Understanding the Contribution of Biometrics to Games User Research

Pejman Mirza-babaei
University of Sussex
Brighton, UK BN1 9QH
pm75@sussex.ac.uk

Sebastian Long, Emma Foley, Graham McAllister
University of Sussex
Brighton, UK BN1 9QH
swl20@sussex.ac.uk, ef66@sussex.ac.uk, g.mcallister@sussex.ac.uk

ABSTRACT
Utilising biometric data has become an increasingly active area in the video games user research community, and a number of academic papers have been published introducing various biometric based analysis techniques in video games research. This paper aims to quantify the value of biometric methods as an addition to traditional observation-based user research methodologies, and their respective contributions to the production of formative feedback during the development of video games. Our results show that observation-based techniques can expose the majority of issues relating to usability, however the biometrics-based approach enabled researchers to discover latent issues in related to players' feelings, immersion and gameplay experience and, in certain categories of issue, reveal up to 63% more issues than observation alone.

Keywords
Biometrics, Evaluation, Video Games, Usability, User Experience (UX), Playtesting.

INTRODUCTION
Within the rapidly growing video game market, researchers are using emerging technologies to optimise players' gameplay experiences. The wide variety of video games make them a popular type of entertainment for a broad range of consumer groups. The interest of this study is to identify the strengths, weaknesses and qualitative differences between the findings of a biometrics-based, event logging approach and the results of a full, observation-based user test study. By gaining an understanding into the contribution of biometric measures to games user research, we aim to explore the integration of methods from across the traditional qualitative/quantitative divide.

Although Human-Computer Interaction (HCI) methods have made progress in understanding the usability of productivity applications and websites, the specific characteristics of video games, such as the addition of 'intentional challenge and emotion', mean that many established methods of user research cannot be applied in the same way. Current methods of evaluating video games user experience (UX) commonly include subjective self-reports through questionnaires, interviews, and focus groups (Fulton & Medlock, 2003) as well as objective reports through observational video

Proceedings of DiGRA 2011 Conference: Think Design Play.

analysis (Lazzaro, 2004). Many of these methods have limitations when applied to video games.

Traditional User Research Methods

Observation

Behavioural observation logs are industry's strongest analysis tool in video games user research. They can provide a basis for a detailed analysis of usability (Pagulayan et al 2002), and fun and game experience (Poels et al 2007). Observation involves watching the player interact with the game and picking up cues from their facial expressions and body language. A major benefit to using observation sessions is that they are relatively easy to conduct, and can potentially provide a rich source of data. However, whilst analysis can be performed 'live', understanding behaviour requires precise interpretation and, unless the video data is captured and reviewed, important events can be missed by researchers. Studying observational data as an indication of human experience is a lengthy and difficult process which must be undertaken with great care to avoid biasing the result (Marshall & Rossman, 1999).

Think-aloud

A commonly used extension to observation is think-aloud or verbal reporting which involves the player describing their actions, feelings and motivations during gameplay. The aim is to get inside the players' thinking processes 'in the moment', potentially revealing unobservable details and providing researchers with immediate feedback. However, it is unprompted and many participants find it unnatural, which can affect the gameplay experience. Furthermore, if the timing aspect of the game is integral to the game mechanic, then such talking will affect this. It is argued that 'think-aloud' techniques cannot effectively be used within game testing sessions because of the disturbance to the player and ultimately the impact they have on game play (Nielsen, 1992).

Heuristics evaluation

Heuristic evaluation provides a formal and accessible usability evaluation method, which can be used even before any code written. There are a number of different heuristic sets created for video game usability evaluation, including PLAY by Desurvire et al (2009), Federoff (2002), Nielsen (1994) and Pinelle et al (2008). Although heuristic evaluation promises to be a low-cost usability evaluation method, it is suffers with problems of subjective interpretation (White et al 2011).

Interviews and Questionnaires

Interviews and questionnaires are frequently used to analyse player experience. They are generalisable, convenient, and amenable to rapid statistical analysis, yet they only generate data when a question is asked, and interrupting gameplay to ask a question is disruptive (Mandryk et al 2006). When users provide information after the playtest, rather than continuously throughout its course, their responses reflect the finished experience and therefore important issues may not be identified (Mandryk & Atkins, 2007). Players will often not recall motives for their actions, potentially leading to post rationalisation. These weaknesses can be addressed to an extent by recording the game video and replaying sections of interest in order to facilitate recall (Gow et al 2010). However, similar to other video analysis techniques, this is highly time consuming and would be less practical for a longer playtest session. Researchers have developed innovative techniques for effective post-session interview video playback, the benefits and challenges of which have been discussed by Mirza-babaei & McAllister (2010) and Kivikangas et al (2011).

