Intelligent Automation And Soft Computing, 2020
Copyright © 2020, TSI® Press
Vol. 26, no. 1, 205–214
https://doi.org/10.31209/2019.100000141

# Detecting Outlier Behavior of Game Player Players Using Multimodal Physiology Data

**Shinjin Kang[1], Taiwoo Park[2]**

[1]School of Games, Hongik University, South Korea
[1]Department of Media and Information, Michigan State University, USA
[2]Department of Engineering and Computer Science, Seattle Pacific University, USA

## ABSTRACT
This paper describes an outlier detection system based on a multimodal physiology data clustering algorithm in a PC gaming environment. The goal of this system is to provide information on a game player's abnormal behavior with a bio-signal analysis. Using this information, the game platform can easily identify players with abnormal behavior in specific events. To do this, we propose a mouse device that measures the wearer's skin conductivity, temperature, and motion. We also suggest a Dynamic Time Warping (DTW) based clustering algorithm. The developed system examines the biometric information of 50 players in a bullet dodge game. This paper confirms that a mouse coupled with a physiology multimodal system is useful for detecting outlier behavior of game players in a non-intrusive way.

KEY WORDS: Physiology multimodal system, Game behavior analysis

## 1 INTRODUCTION

IN recent years, the demand for game player behavior analysis has increased as PC online game and mobile game markets grow. Gaming behavior analysis, specifically for abnormal or outlier behaviors, aims to help gamers to keep engaged in a gameplay by providing gamers with an appropriate level of difficulty. If a player faces a very difficult or unexpected gameplay experience, the player is likely to behave differently than usual. This may not be a problem if the situation occurs at a time intended by a game designer. However, an unexpected player experience is a factor for gamers to stop playing a game so if this behavior occurs at a time not intended by the game designer, the designer needs to determine the timing and reason and fix it to ensure a smooth gameplay experience (El-Nasr et al. 2016).

Game player behavior analysis has been performed to determine these abnormal behaviors and to receive useful feedback from players. Typically, it is done with a variety of usability tests in separate gameplay rooms with a small number of player subjects. This process consists of recruiting and educating the subjects, followed by soliciting the feedback through post-play questionnaires. The post-play questionnaire analysis has advantages, including ease of execution as well as the intuitive nature of survey results (e.g., easy or difficult). However, it comes with a disadvantage in that it cannot provide accurate feedback on a specific event or a certain moment in the gameplay because it relies on the user's short-term memory after the gameplay, which may be incorrect and self-manipulated.

To overcome these weaknesses, a variety of automatic evaluation systems have been suggested: facial expression, brain-computer interface (BCI), electromyography (EMG), galvanic skin response (GSR), heart rate (HR), etc. These methods were designed to find meaningful patterns by classifying or clustering a player's biosignal responses that occur at key events during gameplay. Thus, it is possible to identify instant changes in players' perception without heavily relying on the players' memory afterward. However, it is necessary to have expertise on specialized equipment and analysis methodologies to apply these techniques in practice. The procedural difficulty and the cost of analysis are the burdens of game development teams and companies. They are obstacles to applying state-of-the-art behavior analysis techniques from academia to industry (Scherer et al. 2010). The industry is thus in need of a straightforward methodology to perform a player

CONTACT Taiwoo Park twp@spu.edu

behavior analysis at low cost with common gaming hardware.

This paper proposes a multimodal game player behavior analysis system based on general gaming hardware: a mouse that monitors a gamer's physiological state changes during gameplay in a non-intrusive way. With this system, we recorded changes in galvanic skin response, temperature and motion in relation to specific events during gameplay and applied a clustering algorithm based on dynamic time warping (DTW) for outlier analysis. Ark et al., proposed a system in which a mouse played a significant role in detecting physiological signals (Ark et al. 1999). Sykes et al. investigated the hypothesis that the player's arousal level corresponds to the pressure employed to press buttons on a gamepad (Sykes et al. 2003). However, to the best of our knowledge, there is no previous work that has tried to analyze gamers' experiences by clustering multi-dimensional physiological data gathered from a typical mouse for outlier detection. We evaluate our system with a bullet dodging game with single and multiplayer modes. Our experimental results show that our system is effective in detecting outliers in gameplay.

## 2 RELATED WORK

VARIOUS attempts have been made to evaluate the game player's reaction to game content using biometric information: including brain computing interface (BCI), electromyography (EMG), breathing signals, galvanic skin response (GSR) and representative physiology data. (Scherer et al. 2010) (Ozturk et al. 2018)

BCI is one of the most popular biosignal analysis techniques (Kaplan et al. 2013). This classifies the user's behavior based on the user's brainwave pattern. BCI research related to games include level design issues of Role-Playing Games (Balducci et al. 2017), evaluation of the motivational factors of gamers by EEG (Berta et al. 2013), character control (Finke el a. 2009), and difficulty adaptation and evaluation (Marshall et al. 2013). EMG analyzes the user's muscle movements and responsiveness. EMG research that is related to games includes user experience analysis in first person shooting games (Nacke et al. 2010) and fatigue analysis in a driving simulator game (Balasubramanian et al. 2007). The relationship between breathing and gameplay has also been studied. This includes evaluating a player's pleasure while playing a racing game (Tognetti et al. 2010) and diagnosing the stress level by analyzing the user's breathing pattern (Bernardi et al. 2000). These BCI, EMG, and Breathing Signal analysis methods have the advantage of minimized errors while acquiring data by attaching sensors directly to the subject's body. On the other hand, it involves the common burden of attaching devices to the body (Vliet et al. 2012).

