Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 192 (2021) 2881–2892

Procedia Computer Science

www.elsevier.com/locate/procedia

25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

Face emotional responses correlate with chaotic dynamics of eye movements.

Albert Sledzianowski[a], Krzysztof Urbanowicz[b], Wojciech Glac[c], Renata Slota, Maria Wojtowicz, Monika Nowak[d], Andrzej Przybyszewski[a]

[a]*The Faculty of Information Technology, Polish-Japanese Academy of Information Technology, Warsaw, 02-008, Poland*
[b]*The Faculty of Physics, Warsaw University of Technology, Warsaw, 00-661, Poland*
[c]*The Faculty of Biology, Department of Animal and Human Physiology, University of Gdańsk, Gdansk, 80-309, Poland.*
[d]*The Student Scientific Association "Homunculus", University of Gdansk, Gdansk, 80-309, Poland.*

Abstract

Background and Objective. Paul Ekman has demonstrated that we can estimate emotions on the basis of individual facial muscles movements. However, there is a lack of information in the literature about correlation between eye movements and facial emotions. Our objective was to find out whether emotion could be also visible in the dynamical properties of the eye movements.

Methods. We have performed our experiment in two sessions related to different video presentations: 1) a reference session with all emotions presentation, and 2) the main session showing presentation of some specific emotions. During both sessions, we have video recordings of the subjects' face expressions (FE) and eye movements (EM). We have calculated parameters of FE and EM dynamical changes and content of the noise. On this basis, by using time changes of EM, we have predicted 6 different face emotions.

Results. We have recorded face expressions of 49 subjects who had the strongest responses to Happiness and Contempt facial emotions. We found statistically significant differences in parameters' values describing FE and EM between the reference and the main sessions trials. Parameters connected to the Chaos showed highly positive correlations with Happiness, while both the Linear and the Noise components were mostly negatively correlated with this emotion. We achieved highest results with help of the K-Nearest Neighbors algorithm obtaining: Accuracy of 0.89 (+/- 0.01) with the ROC-AUC score of 0.88 (F1 = 0.89), Precision = 0.85, Sensitivity / Recall = 0.93, Specificity = 0.82.

Conclusions. We have observed that when the intensity of the Happiness increases, the eye movements become more chaotic and behave less noisy. We see possibilities for use of presented methods as a support for predictions in face expressions, when the lower part of the face is partially hidden, e.g. by a protective mask worn during the COVID epidemic. This could also be a method of confirming the authenticity of the Happiness mimicry, because it would be difficult to voluntarily correlate eye movements with certain levels of the Chaos.

Keywords: machine-learning; eye moves; empathy; face emotions; happiness; noise; chaos; nonlinear dynamical systems

1877-0509 © 2021 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of KES International.
10.1016/j.procs.2021.09.059

1. Introduction

P. Ekman and W. Friesen developed The Face Action Coding System (FACS) that became a standard to objectively describe facial expressions, for the first time presented in 1978 in "Facial Action Coding System: A Technique for the Measurement of Facial Movement". The FACS interprets facial expressions as changes in the groups of Action Units (AUs), which represent the activity of specific groups of facial muscles. The activity of specific AU groups corresponds with the facial mimicry shown during specific emotions. Those relations are used in many implementations to estimate facial emotions, also by analytical libraries like the Open Face (OF). The OF library allows one to detect the face in a video, set and track landmarks describing the shape of face elements and evaluate the presence and the intensity of specific AUs [1]. The OF also provides information about eye direction expressed as the gaze vectors [1]. The webcam can give simple access to the video-oculography (VOG) and eye tracking are used in many applications. Analysis of eye movement is the subject of research in neurodegenerative diseases such as Parkinson's or Aizhaimer's disease, where damage to pathways and brain regions affects eye movement performance [2], [3], [4]. It is also used in investigation of the influence of specific actions on eye fatigue or even in hazard detection, such as fatigue while driving a car [5]. In our research we used a time series of eye coordinates to calculate noise or chaos and intensity of AUs to determine the emotional state of the subject at the given moment. In general there are many evidences regarding non-linear behaviour of human body, which features can be determined by mathematical (deterministic) calculations, it is also very sensitive to initial conditions, moreover, elements of human behavior never repeat exactly and show the effects of any disturbances, just as with elements behaving according to the chaos theory [6]. Analysis of nonlinear dynamics are used in i.e. estimations of heart rate variability, where low-dimensional chaos was found as associated with health [6]. In this research we decided to use non-linear methods to test eye movement to investigate if we can describe its dynamics by non-linear parameters during specific facial emotions. In the literature there are already papers trying to apply chaos-like methods on eye tracking data. C. Aştefănoaei et al. [7], tried to estimate correlation dimensions (the Lyapunov exponent and the Hurst exponent) based on the eye movements. These methods are well known in chaos theory as the base methodologies for chaotic motion estimations. Authors claim that data they have analyzed have properties of chaos. Nevertheless, there are several problems with using the embedding dimension. It might be inefficient for two reasons: at first, the correlation dimension does not reach the saturation point when the embedding dimension is increasing. Second one, because the embedding dimension should be larger than $3D+1$, where D is the dimension of the attractor, here the topology dimension. The topology dimension is smaller than correlation dimension, but in most cases very similar. S. S. M. Chanijani analyzed 3 types of students: novice in knowledge, medium and experts [8]. His paper claims that information entropy of eye movements is higher for novice people. This could be very intuitive, thus novice students are rather uncertain of answers to questions (in more anxiety) producing larger entropy (entropy measures level of uncertainty). Harezlak et al. [9] writes about nonlinear signal analysis methods such as entropy, correlation dimension, correlation integral, while using Low Pass Filter (LPF) which is a linear method. If we denoise the signal with the Fourier transform, the nonlinear dependencies will be eliminated, thus a question should be asked about the reliability of the simultaneous use of the LPF and Nonlinear Methods.

