ISSN: 2451-8611
ISBN: 978-1-910172-00-1

Exploring Facial Expressions for Human-Computer Interaction: Combining Visual Face Tracking and EMG Data to Control a Flight Simulation Game

Saeedeh Bayatpour, Ulysses Bernardet, Steve DiPaola, Alexandra Kitson, Bernhard E. Riecke

School of Interactive Arts and Technology, Simon Fraser University
Surrey, BC, Canada
sbayatpo@sfu.ca, Ulysses_bernardet@sfu.ca, sdipaola@sfu.ca, akitson@sfu.ca, ber1@sfu.ca

Abstract

In many affective computing paradigms a user's internal state is used as an implicit control signal in an interaction. In the work presented here, we are exploring the utilization of two measurement techniques commonly used to assess a user's affective state as an explicit control signal in a navigation task in a virtual environment. Concretely, we are investigating the feasibility of combining a real-time emotional biometric sensing system and a computer vision system for human emotional characterization and controlling a computer game. A user's "happiness" and "sadness" levels are assessed by combining information from a camera-based computer vision system and electromyogram (EMG) signals from the facial corrugator muscle. Using a purpose-designed 3D flight simulation game, users control their simulated up-down motions using their facial expressions. To assess if combining visual and EMG data improves facial tracking performance, we conduct a user study where users are navigating through the 3D visual environment using the two control systems, trying to collect as many tokens as possible. We compared two conditions: Computer vision system alone, and computer vision system in combination with the EMG signal. The results show that combining both signals significantly increases the users' performance and reduces task difficulty. However, this performance increase is associated with a reduced usability due to the need to have EMG sensors on one's forehead. We hope these results from our study can help in future game designs, aid the development of more immersive virtual environments, and offer for alternative input methods where traditional methods are insufficient or unfesasible.

Keywords

Facial Expression, Facial EMG signal, Emotion, computer vision system.

Introduction

Emotion is one of the most important concepts in psychology, human computer interaction and many other areas. Emotions can be modeled either by a *dimensional approach* where emotions are coordinates in a space or a *classification approach* that categorizes emotion in few descriptive words [1]. In general, emotion can be described by four components: behavioural reactions, expressive reactions, physiological reactions and subjective feelings [2]. Gokcay *et al*. [3] categorized the methods of measuring emotions into psycho-physiological approaches such as: fMRI, EEG and EMG and qualitative approaches such as self-reporting, observations and non-verbal behaviours. Facial expressions are one of the most important ways of conveying emotion, and are controlled by facial muscles and skin movements. Facial expression can happen both voluntarily or involuntarily [7]. Hess et al. [4] investigated facial reactions to the emotional facial expressions of people as either affective or cognitive. Erickson and Schulkin (2003) [5] described the relationship between perception and presentation of facial expressions as cognitive processes. Facial expression can be analyzed through image processing and computer vision techniques [6]. Also, facial electromyography (fEMG) is another common approach to detect and measure the facial muscle activities. Of particular importance here the corrugator muscle that is associated with frowning, and the zygomaticus muscle that is associated with smiling [7]. These are commonly used to characterize happiness and sadness emotions. For example, facial expression is used to test the audience responses to new products and computer games. In game design facial expression can indicate the emotional response of the game player to different features of a game and thus be used to provide valuable feedback in the game design process [8, 9, 10, 11, 12]. In the current work, however, we use a different approach, in that we explore the usage of facial expressions as an input parameter. That is, participants actively control a 3D flight simulation through their facial expressions. Using facial expression in addition to traditional input methods such as keyboard, mouse, track pads or joysticks opens up new ways of interaction with a system, especially in situations where hand-based input is unfeasible. Examples include situations were hands cannot be used because they are tied up with other tasks, or for users with disabilities or special needs who cannot easily use input devices such as keyboards or mouse. We do not intend to imply that facial expressions should be used to replace traditional input methods, but rather explored how they might augment them.

