

Objectively Evaluating Entertainment Technology

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

ABSTRACT

Emerging technologies offer new ways of using entertainment technology to foster interactions between players and connect people. Evaluating entertainment 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 robust. This research uses previous literature and empirical results to create a methodology for objective evaluation of entertainment technology. By gathering physiological data in the context of game play, we intend to correlate physiological responses with subjective reports and with game events. This framework would be a powerful tool used by designers, developers, and researchers to inform their design and evaluate their decisions.

Categories & Subject Descriptors: H.5.2 [Information Interfaces and Presentation] User Interfaces-Evaluation/methodology.

General Terms: Design, Human Factors.

Keywords: Evaluation methods, collaboration, games, physiological measurements, fun.

INTRODUCTION

Emerging technologies in ubiquitous computing and ambient intelligence offer exciting new interface opportunities for entertainment technology, as evidenced in related conference workshops and research articles [1].

Our research team is interested in employing new technologies to foster interactions between users in co-located, collaborative entertainment environments. For example, we created a hybrid board-video game system to enhance player interaction [4]. We also created a collaborative game environment on handheld computers where players work together but individually access the shared game space, enhancing the collaboration [3].

Evaluating the success of our new interaction techniques and environments has been challenging. Traditional objective measures used for productivity environments,

such as time and accuracy, are not relevant to collaborative play. We 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.

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 [3-5]. 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 [6, 7]. Psychologists have been using physiological measures as unique identifiers of human emotions such as anger, grief, and sadness [2]. Physiological data have not been employed to identify human experience states of enjoyment and fun. My doctoral research focuses on using physiological data as objective indicators of challenge, fun, boredom, and engagement in electronic entertainment environments.

This research records users' physiological, verbal and facial reactions to game technology, and applies post-processing techniques to correlate their physiological data with their subjective reported experience and events in the game. Our goal is to create a methodology for objective evaluation of entertainment technology, as rigorous as current methods for productivity systems.

INITIAL EXPERIMENT

We manipulated the difficulty of a game environment, hoping to elicit varying levels of boredom, challenge, frustration, and fun. The 8 subjects were experienced male gamers, age 21-26. We recorded video of their face and the controller, the game screen, and the physiological data while they played NHL2003™ in four difficulty levels (Figure 1). We made an informed choice to collect galvanic skin response (GSR), heart rate (HR) and inter-beat interval (IBI) from electrocardiography (EKG), electromyography (EMG) of the jaw, respiration rate (RRate), and respiration amplitude (RAmp). After each condition, the subjects rated their experience in terms of fun, challenge, boredom, and frustration on a five-point scale. Free form explanations were also solicited.

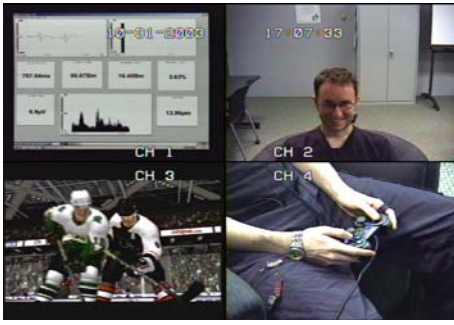


Figure 1: The physiological data, the participant's facial expression, the controller, and the game were all recorded into a single video signal.

Physiological data for each condition were processed into descriptive statistics including the mean, peak, and standard deviation. Spectral analysis of IBI was used to obtain heart rate variability (HRV), an indicator of mental effort [6].

Results and Challenges in Experiment One

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. For example, when playing in the beginner condition, one player set up fancy plays to score pretty goals to make the game interesting. These activities changed 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. We annotated the physiological data with game events, which showed us distinguishable game events associated with peaks, declines, and increases in the physiological data. Our next experiment aims to measure subjective experience continuously.

NEXT STEPS

In the next experiment, participants will play either against the computer, against another co-located player, or against another player on-line. Based on prior video analysis of teenagers playing a math game under similar conditions [5], we have a baseline for how players might respond. This manipulation also addresses our original questions on how entertainment technology can promote social connections.

To match the subjective data to the time series physiological data, we will use a retrospective think-aloud technique. A traditional think-aloud technique would not

work since verbalizing the experience would significantly affect game play. Having subjects verbalize their experience while replaying the game after each condition creates a retrospective, continuous, subjective report.

CONCLUSIONS AND FUTURE WORK

After correlating subjective report with physiological data, we hope to use our technique to provide a baseline for using physiological state to index human experience. This will provide an objective means of evaluating technology built for entertainment purposes. We will then validate our technique by comparing it to the standard method of evaluation using video analysis. Our technique could then be converted into a tool for researchers and developers to use to evaluate their entertainment products. This would be a powerful tool, more salient than information gathered through focus groups. Game developers and designers could pinpoint how events and characters in their game impact the user, and adjust their design, thus reducing the risks associated with development.

Finally, modeling a user's engagement and fun is beneficial for feeding physiological information into a game as real-time input. Although it is straightforward to map biosensor output to game input, our results would yield a *meaningful* framework for using biometrics as input into a game.

ACKNOWLEDGMENTS

Thanks to EA Canada, NewMIC, SFU Surrey, my advisor Kori M. Inkpen, and my committee.

REFERENCES

1. Björk, S., J. Holopainen, P. Ljungstrand, and R.L. Mandryk. (2002). *Introduction to Special Issue on Ubiquitous Games*. Personal and Ubiquitous Computing, **6**, 358–361.
2. Ekman, P., R.W. Levenson, and W.V. Friesen. (1983). *Autonomic Nervous System Activity Distinguishes among Emotions*. Science, **221**(4616), 1208-1210.
3. Mandryk, R.L., K.M. Inkpen, M. Bilezikjian, S.R. Klemmer, and J.A. Landay. (2001). *Supporting Children's Collaboration Across Handheld Computers*. in Extended Abstracts of CHI 2001, 255-256.
4. Mandryk, R.L., D.S. Maranan, and K.M. Inkpen. (2002). *False Prophets: Exploring Hybrid Board/Video Games*. in Extended Abstracts of CHI 2002, 640-641.
5. Scott, S.D., R.L. Mandryk, and K.M. Inkpen. (2003). *Understanding Children's Collaborative Interactions in Shared Environments*. Journal of Computer Assisted Learning, **19**(2), 220-228.
6. Vicente, K.J., D.C. Thornton, and N. Moray. (1987). *Spectral Analysis of Sinus Arrhythmia: A Measure of Mental Effort*. Human Factors, **29**(2), 171-182.
7. Wilson, G.M. (2001). *Psychophysiological Indicators of the Impact of Media Quality on Users*. in Doctoral Consortium at CHI 2001, 95-96.