J Multimodal User Interfaces (2015) 9:43–54
DOI 10.1007/s12193-014-0153-4

ORIGINAL PAPER

An analysis of player affect transitions in survival horror games

**Vanus Vachiratamporn · Roberto Legaspi · Koichi Moriyama ·
Ken-ichi Fukui · Masayuki Numao**

Received: 22 January 2014 / Accepted: 21 April 2014 / Published online: 11 December 2014

Abstract The trend of multimodal interaction in interactive gaming has grown significantly as demonstrated for example by the wide acceptance of the Wii Remote and the Kinect as tools not just for commercial games but for game research as well. Furthermore, using the player's affective state as an additional input for game manipulation has opened the realm of affective gaming. In this paper, we analyzed the affective states of players prior to and after witnessing a scary event in a survival horror game. Player affect data were collected through our own affect annotation tool that allows the player to report his affect labels while watching his recorded gameplay and facial expressions. The affect data were then used for training prediction models with the player's brainwave and heart rate signals, as well as keyboard–mouse activities collected during gameplay. Our results show that (i) players are likely to get more fearful of a scary event when they are in the suspense state and that (ii) heart rate is a good candidate for detecting player affect. Using our results, game designers can maximize the fear level of the player by slowly building tension until the suspense state and showing a scary event after that. We believe that this approach can be applied to the analyses of different sets of emotions in other games as well.

Keywords Affective gaming · Physiological signal · EEG · EKG · Anxiety · Suspense · Fear

V. Vachiratamporn (✉) · K. Moriyama · K. Fukui · M. Numao
The Institute of Scientific and Industrial Research,
Osaka University, Osaka, Japan
e-mail: vanus@ai.sanken.osaka-u.ac.jp

K. Moriyama
e-mail: koichi@ai.sanken.osaka-u.ac.jp

K. Fukui
e-mail: fukui@ai.sanken.osaka-u.ac.jp

M. Numao
e-mail: numao@ai.sanken.osaka-u.ac.jp

R. Legaspi
The Institute of Statistical Mathematics, Research Organization
of Information and Systems, Tokyo, Japan
e-mail: legaspi.roberto@ism.ac.jp

1 Introduction

Over the past 15 years, the survival horror game genre has brought the conventions of horror films to interactive media that create a new realm of horror genre by giving audience the power to control the protagonist and interact with the horror environment themselves [14,26,29]. In doing so, the games give audience the first hand experience of the protagonist's fate. However, many survival horror games still stick with horror film strategies that utilize the series of linearly scripted events to lead the player to the target emotion (i.e., fear) [1, 29], rather than emphasize the interactive aspect of the game. Even though a good story and well-crafted scary events can keep players interested throughout the entire game, horror elements in pre-scripted events diminish the scare experience as the players play the game again because the anticipation of what is going to happen is not there anymore.

There are some recent indie-developed games, such as the "Amnesia: The Dark Descent" (Frictional Games, 2010) and "Slender: The Eight Pages" (Parsec Productions, 2012), that loosely create unpredictable scary moments for players repeatedly by utilizing semi-scripted scary events wherein most of the game elements are still pre-scripted but the action of monsters changes based on player actions and some random parameters. As a result, the games increase the unpredictability of scary events and keep players on their toes all the time because the players do not know exactly how the scary game elements are coming out even when they repeat the game many times. Obviously, however, there comes a time when this kind of approach does not work well anymore. Unlike the pre-scripted approach where the sequence of events carefully leads a player from one emotion to another, it is more difficult for game designers to create such kinds of sequences with this unpredictable approach. Suppose that the monster appears after the player picks up an item. It can suddenly show up in front of the player or it can stay behind and wait for the player to be anxious enough to turn back and look at it. In the former case, the player might expect something more and not get surprised by it, while in the latter case, the player might not decide to turn back and walk away without seeing the monster at all. Hence, trying to maximize fear intensity while keeping scary events unpredictable presents a challenge to survival horror game developers.

Multimodal input gaming, where non-traditional input is used with or without the traditional ones (e.g., keyboard and controller), is not something new and has already appeared in commercial games (e.g., Wii Remote and Kinect). However, it is rather new to go further than simply detecting player actions, specifically, to recognize and leverage player affective states. Physiological signals open up the possibility of continuous affect data detection and recognition in real-time without having to ask players to report their affect which can easily disrupt and ruin their gameplay experience. Many researchers have reported their success of using physiological signals to automatically detect player emotion [7,18,39], which is a crucial step in "affective gaming" that aims to use player affect as an additional input for manipulating or adapting the game scenario [8,11]. However, as far as our knowledge is concerned, there is no research that detects horror media emotions such as *anxiety* and *suspense*. Having a method that can measure such emotions in real time can help increasing the effectiveness of scary events in the survival horror games as the games have more information to predict how a player will react to each game element and can choose the best one for scaring the player.

Therefore, this work tries to construct the measuring method. We first collect physiological signals of players during gameplay and their affective states by letting the players review their gameplays using our own affect annotation tool (AAT). Next, we analyze the transition likelihood of player self-annotated affective states that change continuously during the gameplays. Lastly, we show the potential of the physiological signals in predicting the self-annotated affective states.

This paper is constructed as follows. In Sect. 2, the literature on affect, survival horror games, and affect recognition in games is reviewed. Section 3 describes the method and the tools used for collecting data in this work. Section 4 presents the results. Our conclusion and future work are discussed in Sect. 5.

2 Related works

There are some different usage of the emotion terms associated with horror media research (e.g., fear, horror, terror, anxiety, and suspense). Garner and Grimshaw [5] proposed a framework that incorporates the fear experience with acoustic parameters, which is based on Fanselow's defensive behavior system [4]. They defined four stages of fear scenarios: safe, caution, terror, and horror. *Safe* is when the player is not in a threatening environment. *Caution* is after entering an environment where the threat is expected to appear. *Terror* is after the presence of threat has been confirmed. Lastly, *Horror* is as a fight or fight response after a direct confrontation with threat. Caution, terror, and horror definitions also resemble the definition of *anxiety*, *suspense* and *fear* respectively, as used in the Toprac and Meguid's study [33]. In a game context, these three emotions can be distinguished by the interaction and concreteness of how a threat is perceived. Fear is an emotional response to a specific threat or an attempt to cope with threatening events that have already been seen, whereas anxiety is usually caused by the uncertainty towards an enigmatic or unspecific threat [23,33], and suspense is commonly defined as feeling of uncertainty towards an expected outcome [27]. In this work, we use anxiety, suspense and fear as terms for describing player emotions and self-report labels.

