Face in the Game: Using Facial Action Units to Track Expertise in Competitive Video Game Play

Gianluca Guglielmo
Department of Cognitive Science and Artificial Intelligence
Tilburg University
Tilburg, The Netherlands
G.Guglielmo@tilburguniversity.edu

Paris Mavromoustakos Blom
Department of Cognitive Science and Artificial Intelligence
Tilburg University
Tilburg, The Netherlands
P. Mavromoustakosblom@uvt.nl

Michal Klincewicz
Department of Cognitive Science and Artificial Intelligence
Tilburg University
Tilburg, The Netherlands
M.Klincewicz@tilburguniversity.edu

Boris Čule
Department of Cognitive Science and Artificial Intelligence
Tilburg University
Tilburg, The Netherlands
B.Cule@tilburguniversity.edu

Pieter Spronck
Department of Cognitive Science and Artificial Intelligence
Tilburg University
Tilburg, The Netherlands
P.Spronck@tilburguniversity.edu

Abstract— In this study, we extracted facial action units (AUs) data during a Hearthstone tournament to investigate behavioural differences between expert, intermediate, and novice players. Our aim was to obtain insights into the nature of expertise and how it may be tracked using non-invasive methods such as AUs. These insights may shed light on the endogenous responses in the player and at the same time may provide information to the opponents during a competition. Our results show that player expertise may be characterised by specific patterns in facial expressions. More specifically, AU17 (chin raiser), AU25 (lips apart), and AU26 (jaws drop) intensity responses during gameplay may vary according to players' expertise. Such results were obtained by training a random forest classifier to test whether we can use these three AUs alone to accurately detect players' expertise. The classifier reached 0.75 accuracy on 5-fold cross-validation, after balancing the class weights, and 0.85 after having applied the Synthetic Minority Over-sampling Technique (SMOTE) function. These results suggest that AUs can be effectively used to discriminate different levels of expertise in competitive video game players.

Keywords—video games, player modelling, expertise, machine learning, facial expressions, decision-making

I. INTRODUCTION

Video games are not just entertainment. They are also being used by industries to introduce new concepts of business [1] and to train new employees [2]. Video games have been used in research to investigate group behaviour [2], cultural differences [3], crisis management [4], and moral decision-making [5, 6]. Those seeking entertainment are attracted to video games on account of the experience they provide and by their facilitating social interactions [7]. Video games are one of the preferred mediums of entertainment not only among teenagers but also among adults [8]. Serious players in competitive games spend hours in play but also in developing particular skills. Tournaments are venues for testing these skills against other competitive players. In general, video game tournaments have become a major business for video game companies seeking to promote a title or increase a game's player base [9]. For players, they are an opportunity to test and develop their expertise and to engage in high-level play; this makes tournaments an ideal environment to study video game expertise. In this study, we used player data from a Hearthstone tournament [10]. This type of data is not generally easy to access in non-lab settings and affords an opportunity to study how players behave in a real video game competition. We focused on facial expressions, since they can be measured with non-invasive methods and assumed that they are a proxy for the underlying psychological states of the players. A Random Forest Classifier (RFC) was used to discriminate between levels of expertise using information extracted from facial expressions.

II. BACKGROUND AND RELATED WORK

A. Video games, Expertise, and Decision-making

Experts behave differently from novices. For example, expert traders show a higher high-frequency heart rate variability when trading compared to novices [11] and expert dermatologists have different fixation patterns compared to less experienced colleagues when evaluating dermatological images [12]. Expert video game players also behave differently from novices. Anecdotally, experts are calmer, more focused, and

have more of a "poker face" than novices, just as experts in competitive card games, such as poker [13]. Given this, facial expressions may offer an interesting method to study expertise in video gameplay and particularly competitive card gameplay.

Gray and colleagues [14] used Tetris to identify qualities of video game expertise that go beyond reaction time and manual dexterity: game-specific techniques and strategies for overcoming performance plateaus. Identifying specific techniques allows players to develop personalised training regimens that can push them to the next level of performance. Expertise in Tetris, and perhaps also in other video games that do not depend predominantly on manual dexterity, is a matter of identifying these techniques and exploiting them (in the case of Tetris: planning efficiency, pile management, zoid control, pile uniformity, minimum line clears, and rotation corrections). Importantly to our work with Hearthstone, Gray and colleagues identify three classes of expertise, each connected to mastery over these techniques: novices, intermediates, and experts [15]. While we did not perform a detailed analysis of techniques and performance plateaus here, future work with Hearthstone should look for similar fine-grained distinctions in technique mastery to identify expertise. For example, our methods could be applied to card placement, deck composition, or other techniques that compose expertise in Hearthstone. Once these are identified, their exercise during gameplay could then be correlated with videos of players' faces and AUs.