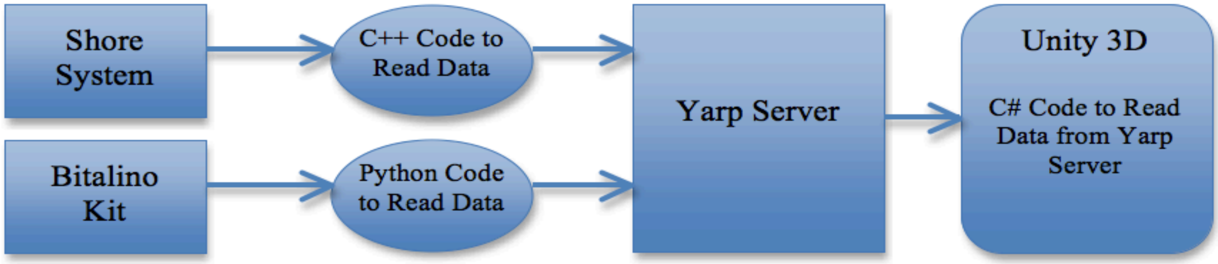

Figure 1. A schematic representation of the connection between the Shore System, Bitalino and Unity 3D environment through yarp server.

After explaining the implementation of the prototype system, this paper presents a feasibility study to evaluate the system, followed by a general discussion and outlook.

Implementation of the Prototype

The main design goal was to develop a novel user interface system for emotional interaction of the user in an immersive 3D environment where users get direct feedback about the effectiveness of their control in a game-like setup. We achieved this goal by user's facial expressions directly controlling visual changes in the environment, namely the user's locomotion through the environment created in Unity 3D. A 3D sky environment gives users the ability to have a feeling of flight in the sky and moving upwards if the user displays happy facial expressions and moving downwards if the user displays sad expressions, thus controlling one's locomotion through a path defined by a series of hoops in the air. User's happiness and sadness levels were extracted by combining information of a camera-based computer vision Fraunhofer Shore system [6] and facial EMG signal using a Bitalino kit with three electrodes placed on the user's head [13]. Using a Yarp server [14] as a middleware provides the ability to send real-time data from the Fraunhofer Shore system and the Bitalino kit to the Unity project. These sub-systems are explained in more details in the following sections.

3D Virtual Environment and Motion Control

To assess the ability of the user in controlling and moving in a 3D environment through facial expressions, we designed as simple virtual reality flight simulator where participants were tasked to use their facial expressions to control the simulated flight along a path defined by a number of hoops to fly through as shown in Figure 2. The 3D environment consists of a blue sky (skybox in Unity) with a large number of drifting clouds to provide strong optic flow. In addition, a route was defined by 30 golden hoops that player are asked to fly through, similar to a tunnel-in-the-sky display used in actual airplane flight. Using a game-like paradigm, the aim of the user is to fly through as many hoops as they can to maximize the number of collected hoops. A hoop is counted as collected if the user manages to fly through it as opposed to miss it. The hoops are spaced 25m equally apart on a forward z-axis with their vertical location in a range (-10m to 10m). The simulation was explicitly designed to be simple, such that users only had to control one degree of freedom, their vertical motion. Forward motion was controlled by the simulation. By starting the game, the user starts a flight through a route of hoops with constant gradual forward acceleration. This constant acceleration results in the user starting to fly forward with a low speed that gradually increases to make the game increasingly challenging over time. The hoops are located in different vertical positions, and users task is to use their facial expressions to control the vertical motion such that they fly through the next hoop. The user needs to show enough level of happiness to move in the upward direction and show sadness to move in the opposite, downward direction. For example, depending on the height of the hoops, sometimes the user needs to show only a close-lip smile and sometimes a big obvious smile. Note that future work would be needed to distinguish between users displaying authentic emotions of sadness and happiness versus just making facial expressions – the current prototype and study was not designed to disambiguate between them.

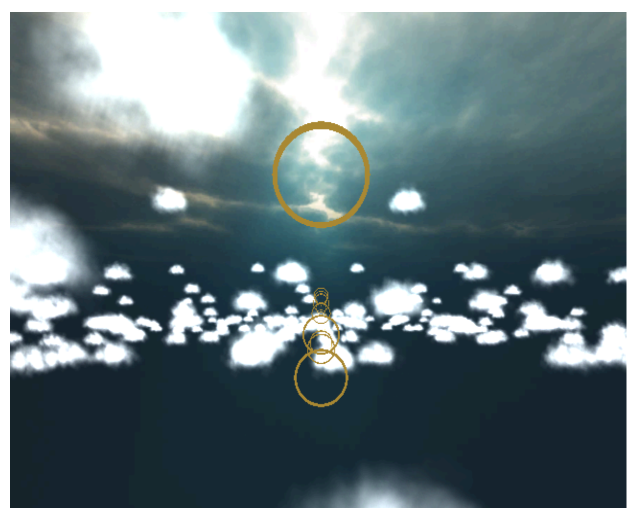

Figure 2. The 3D environment was designed in Unity 3D and consists of a sky simulation with a series of golden hoops to fly through.

