Scientific Research

Individual Classification of Emotions Using EEG

Stefano Valenzi[1*], Tanvir Islam[1*], Peter Jurica[1], Andrzej Cichocki[1,2]

[1]Advanced Brain Signals Processing Laboratory, RIKEN Brain Science Institute (RIKEN BSI), Wako, Japan
[2]Systems Research Institute Polish Academy of Science, Warsaw, Poland
Email: stefano@brain.riken.jp

Abstract

Many studies suggest that EEG signals provide enough information for the detection of human emotions with feature based classification methods. However, very few studies have reported a classification method that reliably works for individual participants (classification accuracy well over 90%). Further, a necessary condition for real life applications is a method that allows, irrespective of the immense individual difference among participants, to have minimal variance over the individual classification accuracy. We conducted offline computer aided emotion classification experiments using strict experimental controls. We analyzed EEG data collected from nine participants using validated film clips to induce four different emotional states (amused, disgusted, sad and neutral). The classification rate was evaluated using both unsupervised and supervised learning algorithms (in total seven "state of the art" algorithms were tested). The largest classification accuracy was computed by means of Support Vector Machine. Accuracy rate was on average 97.2%. The experimental protocol effectiveness was further supported by very small variance among individual participants' classification accuracy (within interval: 96.7%, 98.3%). Classification accuracy evaluated on reduced number of electrodes suggested, consistently with psychological constructionist approaches, that we were able to classify emotions considering cortical activity from areas involved in emotion representation. The experimental protocol therefore appeared to be a key factor to improve the classification outcome by means of data quality improvements.

Keywords

EEG, Human Emotions, Emotion Classification, Machine Learning, LDA

1. Introduction

Human emotion plays a critical role in perception, cognition and (social) behavior. Good management of emo-

*These authors equally contributed to this work.

tions of oneself and of others may improve the quality of life. A good management of emotions is possible only if individuals recognize emotions reliably. This is also true for the realization of practical applications: every application has to rely on either consistent offline or online emotion recognition. Offline applications may be used to study individuals' responses to different stimuli (e.g., neuromarketing) and to understand the brain basis of decision making (e.g., rehabilitation purposes). Online applications may contribute in improving people's emotional intelligence (facilitating social and emotional learning) and, in case of impaired patients or babies, monitoring their states (e.g. happy, relaxed, stressed, or sad). Further, online applications can exploit feedback from users' emotional states to customize computerized services: a simple discrimination between pleasant and unpleasant emotions would allow an algorithm to learn how to improve provided performances.

Considering that the primary objective of this study is to perform computer aided classification of some human emotions, the initial question is: what is emotion? There is no universally accepted definition of emotion, and ultimately the main differences in definitions relate to the approaches adopted to define them. Two widely accepted approaches are those of psychology and neuropsychology. From psychology perspective, the study of emotions is the study of the conscious representation of the emotional experience: what individuals feel. From neuropsychology perspective, emotion is seen as a set of coordinated responses that take place when an individual faces a personally salient situation [1], in other words, how our bodies react to emotional stimuli (even when the emotion perception is unconscious). In the present study we assume a middle position by trying to define how physiological changes occur when our feelings change. We adopt such approach assuming that, when participants recognize their emotions well, the association between physiological data and perception of different feelings will be reliable.

The reliability of data is an important matter in case of classification tasks. So the next question is: what kind of data should be collected? Many researchers have proposed computer aided emotion recognition methods based on behavioral and physiological data. Considering that even humans cannot univocally recognize emotions using nonverbal communication cues such as facial expressions [2] and/or human voices [3] behavioral data seem to be insufficient to objectively perform such classification.

Among different physiological data that originate both from peripheral and central nervous system, currently EEG signals appear to provide the most suitable information to study human emotion [4]. EEG data recorded with appropriate procedures may reveal information on positive and negative emotional states (approach-withdrawal model as in [5]) as well as information belonging to emotional components related to the emotion representations (psychological constructionist approaches as in [6]).