B. Facial Expressions and Machine Learning

Facial expressions, coded in the facial action coding system (FACS), introduced by Friedman and Ekman [16], are effective in the identification of depression [17], pain in patients who cannot communicate verbally [18], and in predicting the popularity of YouTube videos [19]. AUs combined with machine learning models, such as support vector machine and K-nearest neighbour, have been used for automatic stress detection [20] and to discriminate between expressions of "pain" and "no pain" [21]. The usefulness of machine learning in this context is confirmed by other studies using AUs combined with neural networks. For example, the convolutional neural network implemented by Liliana obtained 0.92 accuracy in discriminating 8 different emotions [22]. More recently, AUs and machine learning models were effectively used to track players' behaviour in video games. For example, in a study by Guglielmo, Peradejordi, and Klincewicz [23], a convolutional neural network was successfully implemented to detect which AUs are relevant to discriminate between baseline and decision-making processes when players made decisions. Such results seem to provide evidence that AUs, combined with machine learning models, can be effective in detecting differences in expertise levels in video games players.

The evidence provided by previous studies supports the hypothesis that AUs contain relevant information for classification purposes. For this reason, we extracted features from AUs using descriptive statistics and the matrix profile (MP). MP is an algorithm that can be used for anomaly detection and motif discovery [24] in time series data, such as electrocardiogram data [25], or for exploring similarities in DNA strings [26]. However, up to date, MP has not been used to find patterns in facial action units.

III. METHODS

A. Hearthstone

The competitive game used in this study is Hearthstone, an online, two-player collectible card game (in its "standard" competitive format). The goal of the game is to reduce the opponent's health points to zero, using a deck of 30 cards selected from a large card collection. These cards can either be monsters, which are placed on the game board, or spells, which may interact with both the game board and the players' health points directly. Also, each deck is strictly associated with a unique player "class" (e.g. warrior or mage class). Hearthstone is a turn-based game; on each turn, a player draws a card and uses a limited resource called "mana" to play cards from their hand into the game board. The winner is declared when one of the players' health points reaches zero.

B. Competition and experimental setup

We set up a best-of-three competitive tournament, which means that a player had to win two games against their opponent in order to win a match. After losing a match, players were placed in the tournament's lower bracket; losing a match in the lower bracket meant elimination from the competition. As a result, even players who did not win any games throughout the tournament played a minimum of four games. The semifinal and the final of the competition were played in a best-of-five format (three game wins required to win a match). In total, 31 matches (78 games) were recorded, after a match was excluded from the dataset due to recording failure.

In order to strengthen the competitive nature of the tournament, we offered commercial rewards to the top three ranked players. The tournament champion received a gaming mouse and a handmade Hearthstone souvenir card. Runner-up and third place winners received a Blizzard (the company that developed and published Hearthstone) gift card. The tournament was played in conquest format; each player declared three (in best-of-three matches) or four (in best-of-five matches) decks of unique classes. Each deck that won a game within a match was not allowed to be used for the remainder of that match.

C. Participants and Data Collection

Before their first match, participants signed an informed consent and filled in a short demographic questionnaire. This questionnaire included players' age and gender information, as well as the amount of hours per week spent on Hearthstone and a subjective self-assessed score of Hearthstone gaming experience on a 1 to 5 scale. Opponent players were positioned on computers facing each other (see Figure 1) and their respective monitors were set to minimum height to ensure eye contact between the players. Before starting the tournament, the players signed an informed consent stating that their faces would be recorded.
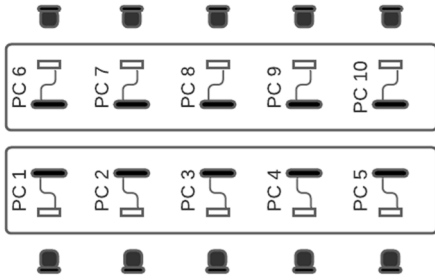
Fig. 1. Setup of the lab space where the tournament was held.