Most research on survival horror games tried to find a way to elicit greater and heightened fear intensity from the player using various approaches. Dekker and Champion [1] tried to increase cinematically augmented horror by using the player's biofeedback to dynamically modify the gameplay (e.g., movement speed), visual (e.g., screen shader), and sound (e.g., sound volume) features. The results from their post-experiment interview indicated that more players preferred the biofeedback-enhanced version to the non-enhanced version. Biofeedback data, however, was used only for as an additional game control and the quantitative evaluation on the data was outside the scope of their work. Garner et al. [6] tried to relate different sound properties (e.g., pitch, loudness, and 3D positioning) to the player's fear intensity response, using the player's in-game action and real-time vocal response for evaluation. To strike the balance between getting the immediate evaluation from the player and not interrupting the flow of the game, the player had to speak or shout the appropriate numbers (1–5) to rate the "emotional impact" in response to each key sound that he had just heard. However, no conclusive evidence was found as the sound properties might have been too conservative to trigger significantly different fear intensity. Toprac and Meguid [33] also tried to determine how to cause fear, suspense and anxiety in players by using different sound properties (e.g., volume, timing, and source). They collected self-report surveys and interviews for analyzing both quantitative and qualitative aspects. The results suggested that the best sound design for causing fear is high volume sound effects that are well-timed with the visual element, and for causing anxiety and suspense is the untimely medium volume and acousmatic sound effects. Nevertheless, there is no research on survival horror games that investigates player affective responses as a continuous experience and analyzes how the player affective states transition during the gameplays.

Several researchers have been trying to measure the players' gameplay experience, which is normally derived from self-report questionnaires using various tools and methods. Nacke and Lindley [21] created three separate levels of a first-person shooter (FPS) game which were designed to assess boredom, immersion, and flow experiences. They used game experience questionnaires (GEQ) for subjective measurement of various gaming experiences (e.g., immersion, flow, challenge, and tension), and used facial electromyography (EMG) and galvanic skin responses (GSR, also known as skin conductance) as objective valence and arousal measures, respectively. The result showed that the responses from EMG and GSR are significantly different over the three levels, whereas GEQ showed significant differences only on challenge and tension. Drachen et al. [3] collected participants' heart rates (HR), GSR and GEQ across three different FPS games. The results showed that a high HR is indicative of tensed and frustrated emotions while a low HR is indicative of positive affect, flow, feeling of competence, immersion and low levels of challenge. GSR, on the other hand, showed the correlation with negative affect. Similarly, Martinez et al. [19] also made use of HR and GSR features to predict reported affective states across two dissimilar games (predator/prey and racing games) using artificial neural network models. The results showed the affective models trained on one game can predict the reported affective states of the other game, and also suggested using average HR and two-step GSR variation features for affect prediction. Jennett et al. [12] investigated player immersion in games and whether immersion can be measured quantitatively through a series of experiments. Their overall findings showed that immersion can be measured subjectively, i.e., by questionnaires, and objectively, i.e., by completion time and eye movements, and also that immersion can also happen even when players are in negative emotions.

Using self-report questionnaires at the end of a session can only assess player affect over a whole game experience. Furthermore, using the questionnaires, the game has to be limited to a short duration for the player to effectively remember his affective experience and only one specific emotion can be effectively assessed at a time. For example, as mentioned above, Nacke and Lindley [21] needs three game levels for assessing boredom, immersion, and flow experience. These constraints do not sit well with the dynamic nature of interactive gaming experience. There are only few researchers who are interested in modeling players' emotions continuously throughout the game. Mandryk and Atkins [18] used a fuzzy logic model to transform physiological data (HR, GSR, and EMG) to arousal-valence space and then to game-related emotions during play experiences. The fuzzy rules were created based on psychophysiology literatures (i.e., arousal correlated with GSR and HR, and valence correlated with EMG and HR) and the model was then compared with subjective reports. With the continuous measurement of emotions in real-time, the transition of player affect caused by game elements (e.g., sound and images) can be analyzed individually. More importantly, it also opens up the possibility of adapting those elements in real-time based on the player's current affective state.

3 Experiment methodology

We introduce in this section the game environment used to elicit affective responses from players and show how continuous self-reported emotions were collected using our own AAT. Then, we introduce the tools and techniques used for getting the features from electroencephalography (EEG) and electrocardiography (EKG) signals, and keyboard–mouse activities (KMA). Lastly, the experiment procedure and the participants' information are shown.

3.1 Game environment

We had the choice between using an existing game or creating an entirely new game for our experiments. We reflected on the benefits and drawbacks of each. The benefits of using an existing game are that the quality of the game is guaranteed since it passed through various reviews and that the game is immediately usable for experimentation. The problem, however, is that we do not have any control over the games's design. For example, we cannot modify the game even if it is not suited well with our experiment setup, the gameplay information like event logs cannot be extracted automatically, and it will require a time-consuming manual event-annotation task [37]. On the other hand, we can do such modifications and in-game analyses with a self-developed game. Then again, the quality can be questioned whether it can credibly elicit target emotions from players or not. We finally decided to use an existing game because the main objective of our work is to analyze the player's affective states which does not necessitate modifying any parts of the game.

The survival horror game we used for the experiment is "Slender: The Eight Pages" (STEP),[1] a free downloadable indie-developed game that is based on the internet mythos of the Slender Man, a faceless creature who passively stalks his targets until they mentally break down and disappear with

[1] http://parsecproductions.com/slender/.

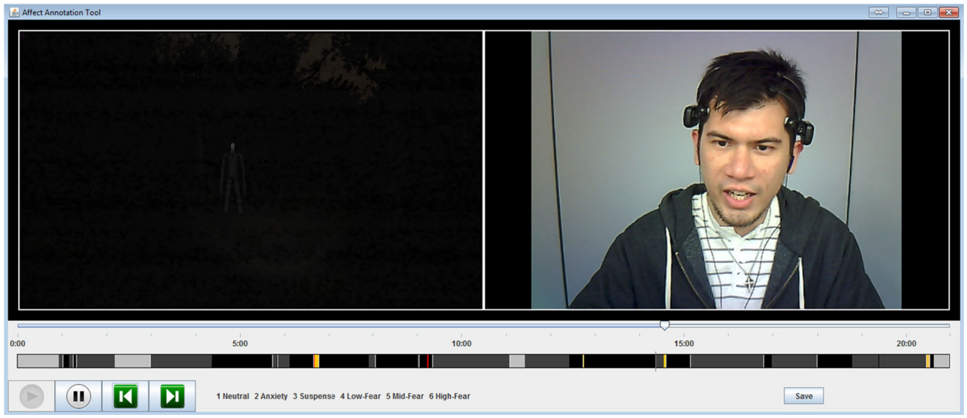

Fig. 1 Screenshot of the affect annotation tool

him. STEP was first released in June 2012. The updated version (0.9.7) is used in this experiment. STEP was praised for its simple yet effective horror approach; for example, a famous gaming website IGN called this game "pure horror" [24]. The player is situated in the woods with the objective of collecting eight pages of paper scattered randomly around ten areas while avoiding being captured by the Slender Man. Without using any violent visual effects, STEP makes use of sound to create an atmosphere of being followed by the Slender Man. The game ends when the Slender Man is too close or when the player looks at him for too long. Hence, the player always has to pay attention to the appearance and presence of the Slender Man.

