Hindawi
International Journal of Computer Games Technology
Volume 2018, Article ID 8734540, 14 pages
https://doi.org/10.1155/2018/8734540

Research Article

Automated Analysis of Facial Cues from Videos as a Potential Method for Differentiating Stress and Boredom of Players in Games

Fernando Bevilacqua[1,2] **Henrik Engström**[1] **and Per Backlund**[1]

[1] *University of Skövde, Skövde, Sweden*
[2] *Federal University of Fronteira Sul, Chapecó, SC, Brazil*

Correspondence should be addressed to Fernando Bevilacqua; fernando.bevilacqua@his.se

Received 5 December 2017; Revised 22 January 2018; Accepted 30 January 2018; Published 8 March 2018

Academic Editor: Michael J. Katchabaw

Facial analysis is a promising approach to detect emotions of players unobtrusively; however approaches are commonly evaluated in contexts not related to games or facial cues are derived from models not designed for analysis of emotions during interactions with games. We present a method for automated analysis of facial cues from videos as a potential tool for detecting stress and boredom of players behaving naturally while playing games. Computer vision is used to automatically and unobtrusively extract 7 facial features aimed at detecting the activity of a set of facial muscles. Features are mainly based on the Euclidean distance of facial landmarks and do not rely on predefined facial expressions, training of a model, or the use of facial standards. An empirical evaluation was conducted on video recordings of an experiment involving games as emotion elicitation sources. Results show statistically significant differences in the values of facial features during boring and stressful periods of gameplay for 5 of the 7 features. We believe our approach is more user-tailored, convenient, and better suited for contexts involving games.

1. Introduction

The detection of the emotional state of players during the interaction with games is a topic of interest for game researchers and practitioners. The most commonly used techniques to obtain the emotional state of players are self-reports (questionnaires) and physiological measurements [1]. Questionnaires are practical and easy to use tools; however, they require a shift in attention, hence breaking or affecting the level of engagement/immersion of users. Physiological signals, on the other hand, provide uninterrupted monitoring [2, 3]; however, they are uncomfortable and intrusive, since they require a proper setup in the person's body. Additionally sensors might restrict player's motion abilities; for example, a sensor attached to a finger prevents the use of that finger.

Facial analysis is a promising approach to detect the emotional state of players unobtrusively and without interruptions [4]. The use of computer vision for player experience detection is feasible and visual inspection of gaming sessions has shown that automated analysis of facial expressions is sufficient to infer the emotional state of players [5, 6]. Automatically detected facial expressions have been correlated with dimensions of game experience [7] and used to enhance player's experience in online games [8, 9]. Automated facial analysis has become mature enough for affective computing; however, there are several challenges associated with the process. Facial actions are inherently subtle, making them difficult to model, and individual differences in face shape and appearance undermine generalization across subjects [4]. Schemes such as the Facial Action Coding System (FACS) [10, 11] aim to overcome those challenges by standardizing the measurements of facial expression by defining highly regulated procedural techniques to detect facial Action Units (AU).

While previous work explored the use of manual or automated facial analysis as a mean to detect the emotional state of players, aimed at creating emotionally adapted games [12] or tools for unobtrusive game research, they lack an easier

and more user-tailored approach for studying and detecting facial behavior in the context of games. The use of FACS, for instance, is a laborious task that requires trained coders and several hours of manual analysis of video recordings. When automated facial analysis is used, it is often tested on contexts not related to games, or they rely on facial cues derived from models not designed for analysis of emotional interactions in games, such as the MPEG-4 standard [13]. Such standard specifies representations for 3D facial animations, not emotional interactions in games. Automated facial analysis is also commonly performed on images or videos whose subjects are acting to produce facial expressions, which are likely to be exaggerated in nature and not genuine emotional manifestations. Those are artificial reactions that are unlikely to happen in a context involving subjects interacting with real games, where emotional involvement between subject and game is stronger. Another limitation of previous work is the common focus on detecting facial expressions per se, for example, 6 universal facial expressions [14], not necessarily detecting isolated facial actions, for example, frowning, associated with emotional reactions in games. Finally people are different and elements as age and familiarity with a game influence the outcome of automated facial analysis of behavioral cues [15], and different games might induce different bindings of facial expressions [7]. As a consequence, a more user-tailored contextualization is essential for any study involving facial analysis, particularly involving games. Empirical results of manual annotations of facial behavior in gaming sessions have indicated more annotations during stressful than during boring [16] or neutral [17] parts of games. Further investigation of such findings using an automated analysis instead of a manual approach is a topic of interest for game researchers and practitioners, who can benefit from improved tools related to facial behavior analysis.

In this paper, we introduce our method for automated analysis of facial cues from videos and present empirical results of its application as a potential tool for detecting stress and boredom of players in games. Our method is based on Euclidean distances between automatically detected facial points, not relying on prior model training to produce results. Additionally the method is able to cope with face analysis under challenging conditions, such as when players behave naturally, for example, moving and laughing while playing games. We applied our method on video recordings of an experiment involving games as emotion elicitation sources, which were deliberately designed to cause emotional states of boredom and stress. During the game session, subjects were not instructed to remain still, so captured corporal and facial reactions are natural and emerged from the interaction with the games. Subjects perceived the games as being boring at the beginning and stressful at the end with statistically significant differences of physiological signals, for example, heart rate (HR), in those distinct periods [18]. This experimental configuration allows the evaluation of our method in a situation involving game-based emotion elicitation, which contextualizes our automated facial analysis in a more game-oriented fashion than previous work. Our main contribution is twofold: firstly we introduce a novel method for automated analysis of facial behavior, which has the potential to be used to differentiate emotional states of boredom and stress of players. Secondly we present the results of an automated facial analysis performed on subjects of our experiment, who interacted with different games under boring and stressful gameplay conditions. Our results show that values of facial features detected during boring periods of gameplay are different from values of the same facial features detected during stressful periods of gameplay. Even though the nature of our games, that is, 2D and casual, and the sample size ($N = 20$) could be limiting factors for the generality of the evaluation of our method, we believe our population of experimental subjects is diverse and our results are still promising. Our study contributes with results that can guide further investigation regarding emotions and facial analysis in gaming contexts. It includes information that can be used to create nonobtrusive models for emotion detection in games, for example, fusion of facial and body features (multimodal emotion recognition) which is known to perform better than using either one alone [19].

The rest of this paper is organized as follows. Section 2 presents related work on manual and automated facial analysis focused on emotion detection. Section 3 presents our proposed facial features, the experimental setup, and the methodology used to evaluate them. Sections 4 and 5 present, respectively, the results obtained from the evaluation of the facial features and a discussion about it. Finally, Sections 6 and 7 present the limitations of our approach, a conclusion, and future work.

2. Related Work

The analysis of facial behavior commonly relies on data obtained from physical sensors, for example, electromyography (EMG), or from the application of visual methods to assess the face, for example, feature extraction via computer vision [20]. The approach based on EMG data uses physical sensors attached to subjects to measure electrical activity of facial muscles, such as the zygomaticus, the orbicularis oculi, and the corrugator supercilii muscles (Figure 1), associated with smiling, eyelids control, and frowning, respectively. Hazlett [21] presents evidence of more frequent corrugator activity when positive game events occur. Tijs et al. [3] show increased activity of zygomatic muscle associated with self-reported positive emotions. Similarly, Ravaja et al. [22] show that positive and rewarding game events are connected to increase in zygomatic and orbicularis oculi EMG activity. Approaches based on EMG are more resilient to variations of lighting conditions and facial occlusion; however, they are obtrusive since physical sensors are required to be attached to the subject's face.