OBS software was used to capture players' webcam feeds and game screens, recorded at 30 frames per second. The webcam feed can be found at the bottom left of each recorded file, while the game screen is placed in the top right (see Figure 2).

Fig. 2. Snapshot of player webcam and game board recording.

We ensured that the overlap between the webcam and game recording did not occlude any important game or facial information. The recording was started on each computer 15 minutes prior to the tournament's start and ended after the final game was played. For each individual game, the start and end timestamp as well as the participating players and game results were manually annotated by the authors after the tournament was completed. We defined the start of each game as the moment when the player classes (decks) are announced and shown on players' screens; the end of a game is defined as the moment when the result of the game is announced and shown on the game screen.

For this study, 17 players (all male, age $M = 22.7$ years, $SD = 3.6$ years) voluntarily signed up and participated in the competition. The average experience of players was $M = 3.4$ ($SD = 1.2$) and the average hours per week (hr/w) reported was $M = 9.4$, ($SD = 6.8$). Furthermore, a unique participant ID was assigned to each player, for anonymity purposes.

Our recordings dataset consists of 156 separate gameplay videos (78 games x 2 players). The total duration of the recordings was 26.5 hours, with an average per-game duration of 10.4 minutes ($SD = 5.4$ minutes). Each participant played 9.1 games on average ($SD = 5.1$ games).

D. Definition of Expertise

In order to define the level of player expertise, a k-means algorithm was used to cluster the participants in groups. The clustering was based on the self-assessed average experience of the players (expressed with a number ranging from 1 to 5) and their self-reported average hours of playing per week (hr/w). The optimal number of clusters, 3 (see Figure 3), was found using the elbow method and the silhouette score (0.58). The quality of the clusters was subsequently evaluated by looking at the correlation coefficient between the 3 assigned clusters and the information used to create them.
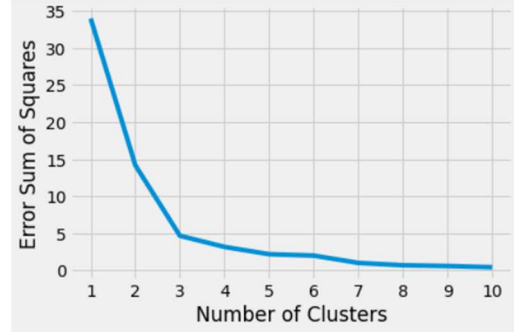

Fig. 3. Results for the optimal number of clusters according to the elbow method.

The 3 clusters correlated positively with both the average hr/w and the self-assessed average experience score, thus yielding both correlation coefficients (r) of 0.90 ($p < 0.001$). The average hr/w and the self-assessed experience had a correlation of 0.77 ($p < 0.001$). Given the presence of 3 clusters positively correlated with the variables used to create them, three levels of expertise ("Novices", "Intermediates", "Experts") were derived. 4 novice participants played the game for 1.75 hr/w ($SD = 1.30$) and their mean self-assessed experience score was 1.5 ($SD = 0.5$). 9 intermediate players played for a mean of 8.28 hr/w and they had a self-assessed mean experience of 3.78 ($SD = 0.41$). Finally, 4 expert players played a mean of 19.75 hr/w ($SD = 3.90$) and had a self-assessed experience of $M = 4.75$ ($SD = 0.43$). For clustering, we decided to use the self-assessed experience and the hr/w since the final score might be influenced by the order of the opponents. For example, 2 experts may face each other during the early phase of the tournament resulting in a poor final score of the eliminated one.

E. Action Units Extraction and Selection

AUs were extracted from player videos using OpenFace; an open-source software that allows to extract 17 AUs (1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, and 45). OpenFace provides information about the presence of an AU (0 or 1) and about its intensity (ranging from 0 to 5). For this work, the intensity alone was taken into account since it already provides information about the absence of activity (corresponding to 0) and the maximum activity (corresponding to 5). OpenFace also gives a confidence estimation for each estimation of AU intensity. Players might touch their faces during the gameplay or keep their faces turned away from the camera; this may result in lower confidence in the estimation of AU intensity. Data for each match and the players who took part in it were stored in CSV files.