STEP uses an FPS control style where a player uses the mouse with their right hand to rotate the character's view while pressing keyboard on their left hand to move the character. `W`, `S`, `A`, and `D` keys on the keyboard are used for moving a character forward, backward, sidestep to the left, and to the right, respectively. The player can press the `Shift` key to activate running which consumes stamina but he cannot run when exhausted. Stamina regenerates slowly while the player is walking and fast while he stands still.

The game takes about 5–20 min regardless of whether the player collects all eight pages or gets captured by the Slender Man. The short gameplay time has a twofold advantage. First, it makes possible for participants to play the game multiple times. Secondly, it takes less time for participants to re-watch the gameplay video and do the annotation. The random placement of the pages, some random behavior of the Slender Man, and changes in the player's decision provide the player with various unpredictable encounters with the Slender Man and the game can keep players on their toes even though they have already played the game many times. These make STEP best suited for our experiments.

3.2 Affect annotation tool

The player's affective state changes dynamically as the game progresses; hence, it will be inaccurate to describe the game experience with a single affective state value. However, unlike in music perception where the self-reported annotation can be done in real-time [16], the concurrent annotation is impossible in an interactive environment because it imposes an additional cognitive load upon the user while the user interacts with the environment [17]. One way to handle this problem is to let the subject watch the recorded video of his activities and annotate his affective states continuously [20,34].

We developed an AAT where the player can annotate his affect continuously by watching his recorded gameplay and facial expression videos simultaneously.[2] MSI Afterburner[3] (v2.3.1) and Window Live Movie Maker[4] were used for recording gameplay and the player's facial expression videos, respectively. Figure 1 shows a screenshot of AAT wherein the gameplay video is playing at the left side and the player's reaction is shown at the right. The bars below the videos display the timeline showing the current game time and the annotation bar showing the self-reported affective states.

Our focus is on player affective responses prior to and after the appearance of the Slender Man because it is the main scary event of the game. It can appear multiple times during the game. An affective state is labeled as

– *Neutral* if there is no feeling of uncertainty towards the appearance of Slender Man,
– *Anxiety* if the player thinks that the Slender Man is near or is going to appear soon but does not know or cannot imagine how the Slender Man is going to show up, and
– *Suspense* if the player has a strong feeling of how the Slender Man is going to appear.

[2] While there are available labeling tools specific for the purpose, e.g., FEELTrace [30], we did not use such tools because they commonly play only one video at a time which requires an additional step of combining the videos before the annotation process.

[3] http://event.msi.com/vga/afterburner/.

[4] http://windows.microsoft.com/en-us/windows-live/movie-maker.

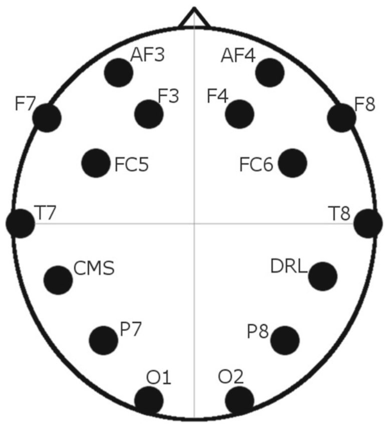

Fig. 2 Electrode placement positions of Emotiv EPOC

We call these three emotions *pre-fear affects*. After the Slender Man is seen, the player experiences *post-fear affects* which are categorized as levels of fear experienced, namely:

– *Low-Fear* if the player experiences no fear or a very low level of fear,
– *Mid-Fear* if a normal level of fear is experienced, and
– *High-Fear* if the player thinks the event is really scary.

The player annotates with these labels while watching the videos. We assume that the player stays on post-fear affects until a new scary event is anticipated and goes back to annotating the situation with pre-fear affects. AAT recorded the player affective states at one sample per second.

3.3 Electroencephalography

Electroencephalography is the recording of electrical activities of the brain and has been used in several literatures on emotion recognition [8,31,36]. We used the Emotiv EPOC,[5] which is a high resolution, multi-channel, wireless portable EEG system for recording EEG data.

Emotiv EPOC has 14 electrodes that are placed on the scalp based on the international 10–20 system at AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4 as shown in Fig. 2, and records electrical activities with a 128 Hz sampling rate.

To extract features from EEG signals, we calculated their fractal dimension (FD) values that reveal the complexity of the signals and quantify the concentration level of brain state [36]. FD has been used in several EEG-based emotion recognition researches using music or games as stimuli to elicit emotions [31,36]. We calculated the FD values using the Higuchi method [10] based on the algorithm proposed by Wang et al. [36]. For a time series $x(i)$ where $i = 1, \ldots, N$, a new series $X_m^k(i)$ is constructed as follows:

$$X_m^k \;:\; x(m),\; x(m+k),\; x(m+2k),\; \ldots,\; x\left(m + \left\lfloor \left(\frac{N-m}{k}\right) \right\rfloor k\right) \qquad (1)$$

where $k = 1, 2, \ldots, 2^{(\lfloor \log_2 N \rfloor - 4)}$ is the interval time and $m = 1, 2, \ldots, k$ is the initial time. Then the length of the series X_m^k is defined as:

$$L_m(k) = \frac{1}{k}\left[\left| \sum_{i=1}^{\lfloor \frac{N-m}{k} \rfloor} \left| x(m+ik) - x(m+(i-1)k) \right| \right. \times \frac{N-1}{\lfloor (\frac{N-m}{k}) \rfloor k}\right]. \qquad (2)$$

The total length $L(k)$ for $x(i)$ is obtained by averaging all the sub-series lengths $L_m(k)$ for the given k and then $L(k) \propto k^{-\mathrm{FD}}$. FD is obtained by drawing a logarithm plot between $L(k)$ and k and calculating the slope.

Different affective states might cause changes in EEG signals over different lengths of time. For example, the player might have confirmed that the Slender Man was in the position he expected and he went into the low-fear state for a short time, or that the Slender Man suddenly showed up and led the player to high-fear state for a long time. In order to reveal these EEG activities, FD values were calculated over the window size 1, 2, 5, and 10 s with 1-s sliding windows for each EEG signal obtained from each electrode. It means that we created a new time series of FD values with the same 1 Hz sampling rate with the affect data from AAT.