A number of industry-standard usability approaches use these traditional methods in combination; for example, RITE (Rapid Iterative Testing and Evaluation) (Medlock et al 2002) which employs observation and think-aloud techniques with the addition of an attending software engineer to rapidly alter the design, based on the findings of the usability testing. Changes can be made after observing as few as one participant, with altered designs subsequently tested on the remaining participants. Other variations include open-ended usability tasks, paper prototypes, and empirical guideline documents (Pagulayan et al 2002).

Biometrics in Video Game HCI

In partial response to the weaknesses of typical HCI methods for video game user research, biometric technology is being increasingly used to enable UX researchers to capture physiological measurements for analysis. This has been discussed with relation to traditional methodologies at a Future Play panel (Nacke et al 2009).

Psychologists use physiological measures to differentiate between human emotions such as anger, grief and sadness (Ekman et al 1983). Some researchers in UX have used physiological measurements to evaluate emotional experience in video games. Hazlett (2008) describes the use of facial Electromyography (EMG) as a measure of positive and negative emotional valence during interactive experiences. Ravaja (2006) measured facial EMG and cardiac interbeat intervals (IBIs) in addition to self-report ratings to index physiological arousal and emotional valence. Mandryk et al (2006, 2007, 2008) described experiments designed to test the efficiency of physiological measures as evaluators of collaborative entertainment technologies by examining physiological responses to different interactive play environments. Nacke and Lindley (2008) created a real-time emotional profile (flow and immersion) of gameplay by measuring Electroencephalography (EEG), Heart Rate (HR), EMG, Galvanic Skin Response (GSR) and using participant eye-tracking. Their results demonstrate correlation between subjective and objective indicators of gameplay experience (Nacke et al 2008), showing the potential to provide real-time emotional profiles of gameplay that may be correlated with self-reported subjective descriptions (Nacke, 2009). In their work with Grimshaw they looked at the effects of sound and music in a video game on players' EMG and GSR (Nacke et al 2010). The study by Yannakakis et al (2008) statistically correlated psychophysiological and subjective measures of emotional components of player experience.

Some researchers have used event-based biometric analysis to construct a player's emotional profile. Nacke et al (2008) created an automated system that allows reporting of phasic psychophysiological responses at game events. Ravaja et al (2006) assess specific game events based on different or contradictory physiological responses trigger by the game events. Mirza-babaei & McAllister (2011) have used player biometric responses, in conjunction with player self-reporting and structured post-session interview, to visualise player experience in game events as a Biometric Storyboard.

Cacioppo et al (2007) show that using a response profile for a set of physiological measures could enable researchers to perform a more detailed analysis, and that it allows response profiles and psychological events to be correlated. Tognetti et al (2010) have used physiological data to recognise user enjoyment in a car racing game. Drachen et al

(2010) reported a case study on GSR and HR correlations with player gameplay experience in a First-Person Shooter (FPS) game.

Generally, researchers using biometric approaches may find it difficult to match the obtained quantitative data to the player's emotional experience during play (Gow et al 2010). It is also possible to consider that a player was emotionally aroused not because of specific in-game elements but as a response to an external activity, anticipation, or as a result of something not otherwise observed (Gow et al 2010). Moreover, specific types of measurement of different responses (such as GSR, EMG, HR and EEG) are not trustworthy signs of well-characterised feeling (Cacioppo, 2007). The often described 'many-to-one' relationship between psychological processing and physiological response (Cacioppo et al 2007) allows for physiological measures to be linked to a number of psychological structures, for example; attention or emotion. Ambinder states; "Some responses or measurements are difficult to correlate with something specific that happened in the game" (cited in Onyett, 2009).

Physiological measures are continuous and involuntary, thereby revealing players' true physiological state for analysis, removing the subjectivity and some need for interpretation by researchers. A recent review of the current state of physiological game research has been provided by Nacke (2011) and Kivikangas et al (2010).

Galvanic Skin Response

Of the biometric measures adapted for use in video games user research analysis, we chose to collect Galvanic Skin Response (GSR) as it provides both low invasiveness and ease of use. Through the use of small, finger-mounted sensors, measures of a participants' arousal can be recorded and visualised in real time, which is therefore suited to both during and post-gameplay analysis. However, the use of a game control pad may introduce movement artefacts, particularly during circumstances in which the sensors are not attached securely.

