Using Deep Learning to Detect Facial Markers of Complex Decision Making*

No Author Given

No Institute Given

Abstract. In this paper, we report on an experiment with The Walking Dead (TWD), which is a narrative-driven adventure game where players have to survive in a post-apocalyptic world filled with zombies. We used OpenFace software to extract action unit (AU) intensities of facial expressions characteristic of decision-making processes and then we implemented a simple convolution neural network (CNN) to see which AUs are predictive of decision-making. More specifically, this study aims to identify the facial regions that are predictive of decision-making. Our results provide evidence that the pre-decision variations in action units 17 (chin raiser), 23 (lip tightener), and 25 (parting of lips) are predictive of decision-making processes. Furthermore, when combined, their predictive power increased up to 0.81 accuracy on the test set; we offer speculations about why it is that these particular three AUs were found to be connected to decision-making. Our results also suggest that machine learning methods in combination with video games may be used to accurately and automatically identify complex decision-making processes using AU intensity alone. Finally, our study offers a new method to test specific hypotheses about the relationships between higher-order cognitive processes and behavior, which relies on both narrative video games and easily accessible software, like OpenFace.

Keywords: Video Games · Decision-Making · Facial Expression Machine Learning

1 Introduction and Related Work

1.1 Decision-making in Video Games

Decision-making has been studied extensively in social psychology and economics with paradigms such as the prisoner dilemma, the ultimatum game, and the dictator game [3]. These paradigms are largely grounded in game theory, which assumes idealizations about rationality, utility, and often ignores the unique ways in which people make decisions in different contexts. Video games provide an alternative to game theory paradigms in the study of decision-making precisely because they provide a rich context for decisions in the form of a narrative, including in-game mechanics, and non-player characters (NPC) [25].

* xxx

NPCs are important in moving video-game narratives forward and also in framing the decisions players make while playing. This framing typically involves consequences in the narrative of the game and expressions of emotions on the part of the NPCs. In this sense, decisions made in video games may involve similar cognitive and affective mechanisms that are at work during decision-making in real life, where meaningful decisions happen in a rich context with consequences that affect other people. The important difference, of course, is that consequences in video games affect the game world and NPCs, while decisions out-of-game affect the real world and real people. This difference, while a limitation, is also what makes video games especially promising in the study of complex decision-making, in that they provide a safe environment to experience new forms of agency without worries about the consequences [17]. This is also why video games are particularly useful in education, where mistakes do not have the sort of consequences that analogous real-world decisions would [1]. Considering the aforementioned advantages, we decided to use TWD for our study, since its rich narrative presents scenarios that, to a certain extent, can be compared to the ones presented in real life.

1.2 Facial Expressions and Machine Learning

It is an old idea that the face is the window to the soul. Facial expressions have been systematically studied and linked to what has been understood to be a set of basic emotions at least since Darwin [4], but have recently also been found to vary depending on the cultural context [13]. Emotions typically evoke a sympathetic system response. Being exposed to a stimulus, including making a decision, can also sometimes elicit a sympathetic response, which in turn changes heart rate, skin conductance, and facial temperature just as is the case with emotions [18, 8]. Some of these responses, just as is the case with emotions, are accompanied by facial expressions. That said, not as much attention has been paid to the potential links between higher-order processes, such as decision-making, and facial expressions [9].

The range of human facial expressions has been coded in the facial action coding system (FACS) developed by Paul Ekman and colleagues [7]. FACS is now used to measure pain in patients unable to communicate it verbally [15], and even in identifying depression [24]. Facial expressions are also widely used in affective computing, understood to be a research program that aims to use devices and systems to detect emotional states, processes and, responses [21].

Given all this, it is perhaps unsurprising that action units have been used as input for machine learning models. For example, a relatively simple support vector machine (SVM) reached 0.75 accuracy when using AUs as input for automatic stress detection [10]. SVM and k-nearest neighbors (KNN) algorithms can classify expressions of "pain" vs "no pain" and even their intensity [16, 22]. More recently, CNNs have been used to estimate the presence of pain and its intensity [26]. In that last pain classification study, deep learning models showed a higher accuracy if compared to other machine learning models; where the KNN algorithm implemented by Prkachin and Solomon had an accuracy score of 0.86

while the CNN implemented by Semwal and Londhe reached an accuracy score of 0.93 [22, 26]. Deep learning models, and more specifically CNNs, have been successfully used to detect emotions scoring an average beyond 0.92 on 8 classes of emotions [14]. Given all this, it might be not surprising that AUs, combined with other input such as audio, have been used for the detection of depression [27] and to identify micro facial expressions [5]. There is a number of avenues to enrich the input to CNNs, e.g., taking into account the effects of head and face rotation and the spatio-temporal dynamics occurring between AUs [19]. This last approach will probably result in an increment in accuracy for classification all tasks involving AUs and facial expressions, just as the inclusion of audio would. In sum, deep learning models, and in particular CNNs, are effective in detecting patterns in AUs to perform classification in different tasks. For this reason, we used them with AUs obtained during decision-making while playing TWD.