3.4 Electrocardiography

Electrocardiography measures the electrical activities generated by the heart over a period of time. From the EKG data, the HR and the inter-beat interval (IBI) can be calculated using a peak detection algorithm. The HR has been used as an indicator of both valence and arousal [18] and also as a biofeedback for game control [1,22].

We recorded the EKG signals at 2,048 Hz using the EKG-Flex/Pro sensor[6] by attaching three electrodes on the participant's chest and abdomen. EKG features (Table 1) were then calculated after the experiment by the Biograph Infiniti software that came with the EKG-Flex/Pro sensor. Every time a new heart beat is detected, `HR` and `IBI` are updated. `HR` is the number of heart beats for the recent 60 s and `IBI` is the time between the last beat and the current beat. To show the heart beat activity over longer period of time, the mean of `HR` (`HR_epoch_mean`) and the standard deviation of `IBI`

[5] http://www.emotiv.com/.

[6] http://www.thoughttechnology.com/.

Table 1 Features extracted from each signal

Name	Description
EEG	FD values of 14 EEG signals were calculated from {1, 2, 5, 10} s of the raw signals (total of $14 \times 4 = 56$ features), e.g., `AF4_10` is the FD value of AF4 signal calculated over 10 s (see Sect. 3.3)
EKG	Five EKG features were calculated directly by the Biograph Infiniti software (see Sect. 3.4)
HR/IBI	`HR`, `IBI` were calculated every time a new heart beat is detected
Epoch	`HR_epoch_mean`, `IBI_epoch_SD` were calculated from 20 s of data
NN50	Total number of two successive IBIs that differ by more than 50 ms
KMA	The following five KMA data were collected every second. The mean and the standard deviation within {2, 5, 10} s were also used (total of $5 \times 7 = 35$ features) (see Sect. 3.5)
Keyboard	The number of times each button is pressed per second (`move_pressed`, `run_pressed`) and how many buttons are being held over 1 s (`move_hold`, `run_hold`)
Mouse	Distance per second that the mouse has been moved (`mouse_speed`)

(`IBI_epoch_SD`) are calculated from 20 s of raw EKG data. NN50 is the total number of interval differences of successive heart beats that are greater than 50 ms [32]. All of the features were then resampled to 1 Hz.

3.5 Keyboard and mouse activities

Some works have shown that users' typing speed and/or keystroke patterns are different when they have different emotional states [13,35]. Players might show different behaviors when pressing keyboard buttons and/or moving the mouse while experiencing fear as well. Hence, we developed a KMA capturing tool that collects keyboard typing activities and mouse movements assigned to control the game. We are interested in the frequency of button presses that are used to move (`W`, `S`, `A`, and `D`) and run (`Shift`) the player's character. KMA collects the number of times the buttons are pressed per second and the number of buttons that are being held over 1 s. Also, KMA records every second the distance that the mouse has been moved. Similar to the EEG case, in order to observe KMA over different lengths of time, the mean and standard deviation were calculated with the window size of 2, 5 and 10 s with 1-s sliding windows.

3.6 Experiment procedure

All participants were first asked to fill out a questionnaire to obtain profile information such as age, gender, horror preference, experience in FPS game style control and experience in STEP. The participants were then asked to relax for 30 s and then watch the STEP trailer for another 90 s while EEG and EKG signals were recorded. The purpose of the relaxation phase is to capture the baseline values. Showing the trailer was necessary since the game itself has no prologue nor introduction to acquaint the participants with what to be careful and fearful of. Without the trailer, the participants might easily lose the first game by not running away from

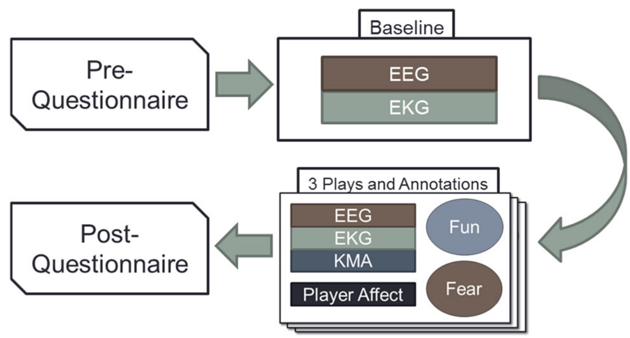

Fig. 3 Experiment procedure

the Slender Man. After recording the baseline values, the participants played STEP while the EEG, EKG, and KMA data were recorded. Immediately after finishing the gameplay, the participants annotated their affect using AAT. They were also asked to rate how much fear and fun they experienced in the gameplay using a 5-point likert scale ($1 = \text{least}$, $5 = \text{most}$). Play and annotation sessions were then repeated for two more times, thereby providing us with three sets of data per player. After finishing all three plays, the participants were asked to answer a post-questionnaire asking them (i) to rate their overall fear, fun, and the scare levels per game element, (ii) whether they felt differently playing the game for the second and/or the third times compared to the first, (iii) which element in the game was scary or fun, (iv) what was the cause of anxiety and suspense feelings, (v) whether they found the physiological devices intrusive, and (vi) to rate how much they remembered their emotions while annotating. Figure 3 summarizes the experiment procedure and the data we obtained from each player.

Table 1 summarizes all the features we used for analyzing the participants' fear transition processes. All features were normalized to a [0, 1] scale using the three gameplay data and the baseline data (except KMA which has no baseline) for each player.

Given a time series $x(i)$ where $i = 1, \ldots, N$ and the minimum and maximum values of $x(i)$, the normalized value of $x(i)$ can be calculated by the following formula:

$$x_{\text{normalized}}(i) = \frac{x(i) - \min_i\{x(i)\}}{\max_i\{x(i)\} - \min_i\{x(i)\}}. \tag{3}$$

3.7 Participants

Eleven participants (six males and five females) aged between 21 and 32 years old (mean = 26.45, SD = 3.11) took part in this experiment. There were three participants who considered themselves as gamers, two participants were not, and the other six participants considered themselves as casual gamers. Four participants did not like horror movies nor games, four participants felt neutral, and three participants liked to watch horror movies and/or play horror games. There were three participants who did not know the FPS type controls and five participants who knew the controls but were not familiar with them. There were two participants who had played STEP before and other two participants who knew STEP but had not played it before.

4 Results

We discuss the results of our experiments from three points of view. First, we examine the questionnaire results that show the participants' overall feelings and impressions towards STEP and this experiment. We then analyze the transition of the participants' affective states using the transition likelihood function. Finally, we show the potential of using physiological signals for automatically predicting player affect.