Contrary to the obtrusiveness of EMG-based approaches, analysis of facial behavior based on automated visual methods can be performed remotely and without physical contact. The process usually involves face detection, localization of facial features (also known as landmarks or fiducial points), and classification of such information into facial expressions [23]. A common classification approach is based on distances and angles of landmarks. Samara et al. [24] use the Euclidean

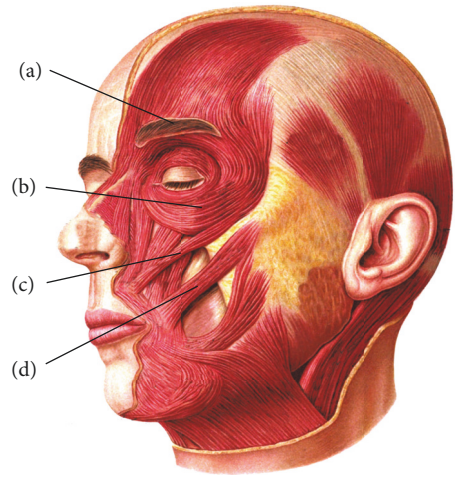


Figure 1: Facial muscles. (a) Corrugator supercilii. (b) Orbicularis oculi. (c) Zygomaticus minor. (d) Zygomaticus major. Adapted from "Sobotta's Atlas and Text-book of Human Anatomy," by Dr. Sobotta (Illustration: Hajek and Schmitson), 1909, in the public domain [35].

distance among face points to train a Support Vector Machine (SVM) model to detect expressions. Similarly Chang et al. [25] use 12 distances calculated from 14 landmarks to detect fear, love, joy, and surprise. Hammal et al. [26] use 5 facial distances calculated from lines in key regions of the face derived from the MPEG-4 animation standard [13], for example, eyebrows, for classification of expressions. Tang and Huang [27, 28] use up to 30 Euclidean distances among facial landmarks also obtained from MPEG-4 based 3D face models to recognize the 6 universal facial expressions. Similarly Hupont et al. [29] classify the same emotions by using a correlation-based feature selection technique to select the most significant distances and angles of facial points. Finally Akakn and Sankur [30] use the trajectories of facial landmarks to recognize head gestures and facial expressions.

Some visual methods rely on manual or automated FACS-based analysis as a standard for categorization and measuring of emotional expressions [31]. Kaiser et al. [17] demonstrate that more AU were reported by manual FACS coders during the analysis of video recordings of subjects playing the stressful part of a game when compared to its neutral part. Additionally authors report lip pull corner and inner/outer brow raise as more frequent AUs during gaming sessions. Wehrle and Kaiser [32] use an automated, FACS-based facial analysis aggregated with data from game events to provide an appraisal analysis of subjects emotional state. Similarly Grafsgaard et al. [33] use an automated, FACS-based analysis to report a relationship between facial expression and aspects of engagement, frustration, and learning in tutoring sessions. Contrary to previous work, Heylen et al. [34] do not rely on FACS, but instead use an empirical, manual facial analysis based on the authors' interpretation of the context. Heylen et al. [34] found that most of the time subjects remain with a neutral face.

The use of facial expressions as a single source of information, however, is contested in the literature. Blom et al. [36] report that subjects present a neutral face during most of the time of gameplay and frustration is not captured by face expressions, but by head movements, talking, and hand gestures instead. In a similar conclusion, Shaker et al. [37] show that head expressivity, that is, movement and velocity, is an indicator of how experienced one is on games. Additionally high frequency and velocity of head movements are indicative of failing in the game. Finally Giannakakis et al. [38] reported increased blinking rate, head movement, and heart rate during stressful situations.

Facial analysis based on physical sensors, for example, EMG, provides continuous monitoring of subjects and is not affected by lighting conditions or pose occlusion by subject's movement. However the sensors are obtrusive and the use of sensors increases user's awareness of being monitored [39–41]. Approaches based on video analysis, for example, FACS and computer vision, are less intrusive. Despite the fact that FACS has proven to be a useful and quantitative approach for measuring facial expressions [31], its manual application is laborious and time-consuming and requires certified coders to inspect the video recordings. The application of FACS also has downsides, including different facial expression decoding caused by misinterpretation in specific cultures [42]. Facial analysis from visual methods, such as the previously mentioned feature-based approaches relying on computer vision, is quicker and easier to deploy. However previous works commonly focus on analyzing images or videos whose subjects performed facial expressions on guidance. Those are artificial circumstances that do not portrait natural interactions of users and games, for instance. When the analysis is performed on videos of subjects interacting with games, usually the aim is to detect a very specific set of facial expressions, for example, 6 universal facial expressions, disregarding head movement and subtle changes in facial behavior.

Our approach focuses on performing facial analysis on subjects interacting with games with natural behavior and genuine emotional reactions. The novel configuration of our experiment provokes two distinct emotional states on

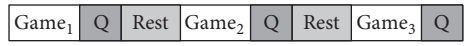


Figure 2: Experimental procedure. $Game_i$ represents the ith interaction of a subject with a game, Q is when the subject answered a questionnaire, and *Rest* is a 138-second period when the subject rested.

(a) (b) (c)

Figure 3: Games used in the experiment. From (a) to (c): Mushroom, where the player must sort bad from good mushrooms by analyzing color patterns; Platformer, where the player must jump over or slide below obstacles while collecting hearts; Tetris, which is a clone of the original version of the game, however without hints about the next piece to enter the screen.

subjects, that is, boredom and stress, which are elicited from interaction with games, not videos or images. Additionally our method focuses on detecting facial nuances from calculations based on the Euclidean distances between facial landmarks instead of categorizing predefined facial expressions. We empirically show that such features have the potential to differentiate emotional states of boredom and stress in games. Our calculated facial features can be used as one of the inputs of multimodal emotion detection models.

3. Method

3.1. Experimental Setup. Twenty adult participants of both genders (10 female) with different ages (22 to 59, mean 35.4, SD 10.79) and different gaming experience gave their informed and written consent to participate in the experiment. The study population consisted of staff members and students of the University of Skövde, as well as citizens of the community/city (see [16] for more information about subjects). Subjects were seated in front a computer, alone in the room, while being recorded by a camera and measured by a heart rate sensor. The camera was attached to a tripod placed in front of the subjects at approximately 0.6 m of distance; the camera was slightly tilted up. A spotlight, tilted 45° up and placed at a distance of 1.6 m from the subject and 45 cm higher than the camera level, was used for illumination; no other light source was active during the experiment.

Participants were each recorded for about 25 minutes, during which they played three different games (described in Section 3.1.1), rested, and answered questions. Figure 2 illustrates the procedure. $Game_i$ represents the ith interaction of a subject with a game. The order of the three games which were played was randomized among subjects. Each game was followed by a questionnaire related to the game and stress/boredom. The first two games were followed by a 138-second rest period, where subjects listened to calm classical music. Before starting the experiment, participants received instructions from a researcher saying that they should play three games, answer a questionnaire after each game, and rest; they were told that their gaming performance was not being analyzed, that they should not give up in the middle of the games, and that they should remain seated during the whole process.

3.1.1. Games and Stimuli Elicitation. The three games used in the experiment were 2D and casual-themed, played with mouse or keyboard in a web browser. When keyboard was used as input, the keys to control the game were deliberately chosen to be distant from each other, requiring subjects to use both hands to play. It reduces the risk for facial occlusion during game play, for example, hand interacting with the face. The games were carefully designed to provoke boredom at the beginning and stress at the end, with a linear progression between the two states (adjustments of such progression are performed every 1 minute). The game mechanics were chosen based on the capacity to fulfill such linear progression, along with the quality of not allowing the player to kill the main character instantly (by mistake or not), for example, by falling into a hole. The mechanics were also designed/selected in a way to ensure that all subjects would have the same game pace; for example, a player must not be able to deliberately control the game speed based on his/her will or skill level, for instance. Figure 3 shows each one of the games.

