Pleasure–arousal–outlier model for quantitative evaluation of game experiences

Shinjin Kang[1] · Soo Kyun Kim[2]

Abstract

This study proposes a pleasure–arousal–outlier (PAO) model to quantify the experiences derived from games. The proposed technique identifies pleasure, arousal, and outlier levels based on the facial expression of a user, keyboard input information, and mouse movement information received from a multimodal interface and then projects the received information in three-dimensional space to quantify the game experience state of the user. Facial expression recognition and distribution, eye blink, and eye glance concentration graphs were introduced to determine the immersion levels of games. A convolutional neural network-based facial expression recognition algorithm and dynamic time warp-based outlier behavior detection algorithm were adopted to obtain numerical values required for the PAO model evaluation. We applied the proposed PAO model for first-person shooter games and consequently acquired evaluation result values that were clearly distinguishable for different players. Such information allows player experiences to be quantitatively evaluated when designing game levels.

Keywords Game play evaluations · Game experiences · Affective gamming · Facial expression

1 Introduction

Recently, the digital content service market, which includes games, cartoons, movies, and television, shows has been growing rapidly on various platforms. This growing market has increasingly attracted investments in content creation; however, it has

✉ Soo Kyun Kim
  kimsk@jejunu.ac.kr

  Shinjin Kang
  directx@hongik.ac.kr

[1] Hongik University, Sejong 30016, Korea

[2] Jeju National University, Jeju 63243, Korea

also increased loss triggered by the failure of the content to gain popularity after development.

These risks can be minimized by accurately identifying the satisfaction derived from content. Various attempts have been made in the affective computing field to quantify content satisfaction. However, the evaluation of content satisfaction entirely depends on the voluntary feedback of individuals. The challenges posed by such subjective and personalized evaluation elements in the content industry limit the introduction of automated quantitative evaluation techniques. Particularly, in the game industry where long development periods are required, the demand for quantitative satisfaction evaluation is significantly high; however, such an evaluation is yet to be practically introduced in the game industry, owing to the absence of reliable evaluation models and the inaccuracy of evaluation results.

This study proposes a satisfaction or fun quantification model that is practically beneficial in developing games and verifies its effectiveness by actually using game play feedback. The objective of this study is to identify pleasure, arousal, and outlier levels that appear to be meaningful in the game feedback during the game development stage, as well as evaluate the game play based on the identified levels. The proposed model was obtained by expanding a pleasure–arousal–dominance (PAD) model [31] that was used in conventional affective evaluation models and suitable for games. The "dominance" in the PAD model could be suitable for affective evaluation in daily life. However, it is not significantly meaningful and is a difficult parameter to measure. Therefore, we excluded this parameter and instead introduced an outlier parameter that can be measured and viewed as information required for actual game development. To evaluate the experience of a game player in an ordinary PC gaming environment, we used three ordinary gaming interfaces (keyboard, webcam, and mouse); we did not employ any other equipment, because professional biometric signal equipment (for example, galvanic skin response and brainwave monitors) can elucidate the inherent reaction of a user but cannot be practically introduced in the industry owing to the inconvenience of wearing the equipment and difficulty in analysis of the generated data. The proposed methodology can collect data required for evaluation while the user naturally plays the game in an ordinary gaming environment, which has the advantage of increasing the evaluation accuracy by preventing the user from recognizing the situation as that for measurement and thus responding differently. In addition, continuous machine learning training is possible because data can be easily collected. Consequently, the accuracy of evaluation results will be improved in the long term.

The proposed approach differs from conventional game emotion evaluation methods because the numerical measurement accuracy of evaluation of each game component is increased by introducing a previously verified machine learning algorithm in each component. Compared to previous works, the features of our approach are presented as follows.

– **PAO model**: We proposed a novel evaluation model specialized in game play evaluation. The model can easily measure data better than the conventional PAD model and extract various additional derivative indices. A more beneficial insight can be provided in the game development stage.

– **Machine learning-based multimodal evaluation system**: To integrate the evaluation parameters, we proposed a framework that stably extracts the parameters required to evaluate the satisfaction derived from games from three interfaces (webcam, keyboard, and mouse) that are essential in the ordinary gaming environment.

2 Previous works

The methods for analyzing and incorporating game user data into games have been improving steadily. The games available in the game market and currently enjoyed by users have been developed through such game user data analyses. Game user data can be collected in two ways. The first is the survey-based opinion data, and the second is the user play log data obtained during game play. Particularly, the survey-based methodology, which indicates the state of a user by the responses of the user, is mostly adopted in fields where subjective decisions are required.

Hudlicka [1] used the term "affective gaming," to demonstrate the importance of the adaptation of the game to the emotion of a player, the ability of the game to detect the emotion of the game player, and adjust the game reaction to the emotion. Lee [2] introduced a method for predicting whether and when a player will stop playing a commercial game during its free play period using a log-based analysis scheme in the game and subsequently applied the method to measure user departure [3]. In addition, the game behavior of the player is observed and behavior similarities are grouped using game play behavior analysis via game bot detection [4].

The game behavior analysis of a player is also connected with emotion recognition, which is a very important element and subject in the affective computing field. Research on emotions in the laboratory has various problems triggered by technical issues. Banziger et al. [5] suggested several ways to motivate the creation of the Geneva multimodal emotion portrayal, described the overall advantages and limitations of emotion research, and defined the emotional behavior of a user in multiple modes. Petrantonakis et al. [6] addressed emotion measurement problems posed by quantitatively estimating the emotion-inducing level of a subject under a proper stimulus and by introducing new feature vectors to improve the performance of an electroencephalography-based emotion recognition system. Wang et al. [7] analyzed the cardiovascular reaction measured using electrocardiogram (ECG) equipment to recognize emotions, classified the elements that were extracted from the ECG and reflected emotions, and recognized two types of emotions, pleasure and sadness [7]. Wagner et al. [8] used a k-NN, linear discriminant function, and multilayer perceptrons to identify anger, sadness, happiness, and pleasure, which are physiological data, to achieve a recognition rate of 80%. Leon et al. [9] measured and adopted heart rate, skin resistance, and blood pressure to recognize neutral, negative, and positive emotions, which can be considered as three types emotions. McDuff et al. [10] adopted remote detection of cognitive load used in stress measurement to measure heart rate and heartbeat alterations; using this detection scheme, they measured the stress recognition level of a user. Their system could predict stress more accurately than conventional systems using physiological parameters (heart rate,