Given the naturalistic conditions of data collection, we used a threshold to select videos with a tolerable number of frames with low confidence. 0.75 was our confidence baseline and

videos with more than 210 frames (corresponding to 7 seconds) lower than the mentioned threshold were not used. After this filtering process, a total of 83 videos out of the original 156 were kept for analysis purposes. Eventually, given the below the threshold confidence, videos conveying the gameplay of one participant were completely dropped while all other participants had at least one video containing their gameplay.

In order to find the most relevant AUs, the highest standard deviations in AUs' intensities were calculated for each of the 83 videos with the assumption that they are indications of significant variability of AU intensity over time. Eventually, we selected the 3 AUs visually presenting a standard deviation higher than the others across 83 videos. Such AUs were AU17, AU25, and AU26 (see Figure 4).
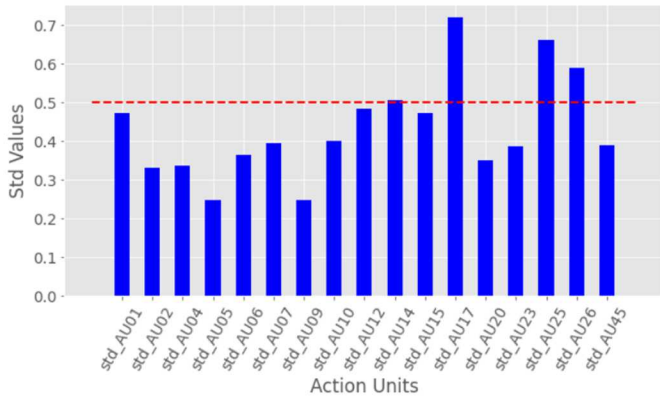

Fig. 4. Standard deviations of the 17 AUs extracted with OpenFace.

F. Feature Extraction from AUs of interest

The features were extracted from AU17, AU25, and AU26 using descriptive statistics and the MP. The MP provides a vector containing the z-normalised Euclidean distances between subsequences in a time series. More precisely, for each subsequence, MP computes the distance to its nearest neighbour (i.e. the most similar subsequence) [26]. A re-occurring subsequence would therefore result in a value of 0 being added to the vector. More generally, regular behaviour would result in low values. Therefore, intuitively, an MP vector containing, on average, high values could be indicative of irregular or unstable behaviour. For this study, we used a subsequence window length equal to 6 frames (corresponding to 200 milliseconds of recording). Such a window can fit facial micro-expressions, which could last up to 200 milliseconds [27].

From the MP, obtained with a 6 frames window, the mean, standard deviation, maximum, and median were extracted as features using the Python Stumpy library [28]. These features together with the mean, standard deviation, maximum, and median obtained from the descriptive statistics were extracted from every single video and constitute the features of the final dataset.

Afterwards, each row of the dataset was labelled according to the 3 previously defined levels of expertise. Finally, the dataset used for classification purposes consisted of 83 rows (instances) and 25 columns (8 features times the 3 AUs used plus 1 label column). The final dataset contained 20 instances

labelled as "Novices", 49 labelled as "Intermediates", and 14 labelled as "Experts".

G. Features Selection and Classifier

The classification task was performed using a Random Forest Classifier (RFC), which is an ensemble classifier based on multiple decision trees. RFC provides high explainability and good performance when classifying players' facial expressions [29]. To optimise the performance of the classifier we also ran a random search of possible hyperparameter values, the results of which can be found in Appendix A. Afterwards, class weights balancing was performed to reduce effects due to the strong imbalance present in the dataset.

To reduce overfitting, and obtain insights about the most relevant features and AUs that track expertise we used the extra trees method [30] for feature selection. Using such a method, we extract the main 6 features (Matrix median AU17, Std AU25, Std AU17, Mean AU17, Std AU26, Mean AU25) that provide higher accuracy on this classification task (see Figure 5).
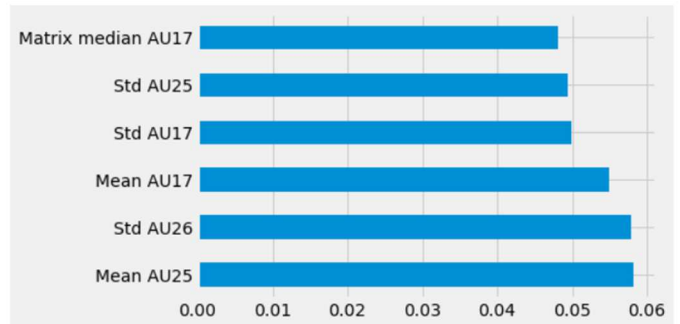