However, in the literature we didn't find the use of chaosity calculations to examine the eye movements while feeling a specific emotion and this is what we wanted to test during our experiment. We decided to trigger emotions using recordings of other people's faces, so one might say we examined the eye movements while respondents felt the empathy expressed by the feedback on face emotion on the presented video stimuli. Empathy or making inferences about the mental states of others is a key requirement for communicative interactions. Those interactions are essentially related to the social brain exchanging feelings and beliefs from one mind to another [10]. The cognitive processes are also related to the intentions, firstly was called "mentalazing" by Kampe et al. (2003) [11]. Over the years functional Magnetic Resonance Imaging (fMRI) experimenters like Morelli and Lieberman found evidence for a core set of neural regions that support an empathetic state [12]. These researches suggest that regions related to mentalizing, dorsomedial prefrontal cortex, medial prefrontal cortex and temporo-parietal junction, are central to the experience of the empathy, potentially helping us in understanding of varied emotional images and everyday experiences of other people [12]. Experiments with the fMRI also have shown that the following brain regions: the paracingulate cortex and the temporal poles, bilaterally are activated by the eye contact or similar actions triggering the "ostensive" signal [11]. In our experiment, we have assumed that this and other similar processes could influence the EM by focusing observation of the subject on particular elements of the model's face presented in the video stimuli. We have taken into account that different facial emotions are assembled from different numbers of AU. Thus, the method used to calculate measurement and dynamical noise was insensitive to this aspect.

2. Experimental methods

We have recorded subjects' faces expressions consecutively watching two types of video stimulus in two sessions. In the first "reference" session the video has showed a series of different facial emotions composed of 7 video-clips. The sequence of the emotions was: Disgust, Happiness, Surprise, Disgust, Surprise, Fear. It was a reference test about how subjects respond to the same set of stimuli (independent variable). The videos came from the Human Identification Project database, containing classified recordings of face emotions presented by models [13]. During the second main session, we randomly presented one of 3 videos:

- Presenting a neutral image (empty chair / no person) and a soundtrack with audience laughter stimulating Happiness (only an audio stimulus).
- Presenting a recording of a crying woman's face stimulating the Sadness and the silence (only an image stimulus).
- Presenting a recording of a crying woman's face and an audio track with the audience laughing (a contradictory emotional message that can be perceived in a subjective way).

In these three video presentations, we have tried to recreate the situation found in the popular Reaction-Videos (RV) presenting reactions of the person watching a specific video material [14]. We considered this type of video as a good stimulus that can evoke emotions in the viewer. Therefore, in our experiments we decided to use a similar recording scheme. During recordings the camera was positioned behind the monitor, to create a similar plan as on the RV in which you can see that a person is watching something, but the actual video only shows his reactions as a viewer. In our experiment, we also wanted to check how different schemas of audio and video stimuli affects the subjects. The RV usually mixes the audio track from the original video and from the actual video showing the reaction. We were interested in the influence of such audio-dissonance on the emotional reactions, this is why we included a video containing incoherent stimuli - a crying woman with the background sound of the laughing audience.

In the experiment we have used the IBM ThinkVision 6636 15" square LCD monitor (1024 x 768) and the Logitech C922 Pro Stream camera in frequency of 60Hz and resolution of 720p. All recordings were shot with white background, the same camera frame, the same lighting conditions. Camera based VOG can provide similar accuracy in eye tracking to that of a head-mounted IR-based VOG systems [5]. We decided to use a video camera instead of an eye-tracker, because this way we could simultaneously estimate face emotions. Implementation of this type of method usually consists of steps leading to face recognition followed by locating facial landmarks which can be used for both, to locate eyes and to estimate the AUs intensity.

Before the recordings, subjects were placed at a distance of 80 cm from the monitor in the center of the camera frame. Each participant was asked to watch a video material and behave naturally. When the experiment started, the subject watched the video sequence showing different facial emotions. The reference recording lasted 35 seconds (2100 frames) and that length was determined experimentally as the optimal time from the view point of maintaining focus on different stimuli by the subjects. The main video was a little longer and lasted one minute (3600 frames). After the video session, each of the respondents completed a questionnaire in which we asked about felt emotions and observed them in the video.

In the next part of the experiment, recordings of subjects watching the stimuli videos were classified by algorithms predicting the occurrence and the intensity of facial emotions and also parameters of eye movements expressed in world coordinates: x, y, z.[1] The OF classifies eye, face, head and pose features extracted from a video or an image. Figure 1 shows a video-capture of a subject (right side) watching the video-stimuli (video-capture on the left), with face and eye tracking landmarks applied by the OF.