Arousal is commonly measured using GSR or 'skin conductance' (Lang et al 1993). The conductance of the skin is directly related to the production of sweat in the eccrine sweat glands, which act as variable resistors on the surface. As sweat rises in a particular gland, the resistance of that gland decreases (Stern et al 2001) and (Boucsein, 1992). GSR measures this difference, even though the sweat may not reach the surface of the skin and subjects do not have to be sweating to see difference in GSR. Galvanic skin response has a linear correlation to arousal (Lang, 1995) and reflects non-specific emotional responses as well as cognitive activity (Boucsein, 1992) and (Shi et al 2007).

METHOD

A series of experiments were performed in a dedicated game user research laboratory by three evaluators with professional and academic experience in conducting and analysing video game playtest sessions. In this study, two usability and player experience evaluations were conducted separately on identical gameplay footage. A lightweight biometric-based experiment was conducted 'live' during playtest sessions, and a typical observation-based approach was conducted using a dual-expert post-gameplay analysis on recorded video footage of the same session.

The Game

Participants played the first two levels of 'Call of Duty: Modern Warfare 2' (MW2), developed by Infinity Ward in 2009 (Call of Duty: Modern Warfare 2, 2009) and 'Haze', developed by Free Radical Design in 2008 (Haze, 2008). Both games are First Person Shooters (FPS) with Metacritic review scores of 94% for MW2 and 55% for Haze (Metacritic, 2010). These titles were chosen to apply the experimental methodologies to games of different quality, in order to increase the probability of players experiencing a disparate number and type of usability and user experience issues.

The Participants

Potential participants filled out a background questionnaire, which was used to gather information on their previous video game experience, game preference, console exposure and personal statistics such as age. Participants were recruited carefully from this list, ensuring they were casual PC or console gamers, with no previous experience of MW2 or Haze. Six male students, aged 20 to 31, were selected and participated with full informed consent and with no monetary incentive for their involvement.

The Playroom

The experiment was conducted in our game testing laboratory, equipped with a Sony PlayStation 3, a Sony 40" flat screen TV, a Sennheiser wireless microphone to capture participants' verbal comments, a Sony Handycam video camera to capture the player's face, and a BIOPAC system to capture physiological data. All data is synchronized in a single screen for live viewing (Figure 1) in isolated observation room, and also captured for later analysis at the observation-based evaluation stage. The playroom is modelled in the style of a typical living room and participants were seated approximately two meters from the TV and camera.

Figure 1: Example screenshot of the gameplay video.

GSR Recording

GSR data was gathered using the BIOPAC hardware system, sensors and software from BIOPAC Systems Inc. This was measured by using two passive SS3LA BIOPAC electrodes. The electrode pellets were filled with TD- 246 skin conductance electrode gel and attached to the ring and little fingers of the participant's left hand (Figure 2).

Figure 2: Location of GSR sensors on player's left hand.

Experimental Process

After a brief description of the experimental procedure, participants were fitted with physiological sensors (Figure 2) and relaxed for five minutes. Half of the participants were selected at random to play the first and second levels of MW2 on the normal difficulty mode, the remaining half played the first and second levels of Haze with the same difficulty settings. Both games were played on the Sony PlayStation 3 platform.

Analysis Methodologies

The user testing data was subjected to analysis by two approaches, a biometric-based approach and an observation-based approach.

Biometric-Based Approach

In this study we applied a lightweight biometric method, which does not attempt to interpret player emotion, instead using measures of players' phasic physiological data purely to log 'micro-events' on a per-individual basis. These specific moments, identified by peaks in the monitored GSR levels, were noted during the playtest, constructing a log of times during gameplay in which a usability or player experience issue may have been expressed. Micro-events were not analysed or interpreted at the logging stage; and at no time were individual participant's GSR measurements compared to other players, and no numeric analysis of the biometric data was undertaken. Instead, after the gameplay session, the gameplay video footage related to every logged micro-event was played back to players, who were asked to recall these specific moments and inform the experimenter of their thoughts. For example, a biometric micro-event during the use of an all-terrain vehicle in Haze was logged. In post-playtest interview the player was asked "*Can you explain what happened here?*", and the players response was noted and analysed. In this example the player responded: "*I was not sure if I could still drive my buggy or if it was broken. I've started driving it again, but was not sure if it was going to explode soon or not. Eventually, it did*", indicating a usability issue concerning the 'health' of the vehicle. All logged micro-events were addressed in this manner, with usability and player experience issues determined by the players' interpretation of their biological response.