2 Methods

2.1 Data collection and Participants

All participants were asked to play the first episode of TWD while seated in a room with another participant that did the same. All participants signed informed consent forms and were informed about the nature of the study and their rights regarding personal data storage and processing. Participants' gameplay was recorded using screen capture software and their posture and face were recorded using an Open Broadcaster Software (OBS) and an HD Webcam (Logitech C922 Pro Stream); the two recordings were synchronized using a hotkey. The two participants taking part in any session of a recording always used two different computers, while the recordings were started and monitored using another two control computers.

A total of 78 participants took part in the experiment; 51 males with a mean age of 20.11 (SD = 2.63) and 27 females with a mean age of 19.4 (SD = 2.02). 12 participants were excluded since they played TWD before and knew the narrative and decisions presented in the game. One participant decided to quit the experiment because they found the content too disturbing. One participant had to leave due to personal issues and another 5 participants were excluded since they failed to perform the task as instructed. The final lot before data analysis had 52 participants. Game-play recordings were prepared with Sony Vegas Software by being cut into 10 seconds intervals around each decision made in the game. Each participant made 8 decisions during the experimental session, so a total of 80 seconds of video was eventually used to extract the information about AU intensity with OpenFace for each of the 52 participants.

2.2 Decisions selection

All of the decisions we used were particularly important to the narrative of the game and relied on the participant taking into account the context in which they were presented by NPCs (e.g., Figure 1).

Fig. 1. An example of decision presented in the video game

For example, in one of the decisions participants had to decide whether to save a young boy or an older man from zombies. The consequences of these decisions would play out in the narrative of the game and affect NPCs. In general, each decision provided the player with 4 potential alternatives: however, regardless of the decision made by the player, the video game followed a pre-defined course of action, so each participant ultimately ended up playing the same section of the game with the same decisions. Importantly, the 8 decisions that were selected for analysis had more than 30 seconds between them; this in order to avoid potential overlapping effects of between decisions.

3 Data preparation and modelling

3.1 Data extraction

First, we identified the moment a decision was made by referencing the recording of game-play and the recording of the participant. We then used that moment as a representation of the end of the decision-making process and took 5 seconds of the video from before and 5 seconds after. For each of the 52 participants, eight 10 seconds videos were thus obtained, representing the 8 selected decisions made during TWD. The videos were recorded at 30 frames per second leading to a total of 300 frames, where the 150th frame represented the moment in which the decision was made. During this stage, we had to exclude a further 6 participants due to corrupted data or missing frames. Ultimately, 46 participants, with 8 videos each were used to extract AUs.

The AUs used for this work were extracted using OpenFace [2]. OpenFace extracts 17 action units (1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, and

45) that can be described either in terms of their presence (0 or 1) or in terms of their intensity (from 0 to 5). In our work, we extracted just the intensity information since, by itself, it can provide a number ranging from 0, the absence of the activity in the AUs, to 5, conveying the maximum intensity in the AUs. The data obtained were stored in CSV files.

3.2 Data preprocessing

Since our focus was to detect facial AUs related to decision-making processes, we analyzed the 150 frames prior to the actual act of deciding corresponding to the click. This is because we intended to focus on the processes prior to the decision itself. So, we compared the frames belonging to the baseline (0-74) to the frames belonging to the decision-making process (75-149). The 150 frames before making the decisions were equally split considering that the participants read the questions between frame 20 and 75 leaving frame 75-149 as the frames potentially reflecting the decision-making process. This particular split is motivated by the length of the sentences presented in the video game. Considering that the average speed to read 300 words per minutes [23] and the eight sentences introducing the scenario had a number of words ranging from 4 to 10. Reading a 10 words sentence would require around 2 seconds, approximately corresponding to the 55 frames, for this reason, we considered the frames 20 to 75 as reading window that might have slightly varied according to the sentence length and the individual reader speed.

