Entertainment Computing 9–10 (2015) 19–27

Contents lists available at ScienceDirect

Entertainment Computing

journal homepage: ees.elsevier.com/entcom

An annotation tool for automatically triangulating individuals' psychophysiological emotional reactions to digital media stimuli ☆,☆☆

Pedro A. Nogueira [a,*], Vasco Torres [b], Rui Rodrigues [c], Eugénio Oliveira [a]

[a] *Artificial Intelligence and Computer Science Lab, University of Porto, Portugal*
[b] *Faculty of Engineering, University of Porto, Portugal*
[c] *INESC-TEC, University of Porto, Portugal*

ARTICLE INFO

Article history:
Received 15 July 2014
Revised 29 March 2015
Accepted 13 June 2015
Available online 4 July 2015

Keywords:
Emotion recognition
Player modelling
Psychophysiology
Digital media
Triangulation
Emotional reaction identification

ABSTRACT

Current affective user experience studies require both laborious and time-consuming data analysis, as well as dedicated affective classification algorithms. Moreover, the high technical complexity and lack of general guidelines for developing these affective classification algorithms further limits the comparability of the obtained results. In this paper we target this issue by presenting a tool capable of automatically annotating and triangulating players' physiologically interpreted emotional reactions to in-game events. This tool was initially motivated by an experimental psychology study regarding the emotional habituation effects of audio-visual stimuli in digital games and we expect it to contribute in future similar studies by providing both a deeper and more objective analysis on the affective aspects of user experience. We also hope it will contribute towards the rapid implementation and accessibility of this type of studies by open-sourcing it. Throughout this paper we describe the development and benefits presented by our tool, which include: enabling researchers to conduct objective a posteriori analyses without disturbing the gameplay experience, automating the annotation and emotional response identification process, and formatted data exporting for further analysis in third-party statistical software applications.

1. Introduction

For the past two decades, video games have pioneered various breakthroughs in important fields such as computer graphics, animation, artificial intelligence, physics simulation and interaction techniques. These achievements have, so far, produced various improvements, such as more believable virtual environments and interactions. Yet, despite these consecutive improvements, we are now going through a slower evolution consisting mainly of iterative enhancements.

As video games move towards photorealistic graphics, the game research community has started focusing their efforts on promising and yet underexplored areas of the gameplay experience. However, this research avenue entails an objective analysis of what motivates gamers to play and what constitutes a good gaming experience. This question has been previously investigated in the gaming community [1] and one conclusion was that video games are played either to: (a) live out a fantasy, or (b) to relax from the problems of everyday life; at most times a combination of both [1]. In either of these cases, a common trait is found: *video games that are able to provide an engrossing experience are always emotionally evocative to some degree.*

Due to their high emotional elicitation potential, digital games have been increasingly used in many tangent research areas. For example, digital games are a suitable alternative to real life studies or dangerous, expensive or logistically challenging studies, such as phobia treatment [2]. Common methods used in these types of research mostly include behavioural observation, psychometric questionnaires or psychophysiological data annotation. Most studies still use the first two former methods, which offer rather vague insights into the emotional alterations that the presented stimuli elicited on each participant. More recent studies have recently

☆☆ This paper has been recommended for acceptance by Paolo Ciancarini.

* Corresponding author.
E-mail addresses: pedro.alves.nogueira@fe.up.pt (P.A. Nogueira), vascoptorres@gmail.com (V. Torres), rui.rodrigues@fe.up.pt (R. Rodrigues), eco@fe.up.pt (E. Oliveira).

http://dx.doi.org/10.1016/j.entcom.2015.06.003
1875-9521/© 2015 Elsevier B.V. All rights reserved.

started exploring the potential of psychophysiological data [1–4], but usually require manual sensor calibration and annotation of each reaction. This means there is a current need for a method capable of automating the classification of this psychophysiological data. Also, due to the high workload involved in the required data annotation (e.g. logging game-related events and identifying relevant emotional responses), this process should also be automatized.

Our proposed solution consists of a tool to automate the analysis of psychophysiological-measured emotional reactions to in-game stimuli. This tool allows users to load gameplay session videos and synchronise them with the corresponding physiological recordings. Rather than displaying the raw physiological readings, as with more traditional approaches, the tool then interprets these readings as a continuous prediction of arousal and valance ratings [5] using our previously proposed method [6]. This continuous prediction is built-in as a standalone module that can be replaced or parameterised *on-the-fly* by users so that the tool is useful in a wide range of application scenarios. The tool then allows users to replay the game session videos and manually annotate gameplay events using a simple GUI. Alternatively, users can also import text-formatted game event logs in case these are available, thus bypassing (or augmenting) the manual annotation phase altogether.