The *Mushroom* game, shown in Figure 3(a), is a puzzle where the player must repeatedly feed a monster by dragging and dropping mushrooms. Boredom is induced with fewer mushrooms to deal with and plenty of time for the task, while stress is induced with increased number of mushrooms and limited time to drag them. The *Platformer* game, shown in Figure 3(b), is a side-scrolling game where the player must control the main character while collecting hearts and avoiding obstacles (skulls with spikes). Boredom is induced

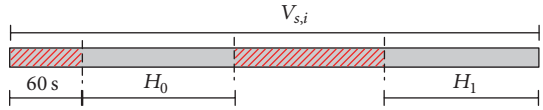


Figure 4: Extraction of video segments H_0 and H_1 containing boring and stressful game interactions, respectively. Initial 60 seconds of any video $V_{s,i}$ are ignored and the remaining is divided into three pieces, from which the first and the last ones are selected. Stripes highlight discarded video segments.

with a slow pace and almost no hearts or obstacles appearing on the screen, while stress is induced with a faster pace, several obstacles, and almost no hearts to collect. Finally the game *Tetris*, shown in Figure 3(c), is a modified version of the original Tetris game. In our version of the game, the next block to be added to the screen is not displayed and the down key, usually used to speed up the descendant trajectory of the current piece, is disabled, preventing players from speeding up the game. Boredom is induced by slow falling pieces, while stress is induced by fast falling pieces. All games used the same seed for random calculations, which ensured subjects received the same sequence of game elements, for example, pieces in Tetris. For a detailed description of the games, refer to [16].

Previous analysis conducted on the video recordings of the experiment [18] supports the use of three custom-made games with linear and constant progression from a boring to a stressful state, without predefined levels, modes, or stopping conditions as a valid approach for the exploration of facial behavior and physiological signals regarding their connection with emotional states. Previous results confirm with statistical significance that (1) subjects perceived the games as being boring at the beginning and stressful at the end; (2) the games induced emotional states, that is, boredom and stress, and caused physiological reactions on subjects, that is, changes in HR. Analyses of such changes indicate that HR mean during the last minute of gameplay (perceived as stressful) was greater than during the second minute of gameplay (perceived as boring). An exploratory investigation suggests that HR mean during the first minute of gameplay was greater than during the second minute of gameplay, probably as a consequence of unusual excitement during the first minute, for example, idea of playing a new game. Finally manual and empirical analyses of the video recordings show more facial activity in stressful parts of the games compared to boring parts [16].

Our experimental configuration and previous analysis provide a validated foundation for the application and evaluation of our method for automated analysis of facial cues from videos. Our intent is to test it as a potential tool for differentiating emotional states of stress and boredom of players in games, which can be evaluated with our experimental configuration, since such information can be categorized according to the induced (and theoretically known) emotional states of subjects.

3.1.2. Data Collection. During the whole experiment, subjects were recorded using a Canon Legria HF R606 video camera. All videos were recorded in color (24-bit RGB with three channels × 8 bits/channel) at 50p frames per second (FPS) with pixel resolution of 1920 × 1080 and saved in AVCHD-HD format, MPEG-4 AVC as the codec. At the same time, their heart rate (HR) was measured by a TomTom Runner Cardio watch (TomTom International BV, Amsterdam, Netherlands), which was placed on the left arm, approximately 7 cm away from the wrist. The watch recorded the HR at 1 Hz.

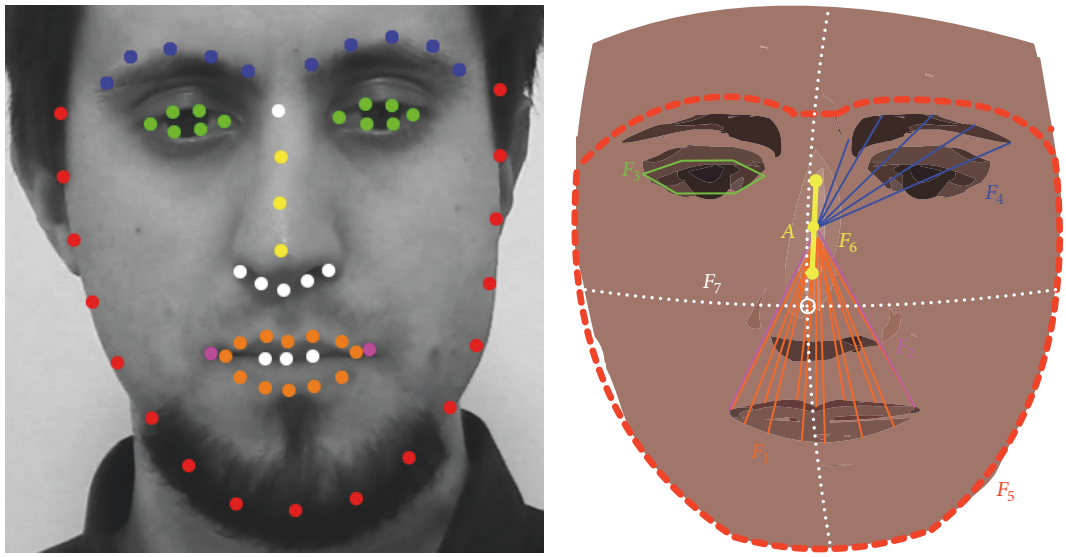
3.2. Data Preprocessing. The preprocessing of video recordings involved extraction of the parts containing the interaction with the games and the discard of noisy frames. Firstly we extracted from the video recordings the periods where subjects were playing each one of the available games. It resulted in three videos per subject, denoted as $V_{s,i}$ where s is the sth subject and $i \in \{1, 2, 3\}$ represents the game.

As previously mentioned, the games used as emotional elicitation material in the experiment induced variations of physiological signals on subjects, who perceived them as being boring at the beginning and stressful at the end. Since our aim is to test the potential of our facial features to differentiate emotional states of boredom and stress, we extracted from each video $V_{s,i}$ two video segments, named H_0 and H_1, whose subject's emotional state is assumed to be known and related to boredom and stress. In order to achieve that, we performed the following extraction procedure, illustrated in Figure 4. Firstly we ignored the initial 60 seconds of any given video $V_{s,i}$. The remaining of the video was then divided into three pieces, from which the first and the last were selected as H_0 and H_1, respectively.

The reason why we discarded the initial part of all game videos is because we believe the first minute might not be ideal for a fair analysis. During the first minute of gameplay, subjects are less likely to be in their usual neutral emotional state. They are more likely to be stimulated by the excitement of the initial contact with a game soon to be played, which interferes with any feelings of boredom. Additionally subjects need basic experimentation with the game to learn how to play it and judge if it is boring or not. Such claim is supported by empirical analysis of the first minute of the video recordings that show repeated head and eye movements from and towards the keyboard/display. As per our understanding, the second minute and onward in the videos is more likely to portrait facial activity related to emotional reactions to the game instead of facial activity connected to gameplay learning. Regarding the division of the remaining part of the video into three segments, from which two were selected as H_0 and H_1, we followed the reasoning that the emotional state of subjects was unknown in the middle part of $V_{s,i}$. Based on self-reported emotional

Table 1: Information regarding calculated facial features.

Name	Notation	Description
Mouth outer	F_1	Sum of the Euclidean distance between the mouth contour landmarks and the anchor landmarks. It monitors the zygomatic muscle.
Mouth corner	F_2	Sum of the Euclidean distance between the mouth corner landmarks and the anchor landmarks. It monitors the zygomatic muscle.
Eye area	F_3	Area of the regions bounded by the closed curves formed by the landmarks in contour of the eyes. It monitors the orbicularis oculi muscle.
Eyebrow activity	F_4	Sum of the Euclidean distance between eyebrow landmarks and the anchor landmarks. It monitors the corrugator muscle.
Face area	F_5	Area of the region bounded by the closed polygon formed by the most external detected landmarks.
Face motion	F_6	Average value of the Euclidean norm of a set of landmarks in the last N frames. It describes the total distance the head has moved in any direction in a short period of time.
Facial COM	F_7	Average value of all detected landmarks. It describes the overall movement of all facial landmarks.



(a) (b)

Figure 5: Facial landmarks and features. (a) Highlight of 68 detected facial landmarks. (b) Visual representation of our facial features.

states, subjects reported the beginning part of the games as boring and the final part as stressful; additionally there are significant differences in the HR mean between the second and the last minute of gameplay in the games [18]. Consequentially we understand that video segments H_0 and H_1 accurately portray interaction of subjects during boring and stressful periods of the games, respectively.