respiratory organ, and heart rate variability). Google Glass is a wearable computer equipped with an accelerometer, camera, Bluetooth, and transparent display. This device is similar in appearance to eyeglasses, which enables a person to wear it like eyeglasses; it collects the personal information of the person wearing it. Hernandez et al. [11] used Google Glass to introduce a technique for evaluating the current state of the wearer unassisted by other devices. The wearer can check and control their state based on data using their proposed device. Facial expressions are a typical research subject in anthropology, sociology, neuroscience, psychology, and human-computer interface (HCI); in addition, they are studied with considerable interest, particularly in the emotion research field. Ekman et al. [12] objectively and comprehensively described the relationship between what people display on their faces and what they feel using a facial action coding system (FACS). Kapoor et al. [13] analyzed head movements, postures, facial expressions, skin conductivity, and computer mouse pressure to propose a system that senses frustration in a learning environment. Kaliouby et al. used a wearable camera and various sensors combined with a machine recognition algorithm [14] to record and analyze the facial expressions and head movements of a person interacting with another person wearing the camera. Littlewort et al. introduced pain to an expression detection system [15]. Yeasin et al. [16] suggested a spatiotemporal approach to recognize six common facial expressions and calculate interest levels of the subject. In this study, we analyzed multimodal interface-based affective computing in the game industry to suggest a method for quantifying the pleasure derived from games. Nosu et al. [17] proposed a real-time emotion diagnosis system capable of determining emotions from the expressions of a video game player and voice feedback messages. Vachirapanang et al. [18] introduced a system to detect game addiction, whereas Yun [19] introduced stress detection in games. Tan et al. [20] used a first-person shooter game to evaluate player experiences. Hazelett [21] introduced a method that adopts facial electromyography to measure the emotional state of a user. Tan et al. [22] proposed Bro-cam, which provides empathetic feedback based on the physical posture of a player. Zoeller [23] introduced an information collecting system for developers and testers to interact regarding unfinished games, as well as improving their interaction. Martinez [24] used a deep learning algorithm to recognize four types of emotions. Recently, Wang et al. [25] automatically extracted the components of game experiences from online reviews by analyzing texts. Maman et al. [26] examined group cohesion and presented GAME-ON (Group Analysis of Multimodal Expression of cohesiON). Song et al. [27] used a multimodal game frustration database (MGFD) and a long short-term memory recurrent neural network for frustration classification, to analyze facial and vocal expressions and propose automatic frustration recognition. Ringer et al. [28] introduced a method that adopts tensor decomposition for the high-order fusion of multimodal expressions; their system provides a dataset of League of Legends livestreams annotated for streamer influence and game context. Sekhavat el al. [29] investigated the degrees to which the expression of emotions by an opponent could affect the emotions of players and consequently their gameplay behavior. Their findings suggest that expressing the emotional state of an opponent through an emoji in real time affected the emotional state and behavior of players and consequently their playing experience. Kar et al. [30] proposed a hybrid

brain computer interface designed to detect failures in player's gameplay actions. They introduced a simple indoor game, in which a player must pass a ball through a ring fixed on a variable pan-tilt platform.

In this study, we propose various indices capable of stable measurement in the ordinary gaming environment and proposed a novel PAO evaluation model based on the indices. Our approach differs from that of previous studies because it expands the biometric signal and face detection techniques [14–16] to incorporate mouse input into user reaction evaluation and thus improves the accuracy of immersion level measurements in games.

2.1 Evaluation Model

In this study, emotional quantification is achieved via the machine learning of facial expression and game-related interface data measured by a multi-modal device. Existing emotion evaluation processes usually involve surveys and external observation. Despite the participation of experts and inclusion of sophisticated analysis that incurs high costs, the results might be subjective. To address the problems of these emotion evaluation processes, computational cognitive psychology techniques need to be applied to quantify numerical values. Accordingly, we adopted a PAD model [31], which has been tested in the field of cognitive psychology. The PAD model classifies the emotion-individual-content interaction into pleasure, arousal, and dominance, and measures the emotional reactions sensed by the individual. This model proves to be a highly suitable model that may express emotions responding to the content; it exhibits a relatively high level of discrimination when applied to the targets of this study, which include the game, web, and smartphone environments. In addition, the PAD model is suitable for identifying general emotional states related to content feedback; however, to apply machine learning to special content such as games, it is necessary to make an adjustment in the level of measurement. The pleasure axis is an important element especially useful in determining the satisfaction felt in a virtual world. The arousal axis represents the amount of emotion felt in the virtual world. These two axes are slightly more important than the dominance axis. In this study, the dominance axis in the PAD model is excluded, and the pleasure–arousal axis is solely adopted to quantify emotional states within a game. Our goal was to determine the levels of pleasure and arousal from facial expression data and detect outliers from the interaction levels of a mouse and a keyboard, to prevent redundant cases when determining arousal and outliers (e.g., very surprised or sad expressions can always be mistaken as outliers). The outlier index in this study is focused on detecting outlier behavior in the interaction.

3 System

This study proposes a multi-modal interface-based emotion analysis system. The proposed system comprises three modules: (1) a convolutional neural network (CNN)-based facial expression detection module, (2) an outlier behavior detector
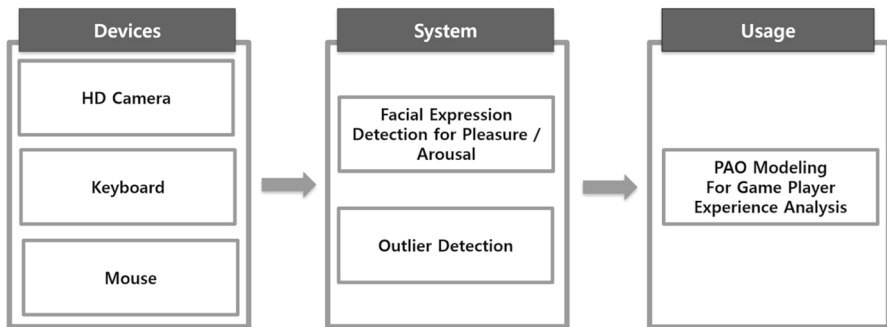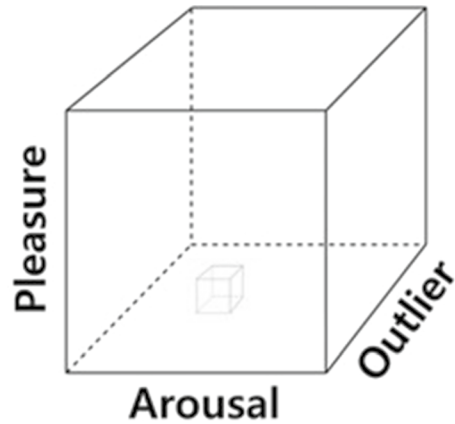
Fig. 1 Proposed PAO model