Our interest was to measure linear, chaos and noise correlation between the eye movement and the facial emotions. The resulting data of the OF classification contained different data series, but for further analysis we chose parameters "AUx_c" and "AUx_r" describing the presence and the intensity of specific AUs and parameters describing the eye

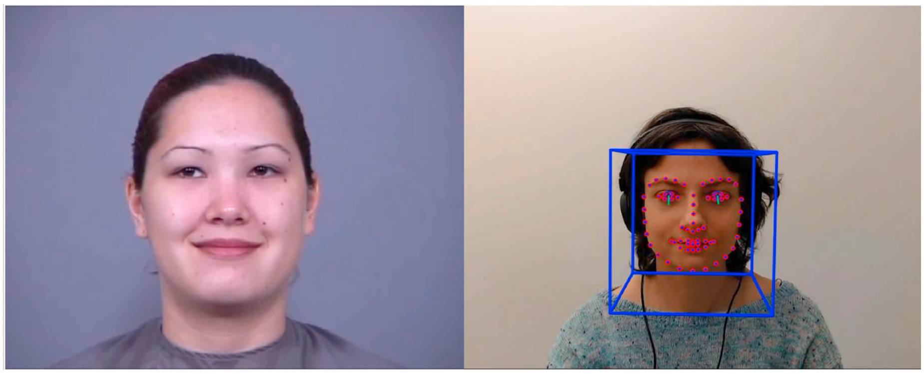


Fig. 1: The video capture of the stimuli (left) and a subject (right side).

movements. The OF describes gaze directions in world coordinates (x,y,z) separately for both eyes, where the eye 0 is the leftmost eye in the image [1]. In order to estimate gaze vectors in camera image, the OF uses a Constrained Local Neural Field (CLNF) landmark detector to detect eyelids, iris and pupils in the camera image. The OF method computes intersection between the camera ray and the eye-ball sphere locating pupils in 3D camera coordinates and then obtaining the gaze coordinates by computing vectors from the 3D eyeball center to the location of the pupil [1]. We did not categorize types of performed eye movements, but it could be assumed that while observing a series of changing faces and facial expressions, subjects performed sequences of fixations and saccades of various lengths. Saccades can be considered as one-directional eye movement with angular velocity exceeding a certain threshold value. In another approach, saccades are characterized by a sufficiently high level of eye angle dispersion, because during fixations the eye moves minimally and the eye angle is relatively constant, in contrast to saccades [15].

In the next step the OF outputs were analyzed by two algorithms. First one computed the measured, additive, static and dynamic noise, also separately for the left and right eye and for each of the gaze vectors. The second one computed emotion intensities from presence and intensity parameters of the AUs. Both methods are described in the next section.

3. Computational methods

Methods derived from chaos theory are used in the case when data are non-regular or when linear methods fail to work. From the first analysis using linear (simple) methods we concluded that the time series of eye movements were visually non-linear, non-periodic and did not create simple patterns and in order to distinguish noise from chaos we needed to apply nonlinear methods. During measurements we described face emotions, especially eye movements as a type of non-stationary systems. We are aware of this fact using the below methodology and selected 3 measurables representing types of system behaviour by the entropy criteria:

- Linear (entropy equal 0).
- Chaos (entropy larger than 0 but finite).
- Noise (entropy infinite).

3.1. Calculations of Chaos

Noise level estimation is based on methodology described in the references [16, 17, 18]. Linear level is computed with a very straightforward base on the main property - the periodicity. Chaos is just a complement to those above (See eq. (1)):

$$Chaos = 100\% - Linear - Noise \tag{1}$$

3.2. Calculations of Linearity

The general idea in Linear level calculations is to subtract the average value for any period (See eqs. (2) and (3)) and compare variance of delinearized signal and the original signal (See eq. (4)). This is just a percent of linearity. The properly constructed equations reads (all codes can be sent to interested upon request):

$$ave_p = \frac{p}{N^2} \sum_{i}^{N-p+1} \sum_{j}^{\frac{N}{p}} x_{(i+j\cdot p)\%N} \tag{2}$$

$$\widetilde{x_i} = x_i - \frac{1}{N} \sum_{p=1}^{N} ave_p \; for\; i = (1, 2, .., N) \tag{3}$$

$$Linear = \frac{Var(\widetilde{x})}{Var(x)} \cdot 100\% \tag{4}$$

3.3. Calculations of Noise

Results presents Noise level as the percentage of stochastic noise to the whole signal eqs. (5) and (6). This reads:

$$Noise = NTS^2 \cdot 100\% \tag{5}$$

$$NTS = \frac{\sigma}{\sqrt{Var\,(x)}}, \tag{6}$$

where σ stands for standard deviation, $Var(x)$ - variance of x, $\{x_i\}$ for $i = (1, 2, ..., N)$ - our raw time series. Having chaotic system one can select 3 types of stochastic noise that can be mixed together with pure deterministic one:

1. Additive noise often called measurement one.
2. Dynamical noise.
3. Parametric noise.

Additive noise does not affect the trajectory of the system and is added afterward to the system trajectory measurement. Dynamical noise has an impact on the system itself and its trajectory, changing it in each step. As the name suggests, dynamical noise is incorporated into the dynamics of the system. Parametric noise also affects the dynamics of the system, changing system parameters. This kind of noise is making the system non-stationary, thus each step of stationarity is disturbed by altering parameters. Parametric noise can be easily transformed to the kind of dynamical noise by set of equations (See eqs. (7) to (9)) (Taylor expansion), where μ is a parametric noise, f - nonlinear function, a - set of parameters:

$$x_{i+1} = f\,(a,\; x_i,\; x_{i-1},\; \ldots,\; x_{i-\tau d}) \tag{7}$$

$$\bar{x}_{i+1} = f\,(a + \mu,\; \bar{x}_i, \bar{x}_{i-1}, \ldots, \bar{x}_{i-\tau d}) \tag{8}$$

$$\bar{x}_{i+1} = f\,(a,\; x_i,\; x_{i-1},\; \ldots,\; x_{i-\tau d}) + \frac{\partial f}{\partial a}\mu + \frac{1}{2}\frac{\partial^2 f}{\partial a^2}\mu^2 + \ldots \tag{9}$$