The preprocessing of the recordings resulted in 6 video segments per subjects: 3 segments H_0 (one per game) and 3 segments H_1 (one per game). A given game i contains $N = 20$ pairs of H_0 and H_1 video segments (20 segments H_0, one per subject, and 20 segments H_1, one per subject). When considering all subjects and games, there are $N = 60$ pairs of H_0 and H_1 video segments (3 games $\times$ 20 subjects, resulting in 60 segments H_0 and 60 segments H_1). Subject 9 had problems playing the Platformer game, so segments H_0 and H_1 from subject 9 in the Platformer game were discarded. Consequentially the Platformer game contains $N = 19$ pairs of H_0 and H_1 video segments; regarding all games and subjects, there are $N = 59$ pairs of H_0 and H_1 video segments.

3.3. Facial Features. The automated facial analysis we propose is based on the measurement of 7 facial features calculated from 68 detected facial landmarks. Table 1 presents the facial features, which are illustrated in Figure 5(b). Our facial features are mainly based on the Euclidean distances between landmarks, similar to some works previously mentioned; however, our approach does not rely on predefined expressions, that is, 6 universal facial expressions, training of a model, or the use of the MPEG-4 standard, which specifies representations for 3D facial animations, not emotional interactions in games. Additionally our method does not use an arbitrarily selected frame, for example, the 100th frame [38], as a reference for calculations, since our features are

derived from each frame (or a small set of past frames). Our features are obtained unobtrusively via computer vision analysis focused on detecting activity of facial muscles reported by previous work involving EMG and emotion detection in games. We believe our approach is more user-tailored, convenient, and better suited for contexts involving games.

The process of extracting our facial features has two main steps: face detection and feature calculation. In the first step, computer vision techniques are applied to a frame of the video and facial landmarks are detected. In the second step, the detected landmarks are used to calculate several facial features related to eyes, mouth, and head movement. The following sections present in detail how each step is performed, including details regarding the calculation of features.

3.3.1. Face Detection. The face detection procedure is performed for every frame of the input video. We detect the face using a Constrained Local Neural Field (CLNF) model [43, 44]. CLNF uses a local neural field patch expert that learns the nonlinearities and spatial relationships between pixel values and the probability of landmark alignment. The technique also uses a nonuniform regularized landmark Mean Shift fitting technique that takes into consideration patch reliabilities. It improves the detection process under challenging conditions, for example, extreme face pose or occlusion, which is likely to happen in game sessions [16]. The application of the CLNF model to a given video frame produces a vector L of 68 facial landmarks:

$$L = [p_0, p_1, p_2, \ldots, p_{67}]^T, \quad (1)$$

where p_i is a detected facial landmark that represents a 2D coordinate (x_i, y_i) in the frame. Facial landmarks are related to different facial regions, such as eyebrows, eyes, and lips. Figure 5(a) illustrates the landmarks of L in a given frame.

3.3.2. Anchor Landmarks. The calculation of our facial features involves the Euclidean distance among facial landmarks. Subsequently the Euclidean distance between two landmarks $a_1 = (x_1, y_1)$ and $a_2 = (x_2, y_2)$ is given as follows:

$$d(a_1, a_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}. \quad (2)$$

Landmarks in the nose area are more likely to be stable, presenting fewer position variations in consecutive frames [38]. Consequently they are good reference points to be used in the calculation of the Euclidean distance among landmarks. In order to provide stable reference points for the calculation of our facial features, we selected 3 highly stable landmarks located in the nose line, denoted as the anchor vector $A = [p_{28}, p_{29}, p_{30}]^T$. The landmarks of the anchor vector A are highlighted in yellow in Figure 5(a).

3.3.3. Feature Normalization. Subjects moved towards and away from the camera during the gaming sessions. This movement affects the Euclidean distance between landmarks, as it tends to increase when the subject is closer to the camera, for instance. Additionally subjects have unique facial shapes and characteristics, which also affect the calculation and comparison of the facial features between subjects. To mitigate that problem, we calculated a normalization coefficient K as the Euclidean distance between the upper and lower most anchor landmarks in A. In other words, K represents the size of the subjects nose line. Since all features are divided by K, their final value is expressed as normalized pixels (relative to K) rather than pixels per se.

3.3.4. Mouth Related Features. Mouth related features aim to detect activity in the zygomatic muscles, illustrated in Figures 1(c) and 1(d), which are related to changes in the mouth, such as lips activity (stretch, suck, press, parted, tongue touching, and bite) and movement (including talking). We calculate two facial features related to the mouth area: mouth outer and mouth corner.

Mouth Outer (F_1). Given vector $M = [p_{48}, p_{49}, \ldots, p_{60}]^T$ containing the landmarks in the outer part of the mouth (highlighted in orange in Figure 5(a)). The mouth outer feature is calculated as the sum of the Euclidean distance among the landmarks in M and the anchor landmarks in A:

$$F_1 = \frac{1}{K} \sum_{i=1}^{12} \sum_{j=1}^{3} d\left(A_j, M_i\right), \quad (3)$$

where A_j and M_i are the jth and ith element of A and M, respectively.

Mouth Corner (F_2). Given vector $C = [p_{48}, p_{54}]^T$, containing the two landmarks representing the mouth corners (highlighted in pink in Figure 5(a)). The mouth corner feature is the sum of the Euclidean distance among the landmarks in C and A:

$$F_2 = \frac{1}{K} \sum_{i=1}^{2} \sum_{j=1}^{3} d\left(A_j, C_i\right), \quad (4)$$

where A_j and C_i are the jth and ith element of A and C, respectively.

3.3.5. Eye Related Features. Eye related features aim to detect activity related to the orbicularis oculi and the corrugator muscles, illustrated in Figures 1(b) and 1(a), respectively, which comprehend changes in the eyes region, including eye and eyebrow activity. We calculated two facial features related to the eyes: eye area and eyebrow activity.

Eye Area (F_3). Given vector $Y_l = [p_{36}, p_{37}, \ldots, p_{41}]^T$ containing the landmarks describing the left eye, highlighted in green in Figure 5(a), and vector $Y_r = [p_{42}, p_{43}, \ldots, p_{47}]^T$ containing the landmarks describing the right eye, highlighted in green in Figure 5(a). The eye area feature is the area of the regions bounded by the closed curves formed by the landmarks in Y_l and Y_r, divided by K. We calculated the area of the curves using OpenCV's `contourArea()` function, which uses Green's theorem [45].

Eyebrow Activity (F_4). It is calculated as the sum of the Euclidean distances among the eyebrow landmarks and the anchor landmarks in A. Given the vector $W_l = [p_{17}, p_{18}, \ldots, p_{21}]^T$ containing the landmarks describing the left eyebrow, highlighted in blue in Figure 5(a), and the set $W_r = [p_{22}, p_{23}, \ldots, p_{26}]^T$ containing the landmarks describing the right eyebrow, highlighted in blue in Figure 5(a). The eyebrow activity feature is calculated as follows:

$$F_4 = \frac{1}{K}\sum_{i=1}^{5}\sum_{j=1}^{3}\left[d\left(A_j, W_{l,i}\right) + d\left(A_j, W_{r,i}\right)\right], \quad (5)$$

where A_j, $W_{l,i}$, and $W_{r,i}$ are the jth, ith, and ith element of A, W_l, and W_r, respectively.

3.3.6. Head Related Features. Head related features aim to detect body movements, in particular variations of head pose and amount of motion that the head/face is performing over time. We calculated three features related to the head: face area, face motion, and facial center of mass (COM).

Face Area (F_5). During the interaction with a game, subjects tend to move towards (or away from) the screen, which causes the facial area in the video recordings to increase or decrease. Given vector $F = [p_0, p_1, \ldots, p_{16}]^T$ containing the landmarks describing the contour of the face, highlighted in red in Figure 5(a). The face area feature is the area of the region bounded by the closed curves formed by the landmarks in $F \cup W_r \cup W_l$, divided by K. Similar to the eye area, we calculated the area under the curves using OpenCV's `contourArea()` function.