Fig. 2 Process of the proposed system

based on a dynamic time warp (DTW) algorithm, and (3) a PAO modeling system (see Fig. 1). This system synchronizes data gathered from three devices (camera, keyboard, and mouse) on the same time axis and records the data in CSV and face image files. Subsequently, the CSV files are imported to detect facial expressions and outliers. The proposed system is obtained by extending the facial expression analyzer and outlier detectors of previous studies [32]. Although previous studies focused on detecting individual parameters stably, we gathered information on multiple parameters and modeled data to be used in the industry. Figure 2 presents the process of our proposed system.

We acquired the numerical values required for the final PAO model in two stages. Data collected in each terminal were sent to the facial expression and outlier detection modules and were converted into values on the three axes of the PAO model. The pleasure and arousal levels were determined by recognizing the facial expressions from the webcam, whereas the outlier level in the game interaction was determined from the keyboard and mouse via time-series data analysis. The facial expression detection and PAO evaluation modules are described in this section. Figure 3 illustrates the relationship between each module and the parameters.
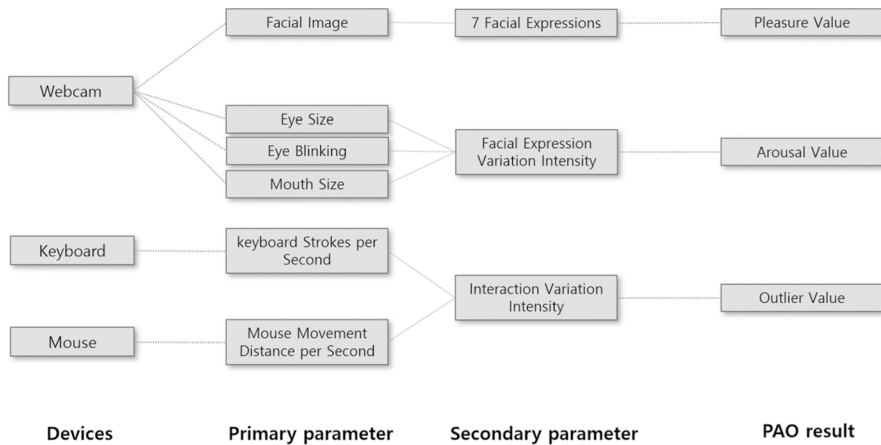
Fig. 3 Relationship between the parameters used in each step and the final PAO model

3.1 Facial expression detection

The webcam captures face images of a user in high definition. Facial expressions are input via the webcam device, which enables non-obtrusive contact. The input image is transferred to the CNN-based facial expression detection module first, which classifies the facial expression as one of eight expressions. The proposed CNN-based facial expression classifier is an extension of the facial expression classifier used in Ho's study (2016) [33]. The dataset used for training the model is from a past Kaggle Facial Expression Recognition Challenge (FER2013). It comprises a total of 35887, 48 × 48-pixel grayscale images of faces, each labeled with one of the seven emotion classes (anger, disgust, fear, happiness, sadness, surprise, and neutral). The model comprises three sequential convolution layers (32, 64, and 128 feature maps). Softmax activation function is adopted at the output layer. The model performs well in classifying positive emotions resulting in relatively high precision scores for the "happy" emotion (76.7%) and low precision for the "sad" emotion (39.7%). Our goal is to map the main facial expressions detected while playing a game onto the one-dimensional pleasure axis. This requires seven detected expression results to be mapped onto the pleasure axis. We assigned +0.1 (maximum of 1) to two emotions (happy and surprise) and –0.1 (minimum of –1) to the remaining four emotions (anger, disgust, fear, and sadness). The purpose of this assignment was to obtain the overall similar data distributions. No value was incorporated in the neutral case. We set the unit time of emotion evaluation to 1 min. Because images are sampled at 6-s intervals within 1 min, 10 evaluations take place in 1 min. The total evaluation value was finally determined as the pleasure value of the corresponding period. This process generates a pleasure graph where the facial expressions of the user are assigned values between 0 and 1. The tension graph is in arbitrary time intervals (e.g., 1 min) and can be discretized with the average of the data collected during the corresponding time. This reflects the requirement of the industry that feedback should be received in main game play interval units. Equation 1 presents the equation used in

the pleasure-axis calculation. Six emotion evaluation values (S_happy, $S\_surprise$, $S\_sad + S\_disgust + S\_fear + S\_angry$) were summed, and positive and negative weights (W_1 and W_2) can be assigned as required.

$$\mathrm{Pleasure Value} = W_1(S_{\text{happy}} + S_{\text{surprise}}) + W_2(S_{\text{sad}} + S_{\text{disgust}} + S_{\text{fear}} + S_{\text{angry}}) \quad (1)$$

Although a CNN-based algorithm was adopted for the facial expression recognition, a facial feature-based algorithm was used for the arousal detection. We adopted dlib [34] to extract a total of 68 feature points from a face and separately selected the angles between main feature points that are considered to be required for the outlier behavior analysis. We used angles, instead of lengths, to eliminate the effect of the distance between the camera and face. Various feature angle definitions are available; however, we extracted the corresponding angles of only two portions (mouth and eyes) from an experimental perspective. Observing the face of a player while the player plays a game shows that the player's mouth shape changes drastically. Because the mouth is the largest among the main components of the face, it makes up the largest part of the overall face change. The results of the pre-experiment confirmed that this input dimensionality reduction process slightly improved classification accuracy. We defined two vertices (left and right vertices) on the mouth and used the angles at the vertices to detect mouth changes. The unit of the angles is radians. We evaluated the difference between the measured amount of change in the eyes and mouth ($Angle\_eye$, $Angle\_mouse$) and the average amount of change ($Angle\_avgeye$, $Angle\_avgmouse$ as the difference in arousal.

$$\mathrm{Arousal Value} = ||Angle_{\text{eye}} - Angle_{\text{avgeye}}|| + ||Angle_{\text{mouse}} - Angle_{\text{avgmouse}}|| \quad (2)$$

Another objective of this study was to detect eye alterations. Our experience indicated that while playing a game, the eyes blink rapidly or become wider before and after encountering new interesting events. The pupils hardly change when playing the game in a monotonous manner. We interpreted intervals during which the pupils change as situations where the player pays attention owing to stimulations from the game. The shape changes of the eyes and mouth vary significantly depending on the physical features of each person. Consequently, we decided to determine the change information of the user using differences from the overall average values of the two parameters by calculating the corresponding averages after each experiment ends. The level of arousal may vary depending on individuals and game genres. It is difficult to evaluate a player who plays a multi-player game and actively communicates with other users based on the same reference value of arousal that is used to evaluate players who play single-player games. We recognized that such change rates are difficult to evaluate absolutely; therefore, we decided to perform relative evaluations and compare the change rates between experiment participants.