The most difficult and complex part are the calculations of noise level. We use an approach which counts all types of noise together as the sum [16]. Our methodology cannot distinguish separately levels of each 3 types of noise mentioned above. Noise can only be estimated as oneness. On the other hand the methodology estimates the whole range of noise levels, from 0 to 100%. To this end, we use embedding in the Taken's space. This embedding method can transform a single time series (column of measurements consecutive in time) to multidimensional time vectors. Taken's theory proves that such an embedding fully reconstructs the topology of any system with embedding parameters properly set and with enough data. In the Takens space the state of the system is represented by a vector eq. (10):

$$X_i = \{x_i,\; x_{i-1}, \ldots, x_{i-\tau m}\} \tag{10}$$

where τ - embedding delay, m - embedding dimension. Next the method goes through calculations of correlation integral and through an average length of lines in the recurrence plots. These lines in the recurrence plots just count the number of steps of neighbor trajectories staying close together. The average line in recurrence plots was calculated against sizes of bins by which we discretized the space. After all, one can derive equations for correlation entropy K_2 of the system perturbed by noise. However, from correlation integral not only the correlation entropy K_2 can be derived, but also the correlation dimension D_2. Combining this statement into one eq. (11) which reads:

$$\lim_{n\to\infty} C^n(\varepsilon) = D_2 \ln\varepsilon - m\tau K_2(\varepsilon) \tag{11}$$

After algebraic transformations we can derive the equation for σ (noise standard deviation) and how it is affecting entropy of the system $K_2(\varepsilon)$ in theory. On the other hand, from the time series we can calculate correlation entropy based on a given range of ε (size of bins in discrete space). Fitting the theoretical function to empirical calculations, gives values of σ. The Noise levels as well as the Linear ones were calculated on 300, 400 and 1000 frames windows shifted by 1 frame. Depending on the window size, it gave us 1800, 1700 and 1100 results of noise levels consecutively in time. Window width of 1000 is optimal for our method [18, 19]. However, we used this size only as reference, as in our case the optimal window size of the stationary data was around 400 points. 300 is equal to 5 seconds, that is the length of the recording presenting one emotion in the reference video. We decided to also check the window width of 400 frames, because we noticed that the response of the subjects is often delayed relative to the video-stimulus. In order to reveal the optimal window size we performed an estimation of the Noise, the Linear and the Chaos, separately for window width 300 and 400, and to make sure that the number of points is not too small for our method, we counted cross-correlations to the reference window (1000). After this analysis, the average correlation for a window of 300 points turned out to be insignificant, from which we concluded that the window of 300 points was too small as a sample. In contrast, the window width of 400 showed very good results. An additional conclusion from the cross-correlation analysis between 400 and 1000 width, was that the center of the window is the optimal reference point for the entire estimation window (relative to the happiness intensity point).

3.4. Calculations of statistical data

In the next step we examined statistical data including differences in emotional intensities between the reference and the main recording sessions. We used the sliding window method with the window width of the same size as in the chaos algorithm calculations. But, to obtain the higher statistical significance, windows were shifted by its full length (400 frames). Obtained results allowed us to select data related to the Happiness emotion, which is best suited for further analysis due to the amount and the quality of the data. We also examined correlations between the intensity of Happiness and particular chaos parameters using tools implemented in the Sci-Kit Learn library [20]. We calculated correlations using Pearson's method and their statistical significance expressed in the p-value. Additionally, to determine the scale of the errors arising during the calculation of the correlation coefficient, we conducted a permutation tests during which correlations were calculated on randomly selected data. We conducted 1000 such trials for each of the parameters and results (the highest collected values obtained from all trials) were compared with the correlations calculated on the real data. These analyzes described and helped us to understand dependencies in our data.

3.5. The FACS Model

In the FACS model basic face emotions are composed from the set of AU representing activity of specific facial muscles [21]. Types of emotion expressed by the face are always associated with the activity of the specific AU. We used the standard simple mappings to predict the face emotions, i.e. prediction of the Happiness was based on the activity of AU06 ("cheek riser"/orbicularis oculi) and AU12 ("lip corner puller"/zygomaticus major). Because the OF's intensity and presence classifiers have been trained separately and on slightly different data sets, the predictions of both parameters are not always consistent. The presence model can predict AU as not being present, while the intensity model could predict its value [1]. To avoid such occurrences, in our algorithm, at first, we have checked the Emotion Presence (EP) (See eq. (12)) by checking if all associated AU were active (> 0). If so, we estimated Emotion

Intensity (EI) (See eq. (13)) as an average from intensities of AU associated with the emotion, otherwise EI was 0, as follows:

$$EP = P(AU[1]) \cdot P(AU[2]) \cdot ... \cdot P(AU[n]) \quad (12)$$

$$EI = \frac{I(AU[1]) + I(AU[2]) + ... + I(AU[n])}{n} \cdot EP \quad (13)$$

where P stands for a function which expresses the presence of emotion in AU, I - intensity of given AU, n is the number of AU's defined for happiness.

3.6. Machine Learning Methods

The final analysis started from preprocessing, where intensities of the Happiness were discretized into 2 bins, equal to zero or larger. The remaining parameters were normalized using the "Min-Max" method [20]. The data were then splitted into a train and test sets (80%/20%) using the Stratified Shuffle Split (SSS) method with standard random state = 42 returning stratified and randomized folds with preserved percentage of samples for each class. [20].

After prepossessing we wanted to check the predictions using different ML methods, basing on the collection of EM and its dynamics (chaotic attractor). Before starting, we had a total of 176361 frames of data from the main session trial. The data was not balanced i.e. for the Happiness 143953 (81.62%) frames were registered when subjects did not express this facial emotion, 32408 (18.38%) when its intensity was rated above 0. The data set contained the following attributes calculated separately for left and right eye and x/y/z vectors: the raw eye vector value, the standard data deviation, calculated noise level, the result of linearity equation (standard signal after delinearization / standard signal deviation), the Chaos calculated as 1 - Noise - Linear components and intensity of the Happiness discretized to the binary values.