4.1 Questionnaire result

The participants were asked to rate how much fear and fun they experienced after each game and to rate the overall fear and fun with all games they played using a 5-point likert scale. The average of overall fear ratings was 3.82 (SD = 0.75) and the average of overall fun ratings was 3.46 (SD = 0.82). When looking at after-play ratings, the average of fear ratings was 3.6 (SD = 0.98), while the average of fun ratings was 3.44 (SD = 0.84). It is consistent in overall and after-play ratings that fear ratings were higher than fun ratings. It can be observed that three out of four participants who did not like horror media rated overall fear higher than overall fun; hence, they might not have enjoyed playing the survival horror game as much as the other participants.

The average fear ratings after the first, second, and third play of all participants were 3.5, 3.55, and 3.73, respectively, while the average fun ratings were 2.7, 3.64, and 3.91, respectively. ANOVA test clearly indicated statistical differences among the average fun ratings ($p < 0.001$), but did not show any statistical significance among the average fear ratings ($p > 0.05$). The reason why the participants had considerably more fun in the latter gameplays is probably because they were able to know STEP better and, for example, derived a better way to escape when the Slender Man was near. This allowed them to be more immersed in the game and experience more fun, especially for those who had not seen nor played STEP before. Although the fear ratings remained consistent across the three gameplays, many participants reported that they had felt less fear in the latter games. One participant mentioned that it had taken less time to feel relieved after seeing the Slender Man in the second and third games.

Table 2 Scariness ratings of game elements

Game element	Rating
Overall visual	3
Darkness	3.36
Slender man	3.27
Env. sound	3.72
Shock sound	4.91
Music	4.09
Pages of paper	2.09
Static screen	3.73

We also asked the participants to rate the scariness of game elements with a 5-point likert scale. They were (i) the overall visual of the game, (ii) the increase of darkness over time, (iii) the Slender Man, (iv) the environment sound such as wind sound, (v) the shock sound that was played when the Slender Man was sighted, (vi) the music being more intense as the game progressed, (vii) the messages on the pages of paper, and (viii) the static screen that showed up when the Slender Man was close to the player. Table 2 shows the average scariness ratings of the game elements asked in the post-questionnaire.

The shock sound was considered by most participants to be the scariest game element in STEP, while the music came second. Although the music that imitates the sound of footsteps caused participants to be aware of the Slender Man at first, as the same sound repeated over and over, the participants realized at some point that there is no relation between the music and the appearance of the Slender Man.[7] The shock sound, on the other hand, made participants know that the Slender Man was already in front of them. The feeling that they had to do something to get away from the Slender Man made the sound very scary to them even though they had already heard

[7] In fact, the music is related to the chasing speed which is difficult for participants to observe as the Slender Man stops moving when participants look at it directly.

it many times and not really seen the Slender Man yet. Other than sound, the static screen is a visual element that scared participants with an effect that is almost similar to the shock sound. The difference, however, is that, instead of informing participants that the Slender Man is in front of them, it just alerts them that the Slender Man is near. The probable reason why the Slender Man itself got a slightly low rating was that the Slender Man did not make any explicit action to attack the participants. Most of the time, the participants were just scared of losing rather than being scared of seeing the Slender Man. Rating on the environment sound is also higher than the visual and the darkness, which shows another evidence that the sound elements in this game have more impact to participants' emotions than the visual elements. This results also support the opinion of Parker and Hereema [25] that sound carries more emotional content than any other part of the game.

We also asked the participants to rate how much they remembered their emotions while annotating. The average and standard deviation were 4.27 and 0.47, respectively. This result shows that the participants had confidence in their annotated data.

We obtained no conclusive answer on what caused the participants to feel anxious or suspenseful. Many participants said that the music was the cause of anxiety because they knew that as soon as the music started, the Slender Man also started following them. However, some participants answered that the music made them experience suspense. This indicates that each participant might have his own way of differentiating anxiety and suspense, or it is also possible that the same game element might have had different effects on the participants. Alternatively, as there were many game elements showed through the course of the game, it might be difficult for the participants to recall exactly how they had reacted to certain game elements. It suggests the benefit of using real-time affect detection tools for mapping changes in player affect to game elements.

4.2 Transition likelihood

From the subjective affective states the participants reported using AAT, we can investigate how their affective states transitioned between pre-fear and post-fear affects with the appearance of the Slender Man. To measure the likelihood of transitioning from a pre-fear affective state to a post-fear one, we use the transition likelihood function introduced by D'Mello et al. [2]. The likelihood function, presented in Eq. (4), serves as a better measurement than probabilities because it takes the base rate of transition into account. All pairs of affect transitions from a pre-fear affect F_{pre} to a post-fear affect F_{post} are counted and used to compute the transition likelihood.

Table 3 Transition likelihood from pre-fear to post-fear

From\to	Low-Fear	Mid-Fear	High-Fear
Neutral	**0.06**	**0.22**	−0.26
Anxiety	**0.09**	**0.10**	−0.16
Suspense	−0.05	−0.12	**0.14**

Bold values indicate the positive transition likelihood

$$L(F_{pre} \rightarrow F_{post}) = \frac{\Pr(F_{post}|F_{pre}) - \Pr(F_{post})}{1 - \Pr(F_{post})}. \quad (4)$$

where F_{pre} is *Neutral*, *Anxiety*, or *Suspense* and F_{post} is *Low-Fear*, *Mid-Fear*, or *High-Fear*. $\Pr(F_{post})$ is the transition probability from any F_{pre} to the F_{post} and $\Pr(F_{post}|F_{pre})$ is the conditional probability of transitioning from the F_{pre} to the F_{post}. Transition likelihood L returns a value ranging from $-\infty$ to 1 where $L > 0$ indicates a likely transition with increasing likelihood as it approaches to 1, $L = 0$ indicates that the transition probability is equal to chance, and $L < 0$ means that the transition is less likely to occur compared to the base rate of transitioning into the F_{post} affect.

Table 3 shows the result of transition likelihood values for each pair between F_{pre} and F_{post} with the positive transition likelihood values highlighted. It indicates that the participants were likely to transition into High-Fear if they faced the Slender Man while experiencing Suspense. Neutral and Anxiety had positive transition likelihood values to both Low- and Mid-Fear, especially Neutral had higher likelihood towards Mid-Fear than Anxiety. This shows that the participants were likely to be surprised at the Slender Man when they were in Neutral state but not so much when in Anxiety state. Also, the participants were normally surprised only in the first game because they became more accustomed to the Slender Man in the latter games and were not likely to transition from Neutral and Anxiety to High-Fear.

This result suggests that stimulating player affect to the suspense state is the best way to maximize the effectiveness of producing a scary event. This result also supports Perron's study [26] that showed many successful examples of forewarning techniques used in horror movies and horror games. It suggests that long anticipation of a harmful confrontation (suspense) is more disturbing than short anticipation (surprise). In the case of affective survival horror games, player affect can be used as an input so that the game will detect if the previous forewarning technique was enough to cause the participants suspense. If not, the game can delay the scary event and make use of other game elements such as sound and a distorted vision to try to elicit suspense before actually showing the scary event to participants.