Compared with the BCI, EMG, and breathing signals, GSR signals can be detected in a relatively less-intrusive way. Due to the ease of wearing a GSR sensor and the intuitiveness of GSR analysis results, there have been several attempts to investigate the implications of GSR signal pattern in relation to players' perceived difficulty and abnormal behavior. Moreover, a large number of GSR studies belong to the field of Dynamic Difficulty Adjustment (DDA) research. A DDA-based game has a structure that automatically adjusts the related parameters according to the gameplay progress by linking a part of the game system with the in-game behavior evaluation function (Hunicke et al. 2005).

Mandryk et al. introduced the use of physiological measures as behavioral metrics (Mandryk et al. 2007). Dekker et al. adapted a GSR sensor that clipped on a player's fingers to record and transfer biometric information to make changes in the game content of Half-Life 2, a commercial game engine, and suggested an adaptive horror game with GSR sensors. During gameplay, the game environment was designed to change dynamically based on the player's biometric data to give the cinematically augmented horror experiences (Dekker et al. 2007). Tijs et al. discovered that players' GSR arousals differed significantly when they played Pac-Man at different levels of difficulty (Tijs et al. 2008). Singh et al. found that the number of challenges that a player encountered during gameplay could give an impact on the magnitude of GSR arousals (Singh et al. 2013). Like other physiology signals, GSR is a signal generated by the autonomic nervous system, which has the advantage that the player cannot conceal her reaction intentionally. It has the further merit that it can be measured at a relatively lower cost. However, the experimental results are easily affected by the physiological characteristics of the subjects.

Various multimodal physiology-based analysis techniques have been proposed to accurately infer a player's state. Liu et al. suggested an affect-based DDA mechanism to evaluate an anxiety level of a player. In their experimental game, the difficulty level of the game automatically changes by referencing the player's emotional state in real time (Liu et al. 2009). Ambinder et al. introduced how a commercial game developer could analyze the difficulty level of the First Person Shooting (FPS) by using the player's multimodal physiology signals (Ambinder et al. 2011). Lobel et al. suggested a horror biofeedback game titled Nevermind which employs multimodal interface. In the game, the players' heart rate and eye-movement are continuously monitored and sent to the game system, which is in turn interpreted to the player's level of negative affective arousal. Greater negative arousal causes the game's horror-themed environment to become more disturbing (Lobel et al. 2016). Most of these multimodal studies require the purchase of external emotional analysis hardware specific to each signal and needs to be independently attached to the experimenter. In addition, it is difficult to synchronize all the signals between the signal analyzers after the experiment.

In this paper, we propose a multimodal physiological signal analysis system that collects and analyzes biosignals (e.g., GSR and body temperature) and motion signal data in a non-intrusive way. We developed a small substrate that could be inserted into a typical commercial mouse and made it possible to detect multimodal signals. By using this hardware, the player can participate in the game experiment without attaching any external device. Our system acquires the user's multimodal data via the mouse, and it can analyze the user's sensory difficulty using the acquired data by using DTW and a dendrogram clustering algorithm. The key contributions of this paper can be summarized as follows.

*Contributions:*

1) Design of a mouse-based biosignal detector that detects biosignals of game players in a non-invasive way in typical PC gaming environments

2) Abnormal behavior detection based on dynamic time warping (DTW) and a dendrogram algorithm with multimodal physiology and motion data

## 3 BACKGROUND

THERE were several technical problems and requirements in the application of multimodal analysis technology for outlier detection. We chose the following approach.

***Non-intrusive Physiology Detection System:*** If the player recognizes that the biosignal is being measured, it may affect the accuracy the biosignal measurement. Moreover, additional device attachments are burdensome for both game developers and game testers. To solve this problem, we developed a module integrated in a mouse for non-intrusive detection of players' biosignals. Due to recent technological advances, these modules can be integrated into mouse type hardware. In order to continuously measure various biosignals on the mouse, the user's hand must maintain a constant contact with the position of the sensor in the mouse. Various mouse hardware manufacturers have defined the user's mouse grip method as three types – 1) Palm grip, 2) Claw grip and 3) Tip grip. Figure 1 illustrates the mouse grip types. The palm grip is the most popular type found from more than 50% of gamers. The advantage of the palm grip can be the natural and relaxed way the hand rests on the mouse with the most contact points and support. In the claw grip, the hand is arched up with fewer contact points on the mouse and forms a claw-like shape. The tip grip is the most extreme grip type. It has minimal contact points between the hand and the mouse. This type of grip only uses the tip of the fingers to steer the entire mouse in extremely rapid movements with the least restraint of the hand and wrist (Epic Gear 2018). We set the position of our sensor for the palm grip and claw grip methods, which are used by the most players, to maintain constant contact with the player's hand.