With prepared data sets, we started to train different scikit-learn classifiers. We used scikit-learn's GridSearchCV (GSCV) method for tuning classifiers selected for predictions. The GSCV performs exhaustive search over parameters and values optimized by the cross-validation returning various types of scoring, and in addition the best fitting values of the hyper-parameters [20]. With classifiers parameterized by the GSCV we started to rank their accuracy using different metrics. For this purpose we used cross-validation (CV) scoring method (cross_val_score in the scikit-learn) which divides data set into the folds and performs test and train procedures for each fold with the remaining folds, giving in the result the average of all evaluations. We fitted the model and computed CV score 5 consecutive times (with different splits each time) averaging the accuracy results for each classifier [20].

4. Results

We examined 49 subjects, with the age of 16-68 years in both gender and differing in education. Subjects were randomly selected among visitors of the Brain Awareness Week events organized by the University of Gdansk. We compared statistical results of emotion intensities registered during the reference (102900 frames obtained from the same number of video frames) and the main video trials (176361 frames). These data were from the reference session and from these sections, in which subjects showed non-zero intensity of given emotion. The Table 1 presents the examples of results from this comparison.

The subjects most often responded to the stimulus by showing facial emotions of the Happiness or the Contempt - and to a much lesser extent, the Anger (only 5 respondents). The intensity in about 50% of subjects showed significant statistical differences between results from different trials. For Happiness, 10 out of 18 subjects obtained p-value <0.05, for the Contempt 10 out of 24. In the questionnaire carried out after the experiment, all subjects watching the first video-variant correctly indicated Happiness as the presented emotion. Likewise, all respondents of the second video-variant recognized the Sadness in the movie. The questionnaire also included questions about the perceived emotions before and after the experiment. In the case of the first video variant all respondents, despite various emotions declared before (usually the Contempt), after the broadcasting indicated Happiness as the emotion they felt. In the case of the second video, it was mostly the Contempt and the Sadness. The majority of respondents of the third video, despite the laughter in the background connected it with Sadness (only 2 respondents connected it with Happiness).

Table 1: The statistical results of different emotion intensities

emotion	mean-Ref	n-Ref	mean-Main	n-Main	diff	p-value
Happiness	1.0896 +/- 0.1808	8	1.3768 +/- 0.1773	14	0.2872	0.0017
Happiness	0.3263 +/- 0.1969	7	0.1183 +/- 0.1826	9	0.208	0.0464
Happiness	0.1915 +/- 0.0756	6	0.254 +/- 0.0144	2	0.0625	0.3114
Happiness	0.2479 +/- 0.1949	8	0.235 +/- 0.2103	13	0.0129	0.8897
Happiness	0.418 +/- 0.3655	8	0.419 +/- 0.2773	4	0.001	0.9962
Contempt	1.2323 +/- 0.13	8	0.5525 +/- 0.3885	14	0.6798	0.0001
Contempt	1.6007 +/- 0.201	8	1.7829 +/- 0.1041	14	0.1822	0.0105
Contempt	0.4908 +/- 0.2883	7	0.1668 +/- 0.2654	10	0.3241	0.0302
Contempt	0.7864 +/- 0.4899	8	0.3574 +/- 0.305	6	0.429	0.0848
Contempt	0.2102 +/- 0.1112	6	0.2623 +/- 0.0051	2	0.0521	0.5527

Where columns "mean-Ref" / "mean-Main" presents mean and standard deviation of the intensity of a given emotion for the subject calculated from the data of sliding windows (respectively in the reference and the main trial) and "n-Ref" / "n-Main" presents the number of windows in which the subject indicated the intensity of a given emotion, "diff" shows the difference in means and the "p-value" statistical significance

Table 2: The statistical results in calculated parameters

type	mean_happ==0	mean_happ>0	diff	pvalue
raw_data_gaze_1_z	0.9799 +/-0.0091	0.9677 +/-0.0193	0.0121	0.0094
raw_data_gaze_1_y	0.1526 +/-0.0676	0.2031 +/-0.0777	0.0504	0.0208
raw_data_gaze_0_z	0.9811 +/-0.0106	0.9709 +/-0.0177	0.0102	0.0213
raw_data_gaze_0_y	0.1553 +/-0.0733	0.2018 +/-0.0798	0.0464	0.041
Chaos_gaze_1_y	0.771 +/-0.0632	0.7992 +/-0.046	0.0282	0.0771
Chaos_gaze_0_x	0.7364 +/-0.1312	0.7839 +/-0.0554	0.0475	0.0948
Chaos_gaze_1_z	0.1894 +/-0.1783	0.274 +/-0.2188	0.0846	0.1508

Where columns "type" presents name of the parameter, "mean" the average intensity and deviation for a given parameter, respectively when the Happiness intensity was over or equal 0

Therefore, we concluded that in the case of the dissonance between audio and video tracks, for the majority of respondents the stimulus coming from the image (the facial emotion) had a greater impact on the perception of the emotional message.

At this stage we also noticed limitations of the library used to recognize facial emotions. The OF cannot detect unilateral movements and according to the FACS correct contempt recognition is based on AU12 and AU14, but AU14 (the Dimpler) should only be activated on one side of the face. Although activation of AU14 occurs only during the Contempt, and the AU12 activates only for the Contempt and the Happiness, we concluded that this could bring a large field of uncertainty in our classifications. This is why we decided to exclude Contempt from the next parts of our analysis and concentrated on Happiness. So that in the next step we examined the differences of means in gaze vectors and calculated chaosity parameters between data recorded during the Happiness expressions and when this emotion was not presented. Table 2 presents statistical significance results of individual parameters, separately for the left and right eyes (0/1).