We also investigated the transition likelihood between F_{pre} states (Table 4) to see how STEP built up the participants' emotions before the Slender Man appeared. When the participants were in Neutral, they were likely to transition

Table 4 Transition likelihood between pre-fear affects

From\to	Neutral	Anxiety	Suspense
Neutral	–	**0.71**	−0.13
Anxiety	**0.40**	–	**0.25**
Suspense	−0.25	**0.81**	–

Bold values indicate the positive transition likelihood

Table 5 Transition likelihood between post-fear affects

From\to	Low-Fear	Mid-Fear	High-Fear
Low-Fear	–	**0.84**	−0.19
Mid-Fear	**0.40**	–	**0.31**
High-Fear	−0.30	**1**	–

Bold values indicate the positive transition likelihood

to Anxiety but not likely to transition directly to Suspense. While they were in Anxiety, they were likely to transition to both Neutral and Suspense, and when in Suspense, they were likely to move into Anxiety. These results show that STEP mainly built up the participants' emotions by moving them from Neutral to Anxiety, then to Suspense, and then back to Anxiety and then Neutral.

Unlike the transitions from pre-fear to post-fear states that are the results of the Slender Man's appearance, note that the transitions between pre-fear states can be caused by any events in the game. Sometimes the participants transitioned from Neutral to Anxiety when they collected a page or the music started playing. There were the cases where the participants' affects changed even without any changes in the game as well.

After seeing the Slender Man, participants move to one of F_{post} states and stay there until they start to anticipate a new scary event. There are also transitions between F_{post} states when the participants are experiencing the scary event. Transition likelihood between F_{post} states observed in this experiment is shown in Table 5. Similar to the transitions between F_{pre} states, the participants were likely to transition from Low- to Mid-Fear and from High- to Mid-Fear. From Mid-Fear, they were likely to transition to both Low- and High-Fear. There was no direct transition from High- to Low-Fear in the entire data and there were only few transitions from Low- to High-Fear.

To further investigate this, we looked into two step transitions between post-fear affective states. There were 17 transitions from Low- to Mid- to Low-Fear ($LML = 17$) and 7 transitions from Low- to Mid- to High-Fear ($LMH = 7$). On the other hand, there were 16 transitions from High- to Mid- to High-Fear ($HMH = 16$) and 10 transitions from High- to Mid- to Low-Fear ($HML = 10$). While one step transitions could not reveal much on how the transitions from Mid-Fear works, these two step transitions showed that the participants were likely to transition back to the previous state that they experienced before transitioning to the Mid-Fear state ($LML > LMH$ and $HMH > HML$). These results suggest that eliciting High-Fear from a player at the start of a scary event is better if the objective is to keep the player in the high fear state along the course of a scary event.

4.3 Classification result

We then tested the potential of physiological signals (EEG and EKG) and game inputs (KMA) in predicting horror-related affect. Although there were 11 participants in the experiment, we could not use EEG data from the first three participants as the recorded data contain a lot of noise from a damaged electrode, which left us with the data of eight participants ($P = 8$) for the classification. We combined and used all of participants data for classification process. Hence, the prediction models we got show the subject independent performance of the models. The techniques used to create the prediction models are C4.5 decision tree [28], multilayer perceptron (MLP) [38], and the simple logistic regression (SLR) [15]. All evaluations were done using 10-fold cross-validation. K-fold cross-validation is the standard way of validation in machine learning that separates data randomly into k set with equal size of data. Then, k−1 set of data are used as training data and other one set is used for testing the model. The process is then repeated for k times so that all the data is used as training and testing data. Prediction models were trained and evaluated separately using EEG, EKG, and KMA to predict pre-fear and post-fear affects. The reason why the models were built separately for pre-fear and post-fear affects is because (i) there is an imbalance between pre-fear affect instances (14,105) and post-fear affect instances (1,594), and (ii) these two sets of affective states can be easily classified by the context of the game (pre-scary and post-scary events).

The default parameters of C4.5 and SLR on Weka (version 3.7.5) [9] were used for classification. For MLP, however, the number of nodes in a hidden layer can greatly influence the performance of the classifier whereas it is sufficient to have only one layer to represent any functions. Hence, we tested how the f-measures change with the number of hidden layer nodes. Figures 4 and 5 show the f-measure of pre-fear and post-fear prediction models trained on MLP with the number of hidden layer nodes from 1 to 80 nodes. In pre-fear prediction models, the maximum f-measure (optimal number of nodes) for EEG, EKG, and KMA are 0.73 (70 nodes), 0.82 (63 nodes), and 0.57 (76 nodes), respectively, and 0.77 (64 nodes), 0.75 (62 nodes), and 0.59 (59 nodes) for the post-fear prediction models. While all of the models converged before the one with 80 nodes, we further tested the pre-fear EKG models by increasing the number of hidden layer nodes to 100, 200,. . ., 500 (with increments of 100). However, we found no further improvement as the f-measure stayed on 0.82.

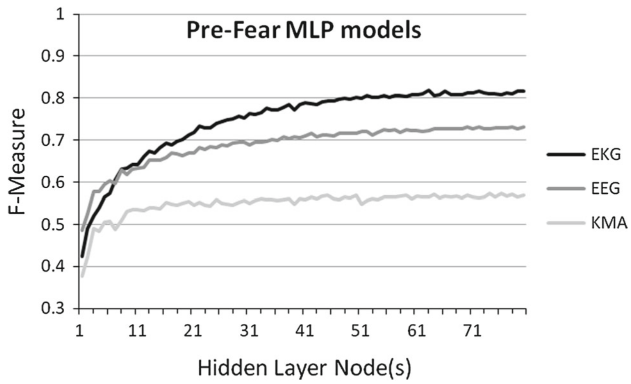

Fig. 4 F-measure of pre-fear prediction models trained with MLP with the number of hidden layer nodes from 1 to 80

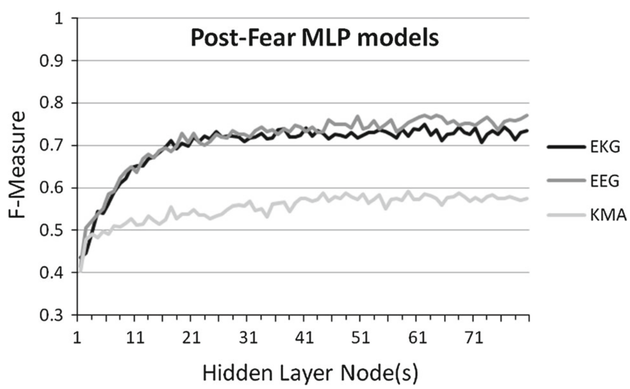

Fig. 5 F-measure of post-fear prediction models trained MLP with the number of hidden layer nodes from 1 to 80