We included the concentration graph to additionally evaluate increased awareness. This is based on our experience that the user blinks less and that the facial expression changes less when the user concentrates on specific tasks. We defined any period during which facial expression changes less than a prescribed level and when eye glance is maintained as immersion intervals. This immersion graph is available as an analysis aid in the content analysis stage.

3.2 Game interaction detection module

The webcam captures face images of a user in high definition. Facial expressions are input via the webcam device, which enables non-obtrusive contact. The input image is transferred to the CNN-based facial expression detection module first, which classifies the facial expression as one of eight expressions.

Log data relating to the movement and usage of the user's keyboard and mouse were significantly beneficial in determining the user's current state. In this study, during system operation, the user's keyboard input and mouse trajectory, as well as input quantity, were tracked. The keyboard input data include all input key information of the user and are stored in the ASCII format. The key input rate and zero interaction time intervals during which no key input takes place can also be identified. The position values within the screen when the left and right mouse buttons are clicked are recorded as the mouse input data. The mouse trajectory information is generated with the mouse input data and indicates whether concentrated interactions of the user have occurred in a certain area. A mouse click heatmap is also available on the screen as mouse movement information. A mouse trajectory heatmap helps identify main click areas on the screen. We can detect a variety of keyboard and mouse input information in detail and harness the information obtained from our analysis. However, in this study, we mostly used the keyboard and mouse data to detect outlier behavior intervals. We calculated the average number of keyboard inputs and average mouse movement distance in 1-min units to separate maximum input value and movement distance values per minute to generate time-series interaction strength graphs. The time-series interaction strength graph is likely to exhibit similar patterns at a certain game level in a single-player game because a corresponding stage can be cleared only if inputs of a certain level or higher take place in the single-player game where scenarios that cannot be modified must be experienced. Any deviation from the input strength or input pattern can be interpreted as a situation where the corresponding player is not performing ordinary actions. Although outlier behavior can be easily determined by examining the mouse and keyboard input strength, we added the shape similarity between interaction graphs as a decision element by applying the DTW algorithm [35]. The DTW algorithm we employed follows the methodology of our previous behavior detection study [32]. The Euclidean distance algorithm, which is often used as a measure of the distance between two time series data, has very low similarity calculation accuracy when the time axis is twisted. In contrast, DTW uses a method of deriving an optimal path by searching the minimum distance between time series data. Through this, it is possible to compensate for the disadvantage of very poor performance due to a different time reference point or a different time axis scale when measuring the similarity of time series data. Because the DTW algorithm can calculate the similarity based on the shapes of graphs, regardless of the length of the time axis, it can compare the differences between the graph shapes. In our previous study, an interval in which similarities between graphs are compared was based on the entire play interval in which the corresponding experiment was conducted. However, in this paper, the time-series intervals in which similarities between graphs are

compared are narrowed to 10 min before and after the corresponding time. The intervals of the graph pattern comparison were narrowed to compare local pattern similarities and improve accuracy. Equation 3 presents the outlier value equation we adopted. The Outlier Value is calculated by adding the difference between the average value (G_i) and the interaction intensity value (G_i_Avg) of the keyboard and mouse, and the difference from the average interaction graph calculated by DTW (G_DTW, G_DTW_Avg).

$$OutlierValue = W_1|G\_i - G\_i\_Avg| + W_2|G\_DTW - G\_DTW\_Avg| \qquad (3)$$

3.3 Pleasure–arousal–outlier evaluation module

The pleasure, arousal, and outlier values calculated from the two modules are gathered and projected onto the three-dimensional axes between 0 and 1. The state of the user at a specific play time can be comprehensively determined from each value. We applied the parameters extracted in the PAO model evaluation process to further visualize a total of six indices for intuitive game play interpretation. The six indices include (1) the radial graph of facial expressions, (2) triangular PAO graph, (3) average keyboard/mouse input strength, (4) facial expression change graph, (5) eye blink change graph, and (6) eye glance concentration graph.

(1) The radial graph of facial expressions presents the distributions of the entire seven facial expressions detected while playing the game. The most derived emotion while playing the game can be identified from this radial graph. (2) The PAO graph shows values between 0 and 1 for the pleasure, arousal, and outliers axes. The higher the pleasure value, the more is the positive emotional expression of the user while playing the game. The higher the arousal value, the stronger the emotional change of the user. The higher the outlier value, the more the player differs from the input patterns of other players. This can be interpreted as the corresponding player's failure to adapt to the corresponding game or a different approach used by the player while playing the game. Here, P, A and O value are processed with the sigmoid function and then rescaled to the unit interval with the min-max function. (3) The average interaction strength graph presents the input strength of the corresponding player in comparison to the average interaction strength of 1. A value lower than 1 means that the input strength declined. (4) The facial expression graph shows the facial expression change adopted in arousal analysis on the time axis, and it can identify the time at which the largest emotional change occurs. We also present the facial expression detected by the facial expression analyzer to illustrate the shape of the facial expression at the time. (5) The eye blink graph shows when the eyes of the user blinks, and provides an indicator of the cognitively unusual time spent by the game player on a certain content. (6) The eye concentration graph shows when the user concentrates and gazes at the monitor. The higher the value, the less change of the eyes observed, which can be interpreted as the time when the player is immersed in the game content. Figure 4 presents the visualization results generated from the proposed system. Our visualization scheme facilitates the intuitive analysis of the game player's state.
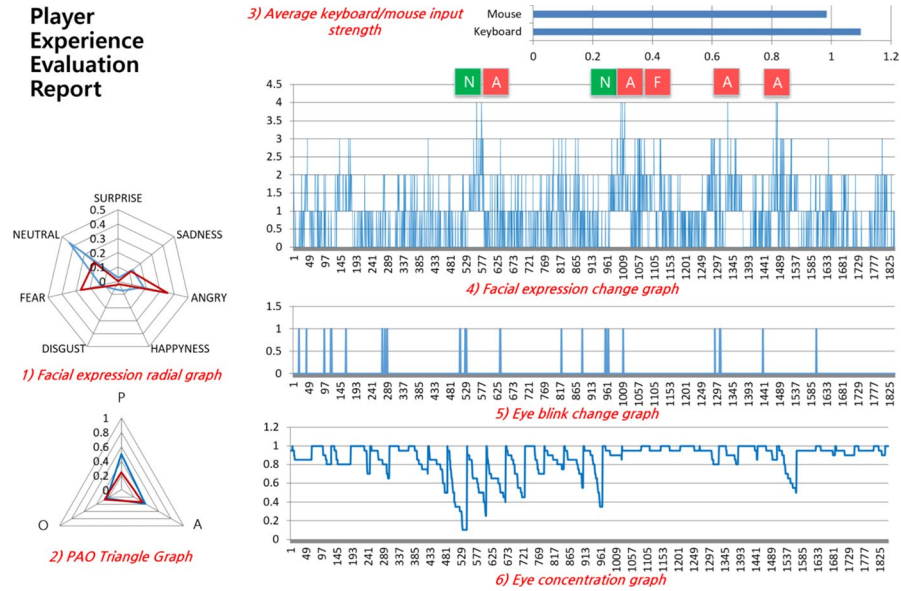
Fig. 4 Player experience evaluation report generated by the proposed system