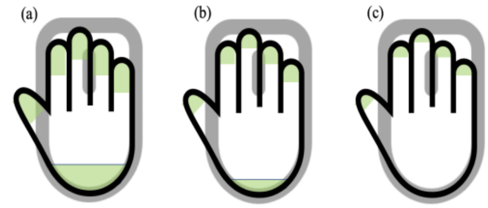



**Figure 1. Three types of mouse grip (a) Palm grip (b) Claw grip (c) Tip grip**

***Outlier Detection Algorithm:*** Game players often show abnormal behaviors when they experience very enjoyable or uncomfortable experiences during gameplay. In a single-player game, the player has a definite behavioral response at the time of the event or game play result that has a direct influence on the game's outcome (i.e., success or failure). During the game play, which demands continuous input (movement, skill, input, etc.) in real time, the user's biosignal reactions do not occur much. However, when a character being manipulated by the player is in danger of imminent to death or fails to clear the game, the related biosignal reaction is significant. We attempted to analyze specific behaviors by detecting biosignal responses at this time. In a competitive game with other players, only a few players can win top rewards. Thus, the relative comparisons among the players are important for determining the leading player. There is a high probability that the user's abnormal behaviors will be stronger. To detect the outlier behavior in single and multiplayer game modes, we developed a clustering algorithm to identify players who show the most unusual behavior at a specific time. We calculated the distance between the multimodal data collected using the DTW algorithm. Based on the DTW distance matrix, we constructed a dendrogram to distinguish the players who were out of the cluster in real time. Then, a client-based visualization tool was developed to make it easier for the development team to check the differences among the biosignals of the players in real time.

## 4 SYSTEM

AS a physiology multimodal system, we developed 1) A mouse module for biosignal detection, 2) a DTW distance calculator, and 3) a dendrogram based outlier detector.

### 4.1 Mouse Module for Biosignal Detection

Our goal in this study was to detect the Galvanic Skin Response (GSR), temperature, and motion variation in a non-intrusive way. The GSR detects the electrical resistance of the skin surface caused by sweating caused by the autonomic nervous system. In general, the GSR sensor measures the change in resistance between two points at a given time. Particularly, it is measured by attaching two electrodes to the index finger. We painted conductive paint on a mouse click button and measured the GSR with the signal received in this area. With this, we were able to measure the GSR signal without the player feeling any physical contact. We attached a GSR, temperature, and motion sensor to an off-the-shelf mouse. The detection module and transmitter were integrated into the mouse and connected to the USB connector. Figure 2 shows our customized mouse hardware. The features of each module constituting this hardware are as follows.

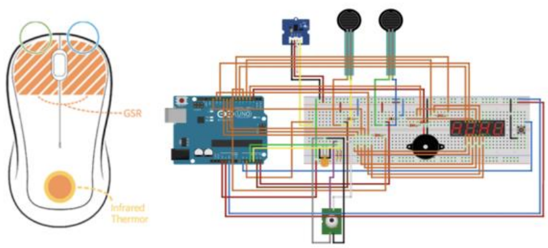


**Figure 2. Mouse module to collect multimodal data (left), and the developed Arduino module (right)**

### 4.2 Biometric information collection device

The biometric information collection device was developed using Arduino. We used WeMos's ESP8266 model, which is small enough to be attached to a player with a WiFi communication function. The built-in Arduino is connected to a mobile battery. The WiFi connection had a WiFi network with a predetermined name, and the biometric information collected through a predetermined internal IP address was transmitted as a message.

### 4.3 Skin conductivity sensor

The Grove GSR Skin Sensor Module v1.2 manufactured by Seeedstudio was used as a skin conductivity sensor as an early stage prototype. As the play progressed, the electrical signals were accepted using two lines electronic painting on two mouse buttons to maintain more stable adhesion to the player's fingers. In order to obtain a stable and smooth graph, Kalman filter was applied for the raw data.

### 4.4 Infrared temperature sensor

To measure the body temperature, we used an infrared temperature sensor. The infrared temperature sensor was modeled by MLX90614 manufactured by Sparkfun Inc. Unlike conventional temperature sensors that measure ambient temperature, this sensor is capable of measuring the temperature at one point. The temperature can be measured from minus 70 degrees to 380 degrees, and the measurement resolution is 0.02 degrees.

### 4.5 Nine-axis IMU (Inertial Measurement Unit)

We used the 9 DoF Sensor Stick model, which was manufactured by SparkFun and supports a 3-axis accelerometer, 3-axis magnetometer, and 3-axis gyroscope (i.e., rotation sensor) to measure the player's mouse movement. The geographical information was used to determine whether the cause of the unclassifiable state was caused by the movement of the player by detecting the movement of the user's hand.