Fig. 5. Relevant features extracted using the Extra Trees method

IV. RESULTS

We trained two RFC classifiers using the 6 main extracted features, namely: an RFC with balanced class weights (RFC-BW) and an RFC combined with the SMOTE function (RFC-SM) to deal with class imbalance. RFC-BW assigns different weights to different classes so that instances of underrepresented classes are given higher weight. As a result, each class, rather than each instance, is given equal weight. RFC-SM, on the other hand, tackles the class imbalance issue by upsampling the number of instances of the underrepresented classes in the data, resulting in a balanced dataset [31]. The performance of these two classifiers was evaluated against a baseline dummy classifier (DC), which uses the "most frequent" strategy. Classification results are shown in Table 1.

TABLE I.   CLASSIFIER PERFORMANCE

	Accuracy	ROC-AUC	Precision	Recall	F1
DC	0.59 (*SD:* .02)	0.50 (*SD:* .00)	0.44 (*SD:* .02)	0.59 (*SD:* .02)	0.35 (*SD:* .02)
RFC-BW	0.75 (*SD:* .03)	0.82 (*SD:* .04)	0.74 (*SD:* .04)	0.75 (*SD:* .03)	0.72 (*SD:* .06)
RFC-SM	0.85 (*SD:* .05)	0.96 (*SD:* .02)	0.86 (*SD:* .04)	0.85 (*SD:* .05)	0.85 (*SD:* .06)

These results show a mean 0.75 accuracy on a 5-fold cross-validation and a mean 0.82 ROC-AUC score for RFC-BW and a mean accuracy and ROC-AUC score of 0.85 and 0.96, respectively, for RFC-SM. In both cases, the scores obtained by the RFC-BW and RFC-SM outperform the baseline score showing that the features extracted from AUs seem to be effective in discriminating between participants having a different level of expertise.

Given the importance of features based on descriptive statistics, we further explored the differences occurring among the 3 groups, in the mean intensity value for AU17, AU25 and, AU26 (See Table 2).

TABLE II.   MEAN INTENSITY IN AU17, AU25, AND AU26

	Novices	Intermediates	Experts
AU17	0.69 (*SD*: .19)	0.80 (*SD*: .20)	0.65 (*SD*: .08)
AU25	0.70 (*SD*: .16)	0.54 (*SD*: .14)	0.55 (*SD*: .10)
AU26	0.53 (*SD*: .21)	0.51 (*SD*: .13)	0.60 (*SD*: .10)

The results in the mean intensity show that "Intermediates" have higher values in AU17 (chin tightener) which is a tightening action unit. This value may represent higher engagement in the "Intermediates" group. Such a hypothesis may be further supported by the low values that this group had in AU25 (lips apart) and AU26 (jaws drop), which are relaxation AUs. On the contrary, experts generally have higher values in the relaxation AUs and lower values in AU17. Such values may represent a lower engagement or a more relaxed approach to the game. For what concerns the novices, they may be triggered just by specific events during the gameplay and this may be motivated by a similar value in AU17 and its direct muscular action counterbalance, AU25.

Given the importance of the "Matrix Median AU17" feature, we decided to further explore the median values in the three action units selected (see Table 3).

TABLE III.   MEAN MATRIX MEDIAN IN AU17, AU25, AND AU26

	Novices	Intermediates	Experts
AU17	0.17 (*SD*: .02)	0.15 (*SD*: .02)	0.15 (*SD*: .02)
AU25	0.17 (*SD*: .04)	0.15 (*SD*: .03)	0.14 (*SD*: .01)
AU26	0.17 (*SD*: .03)	0.15 (*SD*: .03)	0.15 (*SD*: .02)

The results in the MP provide insights into the stability of the patterns presented in the players. The results convey that "intermediates" and "experts", having lower values, showing more stable patterns during the gameplay while novices, having higher values, seem to be characterised by more instability in their facial expressions. These results seem to be connected to the ones obtained in the mean intensities where novices show similar variations in AU17 and AU25. Such results may represent higher engagement in the two more proficient groups if compared to novices.