In order to isolate the biometrically-determined findings of this approach, only usability or user experience issues indicated by the presence of players' biometric arousal and confirmed by the player at the post-session interview were classed as findings. Any other issues noted by the player conversationally during the post-session interview were not included. The process for logging micro-events used only the live feed of the gameplay, and did not involve the review of video footage. Players were monitored remotely by a single evaluator, and as the number of micro-events expressed by the player's GSR is of a fixed number, including more evaluators would not affect the number of findings or their content.

Only negative usability or user experience issues as defined by the player in structured post-session interview were classed as findings. GSR arousal can be indicative of positive gameplay experiences, and number of the micro-events expressed by the players were explained as positive experiences. Positive findings were not included in the analysis of this study; the reasons for this are discussed in the discussion section.

Observation-Based Approach

In this observation-based approach, two evaluators analysed the same gameplay footage that was viewed and recorded during the biometric-based approach. Each evaluator watched and analysed all recorded gameplay videos individually, noting usability and UX issues in a 'double-expert' approach. Biometric readings were not taken into consideration in this post-gameplay analysis. Once each of the evaluators had completed the analysis of each gameplay video, their findings were collated and summarised with identical issues combined, providing a single list of findings. These issues are considered to be representative of those that could be found by the observation-based approach universally, both in content and quantity. For example, a player became lost when attempting to follow a comrade as instructed in MW2, and ended up doubling back through several rooms in their confusion. The player's body language also reflected frustration. This behaviour indicated a usability issue with the location marker prompt to "Follow", which was not visible from the player's original position. Negative usability or user experience issues that were identified from player comments during gameplay were included in the results of this approach, reflecting the content of a typical observation-based methodology.

RESULTS

From the total of 89 issues found, 29 (32.6%) were identified by both approaches. Observation-based user testing established 34 issues (38.2%) that the biometric-based approach did not. Using the biometric-based approach, 26 issues were revealed that were not found in the observation-based user testing methods (29.2%). A total of 58 issues were identified in Haze, with the remaining 31 issues identified in MW2.

In order to gain a better understanding of the nature of the findings, issues were sorted into three categories (CAT1: Gameplay, CAT2: Emotion\Immersion and CAT3: Usability) allowing the strengths and weaknesses of the two approaches to be attributed to certain categories of usability or UX issue. These categories were obtained from Desurvire et al (2009) and are provided in Table 1.

Category 1: Game Play	
1.1 Enduring Play	1.2 Challenge
1.3 Strategy and Pace	1.4 Consistency in Game World
1.5 Variety of Players and Game Styles	1.6 Players Perception of Control
1.7 Goals	
Category 2: Coolness/Entertainment/Humour/Emotional Immersion	
2.1 Emotional Connection	2.2 Coolness/ Entertainment
2.3 Humour	2.4 Immersion
Category 3: Usability & Game Mechanics	
3.1 Documentation/Tutorial	3.2 Status and Score
3.3 Game Providing Feedback	3.4 Terminology
3.5 Burden On Player	3.6 Screen Layout
3.7 Navigation	3.8 Error Prevention
3.9 Game Story Immersion	

Table 1: Issue categories obtained from Desurvire et al (2009)

The categorised observations are shown in figure 3. It is clear that the majority of CAT3 issues were revealed by the observation-based approach, whereas for CAT1 and CAT2 issues, the majority were revealed by the biometric-based approach. This observation is supported by the fact that a chi-square test on the frequency of observations with categories (1, 2 & 3) and approaches (biometric-based, observation-based or both) as factors was highly significant – X^2 (8, N=89) = 26.7, p < .01). This means there is a relationship between approaches and categories that goes beyond what would be expected by chance alone. Chi-square does not describe a relationship, instead it has to be interpreted from the data. It would seem sensible to conclude that the relationship here is CAT1 and CAT2 issues are better revealed by biometrics-based approach than CAT3 issues. This will be discussed further here and in the discussion section.