### 4.6 Data transmission

The measured values from each sensor were transmitted in one string data type. We used a delimiter to identify each sensor value: the skin conductivity value, the filtered skin conductivity value, the 3-axis gyro value, the 3-axis acceleration value, the 3-axis geomagnetism value, the calculated Pitch value, and the Roll value.

## 5 OUTLIER DETECTION

### 5.1 Dynamic Time Warping Algorithm

THE data obtained in this system was time series data. We had to determine at what point in the time series data and at what intervals we would use for specific behavioral analysis. We tried to compare behavioral characteristics among users using multimodal data for a limited time in the game.

DTW is a dynamic programming algorithm that compares two series data and to determine the optimum warping path between them. DTW first creates a matrix D of pointwise distances. The algorithm then runs through D, enumerates all paths w, and finds the optimal warping path Ws. DTWD is mainly used to measure the distance between time series data of different lengths. To calculate DTWD, the starting and ending points must be clearly defined. If the signal is nearly horizontal, it is difficult to determine the starting and ending points. Since we measured the changes of the bio-signals during a specific event (i.e., a short game play), the length of all the signals could be easily defined. We applied a linear interpolation and compared points in one signal to point-point segments in the other to produce better results. We weighted the DTW score by giving more weight heuristically to the conductance level, temperature, and 3-axis motions. This weighting is useful, especially when measurements are less precise for stable parts of the time-serious input data.

### 5.2 Hierarchical Clustering

The similarity value among the players thus obtained should determine which player will react most prominently. This can be solved by classification or clustering techniques. In this case, it is impossible to apply machine learning-based classification techniques because it is very difficult to obtain prior personalized learning data from the player. For the clustering method, it is also difficult to define in advance how many clusters can be detected. We used Dendrogram which is a hierarchical clustering algorithm. The algorithm generates a hierarchy of clusters in which, as the level in the hierarchy increases, clusters are generated by merging the lower level clusters, such that an ordered sequence of clusters is obtained (Steinbach et al. 2000). To decide a way how the merging is performed, a distance measure between clusters needs to be specified, in addition to the one that is used to calculate pairwise similarities. However, a specific number of clusters does not need to be defined for the hierarchy to be created.

To better understand the players, we applied an additional classification technique as post-processing. Our system was able to accumulate players' biomedical signals over time. We classified typical signal patterns with these data. We clustered all accumulated biosignals up to the current time and selected a representative clustering group and extracted the representative multimodal patterns by calculating the mean value of the multimodal data of the selected group. This approach generates presentative biosignal templates which is robust with few data. It potentially requires less training data than machine learning-based or feature-based methods. The generated representative pattern can be used as a separate classification criterion when enough data is acquired.

### *5.3 Data Visualization*

The goal of this study is to give the game developer a quick feedback about the abnormal factors in the game level by detecting the player's outlier behavior in a specific event. We found that we needed a monitoring tool that provides fast feedback in real time. Figure 3 shows our real-time monitoring tool interface. This tool can be run simultaneously with any game running. All the biosignals measured on the mouse are displayed in time series graphs and quantified numbers as well as every frame of the screenshot of the game desktop to create a movie file. In addition, the CSV file is used to export the measured data. By marking the time synchronized with the recorded video in the CSV file, it is possible to easily identify the event in the game. This synchronism is advantageous in enhancing data interoperability with a game to be analyzed with existing biosignal equipment. In addition, it can grasp the change of the biological signal in real time, so that it can instantly grasp an abnormal element in the game level in the test field.


**Figure 3.** **The user interface of the developed monitoring system**

## 6 EXPERIMENT RESULTS

WE verified our anomaly detection performance through two experiments. For this, we developed an experimental game using Unity3D engine. The game is a typical bullet dodge style in which a player tries to avoid bullets flying from four sides for up to 30 seconds. Figure 4 shows the screenshot of the game. We developed two game modes, a single mode and a multiplayer mode. In the single player mode, one player plays a game, and in the multiplayer mode, five players play the game simultaneously. This is to confirm the difference between the reactions in the single and multiplay experiences. The difficulty was created at three levels. We recruited a total of 100 players (10 groups x 10 player) for experiments. All players played the three different levels of difficulty in single mode and multiplayer mode. A total of 600 signal sets were measured (100 players x 3 levels x 2 modes).


**Figure 4.** **The developed bullet dodge game for an experiment**

We first evaluate the overall signal pattern difference according to the level of difficulty in the single player mode. Figure 5 shows an average of 10 group signals at the easiest and the hardest modes. This graph shows that the GSR and temperature signal received through our hardware is reliable. The X-axis of the graph represents the time and the Y-axis the sum of the normalized value of skin conductivity and temperature values. The red line denotes the end of the game. The players' skin conductivity increased as the game difficulty became increased. This is presumably because the players felt more nervous as the game became more difficult. This result re-confirms the strong relationship between the difficulty level and the GSR, which was verified in previous biosignal studies (Tijs et al. 2008).