V. DISCUSSION

This study investigated whether facial AUs can be used to discriminate different levels of expertise between players of a competitive video game. Our results suggest that AU features extracted using the MP method and descriptive statistics can indeed be used to discriminate between experts, intermediates, and novices when used to train a classifier. Furthermore, the features themselves can tell us something about the differences between these groups and how such differences are connected to expertise.

The differences between player expertise were reflected in facial expressions during gameplay and particularly in AU17, AU25, and AU26. These 3 AUs were previously identified as important to detecting stress using machine learning classifiers [20] and in another study with video games that used a convolutional neural network to track decision-making [23]. A question remains whether this behaviour is intentional or not. Players may intentionally refrain from showing emotion during gameplay (keeping a "poker face") in order to hide the quality of their hands. On the other hand, expert players may simply experience less stress, which in turn affects their facial expressions.

In support of the latter interpretation, we note that decision-making has been shown to evoke sympathetic activity [32] due to the attention it requires and the stress it elicits. This is also reflected in our data, where participants in the "Expert" group showed higher mean values in AU25 (lips apart), and AU26 (jaws drop), which represent a relaxation in the lower part of

the face that counterbalances activity in AU17 (chin tightener). We have independent evidence that AU25 and AU26 are dominant in neutral facial expressions [33]. This comports with the platitude that experts typically have more confidence in their decisions, and thus less stress. What complicates this straightforward interpretation is that "Novices" also display higher means in those AUs than "Intermediates." So what is going on?

Our results about the stability of AU intensity over time can make sense of this. On the one hand, "Experts" display stability in the MP for AU25 and AU26 (lower value of MP) and "Novices" display less stable patterns (higher MP values). This suggests that "Novices" engage in more complex, less patterned tension-relaxation behaviour, together with AU17, perhaps caused by distinct game events rather than the state of the player. "Experts" are simply relaxed and stay that way. "Novices" relax after first tensing.

What may be most interesting in our results is the existence of stability in AU patterns together with relative high mean intensity in AU17 in the "Intermediates" group. "Intermediates" do not display complex tension-relaxation behaviour (measured by MP) like "Novices", so in that respect, they are more like "Experts." On the other hand, they are not relaxed like "Experts." Instead, the dominant AU for "Intermediates" is AU17, which corresponds to a tightening of the chin, a possible indication of engagement and effort.

One way to interpret these patterns of differences in expression across groups is the presence of a 'flow' mental state [34] in "Intermediates." Flow is experienced as pleasurable and is marked by a temporal phenomenology in which time appears to speed up and a sense of effortlessness. Typically, flow is experienced when a relatively complex task is challenging enough to demand attention and care, but can at the same time be done with ease. In our case, the "Intermediates" group of players behave in a way that suggests a flow state. They are engaged and stay that way.

In conclusion, we speculate that our results may be tied to the way in which players express (or deliberately refrain from expressing) engagement, stress, and emotion through their facial expressions. However, this activity may also be connected to physiological activity caused by processes in the sympathetic nervous system [32], which is differentially active for experts, intermediates, and novices. There is independent evidence that AUs are an effective method to track sympathetic activity due to, for example, stress [20], so it is perhaps unsurprising that they are also found to be important in a highly competitive environment like a video game tournament. Further work to investigate this question is needed.

We note some limitations of our study. First, our dataset was collected in a natural setting where players were facing each other during a tournament. Such a setting allowed the players to act as they would normally and play the game almost as if they were at home. However, this may have introduced confounds due to, for example, the presence of an opponent's face or noise in the room. Furthermore, some of the players frequently covered their faces with their hands during gameplay. This resulted in lower confidence in a fair proportion of captured frames and loss of data. Furthermore, the dataset had a limited size (17 players who took part in the tournament), all of the participants were white males, some participants occurred more than others, and the 3 clusters found were strongly imbalanced. As a consequence, a bigger sample and a more diverse group of players would strengthen interpretations and confirm the effective presence of 3 clusters of players. Finally, Hearthstone is not an emotionally neutral game by design: it features bright colours, sounds, vibrant animations, etc. It is possible that some of the variations in AU intensities were caused by arousal evoked by the game's graphic-auditory features (especially for novices).

