Biometric Storyboards: Visualising Game User Research Data

Pejman Mirza-Babaei
University of Sussex
Brighton, BN19RH, UK
pm75@sussex.ac.uk

Graham McAllister
Player Research
Brighton, BN29TS, UK
graham@playerresearch.com

Lennart Nacke
University of Ontario Institute of Technology
Oshawa, ON L1h7K4, Canada
Lennart.Nacke@acm.org

Nick Collins
University of Sussex
Brighton, BN19RH, UK
N.Collins@sussex.ac.uk

Geraldine Fitzpatrick
Vienna University of Technology
A-1040 Vienna, Austria
Geraldine.Fitzpatrick@tuwien.ac.at

Gareth R. White
University of Sussex
Brighton, BN19RH, UK
g.white@sussex.ac.uk

CHI'12, May 5–10, 2012, Austin, Texas, USA.
ACM 978-1-4503-1016-1/12/05.

Abstract
Player experience is difficult to evaluate and report, especially using quantitative methodologies in addition to observations and interviews. One step towards tying quantitative physiological measures of player arousal to player experience reports are Biometric Storyboards (BioSt). They can visualise meaningful relationships between a player's physiological changes and game events. This paper evaluates the usefulness of BioSt to the game industry. We presented the Biometric Storyboards technique to six game developers and interviewed them about the advantages and disadvantages of this technique.

Keywords
Biometrics, Storyboards, Video Games, User experience (UX), Visualisation.

ACM Classification Keywords
H.5.2 [Information Interfaces And Presentation]: User Interfaces - *User-centred design*;

Introduction
Most classical user experience (UX) and Human-Computer Interaction (HCI) evaluation techniques do not simply map to player experience (PX) evaluation in games due to the engaging and fluid nature of games. One of the challenges in quantitative player evaluation is to be able to collect data from users (in our case

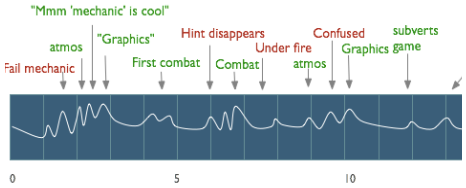

Figure 1. Excerpt from BioSt 1st iteration

Figure 2. Excerpt from BioSt 2nd iteration

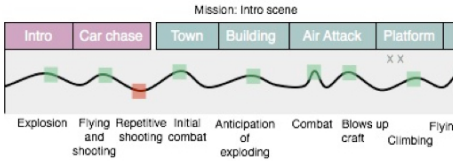

Figure 3. Excerpt from BioSt 3rd iteration

players) without interrupting their gameplay (continuous and unconscious). Since games also thrive on emotional experiences, physiological evaluation is becoming a more popular method together with traditional interviews for player evaluation [3, 5, 9].

Related Literature

Game user research concentrates on evaluating player experience and has borrowed methods from HCI and psychology. Physiological methods (biometrics) are now being integrated in game user research [2, 6] and game development in the game industry [1]. Common physiological measures include Galvanic Skin Response (GSR), facial muscle measures (EMG), cardiac interbeat intervals (IBIs), and Electroencephalography (EEG) [4, 6, 10]. Common approaches distinguish physiological analysis on a temporal dimension: Studying phasic psychophysiological and behavioural responses at game events (points in time) [12] and studying tonic responses to variations in-game variables (time span) [5].

Visualisation is a continuously growing area, with research efforts expanding into many different domains. Visualisation tools address the challenge of analysing and presenting overwhelming amounts of data. On the other side, narratives have always been part of the UX process to communicate how and why a design would work [11]. Storyboards have become popular techniques for visualising interaction, not only in design education, but also in design practice.

Research Topic

One of the challenges is making the interpretation of physiological and player evaluation data meaningful in terms of facilitating design decisions for developers. Steps in this direction are necessary to facilitate the interpretation of these large datasets, possibly creating visual aids for faster navigation and easier interpretation of physiological game engagement. To meet this need we are developing a player evaluation approach called *Biometric Storyboards* (BioSt). The aim of this paper is to evaluate the usefulness of this approach for the game industry.

Iterations of Biometric Storyboards

So far our BioSt design went through three iterations based on three case studies with two game publishers and the feedback from two producers where they used this technique on their unreleased game. The graph itself is drawn based on (1) a player's physiological arousal signals (GSR), (2) player post-session interviews to explain 'why' the change in their GSR occurred [6], (3) players' self-drawn diagrams of their gameplay experience and (4) Observation of player gameplay behaviour (or context). These datasets and the three iterations have been explained in our previous publications [7, 8]. As an introduction to this paper we provide an overview of each:

The first version was divided by time (Fig. 1). Each vertical block is one minute of gameplay, positive comments are in green and negative are in red. The feedback suggested that this design of BioSt was difficult to compare between players. Time is not always meaningful for some games, and beats (or thematic areas) were considered more representative, as in version 2 (Fig. 2).

The main differences in the 2nd design are: 1) each level was divided into thematic areas, this would make the key sections easier to compare; it also shows the

time it took the player to complete that area. 2) Green or red dots shows the positive or negative experience.