Considering that previously proposed classification methods based on EEG signals were methodologically solid and assuming that EEG signals are suitable for emotion classification (and/or components directly related to distinct affective states), it is unclear why the reported classification accuracies did not reach the 90% mark in most cases. A possible explanation is that the used experimental paradigm impaired data quality. If data are poor in quality (e.g., noisy), even the best mathematical tools fail to classify them reliably. Emphasizing the psychological perspective and analyzing the literature, we identified four points that should be improved in the experimental paradigm: 1) environment and equipment setting, 2) emotion elicitation procedure, 3) evaluation of categories of stimuli and 4) evaluation of individual differences.

Different equipment settings [7] and environmental factors may stress participants and consequently affect their immediate emotional state and therefore their emotional perception (see 2.2. for details). Any emotion elicitation procedure chosen, irrespective of its duration, triggers lasting cognitive effects that may affect the subsequent cognitive functions [8]. Such overlap of cognitive effects, called carryover effect, may crucially affect the data [9] [10]. This implies that adjacent emotions in serial elicitations may be confused. To alleviate this effect it is beneficial to have longer inter-stimuli intervals and distractor tasks [11]. Regarding stimuli, we have to be sure that the EEG signals recorded are related to the affective content rather than salient physical features of categories of stimuli. Otherwise, there is a risk of discriminating EEG patterns that reflect physical features, instead of emotions. For instance, video stimuli that are mostly composed of bright scenes and video that are mostly composed of dark scenes should not be the sole representatives of pleasant and unpleasant stimuli (see 2.5 for further details), respectively. Finally, emotions recognition is firstly affected by emotions perception (input) and ultimately by individual ability to describe emotions (output). Also, participants may vary in their sensitivity to stimuli (see 2.6.2 for details). In order to collect physiological data that correspond as much as possible to the expected emotional states we should evaluate such individual differences, aiming to select participants that are not emotionally biased.

To our best knowledge, it appears that all the above mentioned factors were not sufficiently considered in previous emotion classification studies based on EEG data. Unfortunately, the use of validated stimuli does not automatically prevent any of these potentially impairing situations. Therefore, the only solution to reduce these possible risks is to apply well designed controls and precautions.

The present study integrates a standard protocol with the above mentioned additional controls on stimuli and participants. Using well known EEG analysis methods (applied within-subjects) and machine-learning algorithms we obtained excellent results in terms of classification rates. These results suggest that a way to improve computer aided emotion classification reliability is through well controlled experimental protocols and indicate that EEG is suitable to perform online classification of human emotions (or emotion components related to emotion representation). Feasibility of online applications is supported by excellent results obtained by means of 32 electrodes. Results computed by using only 8 electrodes (AF3, AF4, F3, F4, F7, F8, T7 and T8) support the withdrawal model [5]. The consistence of our outcomes with psychological constructionist interpretation of emotions indicates that we classified information related to some components of the emotion representations.

2. Materials and Methods

2.1. Subjects

Ten healthy Japanese participants took part in our study after providing written informed consent. All participants held at least a bachelor degree, five of them were males and average age was 33.1 years (s.d. = 7.0). Screening by Assessing Emotions Scale [12] guaranteed that participants were able to describe and manage their emotions (see 2.6.2.1. for details). For this questionnaire, the higher the score the better the participants were in terms of measured properties. The criterion for participant selection was therefore a score equal or higher than the average indicated in the study mentioned. Participants were paid 1000 Yen per hour and their transportation fee was refunded. Experiments were conducted with approval from RIKEN ethical committee.

2.2. Equipment and Environment Setting

One previous study provided evidences that the sensation of reality of a video stimulus is related to the size of visual field: the larger the visual angle the more intense the stimulus [7]. We therefore used a large display (40 inch) and we placed the participants 1m from the display, obtaining approx. 40˚ horizontal and 30˚ vertical viewing span.

It has been shown that emotional response to video stimuli is affected by environmental parameters such as room lighting [13] and temperature [14]. We set these parameters to be as natural and comfortable as possible to limit possible sources of stress. The room was lit by a dim (low luminance) lighting [15] and room temperature, considering that the experiment was run during winter, was set at 19˚C [16].

2.3. Procedures

Participants ran the experiment individually [17] [18] using a keyboard and a mouse to answer assessment questionnaires and perform distractor tasks (see below). Instructions for these tasks appeared on the same screen used to display the movie clips. After participants finished reading the informed consent, they were given instructions and experimenters mounted EEG electrodes.