VI. Conclusion

Our results provide evidence of the effectiveness of using AUs, combined with machine learning, to identify expertise. Our methods should be used with other video games, including solo games, to confirm this and also test how much of what we found can be generalised. If AUs can be used to identify video game expertise, this may be particularly useful in game development and in esports, where it could help in training and profiling players hoping to become professionals.

References

[1] E. F. T. Buiel., G.C van de Boer-Visschedijk, L. Lebescque, I. Lucassen, "Synchro Mania - Design and evaluation of a serious game creating a mind shift in transport planning," *46th International Simulation and Gaming Association Conference, ISAGA 2015, Kyoto, Japan (18-25 juli, 2015)*, p. 1, 2015.

[2] S. Göbel, S. Vogt, and R. Konrad, "Serious Games Information Center," In *European Conference on Games Based Learning*, pp. 143-146, 2018.

[3] E. Shliakhovchuk and A. Muñoz García. "Intercultural Perspective on Impact of Video Games on Players: Insights from a Systematic Review of Recent Literature." *Educational Sciences: Theory and Practice* 20.1 (2020): pp. 40-58, 2020.

[4] P. Mavromoustakos-Blom, S. Bakkes, and P. Spronck, "Multi-modal study of the effect of time pressure in a crisis management game," in *The 15th International Conference on the Foundations of Digital Games (FDG) 2020*, 2020.

[5] E. Holl, S. Bernard, and A. Melzer, "Moral decision-making in video games: A focus group study on player perceptions," *Human Behav. and Emerg. Tech*, vol. 2, no. 3, pp. 278–287, 2020.

[6] G. Guglielmo and M. Klincewicz, "The temperature of morality: A behavioral study concerning the effect of moral decisions on facial thermal variations in video games," in *The 16th International Conference on the Foundations of Digital Games (FDG) 2021*, 2021.

[7] T. A. Ceranoglu, "Video games in psychotherapy," *Rev. Gen. Psychol.*, vol. 14, no. 2, pp. 141–146, 2010

[8] S. Whitbourne Krauss, S, Ellenberg, and K. Akimoto. "Reasons for playing casual video games and perceived benefits among adults 18 to 80 years old." *Cyberpsychology, Behavior, and Social Networking* 16.12 (2013), pp 892-897. 2013

[9] R. W. Crandall and J. G. Sidak, "Video games: Serious business for America's economy", *Entertainment Software Association*, no. 7, 2020

[10] Blizzard Entertainment. 2014. Hearthstone. Online. (2014)

[11] M. Fenton-O'Creevy, J. T. Lins, S. Vohra, D. W. Richards, G. Davies, and K. Schaaff, "Emotion regulation and trader expertise: Heart rate variability on the trading floor," *J. Neurosci. Psychol. Econ.*, vol. 5, no. 4, pp. 227–237, 2012.

[12] P. Vaidyanathan, J. Pelz, C. Alm, P. Shi, and A. Haake, "Recurrence quantification analysis reveals eye-movement behavior differences between experts and novices," in *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '14*, 2014.

[13] K. G. Schneider, R. J. Hempel, and T. R. Lynch, "That 'poker face' just might lose you the game! The impact of expressive suppression and mimicry on sensitivity to facial expressions of emotion," *Emotion*, vol. 13, no. 5, pp. 852–866, 2013.

[14] W. D. Gray and J. K. Lindstedt "Plateaus, dips, and leaps: Where to look for inventions and discoveries during skilled performance." *Cognitive Science*, 41(7), pp. 1838–1870, 2017.

[15] W.D. Gray and S. Banerjee, "Constructing Expertise: Surmounting Performance Plateaus by Tasks, by Tools, and by Techniques." *Top. Cogn. Sci.*, pp. 610-665, 2021

[16] W. Friesen and P. Ekman, "EMFACS-7: Emotional Facial Action Coding System". Unpublished manual, University of California, 1983

[17] L. I. Reed, M. A. Sayette, and J. F. Cohn, "Impact of depression on response to comedy: a dynamic facial coding analysis," *J. Abnorm. Psychol.*, vol. 116, no. 4, pp. 804–809, 2007.