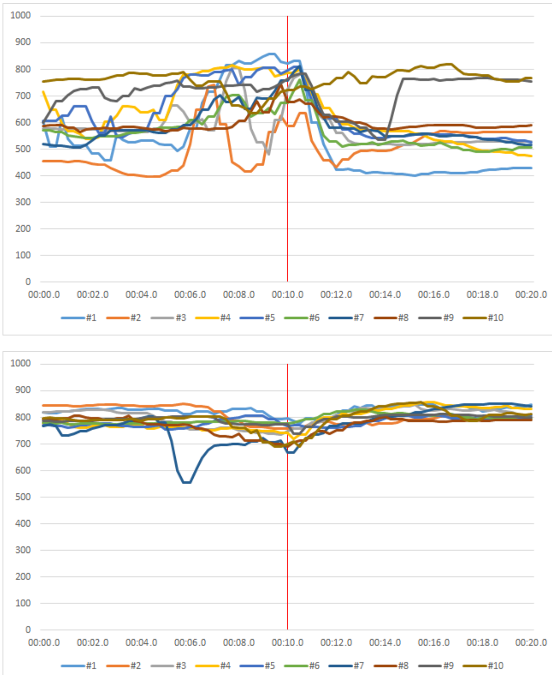



**Figure 5. The GSR signal difference at easy (top) and hard (bottom) game play**

Figure 6 shows the average and standard deviation of the easy, normal and hard level of difficulty for 10 groups. In the graph we find that overall biosignal averages were higher at the hard level than the easy one, and there was almost no difference between the normal and hard level. The higher the degree of difficulty, the higher the biosignal, but it did not increase any more above a certain level.

In the standard deviation graph, we found that the standard deviation at the easy level was much higher than at the other levels, and the standard deviation of the normal level was slightly higher than the one at the hard level. As a result, the biosignals showed very low variability with the relatively high average value at the hard level, and the low average value showed high variability at the easy level.

In our biosignals compared to GSR signals, the temperature was less influenced by the difficulty levels. The temperature showed a pattern in which the overall temperature rose during the hard level gameplay. However, unlike expectations, the rate of temperature change during the game play was not significantly high. It seemed that the temperature sensor did not have a high enough resolution to discriminate between signals from different difficulty levels.

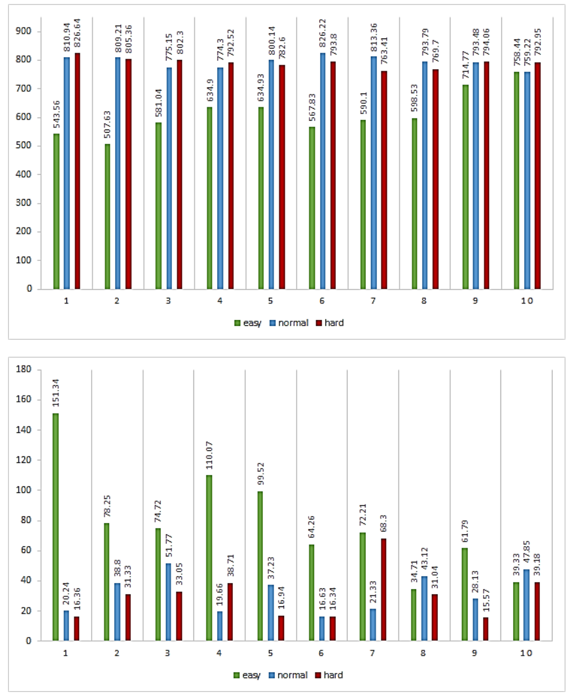



**Figure 6. The average (top) and standard deviation (below) bar chart of three different difficulty level**

Next, we attempted to determine if our system could detect outliers based on the total play data of the game in a single play mode. A total of 14% of the plays were detected as outliers on average for biosignals. The distance was calculated by the DTW algorithm and then clustered with the dendrogram algorithm. During the clustering process, we tried to find representative clustering patterns with accumulated biosignals. We generated a representative pattern of each signal group with mean values of biosignals extracted from the dendrogram cluster results. The most representative clustering cases are shown in Figure 7. These graphs can be typical response patterns of players in the bullet dodge game. (a) Graph decreasing; (b) Graph increasing; (c) Graph maintaining high pitch; and (d) Graph maintaining low pitch. All graphs show a small rise in the initial position of the graph. This shows that a certain amount of multimodal signal changes occurred in common when the players started the game.