User's performance was assessed by counting the number of hoops that the users were able to fly through by controlling and changing their facial expressions. For example, in some part of the travelling route the user needs to show his highest level of happiness and suddenly change it to frown very fast to be able to get two consecutive hoops one at the highest height positions and the other one at the lowest level. The user's measured happiness versus sadness values were applied as upward versus downward forces to a mass-spring system where the first person camera (player) has a mass that is connected to the vertical axis by a spring. As a result, when the user is happy the force is in positive direction and when the user is sad the force is in negative direction, similar to providing an upward versus downward thrust in an actual airplane. When the user's facial expression is neutral the player smoothly goes back to center (medium height where the position on the y-axis is zero). Although the raw sensor data might be choppy, flying up and down and then going back to the center appears smoothly due to the usage of a mass-spring control system. This was done by fine-tuning the spring stiffness and damping values of the mass-spring system in pilot studies.

Facial Expression Information from Fraunhofer-Shore System

The Fraunhofer Shore system is based on a computer vision approach to analyze and characterize the facial expressions collected by a standard web camera on a Windows computer. This system continuously returns values on a scale from 0-100 to represent the four emotional dimensions of happiness, sadness, surprise and angriness of the user as shown in Figure 3. In this work, we focused only on the sadness and happiness values of Shore system, as pilot tests indicated that they were the most reliable emotional indicators.

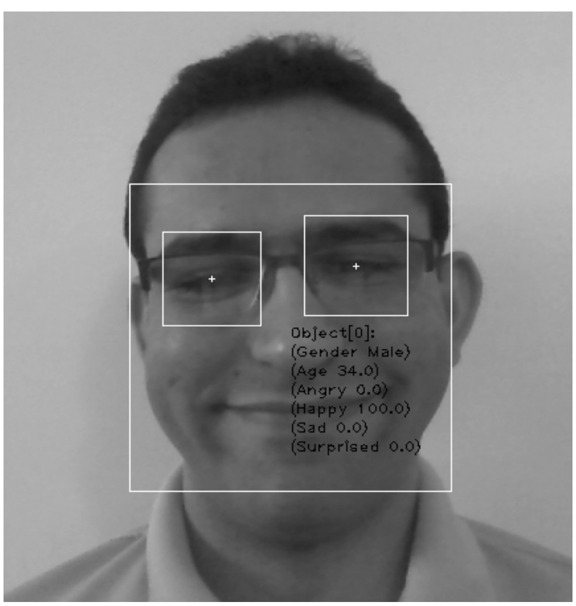

Figure 3. Using Shore system to analyze facial expressions.

In order to transfer real-time data from the Shore system to the Unity project, we used Yarp server [14]. Yarp is a middleware layer that provides the ability of transferring real-time data and communication between different software systems. For example, in this case, data from the Shore system was acquired in C++ and then transferred to the Yarp server. Then we used Unity C# scripting to read data from the Yarp server. For the purpose of using Yarp, one port is opened for sending data to the server (in the C++ code) and one port is opened for receiving data (in the Unity C# script) as shown in Figure 1. The two ports are then connected for real-time data transfer.

Facial Expression Information from Facial EMG Signal

As using facial expression to infer emotional states is inherently noisy, we combined the Shore visual tracking system with a facial EMG Signal. Getting data from the corrugator and the zygomaticus can show muscle contractions in these areas that are associated to frowning (sadness) and smile (happiness) [15, 16, 7].

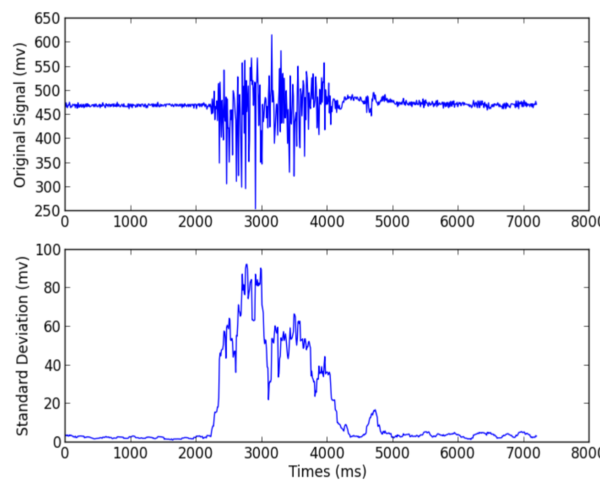

Figure 4. Standard deviation is used on original facial EMG signal to detect contraction in the associated muscles.