[18] A. C. Lints-Martindale, T. Hadjistavropoulos, B. Barber, and S. J. Gibson, "A psychophysical investigation of the facial action coding system as an index of pain variability among older adults with and without Alzheimer's disease," *Pain Med.*, vol. 8, no. 8, pp. 678–689, 2007.

[19] P. Lewinski, "Don't look blank, happy, or sad: Patterns of facial expressions of speakers in banks' YouTube videos predict video's popularity over time," *J. Neurosci. Psychol. Econ.*, vol. 8, no. 4, pp. 241–249, 2015.

[20] G. Giannakakis, M. R. Koujan, A. Roussos, and K. Marias, "Automatic stress detection evaluating models of facial action units," in *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, pp. 728–733, 2020.

[21] K. M. Prkachin and P. E. Solomon, "The structure, reliability and validity of pain expression: evidence from patients with shoulder pain," *Pain*, vol. 139, no. 2, pp. 267–274, 2008.

[22] D.Y. Liliana, "Emotion recognition from facial expression using deep convolutional neural network". In: *Journal of physics: conference series*, vol. 1193, p. 1-5, 2019.

[23] G. Guglielmo, I. F. Peradejordi, and M. Klincewicz, "Using deep learning to detect facial markers of complex decision making," *Lecture Notes in Computer Science (Advances in Computer Games 2021)*, 2021.

[24] Y. Zhu, Z. Schall-Zimmerman, N. S. Senobari, C.C.M. Yeh, G. Funning, A. Mueen, P. Brisk, and E. J. Keogh., "Matrix profile II: Exploiting a novel algorithm and GPUs to break the one hundred million barrier for time series motifs and joins," in *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pp. 739–748, 2016.

[25] R. Wankhedkar and S. K. Jain, "Motif discovery and anomaly detection in an ECG using matrix profile," in *Advances in Intelligent Systems and Computing*, Singapore: Springer Singapore, pp. 88–95, 2021.

[26] C. C. M. Yeh, Y. Zhu, L. Ulanova, N. Begum, Y. Ding, H. A. Dau, Z.Zimmerman, D.F. Silva, A. Mueen, and E. Keogh. "Time series joins, motifs, discords and shapelets: a unifying view that exploits the matrix profile," *Data Min. Knowl. Discov.*, vol. 32, no. 1, pp. 83–123, 2018.

[27] X. B. Shen, Q. Wu, and X.-L. Fu, "Effects of the duration of expressions on the recognition of microexpressions," *J. Zhejiang Univ. Sci. B*, vol. 13, no. 3, pp. 221–230, 2012.

[28] S. Law, "STUMPY: A powerful and scalable python library for time series data mining," *J. Open Source Softw.*, vol. 4, no. 39, p. 1504, 2019.

[29] P. M. Blom, S. Bakkes, and P. Spronck, "Modeling and adjusting in-game difficulty based on facial expression analysis," *Entertain. Comput.*, vol. 31, no. 100307, pp. 1-10, 2019.

[30] J. Sharma, C. Giri, O. C. Granmo, and M. Goodwin, "Multi-layer intrusion detection system with ExtraTrees feature selection, extreme learning machine ensemble, and softmax aggregation," *EURASIP j. inf. secur.*, vol. 2019, no. 1, 2019.

[31] N. V. Chawla, K. W. Bowyer, L. O.Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, 321-357, 2002.

[32] H. Ohira., M. Matsunaga, H. Murakami, T. Osumi, S. Fukuyama, J. Shinoda, and J. Yamada, "Neural mechanisms mediating association of sympathetic activity and exploration in decision-making," *Neuroscience*, vol. 246, pp. 362–374, 2013.

[33] M. A. A. Dewan, M. Murshed, and F. Lin, "Engagement detection in online learning: a review," *Smart Learn. Environ.*, vol. 6, no. 1, 2019.

[34] M. Csikszentmihalyi, "Toward a psychology of optimal experience," in *Flow and the Foundations of Positive Psychology*, Dordrecht: Springer Netherlands, pp. 209–226, 2014.

APPENDIX A

The results of the randomised search for hyperparameters optimisation:

'n_estimators': 2000,
'min_samples_split': 2,
'min_samples_leaf': 1,
'max_features': 'sqrt'',
'max_depth': 20,
'criterion': 'gini',
'bootstrap': True