Entertainment Computing 2 (2011) 11–16

Contents lists available at ScienceDirect

Entertainment Computing

journal homepage: ees.elsevier.com/entcom

Developing a triangulation system for digital game events, observational video, and psychophysiological data to study emotional responses to a virtual character

J. Matias Kivikangas [a,*], Lennart Nacke [b], Niklas Ravaja [a,c]

[a] Center for Knowledge and Innovation Research, Aalto University, P.O. Box 21255, 00076 Aalto, Finland
[b] Department of Computer Science, University of Saskatchewan, 110 Science Place, Saskatoon, Saskatchewan, Canada S7N 5C9
[c] Department of Social Research, University of Helsinki, P.O. Box 54, 00014, Finland

A R T I C L E   I N F O

Article history:
Received 11 October 2010
Revised 11 March 2011
Accepted 12 March 2011
Available online 30 March 2011

Keywords:
Triangulation
Psychophysiology
Experimental psychology
Digital games
Emotional congruency

A B S T R A C T

Game researchers are currently lacking comprehensive data analysis tools that triangulate game events, event-related survey data, and psychophysiological data. Such a tool would allow a comprehensive analysis of player engagement in digital games. The development of this tool was motivated by an experimental psychology study that asked whether emotional reactions to congruent and incongruent emotional stimuli within an intrinsically motivated game task are the same as within the traditional experimental picture-viewing paradigm. To address the needs of our study, we used the Source SDK (Valve Corporation) for creating a system that automates event logging, video management psychophysiological data markup. The system also allowed recording of self-report measures at individual play events without interrupting the game activity.

1. Introduction

Digital games are now employed in many research, science and technology areas. For example, games are suitable stimuli for experimental psychology, since virtual gaming environments allow experimenting with vicariously experienced situations that would pose ethical problems in real life or would be difficult or expensive to carry out. These studies use different methods of investigation, such as behavioral observation (often with video), psychophysiology, and psychometric questionnaires, and an automated solution for combining the different data sources in an easy and functional way is currently lacking.

Our system is an analysis tool created for the study of physiological responses to emotional expressions outside the regular picture-viewing paradigm. The tool creates time-framed video clips of relevant game events and presents them after an experimental session to the participant for recreating their memory of the experience at the event point in the game. The participant is then prompted to answer a questionnaire regarding this specific game event. This allows a triangulation between game events, phasic physiological responses, and self-report measures without interrupting the gaming activity.

To better understand how this experiment motivated the development of this system, we briefly describe the background and related work regarding this experiment. Since this paper focuses on the triangulation system, the experiment is presented as the motivation and as an example, but we do not present results from the study. Finally, we discuss how the system fit our needs for the experiment, and its potential and limitations in a larger context.

2. Background of the experimental study

The research setup for the study had players meet realistic but virtual human non-player characters (i.e., NPCs) in a virtual gaming environment and react to the artificial facial expressions and assumed action tendencies of these characters. Prior psychological research suggests that viewing an affective facial expression elicits – automatically and mostly unconsciously – a similar expression and emotion in the viewer [1]. Furthermore, it has been shown that complex stimuli of congruent (i.e., similar valence) or incongruent (i.e., opposite valence) emotions either intensify or dampen elicited emotions, respectively [2,3].

These basic results are based on controlled experiments where participants looked at pictures without any confounding factors. We wanted to study if the effect can be detected in virtual characters, but also in more ecologically valid situations (e.g., when players meet characters while doing intrinsically motivated tasks). To be comparable to earlier research, our experiment needed to include the recording of physiological responses to, and self-reports of, game events without disturbing gameplay.

* Corresponding author. Tel.: +358 44 0393930; fax: +358 9 431 38391.
E-mail addresses: matias.kivikangas@aalto.fi (J. Matias Kivikangas), lennart.nacke@acm.org (L. Nacke), niklas.ravaja@aalto.fi (N. Ravaja).

1875-9521/$ - see front matter
doi:10.1016/j.entcom.2011.03.006

3. Related work

Incorporating different data streams is most advanced in game evaluation, as making sure a game system works flawlessly is vital to its success on the market [4]. A number of regular software testing approaches are used in the game industry, such as unit testing and bug tracking. Another popular game user research method is direct observation. For example, playtesting, where participants play the game while their gameplay behavior is being observed either by video or by a game user researcher directly [5]. Open-ended tasks allow participants to play the game as if they were at home without any instructions. The experimenter takes notes during the observation procedure and later discusses his observations with the game's designers so they can use this information to shape the design goal of the game [6]. A more research-oriented approach is the extended playtest, which collects survey data at time intervals over several hours of playtime. Recently, playtests have been automated at larger companies such as Microsoft Game Studios, where the TRUE (Tracking Real-time User Experience) instrumentation system makes the triangulation of player feedback data with videos possible [7]. However, this system falls short of integrating physiological data in this analysis, which could provide potentially important data for the evaluation process [8].