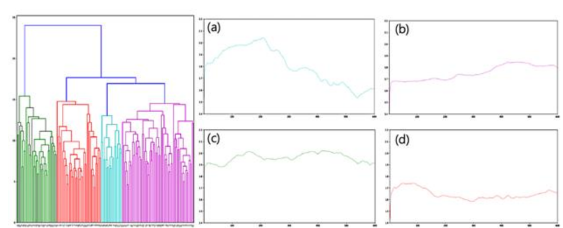



**Figure 7. Dendrogram clustering results and representative graphs**

In graph (a), the graph is very high and then falls off sharply, and in the graph (b), the graph is gradually increasing. However, in graphs (c) and (d), the graph strength is relatively more stable. Here we found that in graph (a) the pattern has some psychological stability, even though the players were somewhat disturbed with the play result. However, in graph (b) the pattern did not decrease the strength of the multimodal signals but gradually increased their intensity, indicating that they maintained tension and excitement during play time. We classify signals with the greatest distance difference as outliers based on the detected representative patterns. We selected 15 signals with the biggest difference and confirmed their gameplay results. As a result, twelve of the detected players failed to clear the game, and three of them cleared the game with a relatively high score. These results show that specific behaviors are more frequently caused by negative results than positive results.

In the second experiment, we tried to confirm the difference in the clusters by the relative comparison among the players in the multiplayer mode. Five players played the same level simultaneously through the online mode, which allowed them to experience the same play at the same stage. Players were able to view their scores in real time, sorted by score and displayed on the screen. As a result, our system classified one outlier as the 1-4 classification pattern at 9% and two outliers as the 2-3 classification pattern at 91% in total. Figure 8 shows the clustering case. Pattern 1-4 occurred mainly from the competitive play among the players. In this pattern, the amount of changes in the vital sign and facial expression changes was relatively larger, and the play classified as outliers was mainly detected from players who scored low at the easy level. Pattern 2-3 occurred mainly in the non-competitive play patterns. This mainly occurred when the difference between the players was relatively small and the play difficulty was high.

We tried to evaluate another outlier detection model with 3-axis motion data. Figure 9 shows the results of the hardest mode and the motion measured by the 3-axis sensor. The bullet dodge game used in this experiment had more bullets moving in four directions as the difficulty increased. At the hardest difficulty level, more elaborate control was needed. Therefore, at the hardest difficulty level, the range of movement became narrower. It can be seen that the hardware proposed in this system normally detects the movement of such players, which shows that our system could detect this characteristic movement of a player in the game. We further utilized motion information detected in three axes to increase the classification accuracy of outliers detected by the GSR. We calculated the mean shift distance and standard deviation obtained from the three difficulty levels and then calculated the difference from the sampled value. When the difference was larger than the threshold value, we classified it as outliers in terms of gameplay. This can be interpreted as an outlier based on the player's control patterns. We used the same clustering method as with the 3-axis signal analysis. We found that play patterns changed very sharply at the hardest difficulty level but moved very little at the easiest.

To evaluate the usefulness of the outlier detection results, a brief survey and interview were conducted after the gameplay. The questionnaire in the survey prompted each player to select the most upsetting moment. 69% of the outliers we detected were included in the levels that the players had marked. After the survey, we conducted a brief interview with the players who showed the most changes in the psychological measures and asked what they felt and why there were psychological fluctuations. The majority of reasons of the mental fluctuations were "not knowing where to move my spaceship", and "feeling nervous because the bullet is faster than I expected". This feedback on psychological reactions