In preliminary studies participants reported the EEG cap montage as annoying. To alleviate the possible negative effect of the headcap montage we decided to mount it as fast as possible (two experimenters were employed in this procedure). Nevertheless, considering that even a "quick headcap" montage could negatively affect participants' condition, we assessed their initial emotional state right after the electrodes were plugged. We used the Pleasure-Arousal-Dominance [PAD] emotion assessment scale [19] to quickly measure participants' emotional state in terms of affective components (pleasant-unpleasant dimension), activation (arousal-nonarousal dimension) and control (dominance-submissiveness dimension). Each combination of these three (numerically assessed) dichotomous dimensions defines distinct emotional states.

After a quick evaluation of participant's answers to the PAD test, we were able to define that all of them were in a neutral state, or nearly so, therefore without further procedures we recorded the EEG baseline (2 minutes in total). Baseline recording was composed of two parts: initially a closed-eye condition (one minute), then an acoustic signal was played to indicate the open-eye condition which lasted for another minute. Ten seconds after

the baseline recording we started the emotions elicitation procedure using video clips (**Figure 1**).

Each participant watched each video clip once with presentation order randomized. The randomization was organized in four sequential blocks; each block consisting of four videos (one for each target emotion). In each block, the first and the third video were randomly chosen among sad and disgusting videos and the second and the fourth among amusing and neutral ones. However, the fourth video displayed in the last block was always an amusing video clip to allow participants to leave the experimental room in a positive emotional state. This counterbalanced randomization was adopted to minimize possible residual carryover effects: knowing that emotional stimuli work additively we avoided to show two similar emotions stimuli contiguously (e.g. a sad and a disgust stimulus are similar in the pleasantness dimension).

Elicited emotions fade or get biased with time after the end of the video and the questionnaire's completion [20]. Taking this into account, we asked participants to assess their emotions directly after each video. We used a post-film questionnaire similar to the one initially proposed in [1] (see 2.6.3. for details). Afterwards, participants were given a distractor task to alleviate the carryover effects between videos presentation (**Figure 1(b)**).

As it has been previously proposed [11], we used a drawing task as distractor task. Participants, using a mouse, could chose different colors and control a brush aiming to reproduce the simple geometric figures displayed on the screen. These figures were presented in 3 different colors and either overlapping or not overlapping. At the end of the distractor task, the participants clicked on a button displayed on the screen. The video-clip that followed was displayed 10 seconds after the button was pressed. The time between two subsequent video stimuli, which was decided by the participants, was on average around 3 minutes.

2.4. Instructions

During the quick EEG cap and electrodes wiring, participants were instructed on how to assess their emotions. We asked them to report: 1) their actual feeling and not their expected emotional reaction to video stimuli (and here we reminded them that responses were anonymous) and 2) how they felt during the video display, avoiding therefore report of their general mood or attitude. These instructions are similar to ones previously used in [21].
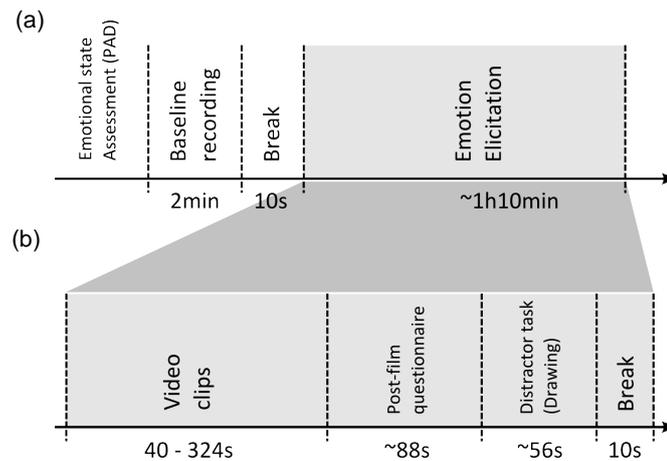

Figure 1. Experimental procedure. (a) Initially we assessed participants' emotional state (using the PAD scale) and then recorded a baseline (2 minute of duration). After 10 s break, we started the emotions elicitation procedure. Duration of the emotions elicitation procedure was on average one hour and ten minutes. (b) For sixteen times (one time for each video-clip) we displayed the video stimulus and right after we assessed the participants' emotional state (using the post-film questionnaire). Once participants filled the post-film questionnaire they had to perform a distractor task. A waiting time of 10 seconds was set between the end of the distractor task and the display of a new video. Duration of video-clips was in the range from 40 s to 324 s; Post-film questionnaire duration time was on average 88 s; distractor task duration was on average 56 s. In total, the inter-stimuli duration time was on average 154 s.