Table 6 F-measure and kappa of general prediction models that predict pre-fear affects and post-fear affects

General	C4.5			MLP			SLR		
	EEG	EKG	KMA	EEG	EKG	KMA	EEG	EKG	KMA
Pre-Fear									
F-measure	0.72	**0.92**	0.65	0.73	0.82	0.57	0.55	0.43	0.38
Kappa	0.54	**0.87**	0.41	0.55	0.70	0.29	0.23	0.09	0.01
Post-Fear									
F-measure	0.68	**0.85**	0.64	0.77	0.75	0.59	0.49	0.44	0.49
Kappa	0.51	**0.77**	0.44	0.65	0.62	0.37	0.23	0.16	0.22

The models were built with C4.5, multilayer perceptron (MLP), and a simple logistic regression (SLR) algorithm using EEG, EKG and KMA features. The default configuration on Weka was used for the classification except for the number of hidden layer nodes in MLP, which were tuned by the preliminary classification
Bold value indicates the best accuracy value in that measurement

Table 6 show f-measure and Cohen's kappa of the general models created from the combined data of all eight participants. For both pre-fear and post-fear affects, the prediction models trained with C4.5 decision tree using EKG features achieved the best result. Pre-fear model trained with MLP performed better with EKG features, while EEG and EKG achieved close f-measure on post-fear MLP models. This result suggests that EKG features have the best potential in predicting player's pre-fear and post-fear affects.

Table 7 Precision and recall of each class from the prediction models trained with C4.5 and MLP using EKG features

Classes (instances)	Precision		Recall	
	C4.5	MLP	C4.5	MLP
N (7,508)	0.95	0.87	0.96	0.94
A (4,030)	0.87	0.74	0.86	0.70
S (2,567)	0.91	0.82	0.90	0.70
LF (624)	0.92	0.86	0.91	0.85
MF (588)	0.80	0.68	0.83	0.80
HF (382)	0.83	0.72	0.80	0.54

N neutral, *A* anxiety, *S* suspense, *LF* low-fear, *MF* mid-fear, *HF* high-fear

To explore the C4.5 models with EKG features further, we first looked at how the decision trees were created. `HR_epoch_mean` and `IBI_epoch_sd` had much higher information gain than the other features and had been chosen to be the top two nodes of the decision tree for both pre-fear and post-fear affects data. This shows that `HR_epoch_mean` and `IBI_epoch_sd` contained more relevant information for classifying horror-related affect than the other features.

Table 7 shows the results precision and recall per class from the C4.5 and MLP models trained with EKG features. For the C4.5 models, the precision and recall are considerably high for all of the classes, even though, the distribution of instances were not balanced. On the other hand, the precision and recall of MLP models were lower and not as equivalent to each other like the C4.5 models. Especially, the recall of anxiety, suspense and high-fear classes became very low when compared to the other classes. This shows that the performance of MLP models might suffer from the problem of imbalanced data, while the C4.5 models show more robustness performance.

We also tested how the models will do when it has different split percentage of training and test set by using different fold (2–15) cross-validations. With twofolds (50 % training data), the f-measure decreased to 0.90 (from 0.92) and 0.80 (from 0.85) for pre-fear and post-fear, respectively. Nevertheless, the overall accuracy did not change much as the number of folds were changed. The average f-measure of pre-fear and post-fear models over 2–15 folds cross-validation were 0.92 (SD= 0.01) and 0.84 (SD= 0.02), respectively.

Although the prediction accuracy of EEG features was not as high as that of EKG features, it is relatively high and shows some good potential in predicting horror-related affect. KMA got the lowest accuracy compared to the other two signals, where f-measure became only 0.65. Note that although KMA might not be as accurate as using physiological signals, it does not require any additional setup and can be used in every game.

These results showed the potential of physiological signals in predicting the participants' self-reported horror-related affect. This type of data collection and prediction can also be applied on any games and any types of emotions, given that participants can recall their emotions by watching their videos. Although EKG got better performance than EEG and KMA, this work only showed that EKG is the best suited for these kinds of affects; the other two modalities might show better performance in other contexts.

5 Conclusion and future work

In this paper, we investigated several aspects of player affect in survival horror games to find out the possibility of integrating the affective gaming concept into the survival horror genre. Brainwaves, heart rate signals and KMA were collected while participants were playing a survival horror game called "Slender: The Eight Pages". The participants were then asked to annotate their affect using our own AAT after finishing each game. The affects were divided into pre-fear affects (Neutral, Anxiety, and Suspense) and post-fear affects (Low-, Mid-, and High-Fear). We analyzed the transition likelihood between the annotated affect and used them as labels for classification. Questionnaire results showed that participants were having more fun to play the game as they were getting used to the system, but getting less fear after repeated the game a few times. The participants also identified that the sound is the main scary element in the game, and it can elicit both anxiety and suspense from the participants. The transition likelihood between pre-fear and post-fear affects revealed that the participants were likely to get more fear by a scary event when they were in the suspense state. Finally, the classification result revealed the potential of EKG signals in predicting both pre-fear and post-fear affects. It should be noted that this work is a pilot study at present because of the small number of participants in the classification, and another experiment with more participants has to be conducted.

As a future work, we want to create our own game so that we can further analyze the relationship between player affect and each game element. It is also possible to learn which game element is the best suited for causing anxiety or suspense to a certain player. Combine all of the findings, our goal is to develop an affective survival horror game where the timing of each game element can be changed in response to player's current emotion, which aims to lead the player to the most fearful experience.

Acknowledgments This work was partly supported by JSPS Core-to-Core Program, A. Advanced Research Networks.