Through categorisation of the results, it is clear that observation-based user testing revealed a significant number of those issues in CAT 3, usability and game mechanics (90.4%). There was an overlap of 40.4% where those issues were also indicated by the biometric approach, with 9.6% of issues being identified only by the biometric approach in this category. Issues in CAT1 and CAT2, concerning players' feelings, immersion and gameplay experience, were more separated. The majority of issues found in these two categories were only indicated by the biometric approach (53.8% in CAT1 and 63.6% in CAT2). Observation-based user testing was less effective than the biometric approach (15.4% for CAT1 and 36.4% for CAT2). 30.8% of the CAT1 issues were found by both methods but there was no overlap in the CAT2 issues. Figure 3 shows breakdowns of issues found in MW2 and Haze in each category.

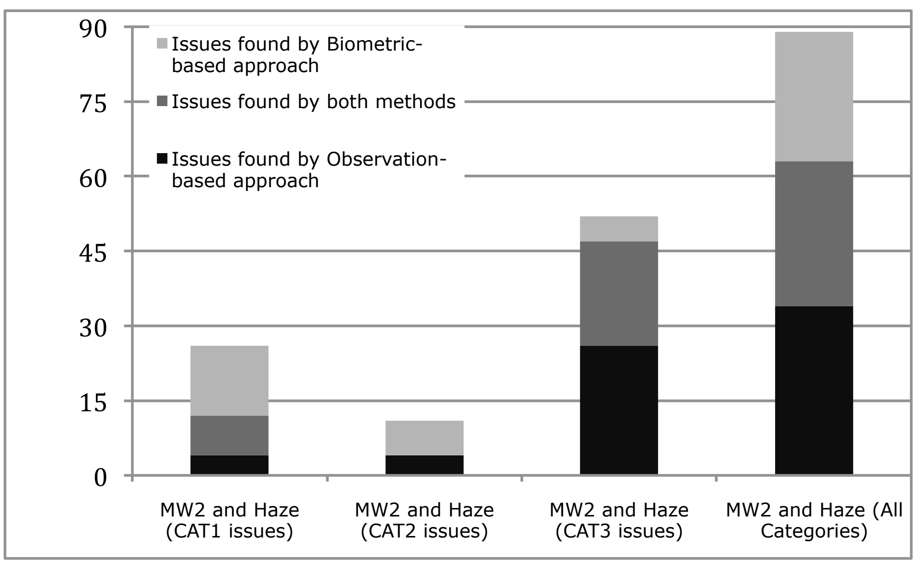

Figure 3: Comparison of number of issues.

DISCUSSION

Overall, observation-based user testing methods distinguished a greater number of issues, however, as literature has suggested, GSR provides only a measure of player arousal, which may not provide a representation of the full player emotion spectrum. Further research into the number of latent usability issues uncovered using differing biometric sensors may suggest that specific sensors, or sensors used in combination, can reveal a yet greater number of issues.

Issue Quantity

The biometrics-based approach revealed a significant number of gameplay issues, many of which were not identified through the observation-based method alone. The results demonstrate the important role that GSR, and potentially other types of biometric measurement, may play in conducting a thorough analysis of video games. Observational methods alone found the majority (90.4%) of issues in CAT3, but just 15.4% of those in CAT1. The addition of just one biometric measure increased the number of findings significantly, providing a valuable contribution to the analysis.

Issue Category

The results also indicate that there is a difference in the type of issue that each of the approaches could reveal. Observation-based techniques can expose the majority of issues relating to usability (CAT3), however the biometric-based approach enabled researchers to discover many more issues in categories related to players' feelings, immersion and gameplay experience (CAT1 and CAT2).

Methodology

This study considered six players with 75 minutes gameplay per player over two video game titles, which can be deemed a reasonable sample size from which valid conclusions can be drawn. Ideally this work would be extended to include video game titles of differing genres, to further investigate the contribution of biometrics across game types. Whilst the study could be conducted with a greater number of participants, the post-session video analysis is highly labour-intensive and therefore a significantly larger sample would be impractical.

During the biometric-based approach, when prompted to recall elements of their gameplay experience, participants were infrequently not able to recall their thoughts or the circumstances shown to them. The experimenter was able to replay more of the gameplay footage if the player was not able to remember the particular moment. If the additional footage was not enough to facilitate recall the experimenter progressed to the following micro-event. These events were excluded from the findings of the biometric-based approach.

The Use of GSR

The study facilitated only the use of GSR, despite the ease-of-access to further biometric measures. The low-intrusiveness of the electrode pellets when connected to the ring and little fingers allows the participant to quickly forget the presence of the sensors and does not severely impact the validity of the experiment. This is especially relevant when considering comparisons to EEG, which, for a reliable reading, would requires 16-32 electrodes attached to the scalp; and facial EMG, which requires electrodes to be adhered to the participants' faces. The sole use of GSR allowed us to minimise participant intrusion.