4 Experiment result

To validate the proposed system, we selected Bioshock: Infinite, a first-person shooter game, as a game for the experiments. The level within the game we used is the first part of the game level stages; it is the initial stage where a novice player receives dramatic tension via the introduction, development, turn of events, and conclusion events. If the corresponding level was played in the same way intended by the designers, the arousal level is expected to increase toward the end. In addition, the game interaction is expected to increase proportionally as the battle with the enemy intensifies. The pleasure part is likely to involve negative facial expressions owing to the frightening events and fierce battles toward the end, although the beautiful background and world introduced in the first part of the level will likely invoke pleasure emotions. We selected this level because the storytelling within the single level is clear, and user reactions were expected to be linearly evident. Figure 5 presents main event screenshots in the Bioshock game used in the experiments.

The experiments were performed with a total of 30 subjects (15 men, 15 women, age group: 23–31, average time: 26.8 min). Most of the experiment participants had more than one year of game experience and spent an average of 5 hours of gameplay a week. For the experiment, the 30 subjects were divided into three groups. Each group was designated different data collection time points of 20, 40, and 60 min. Each subject in the group played the Bioshock infinite for 30 min each.

We obtained various experiment analysis results for different individuals. Among these results, we describe four meaningful results. Figure 6 presents the results of player 1 who exhibited the most similar pattern to the feedback expected for the
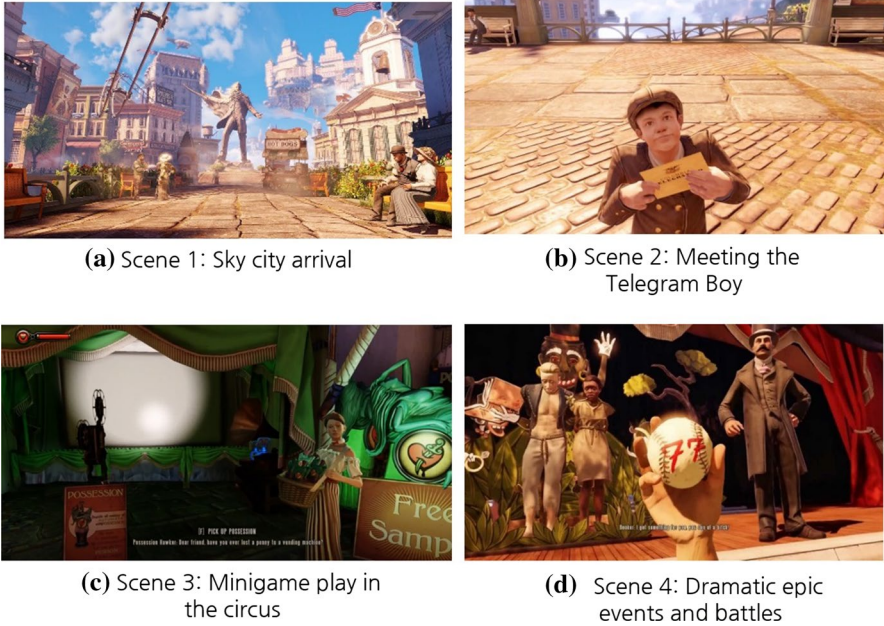
(a) Scene 1: Sky city arrival

(b) Scene 2: Meeting the Telegram Boy

(c) Scene 3: Minigame play in the circus

(d) Scene 4: Dramatic epic events and battles

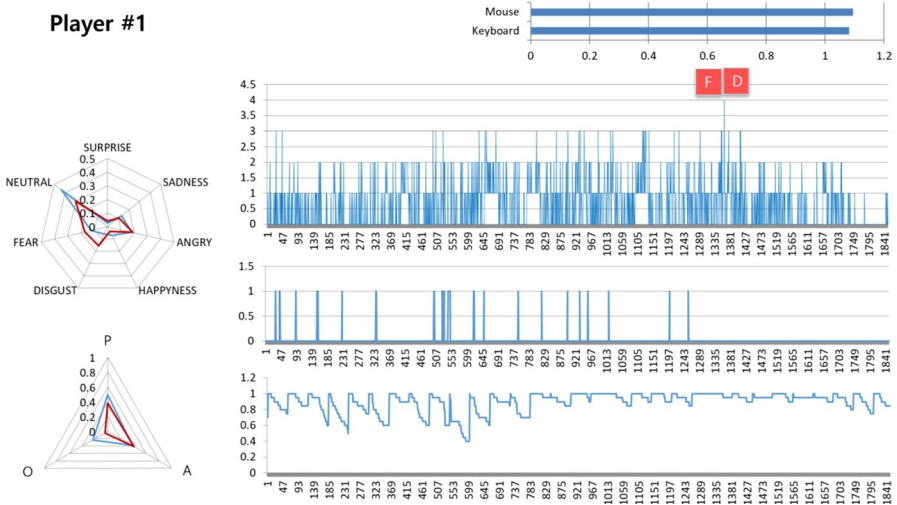**Fig. 5** Main scenes within the Bioshock game level used in the experiments