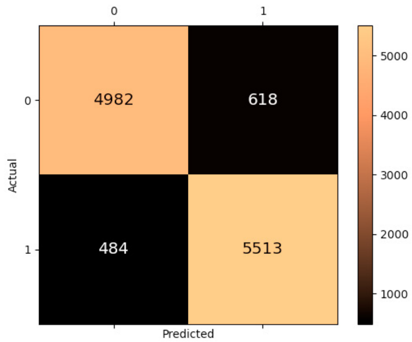
As can be seen in the Table 2, when comparing our data between zero and positive intensity, the most statistically significant are the raw vectors, for the right and then the left eye, and for the 'z' vector (depth of gaze) and the 'y' vector (vertical axis). The next group of parameters with the p-value in the vicinity of statistical significance are 3 parameters from the Chaos group, followed by parameters of other types.

We also examined correlations between the intensity of Happiness and parameters indicating the chaotic EM dynamics by using the permutation methods described in the previous section. According to the results, all parameters connected to the Chaos showed a positive correlation with Happiness. In contrast, both the Linear and Noise were mostly negatively correlated with Happiness (specially the Noise). All of the results were accompanied by p-value <0.05, indicating that correlations were not accidental. Also the results of the permutation tests in all cases gave values <0.01 indicating a very low error rate in the Pearson coefficient calculations.

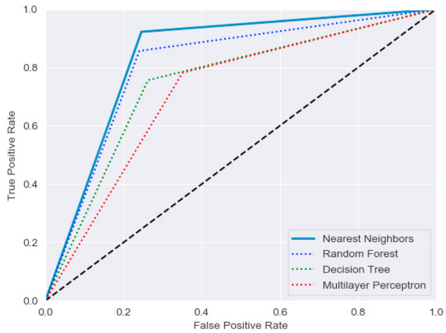
In case of predictions, we achieved the best results by combining the raw parameters determining the values of raw gaze vectors with calculated parameters of the Chaos. With regard to Feature Importance the groups of parameters were presented in the following order: raw vector data, the standard data deviation and next the Chaos, Linear and Noise parameters. We tried to remove and add parameters to the data in different combinations. The data set based solely on raw data gave 95% accuracy indicating the over fitting, while based solely on calculated Chaos, accuracy decreased to 65%. The combination of both groups of parameters gave in our opinion the most reliable prediction

Table 3: The comparison of classifier results

Classifier Name	Accuracy CV Mean	TPR	FDR	ROC-AUC Score
K-Nearest Neighbors	0.89 +/- 0.01	0.94	0.14	0.88
Random Forest	0.88 +/- 0.01	0.89	0.14	0.86
Decision Tree	0.86 +/- 0.01	0.89	0.14	0.86
Multilayer Perceptron	0.66 +/- 0.01	0.73	0.34	0.66



(a) The confusion matrix of the best ranked KNeighbors classifier.



(b) The comparison of classifier ROC curves.

Fig. 2: The results of classifications

results. In addition, we noticed that adding the standard data deviation increases predictions by several percentage points keeping healthy fitting results of the model.

To enrich our rankings we also computed the Receiver Operating Characteristic (ROC) curve with an Area Under the Receiver (AUC). The ROC represents True Positive and False Positive Rate ratio and AUC represents a capability of distinguishing between classes. We compared results of different scikit-learn classifiers and for presentation we selected 4 with different accuracy of classifications. The comparison of prediction results is shown in the Table 3 and in the Figure 2b.

We achieved the highest results using the K-Nearest Neighbors algorithm (metric = 'manhattan', n_neighbors = 3, weights = 'distance') implemented in the SciKit Learn [20]. We obtained Accuracy of 0.89 (+/- 0.01) with ROC-AUC score of 0.88 (F1 = 0.89), Precision = 0.85, Sensitivity / Recall = 0.93, Specificity = 0.82 (False Discovery Rate = 0.14, False Positive Rate = 0.15). Figure 2a presents the "Confusion Matrix" of the prediction model. Random Forest got worse, but still a similar result despite its different specificity, likewise the Decision Tree, which often performs a bit worse than Random Forest because of the lack of voting mechanism. Although all 3 classifiers are model-free, they have a quite different specificity. The Random Forest and Decision Tree belong to deterministic algorithms and are based on attributes and Supervised Learning, while K-Nearest Neighbors is non-deterministic, instance-based and unsupervised [22].It seems that the aforementioned classifiers cope well with nonlinear problems therefore they obtained similar results - in contrast to the neural networks, which explains much worse prediction results obtained by the Multilayer Perceptron.

