiment we focused only on the competitive experience, with the intention of future study looking at cooperation.

Experiment

This experiment was designed to compare the individual events that took place during gaming and to see how the agent and avatar conditions compare. For this we focused only on the competitive gaming scenario.

The study was designed to enable us to analyse the following areas during gameplay; starting the game, initiating combat, kills, deaths, ending the game, winning and losing. These may be experienced

differently and we wanted to see if that is reflected in the physiological measures. It was expected that players will be more aroused in the avatar condition than the agent.

Participants

20 participants aged between 18 and 30 years old, 17 males and 3 females. The level of experience ranged from those having played lots of first person shooting games to those that had played very little or none. The players were sampled on an opportunistic basis. This was done because there was no specific population that was being targeted in this study. No incentive was used for the participants.

Apparatus & Measures

Participants were asked to play the game *Call of Duty: Black Ops 2* in multiplayer mode. This game allows for one on one fighting on a controlled map. This version of COD also has bots in the multiplayer mode, which means that the player can fight one on one with the AI or human opponent. However for this experiment, we used a perceived human opponent, in other words, the player was told that they were fighting against an avatar when, in reality, they were fighting against an agent. This was done to control for behaviour differences so that the behaviour of the opponent itself is not what affects the outcome. The game was played on an Xbox 360 using the standard Xbox 360 controller.

Physiological measurements were taken using a Biopac MP-36 4-channel research system using relevant Biopac GSR and Heart Rate sensors. They connected with the disposable electrodes which were placed on the skin of the participants. For GSR, two electrodes were placed on the little and ring finger of the participant. Although

ideally these should be placed on the index and middle finger [10], these fingers were needed to use the game controller. For heart rate, three electrodes were placed on the skin of the participants: one on the right wrist, and one on either ankle.

This system was connected to the research computer which took all the data in the form of an active graph and outputted into a spreadsheet. For each participant experimental condition a video was made. This video had the onscreen action and webcam footage of the player while they played the game. This computer was located in a separate room to the participant to avoid adding any pressure to them while playing.

A series of three questionnaires were used to supplement the physiological recordings. A pre-experiment questionnaire captured personal details and experience of the participant. In order to capture the participants' evaluation of their gaming experience, they completed a questionnaire immediately following each of the two conditions

Procedure

Participants were first asked to fill in the pre-experiment questionnaire, and then had the electrodes placed on the skin and leads attached and calibrated. Once ready, the player was asked to play one on one with an opponent. The order of play was varied so that one half played the agent first and the other half played the avatar first. Only two different maps were used, these were also randomised. The player played against an agent which was set on normal difficulty. In the agent condition, they were told it was an agent and in the avatar condition they were told it was a human.

Post experiment processing of the spreadsheet data was required. Each video was watched and the data in the spreadsheet was matched up with the type of events that occurred.

Results

There were five pieces of event information that were labelled for the results: 1) Beginning, the recordings at the start of the experiment, 2) Enemy Encounters (EE), the moment that the player sees or hears the enemy, 3) Kills, when the player kills an opponent, 4) Deaths, when the player died, and 5) End, the moment the time ran out and the result appeared on screen. There was a data set for each participant, and for each of the two conditions (agent and avatar). Three participants had issues with the Heart Rate sensors which could not be fixed, so their HR data was not used. One player managed to survive the whole game without dying so no death data was recorded. Five players did not manage a kill in one of their matches so had no kill data.

Event	N	Mean		Std. Deviation	
		Agent	Avatar	Agent	Avatar
Begin GSR	20	30.24	35.10	18.38	17.96
Begin HR	17	87.44	88.46	12.03	11.81
Mean EE GSR	20	54.90	59.92	21.01	17.79
Mean EE HR	17	89.55	92.20	11.95	12.30
Mean Kill GSR	15	63.41	59.75	21.15	18.02
Mean Kill HR	12	95.48	99.97	12.45	13.18
Mean Death GSR	19	59.86	59.66	17.61	20.71
Mean Death HR	16	91.43	95.60	13.97	16.32
Mean End HR	17	87.31	90.60	12.98	12.34

Table 1: Mean average of participant physiological measures for all events

A paired samples t-test comparing the physiological data results, showed that there was a significant difference in Heart Rate when the player encountered the enemy ($t(16)=-2.119$, $p=0.025$), and when the player killed the enemy that was perceived as human, the avatar ($t(11)= -2.744$, $p=0.010$). Deaths did not seem to have the same effect. The results from the GSR data, though in the expected direction, were not significant. A paired samples T-test was used to analyse the results of the questionnaires. The only result that was significant was that of social presence.