The current version (Fig. 3) makes the diagram easier to read and couples behaviour (the text along the bottom) with the associated player experience. This iteration of BioSt was refined based on further comments from the games' producers. Their comments suggested that the experience graph should go down (negative gradient) to indicate negative player experiences to better represent the emotional change, and to better draw attention to and isolate the negative experiences. Secondly, they reported difficulty in pinpointing the exact moments highlighted by the red/green dots, which were key to providing context and establish cause and effect.

The next section reports our further evaluation by taking our three iterations to six game developers who had not seen BioSt before.

Method

We conducted six semi-structured interviews in order to evaluate our three iterations and better understand our target group requirements for future development of the technique. We interviewed six game development professionals (P1: Lead Designer, P2: Creative Director, P3: Designer, P4: Programmer, P5: Animation Designer, P6: Game Director) from midsize UK game design studios. None of the interviewees have seen BioSt before. Each interview took about 30 minutes. Before the interview the participants signed a consent form.

We started the interview by getting their thoughts on user testing (UT) and user test reports. For example, we asked them about their overall experience with UT, what they were hoping for from the report, about its format and how findings were presented (communication). After this opening discussion on UT and reports, we then showed them BioSt. We discussed advantages and disadvantages of this technique while they saw all three iterations.

Results

The interviews suggest the following advantages and possible improvements for BioSt:

- At a glance summary
- Objective credibility
- Location and prioritising of gameplay issues
- Identifying a problem/suggesting a solution
- Clarity/simplicity
- Facilitates the discussion
- Trust/convincing
- Comparison to intended experience

All interviewees had read UT reports before and some of them were involved in conducting UTs in their studio. The main values of UT sessions for P6 are to see: 1) areas of frustration, 2) areas that are difficult to pass (blockers), 3) if the players are having fun, 4) if players understand the game and are using all the game features. They mentioned a need for more visualised data in UT reports. An ideal report would be a process to capture a massive UT data and report it in a way that it is easy to make sense of. For example P2 mentioned in their previous title (a racing game) that they collected game metrics to generate a crash heatmap of each track. He added: "from heatmaps we could see the crashes, but we know they can lead to different experience. Some of them lead to enjoyment

and some lead to frustration, the heatmaps won't show this difference, […] BioSt is somewhere between only seeing the heatmaps and talking to the actual players." [P2] The interviewees suggested the text reports usually cover most of the information they need but "for some issues you won't feel the text report can put them in a right context and time line. For example when interpreting from a report there is no way to see the change of pace and enjoyment." [P6]

The report summary is the section they all read and found most useful. All of the interviewees think BioSt shows an **at a glance summary** of a level.

P6 believes: "using biometrics lends more **credibility** to BioSt. It gives the perception that it includes data that goes beyond what players say but what they feel, stuff they don't realise to vocalise it." This would suggest using biometric measures as one of the data sets for creating BioSt can lend more credibility to it.

Location of issues in each level is an important factor for **prioritising** to fix them. By showing where exactly the issues occurred, BioSt may help developers to prioritise what to fix. "BioSt allows me to see where my good and worst parts are, it helps me to prioritise what to fix." [P6], "this [BioSt] shows me pretty much negative experience happening at the beginning of the game/level, it's concerning." [P1]

The developers don't want UT reports to **suggest a solution** on how to fix an issue. "I just need to know **where the problems are** and how much of the problem it was." [P5]. For them the ideal report would be combination of text to explain what is the problem and short gameplay video of player experiencing the problem. BioSt can show them where the problems are and also by visualising the relationship between issues it may assist developers to come up with a possible solution.

As a visualising tool it is critical that the developers be able to correctly interpret the data from BioSt. P5 said "this is vey **clear**, easy to see the different sections. It is difficult to contrive this to anything else."

Our interviewees were from different positions in game studios, yet they all felt that BioSt would be helpful for them. They also mentioned BioSt would provide a useful information to publishers and studio executives. However P6 mentioned "the issues are usually not small enough to be actionable by single person but this [BioSt] can be use with the whole team to **facilitate the discussion** over a level design." P3, P4 and P5 mentioned that they want to see more data such as players' comments and gameplay video in each indicated events. P3 added: "my view is from a designer's perspective, where we are eager to go into details, like user comments, to see what this guy said about this bit."

Trust is a vital matter for UT reports, "if the designers do not trust the data the problems will stay." [P2]. The interviewees suggested that the most **convincing** case is when the designers personally attend UT sessions and have a face-to-face conversation with players or watch the gameplay video. As for content of UT report P3 said: "It will be wrong if we ignore any statements, but we act on it if many people say same thing is wrong." Specific to BioSt, P2 mentioned: "there are two ways I can trust it, one is for me to totally understand how it is generated, or [second] to see enough

correlation between the results." This is explored further in the discussion section.

The interviewees mentioned they use some sort of storyboards depending on the game they are developing. For example P2 explained that for their previous game they draw a graph of intended emotional states for players. BioSt could make it easier for developers to be able to compare the player's experience with the **experience they intended** to design for. "If BioSt can show the accurate match to our intentional graph that would be a fantastic tool."