5. Discussions

We were wondering why subjects seeing the Sadness related face expressions did not respond appropriately by showing the same emotion - as was in the case of the Happiness. We believe that this phenomenon requires a broader explanation and seems to be very interesting. We assumed that this occurrence could be influenced by social norms, as in our culture we don't openly show Sadness. We react to awkward situations in this respect by masking our emotions e.g. by presenting the Contempt and in the case of our study, declaration of this emotion could be a representation of the Sadness. However, the analysis of this aspect was not the aim of the experiment. The purpose was to show if the dynamics of the oculomotor system depends on the experienced emotion. Our analysis of the EM dynamical changes showed the strong correlation between a certain chaos level and happiness. We tried to substantiate our observations and one explanation may be in increased visual field search, due to rapid and chaotic changes in activity of different AU, tracked by the subject's eyes on the face of the model. There are researches showing similar behavior of eyes e.g. when reading, which like intensive scanning of the available view field, is based on complex sequences of saccades. Experimenters performed by E. J. Paulson has described the reading process by an analogy to chaos theory, demonstrating characteristics of sensitive dependence, fractal self-similarity, and non-linearity [23]. Deliberated and thorough eye movements of a reader are understood as a process focused on scanning page data in looking for attractive parts of text to provide the brain efficient construction of the meaning. The attractor cannot be described here as a predictable fixation patterns, because it varies significantly depending on text difficulty [23]. We think that a similar process may occur during the empathetic reaction, when eyes are looking for elements of the face revealing more information about the emotional state of the person. Similarly to the reading process, chaotic eye movements may support the brain in correction of misunderstandings, explanations and in creating the response to new meanings sent by observed face mimicry. The chaotic EM dynamics during the emotion of the Happiness might also be partly explained as the correction of looking into the screen with the randomly jumping light spot [9]. Only in our case EM are jumping between different AUs in order to confirm that their activity represents Happiness. Other researchers also noticed chaotic dynamics in the heartbeat (as the heartbeat itself is a chaotic signal) which may impact various features of the eyes including microsaccades, changes of the ocular pulse, fluctuations of the pupil size and the eye aberrations [24, 25, 26]. We can speculate that one of the major inputs on the eye movement chaocity is the effect of the heartbeat, as it influences eye features by number of mechanisms including pulsation of fundus and corneal or the blood flow through the ciliary body of the crystalline lens [26]. The nonlinear heartbeat dynamics demonstrated a crucial role in characteristics of the emotional state of a subject, especially in ability to distinguish between arousal and valence, thus allowing for the assessment of four basic emotions [27]. Researches suggest differences in the cardiovascular activity among different emotional states as a result of various physiological adjustments [28]. Those adjustments may be observed as asymmetric frontal EEG responses to emotional arousal and may elicit different patterns of cardiovascular reactions [29]. We were not able to confirm this theory in our research with the pulse-meter used during the recording of the subjects. The recorded pulse data showed no significant differences and were rejected in the early stage of the analysis. Another hypothesis concerns directly trait empathy which seems to influence cognitive and facial responsiveness and make individuals more efficient in processing facial emotion. It was found to improve face detection performance by affecting the number and duration of fixations and reducing response time of eye movements, increasing scanning eye movements in specific areas of interest, strongly associated with activity of cortical event-related potential (increased N200 component amplitude) and facial activation (zygomatic and corrugator) [30].

6. Conclusions

We try to defend the hypothesis that the Happiness as a feedback (empathetic) face emotion can increase chaotic behavior in the eye movements. This means that when the intensity of AUs increases, the eye movements become more chaotic and behave less noisy, which can help in predictions of Happiness. Apart from open discussion about the reasons for chaotic behavior of the oculomotor system during elicited emotion, we can also see a practical application of the machine-learning methodology presented in this text. To detect happiness, the FACS model expects activation of facial muscles around the mouth and eyes (the inner brow raiser). It is enough for a person to cover his mouth with a protective mask or by hand while snorting laughter to cause incorrect predictions in standard classifiers of happiness. We think that in such cases, parameters calculated from the eye movements could be useful to assist in the correct classification of appearing happiness. On the other hand, happiness is one of the most pretended emotions. Just as we are able to learn facial expressions, it would be hard to train eyes so that their movements would correlate with the chaos and not with the noise. We think that our method could support the determination of the authenticity of the happiness declared by facial expressions. We believe that our findings prove the point of further experiments involving signal analysis interpreted as human emotion, by using sensors and combination of different methods in the field of physics for supporting machine-learning classifications.

References

[1] T. Baltrusaitis, A. Zadeh, Y. C. Lim, L. Morency, OpenFace 2.0: Facial Behavior Analysis Toolkit, 2018. doi:10.1109/FG.2018.00019.

[2] E. Pretegiani, L. M. Optican, Eye movements in parkinson's disease and inherited parkinsonian syndromes, Frontiers in Neurology 8 (2017) 592. doi:10.3389/fneur.2017.00592.

[3] A. Śledzianowski, A. Szymanski, A. Drabik, S. Szlufik, D. M. Koziorowski, A. W. Przybyszewski, Combining results of different oculometric tests improved prediction of parkinson's disease development, in: N. T. Nguyen, K. Jearanaitanakij, A. Selamat, B. Trawiński, S. Chittayasothorn (Eds.), Intelligent Information and Database Systems, Springer International Publishing, Cham, 2020, pp. 517–526.

[4] F. W. Bylsma, D. Rasmusson, G. W. Rebok, P. M. Keyl, L. Tune, J. Brandt, Changes in visual fixation and saccadic eye movements in alzheimer's disease, International Journal of Psychophysiology 19 (1995) 33–40. doi:https://doi.org/10.1016/0167-8760(94)00060-R.

[5] J. Naruniec, M. Wieczorek, S. Szlufik, D. Koziorowski, M. Tomaszewski, M. Kowalski, A. Przybyszewski, Webcam-based system for video-oculography, IET Computer Vision 11 (2017) 173–180. doi:10.1049/iet-cvi.2016.0226.

[6] J. P. Higgins, Nonlinear systems in medicine, The Yale journal of biology and medicine 75 (2002) 247—260.

[7] C. Astefanoaei, E. Pretegiani, L. Optican, D. Creanga, A. Rufa, Eye movement recording and nonlinear dynamics analysis – the case of saccades., Rom J Biophys. 23 (2013) 81–92.