Fig. 6 Player 1: Average game play and emotional expression

game. The radial graph of the facial expressions indicates that the facial expressions of player 1 while playing the game include more fear (F) and disgust (D), and the most frequent expression was neutral (blue: average of participants, red: player result values). The PAO graph exhibits slightly lower values of P than the average PAO

reaction, and the O part deviates from the average. The average keyboard/mouse input strength values demonstrate more inputs than the average. The facial expression change graph indicates that the most dramatic facial expression change occurs in the scene 4 part where the most dramatic event takes place, as a cruel scene is described in the corresponding event. The overall facial expression change reaches its peak at the third-quarter level, similar to the introduction, development, turn of events, and conclusion order. The eye blink change and eye concentration graphs show that the overall blinking frequency in the first half was higher than in the second half. This can be interpreted as more immersion in the game toward the second half. In general, the expression of player 1 at the climax of the game was negative. Nonetheless, player 1 appears to have been immersed in the game while playing the game. The O value is high because player 1 actively participated in the interaction. From an interview after playing the game, we found that player 1 did not experience substantial difficulty in the game playing process and expressed satisfaction with the game content. Player 1 also stated willingness in continuing to play the game.

Figure 7 presents the results of player 2, which slightly differs from the average reaction. The radial graph of facial expressions indicates that the facial expression of player 2 while playing the game was anger, surprise, and happiness, which slightly differs from player 1. The PAO graph demonstrates that the P value is slightly higher than the average PAO reaction, and an average value was obtained in the A part. The average keyboard/mouse input strength values show that both the keyboard and mouse interactions did not come close to the average interaction level. The facial expression change graph exhibits more facial expression changes than player 1, and indicates that anger and happiness are mixed in the detected facial expressions. The overall facial expression change reaches its peak at the third-quarter level, similar to the introduction, development, turn of events, and conclusion order. The eye blink change and eye concentration graphs demonstrate that in general, player 2 did not
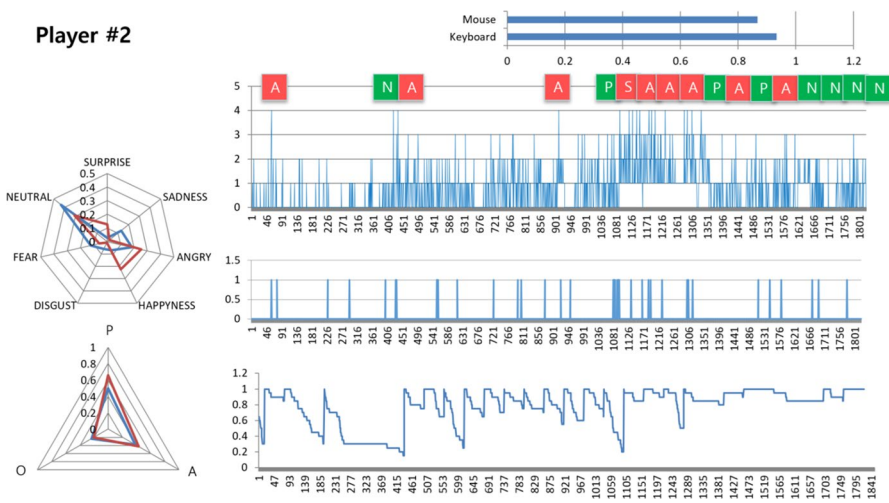

Fig. 7 Player 2: Low level of game play, high level of emotional expression patterns (novice FPS player)

concentrate on the game, and the player's gaze was unstable. Player 2 exhibited a low level of interaction in the game play process, and relatively frequently expressed emotions. During an interview after playing the game, the corresponding player revealed that they were not familiar with the first-person shooter (FPS) game; hence, they were nervous throughout the game, which explains why the overall evaluation indices were unstable.

Figure 8 presents the results of player 3, which is different from the average reaction. The radial graph of facial expressions exhibits fear and neutral facial expression while playing the game. The PAO graph shows that the value of the O part significantly differs from the average. The average keyboard/mouse input strength values indicate that the keyboard interactions did not come close to the average interaction level. The facial expression change graph demonstrates that the facial expression change was not significant, and the detected facial expressions were negative or neutral. The introduction, development, turn of events, and conclusion appear to be irrelevant to the overall facial expression change. The eye blink change and eye concentration graphs indicate that in general, player 3 concentrated on the game while playing. In general, player 3 exhibited a low level of interaction in the game play process and expressed emotions in unexpected intervals. During an interview after playing the game, the corresponding player revealed that they did not find the way between scenes 2 and 3, and thus experienced frustration. Hence, player 3 did not experience the game normally, and did not enjoy the game. This feedback can be regarded as a case where a player did not get to experience the game in the way that the designers intended; hence, the player stops playing the game.

Figure 9 presents the results of player 4, which is also different from the average reaction. The radial graph of facial expressions indicates that player 4 has almost no facial expression while playing the game. The PAO graph shows that
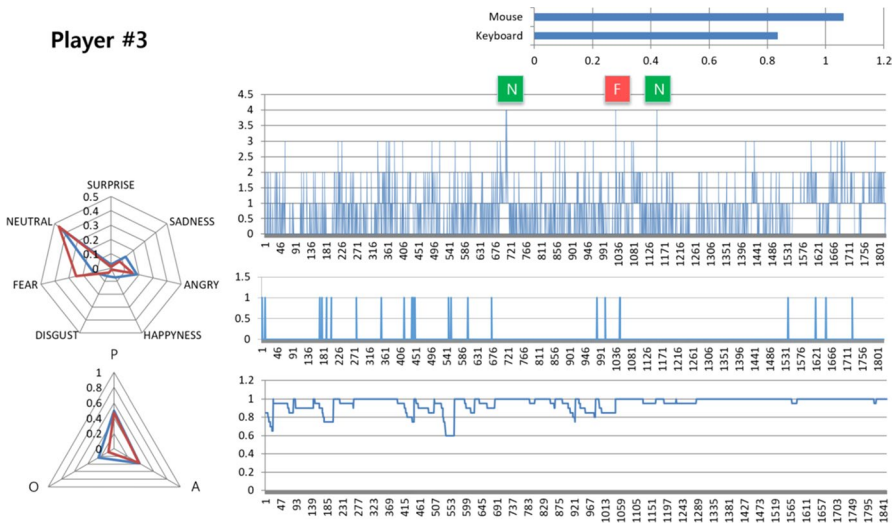