Event	N	Mean		Std. Deviation	
		Agent	Avatar	Agent	Avatar
Valence (AI)	20	6.65	6.90	1.226	1.294
Activation (AI)	20	5.85	6.45	1.268	1.395
Dominance (AI)	20	5.20	4.90	2.238	1.774
Physical Presence (AI)	20	3.60	3.60	1.635	1.188
Spatial Presence (AI)	20	3.30	3.45	1.418	1.538
Social Presence (AI)	20	3.20	4.05	1.704	1.731
Kills	20	5.000	5.350	3.6274	3.7735
Deaths	20	4.200	3.700	2.7261	2.3864

Table 2: Mean average of participant questionnaire responses

The significance of the social presence suggests that the players mostly believed that the avatar was a real human, more so than the agent as was intended. But the effect did not seem to give a greater sense of immersion or emotional impact on the player.

Discussion

The results of the experiment seem to point to the moment when a player encounters the enemy as the one which differs the most between the agent and avatar condition. Higher levels of arousal were expected for the avatar condition over that of the

agent, in line with what has been observed in other experiments [1 and 2]. The fact that this was limited to the one event was unexpected, as it would have been expected that all events relating to the avatar would have been more arousing. It could be the case that when playing against an avatar the player is in a heightened state of readiness and that when the player sees or hears the opponent, they have a quicker response to the stimulus than when they are just looking for an agent opponent. These readings though were not reflected in the GSR data. The atypical result of this experiment could be due to a number of reasons:

The **limitations of the equipment** has an effect on what we were able to measure and how we measured it. The GSR measure though very reactive to arousal has a few issues with it. There is no base value on which we can compare the measure. GSR measures the change of sweat levels on the skin, this means it has to be calibrated and does not give a consistent measure which can be compared [10]. The readings change from one person to the next, so we had to use a calculated nominal value for statistical analysis. However, it means that comparisons from one person to the next is perhaps unreliable or inconsistent.

Another issue we encountered was regarding the **believability of the opponent**. Though the bot was almost indistinguishable from a human judging by the user interface of the game. There are things that may have alerted the player to an inhuman element to the agent masquerading as an avatar. One is the sometimes obviously inhuman behaviour of the on screen opponent, this was reported by a couple of participants who said it made them wonder. Another is

the fact that participants never actually saw a human who they thought they were playing against. Having an actor pretend they were playing against them may have helped.

Then there was the **statistical analysis**, due to the size of the data sampled a consistent method had to be used to analyse the data. This method which involved taking the highest value within the 2 seconds after an event had occurred may not have been the best way to do it. Perhaps taking the mean average in the time afterwards would have been more accurate. Using a 2 second boundary does not take into consideration the length of time of a person's arousal state. This therefore means that the choice of statistical analysis may have had an effect on the result.

Conclusion

In summary, significant differences were found for heart rate measures when players encounter the opponent and for measures of social presence. No significant effects were found for GSR measures, contrary to what was expected. Heart rate measures were also not found to be significant in situations other than the enemy encounter. Possible reasons for these results were put forward in the previous section, and we list below some possible avenues for further exploration based on the issues raised by this study:

- Exploration of the co-operative condition
- Use of more advanced measuring technology
- Use of a real human opponent
- Larger sample size of both males and females
- Different statistical analysis methods

References

1. J. Bailenson and J. Blascovich, "Avatars," in Encyclopedia of Human-Computer Interaction, Berkshire Publishing Group, 2004, pp. 64-68.
2. S. Lim and B. Reeves, "Computer agents versus avatars: Responses to interactive game characters controlled by a computer or other player," IJHCS, vol. 68, pp. 57-68, 2010.
3. S. Fiske and S. Tayler, Social Cognition: from brains to culture, New York: McGraw Hill, 2008.
4. G. Moscovitz, Social Cognition: Understanding self and others, New York: Guildford Press, 2005.
5. V. Y. Yzerbyt, G. Schadron, J. Leyens and S. Rocher, "Social judgeability: The impact of meta-informational rules on the use of stereotypes," J PERS SOC PSYCHOL, vol. 66, pp. 48-55, 1994.
6. P. Mirza-Babaei, L. E. Nacke, J. Gregory, N. Collins, and G. Fitzpatrick. 2013. How does it play better?: exploring user testing and biometric storyboards in games user research. In CHI '13. ACM, 1499-1508.
7. P. Mirza-Babaei, L. Nacke, G. Fitzpatrick, G. White, G. McAllister, and Nick Collins. 2012. Biometric storyboards: visualising game user research data. In CHI EA '12. ACM, New York, NY, USA, 2315-2320. DOI=10.1145/2212776.2223795
8. P. Mirza-Babaei, V. Zammitto, J. Niesenhaus, M. Sangin, and L. Nacke. 2013. Games user research: practice, methods, and applications. In CHI EA '13. ACM, New York, NY, USA, 3219-3222. DOI=10.1145/2468356.247965
9. N. Carlson, Physiology of Behaviour, 10th ed., Boston: Pearson, 2010.
10. M. Dawson, A. Schell and D. Filion, "The Electrodermal System," in Handbook of Psychophysiology, New York, Cambridge University Press, 2007, pp. 200-223.