A total of 736 samples were used as input for the CNN in this work, this number is obtained by considering that the 46 participants had 8 recordings labelled as "baseline" and 8 recordings labelled as "decision-making process". Each of the 736 data point represented a row in the dataset. Being a balanced dataset, 368 rows were labeled as “baseline” and 368 as “decision-making process”. In sum, each of the 17 AUs has its own corresponding file with the same 736 x 75 structure, where 736 is the number of total data points and 75 is the number of frames considered (representing the columns of the dataset), with half of the rows labeled "baseline" and half labeled "decision-making process".

3.3 Model Description

In order to test the predictive value of individual AUs for identifying the decision-making process, we created a 1D CNN, expecting it to serve as a baseline [28]. We decided to use CNNs since they have been successfully used with AUs for prediction and classification tasks [11], as mentioned in the introduction. Furthermore, CNNs were used to perform classification task using a dataset with fewer than 1000 data points, similarly to our own dataset [20]. In the end, our model had 2 convolutional layers, 2 max-pooling layers, and 4 fully connected layers; the structure of the model and its specification is illustrated in Figure 2.

```
Model: "sequential_12"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv1d_24 (Conv1D)           (None, 73, 100)           1000
_________________________________________________________________
dropout_48 (Dropout)         (None, 73, 100)           0
_________________________________________________________________
max_pooling1d_24 (MaxPooling (None, 24, 100)           0
_________________________________________________________________
conv1d_25 (Conv1D)           (None, 22, 32)            9632
_________________________________________________________________
max_pooling1d_25 (MaxPooling (None, 7, 32)             0
_________________________________________________________________
dropout_49 (Dropout)         (None, 7, 32)             0
_________________________________________________________________
flatten_12 (Flatten)         (None, 224)               0
_________________________________________________________________
dense_48 (Dense)             (None, 128)               28800
_________________________________________________________________
dropout_50 (Dropout)         (None, 128)               0
_________________________________________________________________
dense_49 (Dense)             (None, 180)               23220
_________________________________________________________________
dropout_51 (Dropout)         (None, 180)               0
_________________________________________________________________
dense_50 (Dense)             (None, 30)                5430
_________________________________________________________________
dense_51 (Dense)             (None, 2)                 62
=================================================================
Total params: 68,144
Trainable params: 68,144
Non-trainable params: 0
```

Fig. 2. Model specifications

The activation function chosen was Rectifier Linear Unit (ReLU) as suggested in Gudi et al. [11]. The optimizer chosen for our CNN was the Nesterov-accelerated Adaptive Moment Estimation (Nadam). In past studies, Nadam outperformed other optimizers in models aiming to classify different typological data. More specifically, using Nadam resulted in lower convergence time required, lower loss score, and higher accuracy [6]. To minimize overfitting, dropouts were added between the convolutional, the max-pooling, and the fully connected layers to reduce overfitting that might affect results on a small dataset like the one we used [2]. 20 percent of data was used for test purposes, while 10 percent of data was used for validation and to keep track of potential overfitting. The model was trained using a 10 samples mini-batches and 20 epochs. The model used for this study was implemented using Python and more specifically Numpy, Pandas, Scikit-learn, and Keras.

4 Results

Our results suggest that three AUs might be predictive of decision-making processes. As shown in the Table 1 these units all scored above 0.65 (threshold used to select significant AUs).

Table 1. Significant differences in action units across baseline and decisions

	Training		Validation		Test	
AU	**Accuracy**	**Loss**	**Accuracy**	**Loss**	**Accuracy**	**Loss**
17	0.6954	0.5842	0.7627	0.5338	0.7297	0.5279
23	0.6948	0.5854	0.6949	0.5576	0.6824	0.5397
25	0.7240	0.5277	0.6780	0.6214	0.7027	0.5735

These three action units seem useful to discriminate between decision-making processes and baseline even in a simple 1D CNN: AU17 (chin raiser), AU23 (lip tightener), and AU25 (lips part). Other AUs did not reach significance with our model, so were excluded in the reported results, but a more sophisticated model may well find other AUs in the same area of the face predictive.

To further explore the predictive power of these 3 action units, we combined them in a multidimensional input (75,3) to the same network. The final model scored 0.81 on accuracy and 0.50 on loss score on the test set (Table 2).

Table 2. Combined significant AUs across baseline and decisions

Training		Validation		Test	
Accuracy	**Loss**	**Accuracy**	**Loss**	**Accuracy**	**Loss**
0.8144	0.4077	0.7966	0.4188	0.8108	0.4978

For a model this simple, our results suggest that AUs can indeed be used to identify decision-making processes without much modelling. To make it clear that this is not an anomaly, we include the convolution over epochs in Figure 3 below.