GSR provides a reliable source of data, and a fast response rate, reflecting the participants' arousal measurements in 2-5 seconds from the triggering video game event (Lang et al 1993). Measurements of GSR are therefore highly suited to the live logging and also the live video capture procedures demonstrated in this study. GSR also provides a data format that is easily analysed, with clear indications of micro-events visualised from the raw data itself, and requiring no post-session review.

The use of finger-mounted sensors does introduce the problem of movement-induced signal artefacts. Throughout the experiment, a limited number of micro-events (as a result of arousal in biometric reading) were explained as signal noise, due to changing sitting position or stretching of the hands, and therefore did not reflect players biological responses to the game. These events were acknowledged during the event-logging process, and were not shown to the player during the explanatory phase. If a small movement went unnoticed during the logging process, either the video of the player captured during the gameplay session revealed the movement, or players reported that they could not recall the particular moment when prompted, and the experimenter proceeded to the following micro-event. The use of foot-mounted GSR sensors, or the application of an alternative biometric sensor that does not hinder the use of the hands, such as facial EMG or EEG, would reduce the number of movement-induced signal artefacts.

This study begins to highlight categories of player experience and usability issue types which evoke arousal in players' GSR levels, but there are issues discovered by the observation-based approach which remained undiscovered by GSR alone. Further research into the contribution of biometrics using differing biometric sensors, including those related to valence (such as EMG), may allow more of the player emotion spectrum to be represented, which may reveal latent usability issues.

Further Applications

The issues common to the observation-based approach and the biometric approach demonstrate the usefulness of biometrics as a validation tool. An equivocal usability or UX issue can be validated and confirmed by the presence of a biological response and player-reported confirmation of the problem.

Of the biometric findings, players explained many of the micro-events as positive gameplay experiences, but these positive findings have not been included in the analysis for this study. The focus of this study was solely upon the negative issues, since negative usability or UX issues in video game titles are of particular interest to games developers for improvement purposes. Revealing positive events in video games under analysis may provide valuable feedback to game development companies, allowing them to quickly and accurately understand successful elements of their game. Biometrics have been used extensively to identify scenes of high arousal (e.g. Drachen et al 2010); further study may reveal biometrics as an efficient tool for the analysis or validation of positive gameplay experiences. Indicating positive game events may therefore be considered to be as useful as finding negative issues and would further contribute to the understanding of biometrics in games user research.

CONCLUSION AND FURTHER RESEARCH

In conclusion, our study suggests that there is a difference in the type of issue revealed by each of the approaches. The results indicate that observation-based methods would be suitable for video game usability and game mechanic evaluation (CAT 3), as opposed to GSR analysis, which is better suited to the discovery of issues concerning gameplay (CAT1) and player emotional immersion (CAT2). This suggests that some of the more established means of user testing have limited effectiveness when analysing complex entertainment products such as video games. While biometric research continues to require both interpretation and qualified personnel, a combination (mixed-method) of biometric and other typical approaches, such as observation and interview, would provide a more comprehensive set of results.

The primary work highlighted by this study is the need for further research into differing biometric measures. The experimental methodology conducted here has proved functional for determining the strengths, weaknesses and qualitative differences between the findings of a biometrics-based approach (using GSR only) and an observation-based user test study, but the methodology could be extended to include further biometric measures. By performing the biometric-based approach simultaneously or individually for alternative biometric measures, the same analysis procedure documented here can be applied. If required, the observation-based approach can also be conducted to provide yet further benchmarks with which to assess the contributions of the measures.

Of further interest is the differing genres of video game, and the associated types of issue and their biometric expression in participants. For the results of this and further studies to be generalisable to all video game genres, it is important that all game types be represented. Further research may also indicate that certain biometric sensors are better suited to the extraction of usability and player experience issues for certain genres.

Further research into the applications of biometrics for evaluating positive issues in video game usability analysis is also required, potentially allowing video game development companies to quickly and accurately understand successful elements of their game.

The mixed-methods approach generated as a result of combining the two approaches of this study has been constantly evaluated and iterated for over 100 hours of user research sessions on unreleased commercial video games. We have received positive feedback from game developers and producers on how this mixed-method has helped them to optimise the gameplay experience of their titles to high acclaim.