Participants were also instructed that they could skip any uncomfortable videos and take a break if they thought it was required. In the latter case, we asked them to take a break after they filled out the post-video questionnaire (see 2.6.3. for details). Finally, we asked participants 1) to find a comfortable seating position; 2) to be still and not move during the stimulus display to avoid artifacts in EEG recording.

2.5. Video Stimuli and Analysis

We selected 16 video stimuli, four videos for each investigated emotion (sadness, disgust, neutral and amusement). These video clips were classified as very effective stimuli in previous extensive validation studies [22] [23]. Following the instructions in [22] [23] we took the video-clips from the movies indicated in **Figure 2**.

All videos were presented in their original languages (English and French) with Japanese subtitles. "Amputation of Arm" video clip [previously used in [22] was without audio track. "When a man loves a woman" was classified as tender movie clip in [23]. Nevertheless, considering the assessment scores reported in [23], we decided to use it as neutral stimulus. Videos length was variable: from 40 s to 324 s for the shortest and longest clip, respectively.

We adopted video stimuli because they have a very high level of emotional intensity (response strength and breadth), ecological validity, standardization (highly controlled conditions) and attentional capture [1]. Furthermore, since videos possess both auditory and visual modalities, they provide the highest complexity in terms of physical features [1]. Feature complexity may help to reduce the possibility that features of stimuli are classified instead of categories of emotions because of the similarities among salient features of stimuli that belong to the same class of emotion.

To control for this we analyzed static and dynamic visual features of the video stimuli. Results of this analysis are summarized in **Figure 3**, in which we show that the stimuli in the same emotional category are scattered all over the feature space, and stimuli of different emotional categories have large overlaps in feature space, thus making impossible to classify stimuli into their intended emotional categories by using these static and dynamic visual features.

Philadelphia 5:24	The Champ 2:49	City of Angels 4:29	Dangerous Minds 2:13
Amputation 1:05	Seven 1:30	Hellraiser 3:28	Trainspotting 1:44
Blue 1:44	Blue 0:41	The Lover 0:40	When a Man Loves a Woman 0:43
The Dinner Game 1:44	When Harry Met Sally 2:53	Les Visiteurs 2:09	La Cite de la Peur 2:19

Sadness   Disgust   Neutral   Amusement

Figure 2. Video stimuli have been taken from the indicated movies following the instruction given in the relative studies [22] [23]. Length of video-clips minutes and seconds is indicated below titles. As indicated in [23] three videos were taken from the movie "Blue", the two shorter ones were combined into a single video-clip.
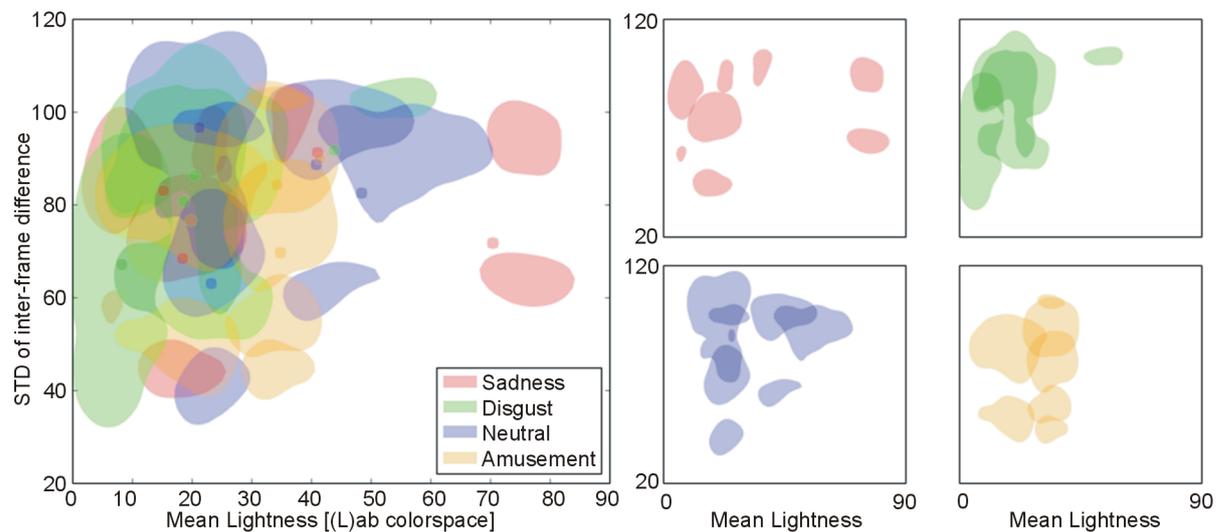
Figure 3. Static and dynamic feature analysis of video stimuli. The main plot (on left) illustrates the relationship and distribution of lightness and motion energy in video stimuli. Colored clouds represent videos according to the emotional category of the stimulus (see legend). The clouds enclose the area of points of which the coordinates represent the values of mean lightness in the center of the screen and motion energy during the whole duration of video clips. Colored circle markers show the expected value. Colored clouds show summary values for all frames of all 16 video stimuli. Four panels on the right shows the results for each category separately.