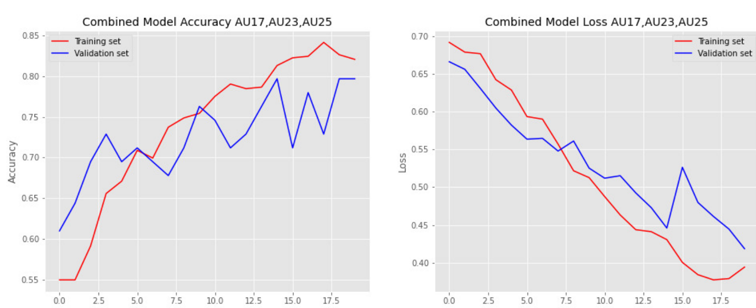

Fig. 3. Convolution of combined AUs 17, 23, and 25

Our results seem to be corroborated by other studies. A study using AUs and SVM found that AU17, AU23, and AU25 intensities are modulated by stress conditions [10]. So, it might be the case that decision-making processes, which sometimes involve sympathetic activity, trigger a response similar to stressful stimuli [27]. Further corroborating our results, a distinct experimental study with these the TWD identified significant variations in temperature of the chin area approximately 20 seconds after decisions that had a moral dimension [12]. In other words, the same area that involves AU17, and AU25, shows a significant variation in temperature after particularly complex and possibly stressful decisions. The AU17 involves a tightening of the muscle mentalis, which is located below the lower lip, while AU25 involves the same muscle relaxation, so one explanation for distinctive variations of temperature in specific parts of the face is the effect of increased blood flow to those areas, which is caused by the engagement of muscles in facial regions [8], as identified in the present experiment with CNNs.

5 Discussion

Given their functional and anatomic connection, the predictive value of AU17 and AU25 might be a result of the tightening of the chin at the beginning of the decision-making process. AU23, on the other hand, is a functional counterbalance to AU25 and logically connected to movements of the mouth and chin. Interestingly, AU26 (jaw drop), is functionally related to AU25, and while not included in the final model due to just-below 0.65 of accuracy, it seems to be engaged during the decision-making process as well. As a consequence, it might be the case that AU17 and AU23 are characteristic of the initial part of the decision-making process while AU25 and AU26 might be peculiar to the end part of the decision-making process when the facial expression returns to baseline. AU17 seems to be counterbalanced by AU26 (jaw drop) while AU23 (lip tightener) might be counterbalanced by AU25 (lips part). In general, we can conclude that there is a tightening of the lip-chin area prior to the decision process and then a relaxation of the chin area after the decision processes.

Further studies might clarify the relationship between sympathetic activity and changes in intensity in specific facial regions. The evidence provided by this study suggests that the cognitive processes of decision-making are in some way connected to muscular activity in the chin area. Such muscular activity might in turn lead to changes of temperature in this area. These effects might be accentuated by a moral dimensions characterizing certain kinds of decisions. Moral decisions, might be more stressful than non-moral ones eliciting a peculiar change in temperature. Future studies might further investigate the effects of decision-making, and moral decision-making, processes using other information extractable from AUs (graphs, videos, frames), and the effect of different stress intensity on the AUs similarly to what has been done for pain detection tasks.