Face Motion (F_6). It accounts for the total distance the head has moved in any direction in a short period of time. For each frame of the video, we save the currently detected anchor vector A, which produces vector $D = [A_1, A_2, \ldots, A_n]^T$, where A_i is the vector A detected in the ith frame of the video and n is number of frames in the video. We then calculate the face motion feature as follows:

$$F_6 = \frac{1}{K}\sum_{j=1}^{3}\sum_{t=1}^{Z-1}\left\|D\left(f-t, j\right) - D\left(f-Z, j\right)\right\|, \quad (6)$$

where Z is the amount of frames to include in the motion analysis, $D(i, j)$ is the jth element of $A_i \in D$, f is the number of the current frames, and $\|\cdot\|$ is the Euclidean norm. In our analysis, we used $Z = 50$ (50 frames, equivalent to 1 second).

Facial COM (F_7). It describes the overall movement of all facial landmarks. A single 2D point, calculated as the average of all landmarks in L, is used to monitor the movement. The COM feature is calculated as follows:

$$F_7 = \frac{1}{K}\frac{1}{N}\sum_{i=1}^{N}\left\|p_i\right\|, \quad (7)$$

where N is the total number of detected landmarks (elements in L) and $\|\cdot\|$ is the Euclidean norm.

Table 2: Mean of differences (±SD) of features between periods H_0 and H_1 ($N = 59$). Units expressed in normalized pixels.

Feature (notation)	
Mouth outer (F_1)	$-20.59 \pm 57.36^{**}$
Mouth corner (F_2)	$-3.90 \pm 10.16^{**}$
Eye area (F_3)	$-0.019 \pm 0.064^{*}$
Eyebrow activity (F_4)	$-15.59 \pm 49.71^{*}$
Face area (F_5)	$-2.60 \pm 7.90^{*}$
Face motion (F_6)	-44.97 ± 326.74
Facial COM (F_7)	-0.029 ± 0.113

$^{*}p < 0.05$; $^{**}p < 0.01$.

3.4. Feature Analysis. The previously mentioned features can be calculated for each frame of any given video; however, facial cues might be better contextualized if analyzed in multiple frames. For that reason, we applied our facial analysis to all frames of all video segments H_0 and H_1. We then calculated the mean value of each facial feature in each video segment. As a result, any facial feature F_i has $N = 59$ pairs of mean values (59 from H_0 and 59 from H_1). From now on, we will refer to the set of mean values in H_0 or H_1 of a given feature F_i simply as feature value in H_0 or H_1, respectively.

Based on a previous manual analysis of facial actions of the video recordings [16] and findings of related work, values of facial features during boring periods of the games are expected to be different than those during stressful periods. Since subjects perceived the games as boring at the beginning and stressful at the end, we assume that values in H_0 and H_1, for all features, are likely to correlate with an emotional state of boredom and stress, respectively. Consequentially we state the following overarching hypothesis: the mean value of features in H_0 is different than the mean value in H_1, for all subjects and games. More specifically, we can describe the overarching hypothesis as 7 subhypotheses, denoted as u_i, where $i \in \{1, 2, \ldots, 7\}$. Hypothesis u_i states that the true difference in means between the value of a given feature F_i in H_0 and H_1, for all subjects, is greater than zero. The dependent variable of u_i is F_i and the null hypothesis is that the true difference in means between H_0 and H_1 for feature F_i, for all subjects and games, is equal to zero.

We tested hypothesis u_i by performing a paired two-tail t-test on the values H_0 and H_1 of feature F_i. We performed 7 tests in total: u_1 (mouth outer), u_2 (mouth corner), u_3 (eye area), u_4 (eyebrow activity), u_5 (face area), u_6 (face motion), and u_7 (facial COM).

4. Results

Table 2 presents the mean of differences of all features between periods H_0 and H_1, calculated for all subjects in all games according to the description in Section 3.3 and analyzed according to the procedures described in Section 3.4. The mean of differences of all features shows a decrease from H_0 to H_1. Comparing the mean difference

of a feature to its mean value in H_0, the decrease from H_0 to H_1 was 10.7% for mouth outer (F_1), 11.8% for mouth corner (F_2), 10.4% for eye area (F_3), 8.1% for eyebrow activity (F_4), 9.4% for face area (F_5), 8.2% for face motion (F_6), and 11% for facial COM (F_7). Changes related to F_6 and F_7 were not statistically significant. All remaining features presented statistically significant changes from H_0 to H_1. The highest decrease with statistical significance was associated with mouth corner, followed by mouth outer, eye area, face area, and eyebrow activity. Those numbers support our experimental expectations that the values for facial features are different when compared between two distinct parts of the games, that is, boring and stressful ones.

The two facial features related to mouth, that is, mouth corner and mouth outer, presented a combined average decrease of 11.24% from H_0 to H_1. The change was the highest compared to all other features. The mean of differences of F_1 and F_2 between periods H_0 and H_1 was $T(59) = -20.59$ (SD 57.36, $p < 0.01$) and $T(59) = -3.9$ (SD 10.16, $p < 0.01$), respectively. Both features had a statistically significant change from H_0 to H_1, which supports the claim that they are different in those periods. Additionally, both features presented SD considerably greater than the mean, which indicates that differences of such features for each subject between periods H_0 and H_1 are likely to be spread out rather than being clustered around the mean value. Features related to eyes, that is, eye area and eyebrow activity, presented a combined average decrease of 9.28% from H_0 to H_1. The mean of differences of F_3 and F_4 between periods H_0 and H_1 was $T(59) = -0.019$ (SD 0.064, $p < 0.05$) and $T(59) = -15.59$ (SD 49.71, $p < 0.05$), respectively. Similar to mouth related features, eye related features had a statistically significant change from H_0 to H_1, indicating that they are different in those periods. Following the same pattern of change of F_1 and F_2, both features F_3 and F_4 also presented a SD considerably greater than the mean, also suggesting that differences of such features for each subject between periods H_0 and H_1 are likely to be spread out rather than being clustered around the mean value.

Finally features related to the whole face, that is, face area, face motion, and facial COM, presented a combined average decrease of 9.52% from H_0 to H_1. The mean of differences of F_5, F_6 and F_7 were $T(59) = -2.60$ (SD 7.90, $p < 0.05$), $T(59) = -44.97$ (SD 326.74, $p = 0.29$), and $T(59) = -0.029$ (SD 0.113, $p = 0.052$), respectively. Face area was the only feature in this category to present a change that was statistically significant between periods H_0 and H_1, supporting the idea that F_5 is different in those periods. On the contrary, F_6 and F_7 lack statistical significance in their differences between periods H_0 and H_1. Similar to facial features related to mouth and eyes, features F_5, F_6, and F_7 presented SD considerably greater than the mean, also suggesting that differences of such features between period H_0 and H_1 are likely to be spread out rather than being clustered around the mean value.

5. Discussion

5.1. Feature Analysis. The overarching hypothesis states that the mean value of features in H_0 is different than the mean value in H_1. The overarching hypothesis is composed of 7 subhypotheses, that is, u_i, one for each feature F_i, where u_i states that the true difference in means between the value of a given feature F_i in H_0 and H_1 is greater than zero. The majority of the calculated facial features, that is, mouth outer (F_1), mouth corner (F_2), eye area (F_3), eyebrow activity (F_4), and face area (F_5), presented statistically significant differences in their mean values when compared between two distinct parts of the games, that is, H_0 and H_1. As previously mentioned, subjects perceived the first part of the games, that is, H_0, as being boring and the second part, that is, H_1, as being stressful. Results support the claim of subhypotheses u_1 to u_5, which indicate that facial features F_1 to F_5 can be differentiated between periods H_0 and H_1 and consequentially have the potential to unobtrusively differentiate emotional states of boredom and stress of players in gaming sessions. Our results refute subhypotheses u_6 and u_7, since features F_6 and F_7 lack statistical significance to be differentiated between periods H_0 and H_1.

