Evaluating Affective Computing Environments Using Physiological Measures

Regan Lee Mandryk
Simon Fraser University
Burnaby, BC, V5A 1S6
rlmandry@cs.sfu.ca
www.sfu.ca/~rlmandry

ABSTRACT
Emerging technologies offer exciting new ways of using entertainment technology to create fantastic play experiences and foster interactions between players. Evaluating collaborative play technology is challenging because success isn't defined in terms of productivity and performance, but in terms of enjoyment and interaction. Current subjective methods of evaluating entertainment technology aren't sufficiently robust. Our research project aims to test the efficacy of physiological measures as evaluators of collaborative user experience with play technologies. We found evidence that there is a different physiological response in the body when playing against a computer versus playing against a friend. These physiological results are mirrored in the subjective reports provided by the participants. This research provides an initial step towards using physiological responses to objectively evaluate a user's experience with collaborative play technology.

INTRODUCTION

Emerging technologies in ubiquitous computing and ambient intelligence offer exciting new interface opportunities for co-located play technology, as evidenced in a recent growth in the number of conference workshops and research articles devoted to this topic [1, 2, 7]. Our research team is interested in employing these new technologies to foster interactions between users in co-located, collaborative play environments. We want technology not only to enable fun, compelling experiences, but also to enhance the interaction and communication between players. These goals are not the traditional goals of productivity enhancement, usually seen in HCI research. We are more concerned with the affective component of computing technologies [12], or generating an emotional response to a play environment.

For example, we recently created two novel collaborative play environments [9, 10] with the goal of enhancing interaction between players and to create a compelling experience. Other researchers have used emerging technologies to create entertainment environments with the same goal in mind [1, 5, 7]. However, evaluating the success of these new interaction techniques and environments is an open research challenge.

Traditionally, human-computer interaction research (HCI) has been rooted in the cognitive sciences of psychology and human factors, and in the applied sciences of engineering, and computer science [11]. Although the study of human cognition has made significant progress in the last decade, the notion of emotion is equally important to design [11], especially when the primary goals are to challenge and entertain the user. This approach presents a shift in focus from *usability* analysis to *user experience* analysis. Traditional objective measures used for productivity environments, such as time and accuracy, are not relevant to collaboration or play.

ISSUES AND CHALLENGES

The first issue prohibiting good evaluation of collaborative play technologies is the inability to define what makes a system successful. We are not interested in traditional performance measures, but are more interested in whether our environment fosters interaction and communication between the players, creates an engaging experience, and is fun. A successful interaction technique should provide seamless access to the game environment and be a source of fun in itself. Although traditional usability issues may still be relevant, they are subordinate to the actual playing experience as defined by challenge, engagement, and fun.

Once a definition of success has been determined, we need to resolve how to measure the chosen variables. Unlike performance measures, such as speed or accuracy, the measures of success for collaborative play technologies are more elusive. We want to increase interaction, enhance engagement, and create a fun experience. The current research problem lies in what metrics to use to measure engagement, interaction, fun, and collaboration.

We have previously used both subjective reports and video coding as methods of evaluating our new technologies although there is no control environment with which to make comparisons [9, 10, 13]. Subjective reporting through questionnaires and interviews is generalizable and convenient, but misses complex patterns. Using video to code gestures, body language, and verbalizations is a rich source of data, but is also a lengthy and rigorous process.

Research in Human Factors has used physiological measures as an indicator of mental effort and stress [14, 15]. Psychologists have been using physiological measures as unique identifiers of human emotions such as anger, grief, and sadness [4]. Physiological data have not been employed to identify human experience states of enjoyment, fun, and interaction. My doctoral research focuses on using physiological data as objective indicators of challenge, fun, boredom, and engagement in electronic entertainment environments.

In our research, we record users' physiological, verbal and facial reactions to game technology, and apply post-processing techniques to correlate an individual's physiological data with their subjective reported experience and events in the game. Our ultimate goal is to create a methodology for the objective evaluation of collaborative play technology, as rigorous as current methods for productivity systems.

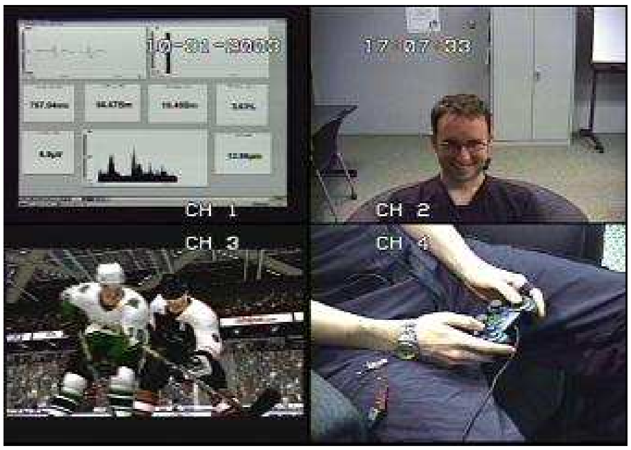

Figure 1: Quadrant display: screen capture of biometrics, video of player's face, video of controller, and screen capture of game.