In order to get the EMG signal, we have used Bitalino device plugged kit. Bitalino can send the data to the computer through Bluetooth wireless technology. In this work, a typical configuration of electrodes is used: a single ground electrode was placed to the top of the forehead which is an inactive place, the other two electrodes were attached above one of the eyebrows with almost one centimeter distance apart [15] as shown in Figure 5. EMG data from the corrugator muscle is processed in a Python script. There are different approached for finding muscle contraction in the EMG signal. Using standard deviation is a common approach for detection of contraction. The contraction happens when the standard deviation of samples in a time window of the signal is more than a threshold (Figure 4.). In fact, if the standard deviation is more than the threshold we concluded that there is a contraction in the corrugator, and the standard deviation is scaled between 0 to100 and is sent to Yarp. Otherwise, there is no contraction and a zero value will be sent.

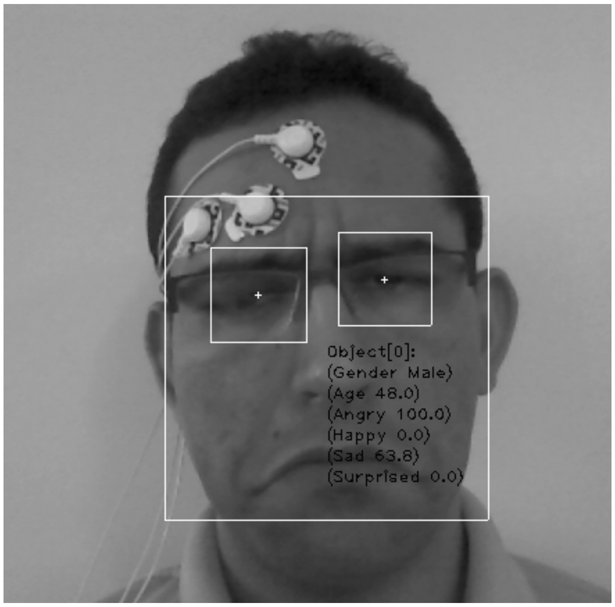

Figure 5. The location of electrodes for getting signal from the corrugator muscle to enhance the performance of the system in detecting sadness.

System Evaluation

After a pilot test of the system using only the Shore facial expression data, we realized that users can easily go in the upward direction by adjusting their happiness level. However, the performance of the Shore system was not as good for sad conditions, and did not work reliably for some participants. Therefore, we decided to test if we can enhance the performance of the system in measuring sadness emotions by combining the Shore signal with the EMG signal of the corrugator muscle. In this first implementation, this was done by simply averaging the sadness values from both Shore system and Bitalino EMG kit. To assess if thus combining the visual and EMG signal can improve facial tracking, we designed a simple user study that assessed user's performance in two conditions:

- Using only the Fraunhofer Shore system
- Using both the Fraunhofer Shore system and facial EMG signal in sadness emotions

We hypothesized that combining the visual and EMG system would allow users to more effectively control the simulated flight via their facial expressions (indicated by travelling through more hoops) and reduce the perceived task difficulty.

Experimental Design

Twelve graduate students between 22-40 years of age took part in this experiment. One of the participants was a researcher and aware of the research hypotheses, the others were naive to the purpose of the study. The participants were selected by volunteer based sampling. Each participants performed the game 4 times, twice with the Shore system an twice with using both Bitalino and Shore system. To avoid biases in learning how to use the system, half of the participants started the experiment with using Shore and then they switched to use both Shore and Bitalino and half of the participants started with using Shore and Bitalino. The experiment took between 10-15 minutes.

Procedure

First, the purpose of the experiment, the procedure, data confidentiality and risks were explained to the participants, and they signed informed consent. Participants were asked to show their emotions through their facial expression (smile or frown) to be able to fly through hoops by controlling the simulated flight in upward or downward directions.