Overall, we evaluated 29 features of the video stimuli. These 29 features were composed of two subgroups. The first subgroup of 19 features was based on either luminance or color (in two color-spaces RGB and Lab) in different portion of the screen. The remaining 10 features were: the mean and standard deviation of inter-frame difference (latter referred to as "motion energy" here), the mean spatial and temporal frequency, their ratio, and the slopes of amplitude spectra (see Appendix for further details).

All luminance measures were highly correlated. For all the video-clips the relationship between the mean luminance [L0] and the feature with the lowest average correlation with all other features is displayed in **Figure 3**. This figure indicates that static and dynamic features of video-clips were similar and well balanced among all presented videos and among emotional categories.

2.6. Measures

2.6.1. Physiological Data

EEG data were continuously collected with a Biosemi Active 2-system with 32 active electrodes and 2 reference electrodes corresponding to 10 - 20 international system. We have connected PC's audio card line output to the Biosemi amplifier. Audio track was used as temporal marker during the video display. All signals were sampled at 2048 Hz sampling frequency.

2.6.2. Assessment of Individual Differences Related to Emotion

The following questionnaires (2.6.2.1. and 2.6.2.2.) were administrated to assess the participants' differences to perceive and manage their emotions. First, the *Assessing Emotions Scale* was used to select participants of this study. We intended to not recruit participants with poor ability to describe their feelings. However, since sensitivity to different modalities of stimuli may vary in participants this screening criterion may be not sufficient. For instance, differences in empathy can elucidate how participants will react to certain stimulus modalities [24]. To test this hypothesis we also measured the participants' empathetic processes using the questionnaire created by Davis [25]. This questionnaire (the Interpersonal Reactivity Index) assesses 4 subscales of empathetic processes. When a study relies on emotion elicitation procedures that involve imagery-tasks or video-stimuli, a measure of participants' empathic process toward fictional scenes (a subscale of Davis' questionnaire) may elucidate whether the participants are suitable for such emotional study. Both surveys, originally written in English, have been translated to Japanese. The assessing emotions scale has been previously translated in other languages such as Hebrew, Polish, Swedish and Turkish version [12].

a) Assessing Emotion Scale

The questionnaire, also known as the Schutte Emotional Intelligence Scale or the Self-Report Emotional Intelligence Test [12], was used to measure the emotional intelligence using a 33 self-report items. Each item is self-descriptive and is rated on 5-point scale from "strongly disagree" to "strongly agree"). Different studies [12] demonstrated that the items, organized in four distinct subscales, are related to four factors. These factors measure perception of emotions, managing emotions in the self, social skills or managing emotions of others, and utilizing emotions. Previous studies [26] have shown that the Assessing Emotions Scale has a very good internal consistency as demonstrated by the Cronbach's alpha evaluated across different samples: on average 0.87. Our criterion was to not accept participants with a score below the average.

b) Fantasy Scale—Interpersonal Reactivity Index (IRI)

IRI [25] is a 28 items questionnaire developed to investigate the empathic process in a multi-dimensional perspective. Each item is self-descriptive and is rated on a 5-point scale (from "strongly disagree" to "strongly agree"). Four factors are evaluated by means of four seven-item subscales. The subscales are the perspective taking scale, the empathic concern scale, the personal distress scale and the fantasy scale. In our study the fantasy scale was administered to measure how much a respondent is empathetic to fictional scenes (theatrical representations, movies or books). For each subscale, Davis found significant differences between men and women. Therefore the internal reliability coefficients were computed separately. Cronbach's alphas of the fantasy scale were 0.78 and 0.75 for men and women, respectively [25].