[8] S. S. M. Chanijani, P. Klein, S. S. Bukhari, J. Kuhn, A. Dengel, Entropy based transition analysis of eye movement on physics representational competence, in: UbiComp Adjunct, 2016, p. 1027–1034.

[9] K. Harezlak, P. Kasprowski, Searching for chaos evidence in eye movement signals, Entropy 20 (2018). doi:10.3390/e20010032.

[10] C. Frith, Role of facial expressions in social interactions., Philosophical transactions of the Royal Society of London. Series B, Biological sciences 364 1535 (2009) 3453–8.

[11] K. K. W. Kampe, C. D. Frith, U. Frith, "hey john": Signals conveying communicative intention toward the self activate brain regions associated with "mentalizing," regardless of modality, Journal of Neuroscience 23 (2003) 5258–5263. doi:10.1523/JNEUROSCI.23-12-05258.2003.

[12] S. Morelli, M. Lieberman, The role of automaticity and attention in neural processes underlying empathy for happiness, sadness, and anxiety, Frontiers in Human Neuroscience 7 (2013) 160. doi:10.3389/fnhum.2013.00160.

[13] A. J. O'Toole, J. Harms, S. L. Snow, D. R. Hurst, M. R. Pappas, J. H. Ayyad, H. Abdi, A video database of moving faces and people, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (2005) 812–816. doi:10.1109/TPAMI.2005.90.

[14] A. L. Swan, Transnational Identities and Feeling in Fandom: Place and Embodiment in K-pop Fan Reaction Videos, Communication, Culture and Critique 11 (2018) 548–565. doi:10.1093/ccc/tcy026.

[15] W. Dai, I. Selesnick, J.-R. Rizzo, J. Rucker, T. Hudson, A parametric model for saccadic eye movement, in: 2016 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), 2016, pp. 1–6. doi:10.1109/SPMB.2016.7846860.

[16] K. Urbanowicz, J. Hołyst, Noise-level estimation of time series using coarse-grained entropy., Physical review. E, Statistical, nonlinear, and soft matter physics 67 4 Pt 2 (2003) 046218.

[17] K. Urbanowicz, J. Hołyst, Noise estimation by use of neighboring distances in takens space and its applications to stock market data, International Journal of Bifurcation and Chaos 16 (2006) 1865–1869. doi:10.1142/S0218127406015763.

[18] K. Urbanowicz, J. J. Żebrowski, R. Baranowski, J. A. Hołyst, How random is your heart beat?, Physica A: Statistical Mechanics and its Applications 384 (2007) 439–447.

[19] G. Górski, G. Litak, R. Mosdorf, A. Rysak, K. Urbanowicz, Dynamics and noise level estimation in two-phase flow through a mini channel, in: Automation and Robotics in Production Engineering, volume 844 of *Applied Mechanics and Materials*, Trans Tech Publications Ltd, 2016, pp. 115–121. doi:10.4028/www.scientific.net/AMM.844.115.

[20] G. V. F. Pedregosa, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in python, Journal of Machine Learning Research 12 (2011) 2825–2830.

[21] P. Ekman, W. V. Friesen, Facial Action Coding System: Manual, 1978.

[22] M. Mohanapriya, J. Lekha, Comparative study between decision tree and knn of data mining classification technique, Journal of Physics: Conference Series 1142 (2018) 012011. doi:10.1088/1742-6596/1142/1/012011.

[23] E. J. Paulson, Viewing eye movements during reading through the lens of chaos theory: How reading is like the weather, Reading Research Quarterly 40 (2005) 338–358. doi:10.1598/RRQ.40.3.3.

[24] S. Ohl, C. Wohltat, R. Kliegl, O. Pollatos, R. Engbert, Microsaccades are coupled to heartbeat, Journal of Neuroscience 36 (2016) 1237–1241. doi:10.1523/JNEUROSCI.2211-15.2016.

[25] D. R. Trew, C. James, S. H. L. Thomas, R. Sutton, S. M. Smith, Factors influencing the ocular pulse — the heart rate, Graefe's Archive for Clinical and Experimental Ophthalmology 229 (1991) 553–556.

[26] K. M. Hampson, E. A. H. Mallen, Chaos in ocular aberration dynamics of the human eye, Biomed. Opt. Express 3 (2012) 863–877. doi:10.1364/BOE.3.000863.

[27] V. Gaetano, C. Luca, L. Antonio, S. E. Pasquale, B. Riccardo, Revealing real-time emotional responses: a personalized assessment based on heartbeat dynamics, Scientific Reports 1 (2014) 2045–2322. doi:10.1038/srep04998.

[28] K. M. Prkachin, R. M. Williams-Avery, C. Zwaal, D. E. Mills, Cardiovascular changes during induced emotion: An application of lang's theory of emotional imagery, Journal of Psychosomatic Research 47 (1999) 255 – 267. doi:10.1016/S0022-3999(99)00036-7.

[29] S. R. Waldstein, W. J. Kop, L. A. Schmidt, A. J. Haufler, D. S. Krantz, N. A. Fox, Frontal electrocortical and cardiovascular reactivity during happiness and anger, Biological Psychology 55 (2000) 3 – 23. doi:10.1016/S0301-0511(00)00065-X.

[30] M. Balconi, Y. Canavesio, Is empathy necessary to comprehend the emotional faces? the empathic effect on attentional mechanisms (eye

movements), cortical correlates (n200 event-related potentials) and facial behaviour (electromyography) in face processing, Cognition and Emotion 30 (2016) 210–224. doi:10.1080/02699931.2014.993306.