Overall, this study has provided evidence that, while observation-based methods identify the majority of usability and user experience issues, biometric-based methods provide a valuable contribution to video games user research. The contributions of biometrics to not only providing a greater number of usability and user experience issues, but also to providing confidence, confirmation, and validation of issues, and the potential of mixed-method approaches. We feel that the most important finding of this paper, however, are the further research questions asked as a result of this study.

ACKNOWLEDGMENTS
We would like to thank Geraldine Fitzpatrick, Lennart Nacke, Sam Hutton and Gareth R. White for their contributions to this paper. We would also like to thank our participants.

BIBLIOGRAPHY
Boucsein, W.: Electrodermal activity. Plenum Press, New York (1992)

Cacioppo, J. T. Handbook of Psychophysiology. Cambridge University Pressn (2007).

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. Psychophysiological Science. *Handbook of psychophysiology* (2007).

Call of Duty: Modern Warfare 2 (2009). Published by Activision. Developed by Infinity Ward.

Drachen, A., Nacke, L.E., Yannakakis, G. & Pedersen, A.L., Correlation between Heart Rate, Electrodermal Activity and Player Experience in First-Person Shooter Games. *In Sandbox* (2010). Los Angeles, 2010. ACM.

Desurvire, H., & Wiberg, C. Game Usability Heuristics (PLAY) For Evaluating and Designing Better Games: The Next Iteration. *Proceedings of the 3d International Conference on Online Communities and Social Computing* (2009). ACM
Ekman, P., Levenson, R.W., Friesen, W.V.: Autonomic nervous system activity distinguishes among emotion. Science. 10, pp.1208 (1983)
Federoff, M. Heuristics and Usability Guidelines for the Creation and Evaluation of Fun in Video Games. *Indiana University Master of Science Thesis* (2002).
Fulton, B., Medlock, M.: Beyond focus groups: Getting more useful feedback from consumers. *In Game Develop Conference* (2003).
Gow, J., Cairns, P., Colton, S., Miller, P., Baumgarten, R., Capturing Player Experience with Post-Game Commentaries. In *CGAT* (2010). Singapore.
Haze (2008). Published by Ubisoft. Developed by Free Radical Design.
Hazlett, R., Using Biometric Measurement to Help Develop Emotionally Compelling Games. In Isbister, K.a.S.N. Game Usability: Advancing the Player Experience. Morgan Kaufmann (2008). pp.187-205.
Kivikangas, J. M., Nacke, L., & Ravaja, N. Developing a triangulation system for digital game events, observational video, and psychophysiological data to study emotional responses to a virtual character. Entertainment Computing (2011), 11-16.
Kivikangas, J. M., Ekman, I., Chanel, G., Järvelä, S., Salminen, M., Cowley, B., Henttonen, P., Ravaja, N. Review on psychophysiological methods in game research. Proceedings of 1st Nordic DiGRA, DiGRA (2010).
Lang, P.J.: The emotion probe. American Psychologist (1995). 85, pp.372
Lang, P.J., Greenwald, M.K., Bradley, M., Hamm, A.O.: Looking at Pictures: Affective, Facial, Visceral, and Behavioral Reactions. Psychophysiology (1993). 73, pp.261
Lazzaro, N.: Why we play games: 4 keys to more emotion. In Game Develop Conference (2004)
Marshall, C., Rossman, G.B.: Designing qualitative research. Thousand Oaks. Sage Publications (1999)
Mandryk, R. "Physiological Measures for Game Evaluation." In *Game Usability: Advancing the Player Experience*, by Katherine and Schaffer, Noah Isbister, (2008). 207-235. Morgan Kaufmann.
Mandryk, R.L. & Atkins, M.S., A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International Journal of Human-Computer Studies* (2007), pp.329-47.
Mandryk, R.L., Atkins, M.S. & Inkpen, K.M. A Continuous and Objective Evaluation of Emotional Experience with Interactive Play Environments. *In CHI. Montreal*, (2006). ACM.
Medlock, M.C., Wixon, D., Terrano, M., Romero, R., and Fulton, B. Using the RITE method to improve products: A definition and a case study. Presented at the Usability Professionsals Association (2002), Florida.
Metacritic. *Call of Duty: Modern Warfare 2 (PS3) reviews at Metacritic.com*. Available at: http://www.metacritic.com/games/platforms/ps3/modernwarfare2?q=%20Call%20of%20Duty%204:%20Modern%20Warfare%202 (accessed Oct. 2010)
Metacritic. *Haze (PS3) reviews at Metacritic.com*. Available at http://www.metacritic.com/games/platforms/ps3/haze?q=haze (accessed Oct. 2010)
Mirza-Babaei, P., McAllister, G., Biometric Storyboards: visualising meaningful gameplay events. CHI 2011 BBI Workshop Proceedings, Vancouver, BC, Canada (2011).