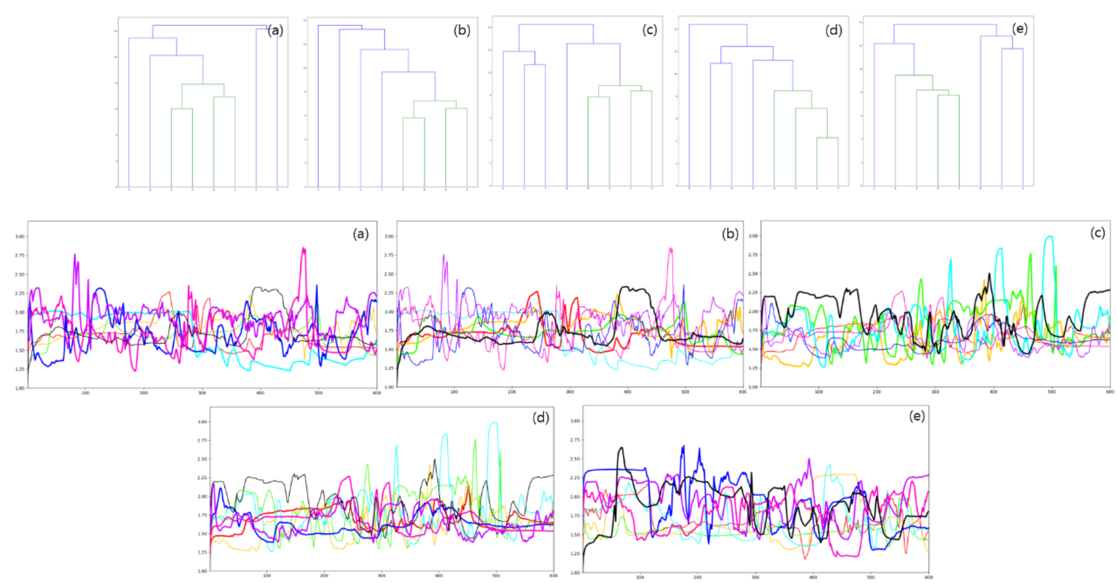



**Figure 8.** **Clustering dendrogram pattern and its time-serious data.**

gave the most negative outlier factors. Through these interviews, we were able to conclude that players often experienced outlier behaviors for negative reasons rather than for positive reasons. About 35% of the subjects said that they did not feel disturbed at the moments detected by the system. The reasons are that "I usually feel a lot of tension, so I sweat a lot.", " I laughed habitually," and "I moved my body frequently because I was nervous throughout the experiment.". Most of these responses were due to the individual habits of the players or intentional behavior to the others.

Table 1 shows the results classified as outliers by the system. About 15% of the subjects in the single mode and about 11% in multiplayer mode were classified as outliers with biosignals. 10% in the single mode and about 8% in the multiplayer mode were classified as outliers with motion data. 8% in the single mode and about 6% in the multiplayer mode were classified as outliers with biosignals plus motion data. This shows that there is a clearer signal difference in the multiplayer mode than in the single player mode. The number of cases classified as outlier by GSR signals were greater than ones by motion data. When classified by the 3-axis sensor, there was a difference in behavior pattern, whereas outlier play classified by GSR signals had only partial differences in the play movement pattern. We conjectured that they were psychologically agitated and that their movements were thus not accurate.

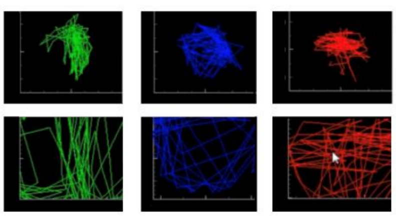


**Figure 9.** 3-axis movement at hard (above) and easy (below) difficulty levels.

**Table 1.** Outlier detection result with proposed system

| Mode | Difficulty | Biosignal Outliers | Motion Outliers | Biosignal & Motion Outliers |
|---|---|---|---|---|
| Single | Easy | 15% | 15% | 11% |
| | Normal | 13% | 9% | 8% |
| | Hard | 18% | 6% | 5% |
| | Ave. | 15% | 10% | 8% |
| Multi | Easy | 8% | 8% | 6% |
| | Normal | 11% | 9% | 7% |
| | Hard | 14% | 7% | 5% |
| | Ave. | 11% | 8% | 6% |

## 7. DISSCUSSION

OUR study had some room for improvement in the following areas. The GSR signal is sensitive to body movements during measurement. If the player does not place the fingers on the mouse button continuously, a disconnected GSR signal is detected. This reduces the accuracy of the detection. This happens when the player has a mouse grip shape that does not attach the palm to the mouse. Most of the players who participated in the game test had a mouse grip shape with the palm of the mouse attached to the mouse. The GSR signal has a different level of perspiration for each player, which can improve the accuracy of the classifier when applying the personalization process. This study did not go through a personalized process. However, if each player has an individual mouse for the experiment, we can expect personalized classified results by accumulating long game play data.

## 8. CONCLUSION

THIS paper suggests an outlier detection technique using a multimodal physiological signal analysis system. We suggested a mouse type detector module, DTW and a dendrogram based clustering algorithm. The proposed system detects the major outliers among the players in a shooting game and gives the various types of useful information to the development team in real time, thereby enabling the level designer to experience a fast feedback process that was not available in the previous development process. We showed how affective computing technology can be applied in the actual game development process. Through this study, we showed that affective computing technology can be utilized as a fun stimulus element in the content development domain.

## 7 ACKNOWLEDGEMENTS

THIS work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.NRF-2019R1A2C1002525)

## 8 REFERENCE

Ambinder, M. (2011). Biofeedback in gameplay: How valve measures physiology to enhance gaming experience. *In game developers conference* (Vol. 2011).

Ark, W. S., Dryer, D. C., & Lu, D. J. (1999). The Emotion Mouse. *In HCI (1)*, 818-823.

Balasubramanian, V., & Adalarasu, K. (2007). EMG-based analysis of change in muscle activity during simulated driving. *Journal of Bodywork and Movement Therapies*, 11(2), 151-158.

Balducci, F., Grana, C., & Cucchiara, R. (2017). Affective level design for a role-playing videogame evaluated by a brain–computer interface and machine learning methods. *The Visual Computer*, 33(4), 413-427.

Bernardi, L., Wdowczyk-Szulc, J., Valenti, C., Castoldi, S., Passino, C., Spadacini, G., & Sleight, P. (2000). Effects of controlled breathing, mental activity and mental stress with or without verbalization on heart rate variability. *Journal of the American College of Cardiology*, 35(6), 1462-1469.