Mouth related facial features, that is, mouth outer (F_1) and mouth corner (F_2), presented statistically significant differences between boring and stressful parts of the games. Both features are calculated based on the distance between mouth and nose related facial landmarks, which presented a decrease in stressful parts of the games. Such decrease could be attributed to landmarks in the upper and lower lips being closer to each other, which could be associated with lips pressing, lips sucking, or talking, for instance. Particularly to the mouth corner feature, a decrease in distance is the result of the two mouth corners being placed closer to the nose area, which could be associated with smiles or mouth deformation, for example, mouth corner pull to left/right. Consequentially, a decrease in the mean value of both features suggests higher mouth activity that involves the approximation of mouth landmarks to the nose area in stressful parts of the games compared to boring parts. Such results are aligned with previous studies that show lip pull corner as a frequent facial behavior during gaming sessions [17] and talking as an emotional indicator [36]. Additionally, stating that our mouth related features were constructed after the zygomatic muscle activity, our results are connected with previous studies that show increased activity of the zygomatic muscle related to self-reported emotions [3] and its connection to changes in a game [22].

Eye related features, that is, eye area (F_3) and eyebrow activity (F_4), also presented statistically significant differences between boring and stressful parts of the games. They presented a decrease in the mean value from H_0 to H_1, which points to landmarks detected in the eyes contour becoming closer to each other in H_1. It suggests that more pixels in the eyes area were detected during H_0 (boring part) and then H_1 (stressful part). Such numbers might indicate less blinking activity or more wide-open eyes during boring parts of the games. Additionally it could indicate more blinking and eye tightening activity (possibly related to frowning) during stressful parts. Both indications are aligned with previous findings, which show increased blinking activity (calculated from eye area) in stressful situations [38]. Regarding the eyebrow feature, its calculation is based on the distance between

facial landmarks in the eyebrow lines and the nose. A decrease in value indicates a smaller distance among eyebrows and nose, which could be explained by frowning, suggesting that subjects presented more frowning action during stressful moments of the game. The mean value of eyebrow activity during H_0 is greater than during H_1, which indicates that the distance between eyebrows and nose was greater during boring parts of the games compared to stressful parts. It could also be the result of more eyebrow risings, for example, facial expressions of surprise, in boring periods compared to stressful periods. Our eye related features were constructed to monitor the activity of the orbicularis oculi and the corrugator supercilii muscles, and our results are connected with previous work that report game events affecting the activity of the orbicularis oculi [22] and the corrugator [21] muscles.

Finally features related to the whole face, that is, face area (F_5), face motion (F_6), and facial COM (F_7), are partially conclusive. Those features are affected by body motion, for example, head movement and corporal posture, so a decrease in value might indicate less corporal movements during H_1 compared to H_0. Face area was the only feature in this category to present a change that was statistically significant. The value of the face area feature is directly connected to subjects' movement towards and away from the camera. A decrease in face area from H_0 to H_1 suggests that subjects were closer to the computer screen more often during boring parts of the games and then during stressful parts. The facial COM feature also presented a decrease from H_0 to H_1. Such feature is connected to vertical and horizontal movements performed by subject's face, being anchored to a fixed reference point and less influenced by head rotations. Despite presenting a change that is not statistically significant ($p = 0.519$), the decrease of facial COM might be an indication that subjects were more still during stressful periods than during boring periods. The face motion feature also presented a decrease from H_0 to H_1 that is not statistically significant ($p = 0.294$). This feature accounts for the amount of movement a subject's face performs in a period of 50 frames (dynamic reference point), which is directly affected by vertical, horizontal, and rotational movements of the head. A decrease could be associated with subjects moving/rotating the head less often during the analyzed 50 frames periods in H_1 than H_0. However, absence of statistical significance suggests the change is not related to subject's emotional state, but other factors such as the inherent behavior associated with game mechanics, that is, head movement caused by observation of cards in the Mushroom game. Our results lack the statistical significance to replicate the findings of previous work, which connect head movements to changes in games, that is, failure [37] and frustration [36], or to stressful situations [38].

It could be argued that the characteristics of each game mechanic influence the mean change of features between the two periods. Such argument is particularly true to features that are calculated based on subject's body movement, that is, face area, face motion, and facial COM. In that case, subjects could move the face as a result of in game action, that is, inspecting mushrooms, rather than being an emotional manifestation. Additionally the mean change of features between the two periods presented SD considerably greater than the mean value, indicating that differences between periods are likely to be spread out. It suggests significant between-subject variations for each feature or game. In order to further explore such topics, we analyzed the changes of all features on a game level. Tables 3, 4, and 5 present the mean, minimum, and maximum change presented by features, in percentages, from period H_0 to H_1, calculated from all subjects in the Mushroom, Platformer, and Tetris game, respectively.

Mouth and eye related features, that is, F_1 to F_4, presented, on average, a decrease from H_0 to H_1 in all three games. However, the decrease does not apply to all subjects, since at least one presented an increase from H_0 to H_1, as demonstrated by the positive values in the Max column in Tables 3, 4, and 5. Comparatively, the mean, minimum, and maximum change

Table 3: Percentage of change of features from period H_0 to H_1 in the Mushroom game ($N = 20$).

Feature (notation)	Mean	Min.	Max.
Mouth outer (F_1)	−12.9	−69.1	22.1
Mouth corner (F_2)	−15.0	−71.6	15.5
Eye area (F_3)	−8.9	−76.9	8.2
Eyebrow activity (F_4)	−8.0	−72.3	9.6
Face area (F_5)	−11.3	−74.5	18.2
Face motion (F_6)	47.2	−61.3	253.8
Facial COM (F_7)	−12.9	−81.0	9.8

Table 4: Percentage of change of features from period H_0 to H_1 in the Platformer game ($N = 19$).

Feature (notation)	Mean	Min.	Max.
Mouth outer (F_1)	−7.4	−54.0	16.9
Mouth corner (F_2)	−8.2	−55.9	15.5
Eye area (F_3)	−6.8	−30.4	20.0
Eyebrow activity (F_4)	−4.9	−31.1	7.8
Face area (F_5)	−5.9	−43.8	14.2
Face motion (F_6)	0.9	−60.2	112.7
Facial COM (F_7)	−3.6	−42.1	23.1

Table 5: Percentage of change of features from period H_0 to H_1 in the Tetris game ($N = 20$).

Feature (notation)	Mean	Min.	Max.
Mouth outer (F_1)	−1.5	−27.8	39.0
Mouth corner (F_2)	−2.1	−26.5	26.9
Eye area (F_3)	−2.6	−19.0	26.1
Eyebrow activity (F_4)	−3.3	−16.2	21.1
Face area (F_5)	−1.4	−24.3	26.7
Face motion (F_6)	−11.3	−85.8	114.3
Facial COM (F_7)	−2.7	−24.7	21.8

of mouth (F_1, F_2) and eye (F_4, F_5) related features are similar in the three games. Consequentially, it is our understanding that features F_1 to F_4 are not affected by the game mechanics; however, they do differ on a subject basis. On the other hand, features related to the whole face, that is, F_5 to F_7, seem to be affected by game mechanics. Both F_5 and F_7 presented, on average, a decrease in the three games. Contrarily F_6 presented, on average, an increase in the Mushroom and the Platformer game. A disproportional mean increase of 47.2% from H_0 to H_1 for feature F_6 in the Mushroom game compared to the Platformer (0.9% increase) and Tetris (11.3% decrease) suggests that the feature is highly influenced by the mechanic of the Mushroom game. In such game, subjects are likely to move the head to facilitate saccadic eye movements used to inspect the cards. As the difficulty of the game increases, the number of cards to be inspected on the screen also increases, which could potentially lead to more (periodic) head movements towards the stressful part of the game.

Finally all features presented changes from periods H_0 to H_1 whose SD is considerably greater than the mean value, as presented in Table 2. The considerable heterogeneous variation of features, as demonstrated by columns *Min* and *Max* in Tables 3, 4, and 5, supports the claim that differences of features between periods are spread out rather than being clustered around the mean. Even though further analysis is required, the high SD and the broad interval of percentage change of all features in the three games, showing decrease of 76.9% and increase of 8.2% for the same feature in the same game, for instance, highlighting the between-subjects behavioral differences. Our interpretation is that a more user-tailored, as opposed to a group-oriented, use of our facial features is more likely to portray such subject-based differences in a context involving emotional detection and games.