Fig. 8 Player 3: Low level of game play, negative emotional expression at an unexpected point (game deserter)
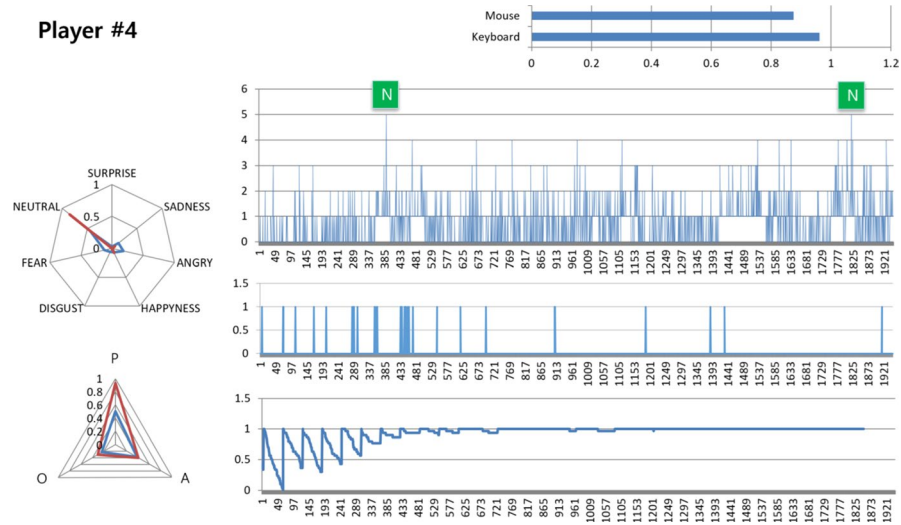
Fig. 9 Player 4: High level of game play, little emotional expression (expert FPS player)

the P part has a much higher value than the average. The average keyboard/ mouse input strength values indicate that the mouse part did not come close to the average interaction level. The facial expression change graph demonstrates that the facial expression change was not large, and the detected facial expressions were all neutral. The introduction, development, turn of events, and conclusion appear to be irrelevant to the general facial expression change. The eye blink change and eye concentration graphs demonstrate that player 4 fully concentrated on the game while playing. In general, player 4 stably played the game in an expressionless condition. During an interview after playing the game, player 4 revealed that they had been playing FPS games for several years, and played such games well. Player 4 really enjoyed the experimental game, and they expressed their willingness to continue playing the game. Such a play pattern corresponds to the pattern of an experienced person who is familiar with a specific game. This pattern can help separate game experts from game novices.

Finally, we attempted to verify whether the experiment subjects' intention to stop playing the game can be estimated. Figure 10 presents the intention to stop playing the game and the PAO graphs of the corresponding players. In general, as the intention to stop playing the game increases, the value of the *P* axis decreases, and the deviation of *O* from the average broadens. This characteristic indicates that the proposed PAO graph can be adopted as a warning index for game deserters. Particularly, players with large PAO graph areas in the P direction and average reactions in the *O* and *A* parts were very satisfied with the corresponding game.
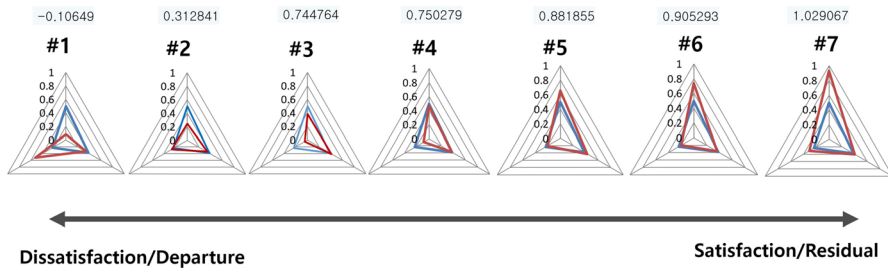
Fig. 10 Intention to stop playing the game and the proposed PAO model graph

5 Discussion

Our research is generally helpful in quantifying gameplay ratings. However, deriving quantitative evaluation results in all cases involves some difficulties. For example, a player who plays the game without expression in a state of high concentration, may be incorrectly classified as having a negative gameplay experience if only habitual negative expressions are considered. In addition, players may be classified as having a positive experience of gameplay based on only habitual positive expressions when they have in fact lost interest in the game. This limitation is a problem that occurs in most emotion classification methodologies based on facial expression recognition. To solve this problem, accurate emotion detection methods must perform a personalization process in advance. However, this issue has not been addressed in the present work. This study assumes that a personalization process would be applied in the quality assurance (QA) stage if a large number of users were tested. Therefore, we note the error caused by individual variations in emotional expression. We discussed the applicability of the proposed methodology in game development processes through an interview with a game development QA professional, and found that the methodology implemented would need to be integrated with the screen recording function to enable the evaluator to verify the results of the analysis. In addition, accuracy may be improved if our research methods were combined with in game playing data (e.g., item acquisition rates or connection retention times). Our methodology may differ in its accuracy depending on the level of game development. In the game development stage, due to cost and security concerns, only a small number of testers attend the game test, which can lead to data imbalance. Therefore, it is advantageous to increase the accuracy to apply only to the partial play section that is expected to have problems rather than the entire game play test. In the game service stage, it can be a condition to receive feedback from large-scale users. In this case, it is possible to collect a large number of data with the user's consent, so a higher level analysis becomes possible in a situation where the data imbalance problem is resolved. The application of the granular computing methodology according to the level of collected data will expand the scalability of the proposed methodology. We plan to apply these complementary measures through collaboration with the game development industry in the future.

6 Conclusion

Various satisfaction evaluation methodologies have been proposed in the affective computing field; however, the game industry is yet to actively adopted these methodologies. This is because the proposed methodologies require complex biometric measurement devices and do not provide indices that can be referenced in the game development stage. To address these problems, this study proposed a novel PAO model that can be applied in an ordinary gaming environment; the study also verified its effectiveness via actual game play evaluations. By introducing a machine learning technique, the proposed methodology allows games to be more objectively evaluated by automating the gaming evaluation part that conventionally relied on survey methods. This methodology indicates that the facial expression recognition technique can be considered to be significantly accurate, and the time-series similar pattern detection algorithm can be utilized in gaming evaluation. Future research would include further precise facial expression analysis for more comprehensive expression interpretations.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest to declare regarding this study.