Once users have finished the annotation of game events, the tool automatically estimates which of the annotated game events prompted emotional reactions. This is done by automatically isolating the emotional responses to each annotated event by applying a parameterised two-dimensional local maxima/minima search algorithm to the emotional classification signal (i.e. the arousal/valence signal result from the predictive model). The system also allows subjective metrics, such as questionnaire responses or user commentaries to be added for each individual event. Sessions can also be saved to a serialized *.eet* data file that can be loaded in a posterior point in time, should a gameplay session require additional analysis. Finally, we included an export feature that writes the identified emotional responses to a structured (tab-delimited) text file for further analysis in third party software tools such as R, Weka or SPSS, as is usual in many of these studies.

Although it is not within the scope of this paper, we succinctly describe the study that led to the tool's development so as to contextualize our motivation and design choices. We do not present the results for the study as it is meant solely as an example. However, we do present a validation on the system's automatic emotional response detection capabilities. This is meant as a measure of the system's overall adequacy for our needs and how it contributes to the typical psychophysiological annotation pipeline. *the emotional classification signal, the system.*

2. Related work

User research methods are usually categorised according to their data source and approach. The data source refers to whether the method measures how its participants act (behavioural) or what they say (attitudinal). On the other hand, the method's approach refers to what type of data is collected (quantitative or qualitative). In game development, although at different stages, virtually all types of user research methods are used. Attitudinal methods (e.g. focus groups, participatory design or desirability studies) are usually applied in earlier development phases, while more concrete (i.e. behavioural and quantitative) methods tend to be used towards the final product delivery deadlines (e.g. beta playtesting periods) [5]. Due to the low sophistication of the available techniques, earlier game user research methods focused more heavily on qualitative methods. In this type of method, a participant would play a specific level or level section while a researcher would observe and annotate his session in real-time, a posteriori, or both [5,6]. These annotations would then be collected for a set of participants representative of the game's target population and later discussed to tune the gaming experience – usually in an iterative design cycle [6].

More recent research-based approaches have focused on being able to track game-play events both in real-time and over a larger time frame than that feasible through manual annotation. The most successful of these approaches is Microsoft Game Studios' TRUE (Tracking Real-time User Experience) instrumentation, which allows logging and annotating (i.e. triangulating) interaction events and player feedback [7]. However, despite its achievements, the system fails to take psychophysiological data into account, which could provide deeper insights into the participant's choices and preferences over time. Complementary to the aforementioned approach, Valve Software has openly announced it is experimenting with biofeedback and psychophysiological user research methods, namely using participants' skin conductance (SC) to measure their relative excitement over several playtesting sessions.

2.1. Data annotation tools & frameworks

Given the often times complex design of experimental studies, several tools have been developed to aid in the data synchronization and annotation process. In this section we will discuss the more popular and relevant ones to our needs, whilst comparing them to our proposed tool.

One of the most popular tools for data annotation is Observer XT (Noldus Information Technology BV, Wageningen, The Netherlands, 2014) [7]. The tool offers a wide range of data annotation and visualization functionalities, from audio-visual data collection to physiological data visualization and an event logging interface. ANVIL is an also popular, general-purpose tool that allows researchers to track uttered words, head movements, body gestures and other similar inputs on audio-visual data [8]. Similarly to the Observer XT, it allows users to augment their annotation with contextual data via an event coding scheme. Being more geared towards speech and body motion analysis, there are also some additional plug-ins that augment its functionality with improved coding schemes [9].

While the aforementioned tools are able to calculate some statistics from the collected data (number of recorded events, variation and distribution of observed events, event latency, etc.), they remain general-purpose ones that are able to contribute little towards physiological or emotional data analysis per se. In other words, they do not provide a way to interpret the physiological data in any meaningful way (other than these generic statistics) and, more importantly, still require the users to manually code events and emotional responses. These issues have been discussed by Gunes et al. in their thorough guide on emotion recognition, which ranges from emotion theories and data modalities to data annotation and interpretation [10] and by Nacke from a game analytics context [11]. In their review, Gunes et al. mention several issues with data annotation, one of which being the lack of a standard for emotion recognition, reporting that *"researchers seem to use different levels of intensity when adopting a dimensional affect approach"*. They also refer the issue of inter-observer variability, stating that *"obtaining high inter-observer agreement is one of the main challenges in affective data annotation, especially when dimensional approach is adopted"* – an issue found by Abrilian et al. in their creation of a large database of emotionally-coded news clips, where the employed subjective coding technique led to low observer agreement levels (0–20% at most at the exception of one emotion) using only 2 observers [12]. The authors conclude that *"(the) development of an easy to use, unambiguous and intuitive annotation scheme remains, however, an important challenge"*.