For getting facial expression through EMG signal we first informed the participant and explained the process. Since we need to attach three electrodes to the participant's face, the researcher described the purpose of using electrodes and assured the participant that using such electrodes on the facial skin was safe and had no side effect. In order to get more accurate signal and reduce the impedance between skin surface and electrode gel, participants were asked to remove makeup or skin oil on their forehead.

Participants then performed two trials in each of the conditions (Shore-only vs. Shore + Bitalino), in counterbalanced order. For each trial the number of collected hoops was recorded as the main performance measure. After each trial, participants were also asked to verbally rate the task difficulty on as scale of 0%=very easy to 100% =very difficult.

Results

As predicted, combining the Shore system with the Bitalino facial EMG signal improved performance (Figure 6), indicated by a higher number of collected hoops (M: 15.75, SD: 2.72) compared to the Shore-only condition (M: 10.29, SD: 1.47), $t(11) = 9.223$, $p < .0001$, $\eta_p^2 = .886$. The effect size η_p^2 of .886 is considered a large effect size [17] and indicates that 89% of the variability in the performance data can be attributed to the independent variable "device", i.e., using the Shore+Bitalino vs. Shore-only to measure facial expressions.

Similarly, combining the Shore system with the Bitalino facial EMG signal yielded reduced task difficulty ratings (M: 31.92, SD: 22.93) compared to the Shore-only condition (M: 53.75, SD: 15.75), $t(11) = 3.987$, $p = .002$, $\eta_p^2 = .591$, see Figure 7. The effect size η_p^2 of .591 is considered a large effect size [17] and indicates that 59% of the variability in the difficulty rating data can be attributed to the measurement device. When asked to indicate which system participants preferred, 10 participants (83.3%) preferred the Shore system alone, one participant (8.3%) preferred using Shore and Bitalino system together, and one participant (8.3%) did not have any preferences

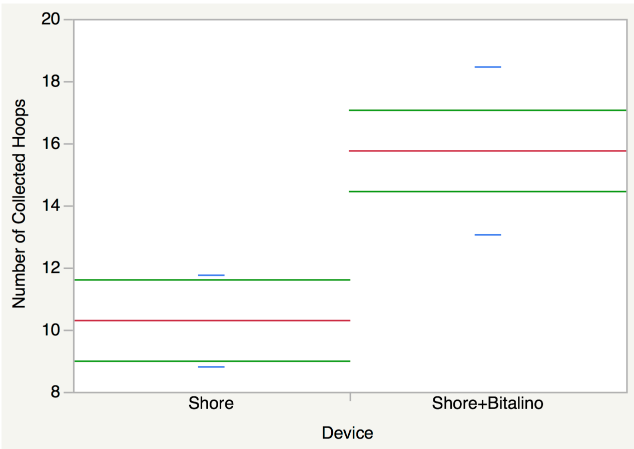

Figure 6. Analysis of the number of collected hoops for two different conditions: using only Shore system for facial expression analysis and using both Shore system and the facial EMG signal. Red lines are the mean values, blue lines are standard deviation and the green lines represent 95% confidence intervals.

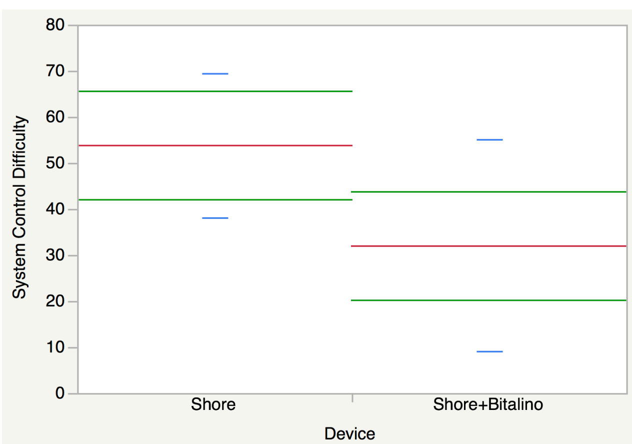

Figure 7. Difficulty in controlling the system for two different conditions: using only Shore system for facial expression analysis and using both Shore system and the facial EMG signal. Red lines are the mean values of difficulty, blue lines are standard deviation and the green lines represent 95% confidence intervals.