Berta, R., Bellotti, F., De Gloria, A., Pranantha, D., & Schatten, C. (2013). Eletroencephalogram and physiological signal analysis for assessing flow in games. *IEEE Trans. Comput. Intellig. and AI in Games*, 5(2), 164-175.

Dekker, A., & Champion, E. (2007). Please Biofeed the Zombies: Enhancing the Gameplay and Display of a Horror Game Using Biofeedback. *In DiGRA Conference*.

El-Nasr, M. S., Drachen, A., & Canossa, A. (2016). Game analytics. *Springer London Limited*.

Epic Gear: https://www.epicgear.com/en/technologies/types-of-mouse-grip

Finke, A., Lenhardt, A., & Ritter, H. (2009). The MindGame: a P300-based brain–computer interface game. *Neural Networks*, 22(9), 1329-1333.

Hudlicka, E. (2008). Affective computing for game design. *In Proceedings of the 4th Intl. North American Conference on Intelligent Games and Simulation*, 5-12.

Hunicke, R. (2005). The case for dynamic difficulty adjustment in games. *In Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*, 429-433.

Kaplan, A. Shishkin, S. Ganin, I. Basyul, I. & Zhigalov, A. (2013) Adapting the P300-based brain-computer interface for gaming: a review. *in IEEE Transactions on Computational Intelligence and AI in Games*, 5(2), 141–149.

Liu, C., Agrawal, P., Sarkar, N., & Chen, S. (2009). Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback. *International Journal of Human-Computer Interaction*, 25(6), 506-529.

Lobel, A., Gotsis, M., Reynolds, E., Annetta, M., Engels, R. C., & Granic, I. (2016). Designing and utilizing biofeedback games for emotion regulation: The case of nevermind. *In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 1945-1951.

Mandryk, R. L., & Atkins, M. S. (2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International journal of human-computer studies*, 65(4), 329-347.

Marshall, D., Coyle, D., Wilson, S., & Callaghan, M. (2013). Games, gameplay, and BCI: the state of the art. *IEEE Transactions on Computational Intelligence and AI in Games*, 5(2), 82-99.

Nacke, L. E., Grimshaw, M. N., & Lindley, C. A. (2010). More than a feeling: Measurement of sonic user experience and psychophysiology in a first-person shooter game. *Interacting with computers*, 22(5), 336-343.

Ozturk, M. U., Arman, A. R., Bulut, G. C., Findik, O. T. P., Yilmaz, S. S., Genc, H. A., ... & Cataltepe, Z. (2018). Statistical Analysis and Multimodal Classification on Noisy Eye Tracker and Application Log Data of children with Autism and ADHD. *Intelligent Automation and Soft Computing*, 24(4), 891-906.

Scherer, K. R., Bänziger, T., & Roesch, E. (2010). A Blueprint for Affective Computing: A sourcebook and manual. *Oxford University Press*.

Singh, N., & Agarwal, R. G. (2013). Arousal Level Determination in Video Game Playing Using Galvanic Skin Response (Doctoral dissertation).

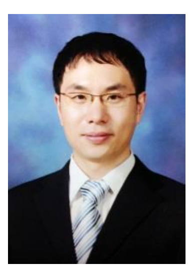
Sykes, J., & Brown, S. (2003). Affective gaming: measuring emotion through the gamepad. *In CHI'03 extended abstracts on Human factors in computing systems*. 732-733.

Tijs, T., Brokken, D., & IJsselsteijn, W. (2008). Creating an emotionally adaptive game. *In International Conference on Entertainment Computing*. 122-133.

Tognetti, S., Garbarino, M., Bonarini, A., & Matteucci, M. (2010). Modeling enjoyment preference from physiological responses in a car racing game. *In Computational Intelligence and Games (CIG) IEEE Symposium on*. pp. 321-328.

Vliet, V.M. Robben, A. Chumerin, N. Manyakov, N.V. Combaz, A. and Van Hulle, M.M. (2012). Designing a brain-computer interface controlled video-game using consumer grade EEG hardware. *in Proceedings of the ISSNIP Biosignals and Biorobotics Conference (BRC'12), IEEE*, 1–6.

## 9 NOTES ON CONTRIBUTORS



**Shin-Jin Kang** received an MS degree from the Department of Computer Science & Engineering from Korea University in 2003. After graduation, he joined Sony Computer Entertainment Korea (SCEK) as a video game programmer. From 2006, He has worked at NCsoft Korea as a lead game designer. He received Ph.D. degree in Computer Science & Engineering at Korea University in 2011. And he is now a professor at the School of Games in Hongik University.

**Taiwoo Park** is an assistant professor at the Department of Engineering and Computer Science in Seattle Pacific University. He received his BS, MS and PhD degrees from the Department of Computer Science in KAIST, South Korea, in 2005, 2008, and 2014 respectively. He had been working as an assistant professor at the Department of Media and Information in Michigan State University from 2014 to 2018. His research interest includes mobile/ubiquitous computing, as well as virtual/augmented/mixed reality.