References

1. Hudlicka E (2008) Affective computing for game design. In: Proceedings of the 4th Intl. North American Conference on Intelligent Games and Simulation, pp 5–12
2. Lee E, Jang Y, Yoon D, Jeon J, Yang SI, Lee SK, Kim KJ (2008) Understanding customer malling behavior in an urban shopping mall using smartphones. In: Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication pp 901–910
3. Ahn J, Hwang J, Kim D, Choi H, Kang S (2008) A survey on churn analysis in various business domains. IEEE Access 8:220816–220839
4. Chung Y, Park CY, Kim NR, Cho H, Yoon T, Lee H, Lee JH (2013) Game bot detection approach based on behavior analysis and consideration of various play styles. ETRI J 35(6):1058–1067
5. Bänziger T, Scherer KR (2007) Using actor portrayals to systematically study multimodal emotion expression: the GEMEP corpus. Lect Notes Comput Sci 4738:476–487
6. Petrantonakis Panagiotis C, Leontios J (2014) EEG-based emotion recognition using advanced signal processing techniques. A Pattern Analysis Approach, Emotion Recognition, pp 269–293
7. Wang Y, Agrafioti F, Hatzinakos D, Plataniotis KN (2008) Analysis of human electrocardiogram for biometric recognition. EURASIP J Adv Signal Process 2008:11
8. Wagner J, Kim JH, Andre E (2005) From physiological signals to emotions: implementing and comparing selected methods for feature extraction and classification, multimedia and expo. IEEE International Conference on Multimedia and Expo, pp 940–943
9. Leon E, Clarke G, Callaghan V, Sepulveda F (2007) A user-independent real-time emotion recognition system for software agents in domestic environments. Eng Appl Artif Intell 20(3):337–345

10. McDuff D, Gontarek S, Picard RW (2014) Remote measurement of cognitive stress via heart rate variability. In: 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)
11. Hernandez J, Picard RW (2014) SenseGlass: using google glass to sense daily emotions, acm user interface software and technology symposium, Honolulu, Hawaii, October 2014
12. Ekman P, Rosenberg EL (1997) What the face reveals: basic and applied studies of spontaneous expression using the facial action coding system (FACS). Oxford University Press, England
13. Kapoor a, Burleson W, Picard RW (2005) Multimodal Affect recognition in learning environments. In: ACM International Conference on Multimedia, 2005; pp 672–682
14. El Kaliouby R, Teeters A, Picard RW (2006) An exploratory social-emotional prosthetic for autism spectrum disorders, wearable and implantable body sensor networks. International workshop on wearable and implantable body sensor networks (BSN'06), 2006; pp 2–4
15. Littlewort GC, Bartlett MS, Lee K (2007) Faces of pain: automated measurement of spontaneous facial expressions of genuine and posed pain. In: Proceedings of the 9th International Conference on Multimodal Interfaces; pp 15–21
16. Yeasin M, Sharma Bullot BR (2006) Recognition of facial expressions and measurement of levels of interest from video. IEEE Transact Multimed 8(3):500–507
17. Nosu K, Kurokawa T, Horita H, Ohhazama Y (2007) Real time emotion-diagnosis of video game players from their facial expressions and its applications to voice feed-backing to game players. In: Proceeding of International Conference on Machine Learning and Cybernetics, 2007; vol 4
18. Vachirapanang K, Tuisima S, Sinthupinyo S, Sirivunnabood P (2012) The classification of the real-time interaction-based behavior of online game addiction in children and early adolescents in Thailand. Int J Adv Res Artif Intell 1(7):7–13
19. Yun C, Shastri D, Pavlidis I, Deng Z (2009) O ' game, can you feel my frustration? Improving user's gaming experience via stresscam. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2009: 2195–2204
20. Tan CT, Rosser D, Bakkes S, Pisan Y (2012) A feasibility study in using facial expressions analysis to evaluate player experiences. Aust Conf Interact Entertain 2012:5
21. Hazelett RL (2003) Measurement of user frustration: a biologic approach. In: Proceedings of the Conference on Human Factors in Computing Systems 2003:734–735
22. Tan CSS, Schöning J, Barnes JS, Luyten K, Coninx K (2013) Bro-cam: improving game experience with empathic feedback using posture tracking. Persuasive technology, lecture notes in computer science. 7822(2013):222–233
23. Zoeller G (2010) Game development telemetry. In: Proceedings of the Game Developers Conference
24. Martinez HP, Bengio Y, Yannakakis GN (2013) Learning deep physiological models of affect. IEEE Comput Intell Mag 8(2):20–33
25. Wang X, Goh DHL (2020) Components of game experience: an automatic text analysis of online reviews. Entertain Comput 33:100338
26. Maman L, Ceccaldi E, Lehmann-Willenbrock N, Likforman-Sulem L, Chetouani M, Volpe G, Varni G (2020) GAME-ON: a multimodal dataset for cohesion and group analysis. IEEE Access 8:124185–124203
27. Song M, Yang Z, Baird A, Parada-Cabaleiro E, Zhang Z, Zhao Z, Schuller B (2019) Audiovisual analysis for recognizing frustration during game-play: introducing the multimodal game frustration database. In: 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), 2019; pp 517–523. IEEE
28. Ringer C, Walker JA, Nicolaou MA (2019) Multimodal joint emotion and game context recognition in league of legends livestreams. IEEE Conference on Games (CoG), 2019; pp 1–8, IEEE
29. Sekhavat YA, Roohi S, Mohammadi HS, Yannakakis GN (2020) Play with one's feelings: a study on emotion awareness for player experience. IEEE Transact Games
30. Kar R, Ghosh L, Konar A, Chakraborty A, Nagar AK (2021) EEG-induced autonomous game-teaching to a robot arm by human trainers using reinforcement learning. IEEE Transact Games
31. Bales RF (2017) Social interaction systems: theory and measurement. Routledge
32. Kang S, Kim D, Kim Y (2019) A visual-physiology multimodal system for detecting outlier behavior of participants in a reality TV show. Int J Distrib Sens Netw 15(7):1550147719864886

33. Gu H, Ji Q (2004) An automated face reader for fatigue detection. 6th IEEE International Conference on Automatic Face and Gesture Recognition, pp 111–116
34. Dlib, http://dlib.net/
35. Weste N, Burr DJ, Ackland BD (1983) Dynamic time warp pattern matching using an integrated multiprocessing array. IEEE Trans Comput 32(8):731–744

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.