References

1. Dekker A, Champion E (2007) Please biofeed the zombies: enhancing the gameplay and display of a horror game using biofeedback. In: Proceedings of digital games research association (DiGRA) 2007 conference, pp 550–558
2. D'Mello S, Taylor R, Graesser A (2007) Monitoring affective trajectories during complex learning. In: Proceedings of the 29th annual conference of cognitive science society, pp 203–208
3. Drachen A, Nacke LE, Yannakakis G, Pedersen AL (2010) Correlation between heart rate, electrodermal activity and player experience in first-person shooter games. In: Proceedings of the 5th ACM SIGGRAPH symposium on video games, pp 49–54. doi:10.1145/1836135.1836143
4. Fanselow MS (1994) Neural organization of the defensive behavior system responsible for fear. Psychon Bull Rev 1(4):429–438. doi:10.3758/BF03210947
5. Garner T, Grimshaw M (2011) A climate of fear: considerations for designing a virtual acoustic ecology of fear. In: Proceedings of the 6th audio mostly—a conference on interaction with sound, pp 31–38. doi:10.1145/2095667.2095672
6. Garner T, Grimshaw M, Abdel Nabi D (2010) A preliminary experiment to assess the fear value of preselected sound parameters in a survival horror game. In: Proceedings of the 5th audio mostly—a conference on interaction with sound. doi:10.1145/1859799.1859809
7. Giakoumis D, Tzovaras D, Moustakas K, Hassapis G (2011) Automatic recognition of boredom in video games using novel biosignal moment-based features. IEEE Trans Affect Comput 2(3):119–133. doi:10.1109/T-AFFC.2011.4
8. Gilleade KM, Dix A, Allanson J (2005) Affective videogames and modes of affective gaming: assist me, challenge me, emote me. In: Proceedings of digital games research association (DiGRA) 2005 conference, pp 547–554
9. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The weka data mining software: an update. SIGKDD Explor Newsl 11(1):10–18. doi:10.1145/1656274.1656278
10. Higuchi T (1988) Approach to an irregular time series on the basis of the fractal theory. Phys D 31(2):277–283. doi:10.1016/0167-2789(88)90081-4
11. Hudlicka E (2009) Affective game engines: motivation and requirements. In: Proceedings of the 4th international conference on foundations of digital games, pp 299–306. doi:10.1145/1536513.1536565
12. Jennett C, Cox AL, Cairns P, Dhoparee S, Epps A, Tijs T, Walton A (2008) Measuring and defining the experience of immersion in games. Int J Hum Comput Stud 66(9):641–661. doi:10.1016/j.ijhcs.2008.04.004
13. Khanna P, Sasikumar M (2010) Recognising emotions from keyboard stroke pattern. Int J Comput Appl 11(9):1–5. doi:10.5120/1614-2170
14. Krzywinska T (2002) Hands-on horror. Screenplay: cinema/videogames/interfaces. Wallflower Press, London, pp 206–223
15. Landwehr N, Hall M, Frank E (2005) Logistic model trees. Mach Learn 59(1–2):161–205
16. Lane JS (2012) The effect of performance experience on vocal music major's perception of musical tension. J Res Music Perform. ISSN: 2326–1722
17. Laurans G, Desmet PM, Hekkert P (2012) Assessing emotion in human–product interaction: an overview of available methods and

a new approach. Int J Prod Dev 16(3–4):225–242. doi:10.1504/IJPD.2012.049835
18. Mandryk RL, Atkins MS (2007) A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. Int J Hum Comput Stud 65(4):329–347. doi:10.1016/j.ijhcs.2006.11.011
19. Martinez HP, Garbarino M, Yannakakis GN (2011) Generic physiological features as predictors of player experience. In: Proceedings of the 4th international conference on affective computing and intelligent interaction, part I, pp 267–276
20. Metallinou A, Narayanan S (2013) Annotation and processing of continuous emotional attributes: challenges and opportunities. In: Proceedings of the 10th IEEE international conference and workshops on automatic face and gesture recognition. doi:10.1109/FG.2013.6553804
21. Nacke L, Lindley CA (2008) Flow and immersion in first-person shooters: measuring the player's gameplay experience. In: Proceedings of the 2008 conference on future play, pp 81–88. doi:10.1145/1496984.1496998
22. Nacke LE, Kalyn M, Lough C, Mandryk RL (2011), Biofeedback game design: using direct and indirect physiological control to enhance game interaction. In: Proceedings of the SIGCHI conference on human factors in computing systems, pp 103–112. doi:10.1145/1978942.1978958
23. Öhman A (2007) Anxiety. Encyclopedia of stress, 2nd edn. Academic Press, New York, pp 236–239
24. Onyett C (2012) Slender is pure horror. http://www.ign.com/articles/2012/07/07/slender-is-pure-horror. Accessed 22 Jan 2014
25. Parker JR, Heerema J (2008) Audio interaction in computer mediated games. Int J Comput Games Technol 2008:178923. doi:10.1155/2008/178923
26. Perron B (2004) Sign of a threat: the effects of warning systems in survival horror games. In: Proceedings of the 4th international conference on computational semiotics, Split
27. Prieto-Pablos JA (1998) The paradox of suspense. Poetics 26(2):99–113. doi:10.1016/S0304-422X(98)00014-X
28. Quinlan JR (1993) C4.5: programs for machine learning. Morgan Kaufmann, San Francisco
29. Roux-Girard G (2011) Listening to fear: a study of sound in horror computer games. In: Grimshaw M (ed) Game sound technology and player interaction. IGI Global, Hershey, pp 192–212
30. Schröder M, Cowie R, Douglas-Cowie E, Savvidou S, McMahon E, Sawey M (2000) 'FEELTRACE': an instrument for recording perceived emotion in real time. Proceedings of the ISCA workshop on speech and emotion. Textflow, Belfast, pp 19–24
31. Sourina O, Liu Y, Nguyen MK (2012) Real-time EEG-based emotion recognition for music therapy. J Multimodal User Interfaces 5(1–2):27–35. doi:10.1007/s12193-011-0080-6
32. Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology (1996) Heart rate variability: standards of measurement, physiological interpretation, and clinical use. Circulation 93(5):1043–1065. doi:10.1161/01.CIR.93.5.1043
33. Toprac P, Abdel-Meguid A (2011) Causing fear, suspense, and anxiety using sound design in computer games. In: Grimshaw M (ed) Game sound technology and player interaction. IGI Global, Hershey, pp 176–191
34. Truong KP, Neerincx MA, van Leeuwen DA (2008) Assessing agreement of observer- and self-annotations in spontaneous multimodal emotion data. In: Proceedings of the 9th Annual Conference of the International Speech Communication Association, pp 318–321.
35. Tsui WH, Lee P, Hsiao TC (2013) The effect of emotion on keystroke: an experimental study using facial feedback hypothesis. In: Proceedings of the 35th annual international conference of the IEEE engineering in medicine and biology society, pp 2870–2873. doi:10.1109/EMBC.2013.6610139
36. Wang Q, Sourina O, Nguyen MK (2011) Fractal dimension based neurofeedback in serious games. Vis Comput 27(4):299–309. doi:10.1007/s00371-011-0551-5
37. Weber R, Behr KM, Tamborini R, Ritterfeld U, Mathiak K (2009) What do we really know about first-person-shooter games? An event-related, high-resolution content analysis. J Comput Mediat Commun 14(4):1016–1037. doi:10.1111/j.1083-6101.2009.01479.x
38. Witten IH, Frank E, Hall MA (2011) Data mining: practical machine learning tools and techniques, 3rd edn. Morgan Kaufmann, Amsterdam
39. Yannakakis GN, Hallam J (2008) Entertainment modeling through physiology in physical play. Int J Hum Comput Stud 66(10):741–755. doi:10.1016/j.ijhcs.2008.06.004