5.2. Comparison with Previous Work. The approach presented by this paper differs from other computer facial expression analysis systems by focusing on the detection of basic elements that comprise complex facial movements rather than on classifying facial expressions. It is aligned with previous work focused on studying the relation between those detected basic elements and emotional states, for example, work by Bartlett et al. [31] and Asteriadis et al. [15]. A direct comparison of our approach with existing facial expression recognition solutions is misleading. Following the direction of Bartlett et al. [31] and Asteriadis et al. [15], this paper intends to investigate facial changes happening in real gaming sessions and the process to detect them. The data related to such changes can then be used to potentially differentiate emotional states, in our case boredom and stress, of players in a gaming context. We present plausible statistical results that support the method and such potential. Previous work focuses on detecting facial expressions per se, including the six universal facial expressions of emotion, typically reporting accuracy rates of machine learning models used to detect those predefined facial expressions. A significant number of those approaches train the models using datasets with images and videos of actors performing facial expressions [24, 26–28, 46], subjects watching video clips [25, 29, 38], or subjects undergoing social exposure [38]. As previously mentioned, those are artificial situations that are significantly different from an interaction with a game. We evaluated our method on a challenging game-oriented context, showing with statistical significance that there are differences between facial activity, not necessarily facial expressions, in two distinct game periods which are associated with particular emotional states, that is, boredom and stress. The process does not rely on a reference point, for example, neutral face, to operate as the majority of previous work. We believe the context of our experiment is sufficiently different from existing work and our results contribute to guide further investigations regarding automated detection and use of basic facial elements as a source of information to infer emotional states of players in games.

6. Limitations

Some limitations of the experimental procedure and analysis should be noted. Firstly our sample size ($N = 20$) is a relatively small number to derive conclusions that can be generalized. A larger sample for the analysis could produce more conclusive results regarding facial activity that could be applied in contexts other than the one presented in our experiment. Our aim, however, is not to standardize the facial behavior of subjects nor to detect particular facial expressions, but to remotely detect basic facial elements and support the claim that they are different in particular moments of the games. As demonstrated with statistical significance, our features present differences at key moments of the games, that is, boring and stressful parts. Those differences were derived from the facial activity of subjects and they do not necessarily rely on the identification of a particular facial expression, for example, joy (smiles). For the context of our experiment, the analysis conducted on those differences shows the potential of our features to differentiate emotional states of boredom and stress. Another limitation is the nature of the games used in the experiment, which are casual and 2D games. Games with different characteristics, for example, 3D games requiring navigation, could produce different results. However we believe our games do have the characteristics expected of a game, such as a sense of challenge and reward, and its 2D nature is not detrimental. The mechanics of the three games are quite different, requiring subjects to perform distinct patterns of eye saccades and head movement to play. The Mushroom, Platformer, and Tetris game require visual attention on the whole screen, on the left side of the screen, and on the top and bottom parts of the screen, respectively. It is our understanding that those elements cover a significant range of different head and eye movement patterns. Those patterns even interfered with some features, for example, face motion (F_6) and facial COM (F_7), as discussed in Section 5.1. Additionally the use of 2D games for studies involving player experience and emotions is recurrent in the literature, for example, an adapted version of Super Mario has been used to create a personalized gaming experience [36], to analyse player behavior [37], and to discriminate player styles based on visual and gameplay cues [15].

7. Conclusion

This paper presented a method for automated analysis of facial cues from videos with an empirical evaluation of its application as a potential tool for detecting stress and boredom of players. The proposed automated facial analysis is based on the measurement of 7 facial features (F_1 to F_7) calculated from 68 detected facial landmarks. Facial features are mainly based on the Euclidean distance of landmarks and they do not rely on predefined expressions, that is, 6 universal facial expressions, nor training of a model nor the use of standards related to the face, for example, MPEG4 and FACS. Additionally, the method does not use an arbitrarily selected frame as a reference for calculations since features are derived from each frame (or a small set of past frames). Features are obtained unobtrusively via computer vision analysis focused on detecting the activity of facial muscles reported by previous work involving emotion detection in games.

The method has been applied to video recordings of an experiment involving games as emotion elicitation sources, which were deliberately designed to cause emotional states of boredom and stress. Results show statistically significant differences in the values of facial features detected during boring and stressful periods of gameplay for features: mouth outer (F_1), mouth corner (F_2), eye area (F_3), eyebrow activity (F_4), and face area (F_5). Features face motion (F_6) and facial COM (F_7) presented variations that were not statistically significant. Results support the claim that our method for automated analysis of facial cues has the potential to be used to differentiate emotional states of boredom and stress of players. The utilization of our method is unobtrusive and video-based, which eliminates need of physical sensors to be attached to subjects. We believe our approach is more user-tailored, convenient, and better suited for contexts involving games. Finally, the information produced by our method might be used to complement other approaches aimed at emotion detection in the context of games, particularly multimodal models.

Currently, work is ongoing to use the proposed method as one of several sources of information in a nonintrusive, multifactorial user-tailored emotion detection mechanism for games. We intend to further investigate the applicability of our method in a new experiment with a larger sample size and the addition of a new game to the experimental setup, for example, commercial off-the-shelf (COTS) game. The facial analysis described here, particularly the differences found in the boring and stressful parts of the games, will be combined to remote photoplethysmographic estimations of heart rate to train a model to identify emotional states of boredom and stress of players. The results presented here will be improved upon and form the basis for a remote emotion detection approach aimed at the game research community.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

The authors would like to thank the participants and all involved personnel for their valuable contributions. This work has been performed with support from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil; University of Skövde; EU Interreg ÖKS Project Game Hub Scandinavia; Federal University of Fronteira Sul (UFFS).

References

[1] E. D. Mekler, J. A. Bopp, A. N. Tuch, and K. Opwis, "A systematic review of quantitative studies on the enjoyment of digital entertainment games," in *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, CHI 2014*, pp. 927–936, May 2014.

[2] P. Rani, C. Liu, N. Sarkar, and E. Vanman, "An empirical study of machine learning techniques for affect recognition in human-robot interaction," *Pattern Analysis and Applications*, vol. 9, no. 1, pp. 58–69, 2006.

[3] T. J. Tijs, D. Brokken, and W. A. IJsselsteijn, "Dynamic game balancing by recognizing affect," in *Fun and Games*, vol. 5294 of *Lecture Notes in Computer Science*, pp. 88–93, Springer, Berlin, Germany, 2008.

[4] J. F. Cohn and F. De la Torre, *Automated Face Analysis for Affective*, The Oxford handbook of affective computing, 2014.

[5] C. T. Tan, D. Rosser, S. Bakkes, and Y. Pisan, "A feasibility study in using facial expressions analysis to evaluate player experiences," in *Proceedings of the 8th Australasian Conference on Interactive Entertainment: Playing the System, IE 2012*, July 2012.

[6] C. T. Tan, S. Bakkes, and Y. Pisan, "Inferring player experiences using facial expressions analysis," in *Proceedings of the 10th Australian Conference on Interactive Entertainment, IE 2014*, pp. 1–8, ACM Press, December 2014.

[7] C. T. Tan, S. Bakkes, and Y. Pisan, "Correlation between facial expressions and the game experience questionnaire," in *Proceedings of the Entertainment Computing-ICEC 2014: 13th International Conference*, vol. 8770, p. 229, Springer, Sydney, Australia, October 2014.

[8] X. Zhou, X. Huang, and Y. Wang, "Real-time facial expression recognition in the interactive game based on embedded hidden markov model," in *Proceedings of the Computer Graphics, Imaging and Visualization*, pp. 144–148, Penang, Malaysia, 2004.

[9] C. Zhan, W. Li, P. Ogunbona, and F. Safaei, "A real-time facial expression recognition system for online games," *International Journal of Computer Games Technology*, vol. 2008, pp. 1–7, 2008.