References

1. Anastasiadis, T., Lampropoulos, G., Siakas, K.: Digital game-based learning and serious games in education. International Journal of Advances in Scientific Research and Engineering (ijasre) **4**(12), 139–144 (2018)
2. Baltrušaitis, T., Robinson, P., Morency, L.P.: Openface: an open source facial behavior analysis toolkit. In: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 1–10. IEEE (2016)
3. Conitzer, V., Sinnott-Armstrong, W., Borg, J.S., Deng, Y., Kramer, M.: Moral decision making frameworks for artificial intelligence. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 31, pp. 4831–4835 (2017)
4. Darwin, C.: The Expression of the Emotions in Man and Animals. University of Chicago Press (2015). https://doi.org/doi:10.7208/9780226220802, https://doi.org/10.7208/9780226220802
5. Davison, A.K., Merghani, W., Yap, M.H.: Objective classes for micro-facial expression recognition. Journal of imaging **4**(10), 119 (2018)
6. Dogo, E.M., Afolabi, O.J., Nwulu, N.I., Twala, B., Aigbavboa, C.O.: A comparative analysis of gradient descent-based optimization algorithms on convolutional neural networks. In: 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS). pp. 92–99. IEEE (2018)
7. Ekman, P., Friesen, W.V.: Measuring facial movement. Environmental psychology and nonverbal behavior **1**(1), 56–75 (1976)
8. Engert, V., Merla, A., Grant, J.A., Cardone, D., Tusche, A., Singer, T.: Exploring the use of thermal infrared imaging in human stress research. PloS one **9**(3), e90782 (2014). https://doi.org/doi.org/10.1371/journal.pone.0090782[
9. Furl, N., Gallagher, S., Averbeck, B.B.: A selective emotional decision-making bias elicited by facial expressions. PLoS One **7**(3), e33461 (2012)
10. Giannakakis, G., Koujan, M.R., Roussos, A., Marias, K.: 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. IEEE (2020)
11. Gudi, A., Tasli, H.E., Den Uyl, T.M., Maroulis, A.: Deep learning based facs action unit occurrence and intensity estimation. In: 2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG). vol. 6, pp. 1–5. IEEE (2015)
12. Guglielmo, G., Klincewicz, M.: 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 (FDG'21). ACM, Motreal (2021). https://doi.org/10.1145/3472538.3472582
13. Jack, R.E., Garrod, O.G.B., Yu, H., Caldara, R., Schyns, P.G.: Facial expressions of emotion are not culturally universal. Proceedings of the National Academy of Sciences **109**(19), 7241–7244 (2012)
14. Liliana, D.Y.: Emotion recognition from facial expression using deep convolutional neural network. In: Journal of physics: conference series. vol. 1193, p. 12004. IOP Publishing (2019)
15. Lints-Martindale, A.C., Hadjistavropoulos, T., Barber, B., Gibson, S.J.: 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 Medicine **8**(8), 678–689 (2007)

16. Lucey, P., Cohn, J.F., Prkachin, K.M., Solomon, P.E., Chew, S., Matthews, I.: Painful monitoring: Automatic pain monitoring using the UNBC-McMaster shoulder pain expression archive database. Image and Vision Computing **30**(3), 197–205 (2012)
17. Nguyen, C.T.: Games and the art of agency. Philosophical Review **128**(4), 423–462 (2019)
18. Ohira, H., Matsunaga, M., Murakami, H., Osumi, T., Fukuyama, S., Shinoda, J., Yamada, J.: Neural mechanisms mediating association of sympathetic activity and exploration in decision-making. Neuroscience **246**, 362–374 (2013)
19. Onal Ertugrul, I., Yang, L., Jeni, L.A., Cohn, J.F.: D-PAttNet: Dynamic patch-attentive deep network for action unit detection. Frontiers in computer science **1**, 11 (2019)
20. Pasupa, K., Sunhem, W.: A comparison between shallow and deep architecture classifiers on small dataset. 2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE) pp. 1–6 (2016)
21. Picard, R.W., Picard, R.: Affective computing, vol. 252. MIT press Cambridge (1997)
22. Prkachin, K.M., Solomon, P.E.: The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain. Pain **139**(2), 267–274 (2008)
23. Rayner, K., Schotter, E.R., Masson, M.E.J., Potter, M.C., Treiman, R.: So Much to Read, So Little Time: How Do We Read, and Can Speed Reading Help? Psychological Science in the Public Interest **17**(1), 4–34 (2016). https://doi.org/10.1177/1529100615623267, https://doi.org/10.1177/1529100615623267
24. Reed, L.I., Sayette, M.A., Cohn, J.F.: Impact of depression on response to comedy: A dynamic facial coding analysis. Journal of abnormal psychology **116**(4), 804 (2007)
25. Ryan, M., Formosa, P., Howarth, S., Staines, D.: Measuring morality in videogames research. Ethics and Information Technology **22**, 55–68 (2020). https://doi.org/10.1007/s10676-019-09515-0
26. Semwal, A., Londhe, N.D.: Automated Pain Severity Detection Using Convolutional Neural Network. In: 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS). pp. 66–70. IEEE (2018)
27. Song, S., Shen, L., Valstar, M.: Human behaviour-based automatic depression analysis using hand-crafted statistics and deep learned spectral features. In: 2018 13th IEEE International Conference on Automatic Face \& Gesture Recognition (FG 2018). pp. 158–165. IEEE (2018)
28. Tang, W., Long, G., Liu, L., Zhou, T., Jiang, J., Blumenstein, M.: Rethinking 1d-cnn for time series classification: A stronger baseline. arXiv preprint arXiv:2002.10061 (2020)