Discussion and Future work

This in progress study aims to improve on the following areas for our future work:

Iterations: after seeing all the three iterations the developers overall feedback on them was:

- All interviewees preferred iteration 3.
- Positive feedback for adding level areas in iterations 2 and 3.
- Positive feedback for having graph annotation and area descriptions in iteration 3.
- Negative comments on iteration 2 as they experienced problem with finding area and arousal explanations. This was fixed for the 3rd iteration.
- Negative feedback for removing time player spent in each area from iteration 3. We will bring this back.

Composite graph instead of individual: Our current design of BioSt visualises how each individual player experienced a game. Based on our interview results this can lead to two problems: 1) Too many individual graphs for developers to look at and 2) Showing how one individual player experiences a game does not convince developers to act on the issues. In order to improve these we need to generalise the individual graphs into a composite graph, showing the correlation of results among players. Our results suggested that the BioSt would be a useful tool if it shows a reasonable correlation between the results of individual players.

Severity: The developers want a tool to help them to prioritise the issues to fix. While BioSt facilitates this in some ways (e.g., location of issue) it could also indicate the severity of each issue, for example this can be achieved by measuring the mean value of GSR arousal among participants at the specific event.

Interactive: The developers want BioSt to be interactive. For example, enabling a mouseover on each point to get the description or to click on each area to zoom in and see comments from different players, or watch a clip of their gameplay video indicating specific problem the developer looking at.

Graph comparison: Developers want to see their intended experience graph in the BioSt, or a graph that can show the difference between intended and actual experience. This would be possible if the developers would be willing to work closely with user researchers and provide them with their intended emotional graph.

Measurement of different experience: Our interviewees suggested the current design just shows green and red points, as high and low arousal experiences. Since these do not actually depict emotional valence, we will add EMG. Adding other

sensors (such as EMG, EEG) will allow approximating wider types of experience.

Conclusion

Our interviews suggest that Biometric Storyboards can be used as a tool to enable discussion. They are easily understandable and use neutral language so that programmers, designers, artists, and producers can all quickly pinpoint areas of the game that are working and those that need refining. The interview results also highlighted areas for further development in future study. The immediate development would be 1) generate a composite graph, 2) measure the mean GSR to indicate issue severity and 3) including new sensors to measure different responses.

Although we developed BioSt on character-based games, the interviewees mentioned they could see how BioSt would benefit them in their current titles including platform, educational, racing and musical games.

Acknowledgements

We would like to thank Mark Knowles and our interviewees for their contributions to this paper.

Reference

[1] Ambinder, M. Biofeedback in Gameplay: How Valve Measures Physiology to Enhance Gaming Experience. Game Developers Conference Vault (2011).

[2] Drachen, A., Nacke, E.L., Yannakakis, G., and Lee Pedersen, A. Correlation between heart rate, electrodermal activity and player experience in first-person shooter games. In *Proc. the 5th ACM SIGGRAPH Symposium on Video Games* (Sandbox '10), ACM, New York, NY, USA, 49-54. DOI=10.1145/1836135.1836143

[3] Hazlett, R. Using Biometric Measurement to Help Develop Emotionally Compelling Games. In Game Usability: *Advancing the Player Experience.* Morgan Kaufmann (2008), 187-205.

[4] Kivikangas, M., Ekman, I., Chanel, G., Järvelä, S., Cowley, B., Salminen, M., Henttonen, P., and Ravaja, N. Review on psychophysiological methods in game research.In *Proc. Nordic DiGRA* (2010).

[5] Mandryk, R., and Atkins, M. A Fuzzy Physiological Approach for Continuously Modeling Emotion During Interaction with Play Environments. *Int. Journal of Human-Computer Studies* 65, 4 (2007), 329--47.

[6] Mirza-Babaei, P., Long, S., Foley, E., and McAllister, G. Understanding the Contribution of Biometrics to Games User Research. *Proc. DIGRA* (2011).

[7] Mirza-babaei, P., and McAllister, G. Biometric Storyboards: visualising meaningful gameplay events. *CHI 2011 BBI workshop* (2011).

[8] Mirza-babaei, P., and McAllister, G. Biometric Storyboards to Improve Understanding of the Players' Gameplay Experience. In *Proc. Videogame Cultures and the Future of Interactive Entertainment* (2011).

[9] Nacke, L., and Lindley, A.C. 2008. Flow and immersion in first-person shooters: measuring the player's gameplay experience. In *Proc. of the 2008 Conference on Future Play*. ACM, New York, NY, USA, 81-88. DOI=10.1145/1496984.1496998

[10] Nacke, L. Directions in Physiological Game Evaluation and Interaction. *CHI 2011 BBI workshop* (2011).

[11] Quesenbery, W., and Brooks K., *Storytelling for User Experience*. Louis Rosenfeld, (2010).

[12] Ravaja, N., Turpeinen, M., Saari, T., Puttonen, S., and Keltikangas-Järvinen, L. The Psychophysiology of James Bond: Phasic Emotional Responses to Violent Video Game Events. *Emotion 8, 1* (2008), 114--120.