Conclusion

This research aims at exploring the use of facial expression in controlling a system. The system gives the users the ability to control a simple flight simulator and navigate through a series of hoops with a force that is a function of the level of user's sadness and happiness. The direction of flying, the speed and acceleration of the movement are based on the force that is applied to a mass-spring system. In this work, we used two different approaches (computer vision and Facial EMG signal) to analyze the facial expression. A user study showed that combining the computer vision system and facial EMG signal allowed for more effective control of the system and also reduced perceived task difficulty; however, it is not vey convenient for the user to use the system by some electrodes attached to the facial skin. This problem can likely be reduced by using miniature electrodes (with less than 5mm diameters) as recommended in [15]. Unlike previous work [8, 9, 10, 11, 12] that used facial expression data for passive evaluation of games, this work shows that such data can be actively used to create a more interactive and game-like immersive environment.

Future Works

This study was limited to obtaining facial EMG signal from the corrugator; in a follow-up study, we plan to add another condition that just relies on the facial EMG signal analysis from both the corrugator and the zygomaticus muscles [7]. Future works could analyze other facial expressions such as anger or fear to provide additional input channels and thus more nuanced interaction and/or additional potential movement and navigation options in a virtual environments. Such method could ultimately be useful in situations where normal hand-based input is unfeasible, for example when hands are busy, or for physically disabled users who cannot easily use traditional input devices for a variety of tasks including movement and navigation in virtual environment. Beyond using facial expressions as a mere input methods, analyzing facial expressions in real-time could also be useful for a wide variety of applications including biofeedback, emotional regulation, affective computing, advertising, user testing, and for augmenting neurogaming applications. For many of these applications, it could become essential to distinguish between users experiencing authentic emotions versus merely performing facial expressions, which might be achieved by triangulating different emotion-sensing methods and in particular including EEG recordings, which are more difficult to "fake" than just facial expressions.

Since we are well aware of to the invasive nature of the EMG and it's potential drawbacks, we would like to test other techniques such as EEG or EDR to find out which psycho-physiological approaches are most suitable for extracting facial expression data and controlling the system. Because changing facial expression rapidly and repeatedly can be tiring, there is also a potential for fatigue limiting usability and long-term usage.

References

1. Machajdik Jana and Allan Hanbury, "Affective Image Classification Using Features Inspired by Psychology and Art Theory." In *Proceedings of the International Conference on Multimedia*, 83–92, New York, NY, USA, 2010.
2. Pieter Desmet, "Measuring Emotion: Development and Application of an Instrument to Measure Emotional Responses to Products," In *Funology*, edited by Mark A. Blythe, Kees Overbeeke, Andrew F. Monk, and Peter C. Wright (Human-Computer Interaction Series 3. Springer Netherlands, 2003), 111.
3. Magy Seif El-Nasr, Jacquelyn Ford Morie and Anders Drachen, "A scientific look at the design of aesthetically and emotionally engaging interactive entertainment experiences," in *Affective Computing and Interaction: psychological, Cognitive, and Neuroscientific Perspective,* Didem Gokcay and Gülsen Yildirim (PA: Hershey, 2011), 281.
4. Ursula Hess, Pierre Philippot, and Sylvie Blairy, "Facial Reactions to Emotional Facial Expressions: Affect or Cognition?," *Cognition & Emotion* 12, (1998): 509–31.
5. Kristine Erickson and Jay Schulkin, "Facial Expressions of Emotion: A Cognitive Neuroscience Perspective," *Brain and Cognition*, Affective Neuroscience 52, (2003): 52–60.
6.http://www.iis.fraunhofer.de/en/ff/bsy/tech/bildanalyse/shore-gesichtsdetektion.html
7. John T. Cacioppo, Louis G. Tassinary and Gary Berntson, *Handbook of Psychophysiology* (*Cambridge University Press*, 2007).
8. Rajava Niklas, Saari Timo, Laarni Jani, Kallinen Kari, and Salminen Mikko. "The Psychophysiology of Video Gaming: Phasic Emotional Responses to Game Events," 2005. In *Proceedings of DIGRA Conference*, http://www.digra.org/wp-content/uploads/digital-library/06278.36196.pdf.
9. Regan L. Mandryk and M. Stella Atkins, "A Fuzzy Physiological Approach for Continuously Modeling Emotion during Interaction with Play Technologies," *International Journal of Human-Computer Studies*, 65, (2007): 329–47.
10. Jing-Kai Lou, Kuan-Ta Chen, Hwai-Jung Hsu, and Chin-Laung Lei, "Forecasting Online Game Addictiveness." In *2012 11th Annual Workshop on Network and Systems Support for Games (NetGames)*, 2012.
11. Richard L. Hazlett, "Measuring Emotional Valence During Interactive Experiences: Boys at Video Game Play," In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1023–26, New York, NY, USA, 2006.
12. Niklas Ravaja, Marko Turpeinen, Timo Saari, Sampsa Puttonen, and Liisa Keltikangas-Järvinen, "The Psychophysiology of James Bond: Phasic Emotional Responses to Violent Video Game Events," *Emotion* 8, (2008): 114–20.
13. http://www.bitalino.com/index.php/plugged-kit
14. http://wiki.icub.org/yarpdoc/index.html
15. T. D. Blumenthal, B. N. Cuthbert, D. L. Filion, S. Hackley, O. V. Lipp, and A. Van Boxtel, "Committee Report: Guidelines for Human Startle Eyeblink Electromyographic Studies," *Psychophysiology* 42, (2005): 1–15.
16. Rebecca D. Ray, Kateri McRae, Kevin N. Ochsner, and James J. Gross, "Cognitive Reappraisal of Negative Affect: Converging Evidence from EMG and Self-Report," *Emotion* 10, (2010): 587–92.
17. Jacob Cohen, *Statistical Power Analysis for the Behavioral Sciences* (Hillsdale, NJ: Erlbaum, 1988).