OUR RESULTS

We have conducted a number of experiments to further our research goal. In our first experiment, we manipulated the difficulty of a game environment, hoping to elicit varying levels of boredom, challenge, frustration, and fun. We analysed both the subjective results and the mean physiological results individually, and also correlated the two data types for each individual.

Strong correlations between subjective ratings and the mean of many physiological measures were present in all players, but these correlations weren't consistent across individuals. One problem was that the subjects enjoyed playing in all of the conditions, even if the difficulty level didn't match their experience (fun median=3.0 for all conditions). The players also created challenges for themselves in the easier levels, changing the nature of the difficulty conditions, confounding the results.

The main challenge with analyzing this experiment was relating single point data (subjective ratings) to time series data (physiology). To match these two types of data, we converted the time series data to a single point through averaging (e.g. mean) or integrating (e.g. HRV) the time series. Although this method has been used in other domains, it erases the variance within each condition. Game design employs variance and reward, thus this approach may not be appropriate.

In the second experiment, to better understand how body responses can be used to create an objective evaluation methodology, we observed pairs of participants playing a computer game. Because this methodology is a novel approach to measure collaboration and engagement, and the results from Experiment One were ambiguous, we used an experimental manipulation designed to maximize the difference in the experience for the participant, so much that they would not be able to compensate with meta gaming activities. They played in two conditions: against another co-located player, and against the computer. We chose these conditions because we have previously observed pairs (and groups) of participants playing together under a variety of collaborative conditions [3, 6, 9, 13]. Our previous observations revealed that players seem to be more engaged with a game when another co-located player is involved.

The results of the second experiment are described in full in a paper at CSCW 2004 [8]. To summarize, we found different mean physiological responses in the body and different subjective reports when playing against a friend versus playing against a computer. Participants found it significantly more fun, engaging, and exciting, and less boring to play against a friend than against a computer. In addition, mean galvanic skin response (GSR), and mean electromyography of the jaw (EMG) were significantly higher when playing against a friend.

Although these results are an encouraging progression towards user experience analysis for collaborative play technologies, they have the same disadvantage as subjective results. They are single points of data representing an entire condition, however, unlike subjective reporting, they represent an objective measure of user experience. Used in concert, these two methods can provide a more detailed and accurate representation of the player's experience.

In order to correlate subjective and physiological responses, we needed to normalize the data. Physiological data has very large individual differences, thus individual baselines have to be taken into account. In order to perform a group analysis, we transformed both the physiological and subjective results into dimensionless numbers between zero and one. For each player, the difference between the conditions was divided by the span of that individual's results. A correlation of the normalized differences would show that the *amount* by which subjects increased their subjective rating when playing against a friend is proportional to the *amount* that the physiological measure increased in that condition. We found that normalized GSR was correlated with normalized fun and inversely correlated with normalized frustration. We also found that normalized respiratory amplitude was correlated with normalized challenge.

In addition to comparing and correlating the means from the two conditions, we investigated GSR responses for small windows of time surrounding game events. The raised mean GSR signals when playing against a friend reveal that players are more aroused when playing against a friend than when playing against a computer. However, we do not know whether this elevated result can be attributed to a higher tonic level or more phasic responses. One of the advantages of using physiological data to create evaluation metrics is that they provide high-resolution, continuous, contextual data. GSR is a highly responsive body signal, it provides a fast-response time-series, reactive to events in the game (see Figure 2).

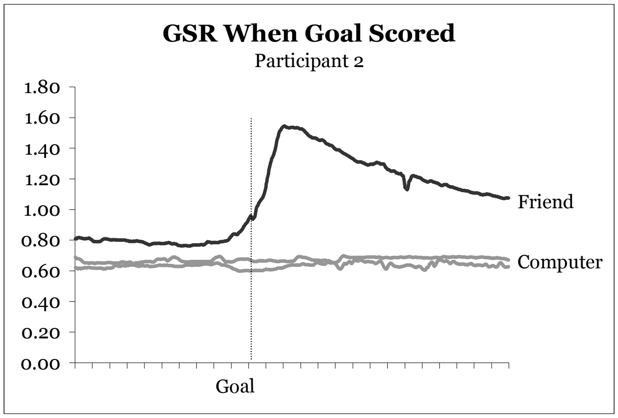

Figure 2: Example participant's GSR response to scoring a goal against a friend and against the computer twice. Note the much larger response when scoring against a friend. Data were windowed 10 seconds prior to the goals and 15 seconds after.

Using methods like the time-window analysis that we conducted, provides continuous objective data that can be used to evaluate the player experience, yielding salient information that can discriminate between experiences with greater resolution than averages alone. In this paper, we graphically represented continuous responses to different game events, and looked at the magnitude of the response using the span of the physiological measure. In our current work, we are taking advantage of the high-resolution, contextual nature of physiological data to provide an objective, continuous measure of player experience.

CURRENT DIRECTIONS

Our initial experiments were designed to correlate physiological data streams with a player's subjective experience. Our current research aims to correlate the physiological data with the actual experience, determined objectively. To do this, we have annotated video data, collected while participants played NHL 2003™ against the computer, against a friend, and against a stranger. We coded not only for game events (e.g. goals for, goals against, hits given, and hits received), but also for interpersonal interactions (e.g. talk, laugh, trash talk).