Mirza-Babaei, P. & McAllister, G., Using Biometrics to better understand the players' gameplay experience. Presented at the Future Game On. ACM Paris, (2010).
Nacke, Lennart. *Affective Ludology. Scientific Measurement of User Experience in Interactive Entertainment*. Game Systems and Interaction Research Laboratory School of Computing Blekinge Institute of Technology, 2009.
Nacke, L.E., Grimshaw, M.N. & Lindley, C.A., More than a feeling: Measurment of sonic user experience and psychophysiology in a first-person shooter game. *Interacting with Computers* (2010).
Nacke, L. & Lindley, C.A., Flow and Immersion in First-Person Shooters: Measuring the player's gameplay experience. *FuturePlay* (2008), pp.81-88.
Nacke, L., Lindley, C., and Stellmach, S. Log Who's Playing: Psychophysiological Game Analysis Made Easy through Event Logging. *Proceedings of Fun and Games, Second International Conference,* Springer (2008), 150-157.
Nacke, L., Ambinder, M., Canossa, A., Drachen, A., Mandryk, R., Pinelle, D., Stach, T. (2009). Future Play Panel: "Game Metrics and Biometrics: The Future of Player Experience Research". May 13, 2009. Future Play Conference 2009.
Nacke, L.E. Directions in Physiological Game Evaluation and Interaction. In CHI 2011 BBI Workshop Proceedings, Vancouver, BC, Canada (2011).
Nielsen, J. Enhancing the explanatory power of usability heuristics. In Proc. CHI '94, ACM (1994) 152– 158.
Nielsen, J. Evaluating the thinking-aloud technique for use by computer scientists. In: Advances in human-computer interaction. Norwood (1992)
Onyett, C.: GDC 2009: Why Are Valve's Games So Polished?, Available at http://uk.pc.ign.com/articles/966/966972p1.html (accessed Oct. 2010)
Pagulayan, R.J., Keeker, K., Wixon, D., Romero, R.L. & Fuller, T. User-centered design in games. The humancomputer interaction handbook 28, 883-906 (2002).
Pinelle, D., Wong, N. & Stach, T. Heuristic evaluation for games: usability principles for video game design. In Proc. SIGCHI, ACM (2008), 1453– 1462
Poels, K., Kort, Y.d., Jsselsteijn, W.: It is always a lot of fun!: exploring dimensions of digital game experience using focus group methodology. In: Future Play. ACM. Toronto (2007)
Ravaja, N., Saari, T., Salminen, M., Laarni, J., and Kallinen, K. Phasic Emotional Reactions to Video Game Events: *A Psychophysiological Investigation. Media Psychology* 8, 4 (2006), 343-367.
Ravaja, N., Timo S., Marko T., Jari L., Mikko S., & Matias K. Spatial Presence and Emotions during Video *Game* Playing: Does It Matter with Whom You Play? Presence Teleoperators and Virtual Environments 15, (2006) no. 4: 381-392.
Shi, Y., Ruiz, N., Taib, R., Choi, E., & Chen, Galvanic skin response (GSR) as an index of cognitive load. In *CHI '07 extended abstracts on Human factors in computing systems* (2007). ACM, New York, NY, USA, 2651-2656.
Stern, R. M., Ray, W. J., & Quigley, K. S. *Psychophysiological recording*. New York: Oxford University Press (2001).
Tognetti, S., Garbarino, M., Bonanno, A.T., Matteucci, M., & Bonarini, A. Enjoyment recognition from physiological data in a car racing game. In *Proceedings of the 3rd international workshop on Affective interaction in natural environments* (AFFINE '10). ACM, New York, NY, USA, (2010) 3-8.
White, G., Mirza-babaei, P., McAllister, G. & Good, J., Weak inter-rater reliability in heuristic evaluation of video games, *CHI* (2011), (pp. 1441-1446).
Yannakakis, G.N., Hallam, J. & Lund, H.H., Entertainment capture through heart rate activity in physical interactive playground. *User Modeling and User-Adapted Interaction* (2008). 18, 1, 207-243.