Authors Biographies

Saeedeh Bayatpour is a PhD student in the School of Interactive Arts and Technology at Simon Fraser University in Surrey, BC, Canada. She has degrees in computer software engineering from Islamic Azad University (B.Sc.) and Simon Fraser university (M.Sc). Her research interests are in Human Computer Interaction, 3D display systems and computer vision.

Ulysses Bernardet is a postdoctoral fellow at the School of Interactive Arts and Technology of the Simon Fraser University, Vancouver, Canada. He has a background in psychology, computer science and neurobiology, holds a doctorate in psychology from the University of Zurich, and was a postdoctoral fellow and lecturer at the Universitat Pompeu Fabra in Barcelona, Spain. Ulysses follows an interdisciplinary approach that brings together psychology, neurobiology, robotics, and computer science. He is the main author of the large-scale neural systems simulator iqr, and the core contributor to the conceptualization and realization of several complex real-time interactive systems. At the center of Ulysses' research activity is the development of models of cognition, emotion, and behavior that are capable of interacting with humans in real-time by means of 3D characters or robots.

Steve DiPaola, is director of the Cognitive Science Program at Simon Fraser University, and leads the iVizLab (ivizlab.sfu.ca), a research lab that strives to make computational systems bend more to the human experience by incorporating biological, cognitive and behavior knowledge models. Much of the labs work is creating computation models of very human ideals such as expression, emotion, behavior and creativity. He is most known for his AI based computational creativity (darwinsgaze.com) and 3D facial expression systems. He came to SFU from Stanford University and before that NYIT Computer Graphics Lab, an early pioneering lab in high- end graphics techniques. He has held leadership positions at Electronic Arts, and Saatchi Innovation. His computational art has been exhibited at the AIR and Tibor de Nagy galleries in NYC, Tenderpixel Gallery in London and Cambridge University's Kings Art Centre. And exhibited at major museums including the Whitney, MIT Museum, and Smithsonian.

Alexandra Kitson became a research intern at Simon Fraser University in January 2013 and started her Masters in Fall 2014. She has a BSc from the University of British Columbia in Cognitive Systems, a multidisciplinary program that combines psychology, computer science, philosophy, and linguistics. Her research interests involve using an interdisciplinary approach to understand human perception and behaviour. In particular, employing technology as a medium to explore the human psyche, create better human-computer interfaces, and provide clinical applications.

Associate Professor Bernhard Riecke joined Simon Fraser University in 2008 after receiving his PhD from Tübingen University and the Max Planck Institute for Biological Cybernetics and working as a postdoctoral fellow at Vanderbilt University and the Max Planck Institute. His research approach combines fundamental scientific research with an applied perspective of improving human-computer interaction. For example, he uses multidisciplinary research approaches and immersive virtual environments to investigate what constitutes effective, robust, embodied and intuitive human spatial cognition, orientation and behaviour as well as presence and immersion. This fundamental knowledge is used to guide the design of novel, more effective human-computer interfaces and interaction paradigms that enable similarly effective processes in computer-mediated environments such as virtual reality, immersive gaming, and multimedia.