In parallel, we extracted certain mathematical features from the physiological time series (e.g. local maxima, local minima, saddle points). We then determined the relationship between the objective video annotation data and the physiological time series data.

This approach completes the triangulation of data sources (subjective data relates to physiological data relates to video annotation data), but also provides an automatic continuous representation of user experience with play technologies. However, this approach is limited by examining each physiological signal (e.g. GSR, EMG) separately.

In our future work, we plan to examine how the combined information provided by the signals can reveal unique emotional responses to play technologies. By using a fuzzy logic model, combined with our previous results, we hope to generate a strong tool for evaluating play technologies. This approach is strong as it is grounded in the data itself, yielding meaningful results.

CONCLUSIONS

The evaluation of computing environments devoted to entertainment and play is ripe for advancement. Subjective data yield valuable quantitative and qualitative results. However, when used alone, they do not provide sufficient information. Physiological measures have previously been used to evaluate productivity systems, especially to reflect a user's stress or mental effort. The application of physiological measurement and analysis to collaborative leisure technology has exciting potential.

Although we do not currently understand how the body physically responds to enhanced interaction, or increased enjoyment, our research project aims to ultimately provide researchers with a methodology for objectively evaluating user experience with collaborative play technologies. We foresee that objective evaluation, combined with current subjective techniques will provide researchers with techniques as rigorous and valuable as current methods of evaluating user performance with productivity systems. In addition, our results can be used to create a powerful tool, used by designers and developers of entertainment technologies.

REFERENCES

[1] Björk, S., Falk, J., Hansson, R., and Ljungstrand, P. (2001). Pirates! Using the Physical World as a Game Board. In *Proceedings of Interact 2001*. Tokyo, Japan.

[2] Björk, S., Holopainen, J., Ljungstrand, P., and Mandryk, R.L. (2002). Introduction to Special Issue on Ubiquitous Games. *Personal and Ubiquitous Computing*, *6*: p. 358–361.

[3] Danesh, A., Inkpen, K.M., Lau, F., Shu, K., and Booth, K.S. (2001). Geney: Designing a collaborative activity for the Palm handheld computer. In *Proceedings of Conference on Human Factors in Computing Systems (CHI 2001)*. Seattle, WA, USA: ACM Press. p. 388-395.

[4] Ekman, P., Levenson, R.W., and Friesen, W.V. (1983). Autonomic Nervous System Activity Distinguishes among Emotions. *Science*, *221*(4616): p. 1208-1210.

[5] Holmquist, L.E., Falk, J., and Wigström, J. (1999). Supporting Group Collaboration with Inter-Personal Awareness Devices. *Journal of Personal Technologies*, *3*(1-2).

[6] Inkpen, K., Booth, K.S., Klawe, M., and Upitis, R. (1995). Playing Together Beats Playing Apart, Especially for Girls. In *Proceedings of Computer Supported Collaborative Learning (CSCL '95)*.

[7] Magerkurth, C., Stenzel, R., and Prante, T. (2003). STARS - A Ubiquitous Computing Platform for Computer Augmented Tabletop Games. In *Proceedings of Video Track of Ubiquitous Computing (UBICOMP'03)*. Seattle, Washington, USA.

[8] Mandryk, R.L. and Inkpen, K. (2004). Physiological Indicators for the Evaluation of Co-located Collaborative Play. In *Proceedings of Computer Supported Cooperative Work (CSCW 2004)*. Chicago, IL, USA.

[9] Mandryk, R.L., Inkpen, K.M., Bilezikjian, M., Klemmer, S.R., and Landay, J.A. (2001). Supporting Children's Collaboration Across Handheld Computers. In *Conference Supplement to Conference on Human Factors in Computing Systems (CHI 2001)*. Seattle, WA, USA. p. 255-256.

[10] Mandryk, R.L., Maranan, D.S., and Inkpen, K.M. (2002). False Prophets: Exploring Hybrid Board/Video Games. In *Conference Supplement to Conference on Human Factors in Computing Systems (CHI 2002)*. p. 640-641.

[11] Norman, D.A. (2002). Emotion and Design: Attractive things work better. *Interactions, 9* (4).

[12] Picard, R.W., (1997). *Affective Computing*. Cambridge, MA: MIT Press.

[13] Scott, S.D., Mandryk, R.L., and Inkpen, K.M. (2003). Understanding Children's Collaborative Interactions in Shared Environments. *Journal of Computer Assisted Learning*, *19*(2): p. 220-228.

[14] Vicente, K.J., Thornton, D.C., and Moray, N. (1987). Spectral Analysis of Sinus Arrhythmia: A Measure of Mental Effort. *Human Factors*, *29*(2): p. 171-182.

[15] Wilson, G.M. (2001). Psychophysiological Indicators of the Impact of Media Quality on Users. In *Proceedings of CHI 2001 Doctoral Consortium*. Seattle, WA, USA.: ACM Press. p. 95-96.