In human-computer interaction (HCI) research, various approaches within the area of affective computing [9] show promising approaches for user sensing and prediction of user emotions. Some of these machine learning approaches are becoming accurate at identifying emotional states, detecting a user's facial expressions [10], analyzing non-verbal behaviors, distinguishing pressure patterns [11], sensing body temperature [12], and discovering emotion from keystroke typing patterns [13]. While these approaches are likely to be helpful for automatically adapting games [14], affective information, especially psychophysiological inference, is seldom used in game user research. In a similar vein, physiological measures have only recently become more popular for assessing game engagement [15–17]. However, physiological measures are employed commonly in basic psychological research [18,19]. Self-report measures, common in game research, are traditionally administered at a different time than playing the game. Thus, they are separated from the experience of playing the game (e.g., in focus groups), happen after an experiment (e.g., interviews), or are administered after a certain time period in the experiment representing an experimental condition (e.g., session questionnaires). Therefore, the capability of self-reports to pinpoint particular responses to events in the game is limited because they rely on recall without visual aids to memory. Since it is difficult not to affect gameplay experience when interrupting the game, researchers have recently proposed game logging triangulation systems that are suitable for real-time, in-game physiological data logging [20] and in-game gaze [21] data logging with an eye tracker. Triangulation in these cases means the cross examination of three or more data sources for being more confident in the experimental result. Based on these prior systems, we designed the tool for our study.

4. Experiment background and game system development

4.1. Congruency experiment

The full details of the experiment and the study can be found elsewhere [22]. To understand how this experiment motivated the system development, it is described here in brief. The experiment was implemented as follows. The experiment had the participants (volunteers, $n = 40$, 21 male, age of 18–31, active game players) rescue captured non-player characters (NPCs). The NPCs had facial expressions that prompted a player to decide whether the NPC was positive (i.e., cooperative) or negative (i.e., hostile, see below). In addition to the facial expression, NPCs had an action disposition towards hostility to or cooperation with the player character, providing the second part of the congruent or incongruent stimulus. The action disposition was made unambiguous by presenting an on-screen marker (i.e., a password of the resistance movement in the game's fictional story frame). Thus, there were two main event types of interest, one where the facial expression was first viewed and the other one where the action disposition was made clear. Regardless of the expression, the participants were supposed (but not forced) to kill the hostile NPC and escort the friendly NPC to the safe place (the game level layout and player decision script, i.e., the anticipated game mechanic, is shown in Fig. 1). The research interest was on the responses elicited in the players by congruent and incongruent expressions and action tendencies of the NPCs.

4.2. Requirements analysis

The physiological responses (EDA, ECG, facial EMG, and EEG) were recorded during the entire experiment, but to get additional insights into participants' subjective motivations and gameplay experience, self-report data was required in the specific context of the relevant events. However, there are currently no methods to pinpoint a specific event in a series of game tasks for self-reporting without relying on interrupting gameplay and administering a questionnaire [7] or interviewing about the event retrospectively. An interruption would defeat the point of a more ecologically valid situation for the experiment and retrospective interviews have no visual aids to help player recall the game experience at the event, which makes them less reliable from a psychological viewpoint. Therefore, the technical requirements for the analysis software were to:

1. Provide a game environment for presenting the stimuli within an intrinsically motivated gaming activity.
2. Present the in-game stimuli (i.e., congruent and incongruent facial expressions and actions of natural human characters).
3. Provide a method for administering self-report measures without interrupting the activity.
4. Provide the ability for the game researcher to compare the self-report measures with psychophysiological responses at the event point (i.e., basic triangulation).

For the first two requirements we used an existing game and development Kit, but requirements 3 and 4 needed a new methodology. We finally settled with stimulated audiovisual retrospective self-reports. Retrospectively showing the participant selected video clips of the relevant events immediately after the game could serve as an audiovisual reminder of the actual experience. The self-report measures could then be administered after each clip to ask about that particular event (similarly to stimulated retrospective think-aloud method employed in usability studies; see e.g., [23]). By using an active reminder, the participants should be able to recall the events more clearly, recall their original experiences more accurately than without the audiovisual reminder. It was not clear, however, whether the video reviews of the game events would actually provide a sufficient reminder, whether the participants would actually imagine the gameplay experience and not use the video itself as a stimulus, or whether the imagined experience would be close enough to the actual experience. By additionally recording physiological responses during the review phase we could then compare the two physiological signals to each other, and to the self-report measures, to see if there were significant differences or if the procedure provided what it was designed to