[10] P. Ekman and W. V. Friesen, 1977, Facial action coding system.

[11] J. F. Cohn, Z. Ambadar, and P. Ekman, "Observer-based measurement of facial expression with the facial action coding system," *The handbook of emotion elicitation and assessment*, pp. 203–221, 2007.

[12] T. Saari and M. Turpeinen, "Towards psychological customization of information for individuals and social groups," in *Designing Personalized User Experiences in eCommerce*, vol. 5 of *Human-Computer Interaction Series*, pp. 19–37, Springer, 2004.

[13] G. A. Abrantes and F. Pereira, "MPEG-4 facial animation technology: survey, implementation, and results," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 2, pp. 290–305, 1999.

[14] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *Journal of Personality and Social Psychology*, vol. 17, no. 2, pp. 124–129, 1971.

[15] S. Asteriadis, K. Karpouzis, N. Shaker, and G. N. Yannakakis, "Towards detecting clusters of players using visual and gameplay behavioral cues," in *Proceedings of the 4th International Conference on Games and Virtual Worlds for Serious Applications, VS-GAMES 2012*, pp. 140–147, October 2012.

[16] F. Bevilacqua, P. Backlund, and H. Engström, "Variations of facial actions while playing games with inducing boredom and stress," in *Proceedings of the 8th International Conference on Games and Virtual Worlds for Serious Applications, VS-Games 2016*, September 2016.

[17] S. Kaiser, T. Wehrle, and P. Edwards, "Multi-modal emotion measurement in an interactive computer game: A pilot-study," in *in Proceedings of the VIII conference of the international society for research on emotions*, pp. 275–279, ISRE Publications Storrs, 1994.

[18] F. Bevilacqua, H. Engström, and P. Backlund, "Changes in heart rate and facial actions during a gaming session with provoked boredom and stress," *Entertainment Computing*, vol. 24, pp. 10–20, 2018.

[19] H. Zacharatos, C. Gatzoulis, and Y. L. Chrysanthou, "Automatic emotion recognition based on body movement analysis: A survey," *IEEE Computer Graphics and Applications*, vol. 34, no. 6, article no. 106, pp. 35–45, 2014.

[20] C. Schrader, J. Brich, J. Frommel, V. Riemer, and K. Rogers, "Rising to the challenge: an emotion-driven approach toward adaptive serious games," in *Serious Games and Edutainment Applications*, pp. 3–28, Springer, 2017.

[21] R. L. Hazlett, "Measuring emotional valence during interactive experiences: boys at video game play," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1023–1026, 2006.

[22] N. Ravaja, T. Saari, M. Salminen, J. Laarni, and K. Kallinen, "Phasic emotional reactions to video game events: a psychophysiological investigation," *Media Psychology*, vol. 8, no. 4, pp. 343–367, 2006.

[23] A. A. Salah, N. Sebe, and T. Gevers, "Communication and automatic interpretation of affect from facial expressions," *Affective Computing and Interaction: Psychological, Cognitive and Neuroscientific Perspectives*, pp. 157–183, 2010.

[24] A. Samara, L. Galway, R. Bond, and H. Wang, "Sensing affective states using facial expression analysis," in *Ubiquitous Computing and Ambient Intelligence*, Lecture Notes in Computer Science, pp. 341–352, Springer International Publishing, 2016.

[25] C.-Y. Chang, J.-S. Tsai, C.-J. Wang, and P.-C. Chung, "Emotion recognition with consideration of facial expression and physiological signals," in *Proceedings of the 2009 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB 2009*, pp. 278–283, April 2009.

[26] Z. Hammal, L. Couvreur, A. Caplier, and M. Rombaut, "Facial expression classification: An approach based on the fusion of facial deformations using the transferable belief model," *International Journal of Approximate Reasoning*, vol. 46, no. 3, pp. 542–567, 2007.

[27] H. Tang and T. S. Huang, "3D Facial expression recognition based on automatically selected features," in *Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops*, pp. 1–8, June 2008.

[28] H. Tang and T. S. Huang, "3d facial expression recognition based on properties of line segments connecting facial feature points," in *Proceedings of the Automatic Face & Gesture Recognition, 8th IEEE International Conference on IEEE*, pp. 1–6, 2008.

[29] I. Hupont, S. Baldassarri, and E. Cerezo, "Facial emotional classification: from a discrete perspective to a continuous emotional space," *PAA. Pattern Analysis and Applications*, vol. 16, no. 1, pp. 41–54, 2013.

[30] H. Ç. Akakn and B. Sankur, "Spatiotemporal-boosted DCT features for head and face gesture analysis," in *Human Behavior Understanding*, pp. 64–74, Springer Nature, 2010.

[31] M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Measuring facial expressions by computer image analysis," *Psychophysiology*, vol. 36, no. 2, pp. 253–263, 1999.

[32] T. Wehrle and S. Kaiser, "Emotion and facial expression," in *Affective Interactions*, vol. 1814 of *Lecture Notes in Computer Science*, pp. 49–63, Springer, Berlin, Germany, 2000.

[33] J. F. Grafsgaard, J. B. Wiggins, K. E. Boyer, E. N. Wiebe, and J. C. Lester, "Automatically recognizing facial expression: Predicting engagement and frustration," in *EDM*, pp. 43–50, 2013.

[34] D. Heylen, M. Ghijsen, A. Nijholt, and R. Op Den Akker, "Facial signs of affect during tutoring sessions," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, vol. 3784, pp. 24–31, 2005.

[35] Wikimedia Commons, *Sobotta's Atlas and Text-book of Human Anatomy 1909*, J. Sobotta, K. Hajek, and A. Schmitson, Eds., 2013, https://commons.wikimedia.org/wiki/File:Sobo_1909_260.png.

[36] P. M. Blom, S. Bakkes, C. T. Tan et al., "Towards personalised gaming via facial expression recognition," in *Proceedings of the 10th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, AIIDE 2014*, pp. 30–36, October 2014.

[37] N. Shaker, S. Asteriadis, G. N. Yannakakis, and K. Karpouzis, "A game-based corpus for analysing the interplay between game context and player experience," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, vol. 6975, no. 2, pp. 547–556, 2011.

[38] G. Giannakakis, M. Pediaditis, D. Manousos et al., "Stress and anxiety detection using facial cues from videos," *Biomedical Signal Processing and Control*, vol. 31, pp. 89–101, 2017.

[39] T. Yamakoshi, K. Yamakoshi, S. Tanaka et al., "A preliminary study on driver's stress index using a new method based on differential skin temperature measurement," in *Proceedings of the 29th Annual International Conference of IEEE-EMBS, Engineering in Medicine and Biology Society, EMBC'07*, pp. 722–725, August 2007.

[40] M. Yamaguchi, J. Wakasugi, and J. Sakakima, "Evaluation of driver stress using biomarker in motor-vehicle driving simulator," in *Proceedings of the Conference Proceedings. Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1834–1837, New York, NY, August 2006.

[41] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 156–166, 2005.

[42] R. E. Jack, "Culture and facial expressions of emotion," *Visual Cognition*, vol. 21, no. 9-10, pp. 1248–1286, 2013.

[43] T. Baltrusaitis, P. Robinson, and L.-P. Morency, "Constrained local neural fields for robust facial landmark detection in the wild," in *Proceedings of the 14th IEEE International Conference*

on Computer Vision Workshops (ICCVW '13), pp. 354–361, Sydney, Australia, December 2013.

[44] T. Baltrusaitis, P. Robinson, and L.-P. Morency, "OpenFace: An open source facial behavior analysis toolkit," in *Proceedings of the IEEE Winter Conference on Applications of Computer Vision, WACV 2016*, pp. 1–10, March 2016.

[45] J. Stewart, "Calculus," 2011, Cengage Learning.

[46] P. Wang, F. Barrett, E. Martin et al., "Automated video-based facial expression analysis of neuropsychiatric disorders," *Journal of Neuroscience Methods*, vol. 168, no. 1, pp. 224–238, 2008.