2.6.3. Emotional Assessment: Post-Film Questionnaire

Our post-film questionnaire was based on a previously used emotion self-report inventory [1]. Our participants retrospectively reported how they felt during the video clip display.

We divided the questionnaire in two parts. The first one (21-item questionnaire) was based on emotion adjectives. At first 19 items were defined (amusement, anger, anxiety, calmness, confusion, contempt, disgust, embarrassment, excitement, fear, guilt, happiness, interest, joy, love, pride, sadness, shame and surprise) and 2 items could be optionally chosen by participants (they could choose one, two or no extra emotion self-report items). Each item was rated on an 8-point scale (from "strongly disagree" to "strongly agree").

The second part was based on items that refer to general emotion dimensions. Following a previous study [27] we decided to use pleasure, arousal, dominance and surprise as dimensions. Each dimension was expressed in an opponent scale (e.g., pleasantness-unpleasantness). Respondents had to rate each dimension using a continuous scale (ranged from 1 to 5) located on a horizontal bar. The bar extremities were marked with emotionally opposite extremes of the dimension. For instance, in the pleasure dimension, a rating of 1 meant the participant felt extremely pleasant, while a rating of 5 meant the participant felt extremely unpleasant. A rating of 3 indicated a neutral state where dimension components were equally balanced.

2.7. Data Analysis

All the data were analyzed using Matlab (by Mathworks Corporation) and LabVIEW (by National Instruments Corporation). We followed the standard data analysis and classification procedure: 1) EEG preprocessing; 2) feature extraction; 3) feature reduction; 4) evaluation of classification rate. One participant asked for many breaks during the experiment and eventually reported to have headache; we excluded this participant from data analysis. The other nine participants on average asked to have two breaks during the experiment. Following the instructions, they asked for breaks during the drawing tasks. Experiment duration was on average two hours including mounting the headcap and breaks. All the results presented in later sections were computed on data collected from nine participants, five of them were men.

2.7.1. EEG Preprocessing

The following preprocessing was conducted before further analysis. At first, EEG data was down sampled to 256 Hz. Then, portions of the data that contain eye blinks or EMG related artifacts were deleted by visual inspection. In our experiment, the participants were instructed to be still during stimuli presentation. Therefore, the EMG related artifacts were very few. EEG data was then filtered by a band-pass filter in the range of 0.16 - 70 Hz. Because of the possible noise from 50 Hz power line, data was then filtered again by a notch filter with rejection of 50 Hz. Finally, we recorded EEG data during all the experiment but just data recorded during each

video display was selected to compute the classification.

Such EEG data was segmented into 6-seconds long epochs. For each video, the first epoch corresponded to the beginning of the display of the video and the last epoch was starting at six seconds before the video end. Further, each epoch was overlapping with its consecutive one by five seconds. Between two adjacent epochs we had therefore just one second of new information. For instance, having a hypothetical video of ten seconds length, we would consider just four 6-seconds long epochs. Selected epochs were individually calculated for further analysis using the Short Term Fourier Transformation [STFT] as described below.

2.7.2. Feature Extraction Method: STFT

For classification of EEG signal into emotion categories, it is necessary to extract features from the signal. The main task of feature extraction is to derive the salient features which can map the EEG signals into consequent emotions. There are several well known categories of feature extraction methods: 1) time domain analysis; 2) frequency domain analysis; and 3) time-frequency analysis. In our case, we chose STFT, which is a well used time-frequency analysis, because STFT can reveal the salient properties of a signal in detail in the time frequency domain with high resolution and is less affected by cross disturbance of multi-component signal.

To extract features, each of the 6 s long epochs was analyzed with STFT, with a Hanning window of 256 samples (1 s long) and a time step of 64 samples. The number of frequency bins was 256, thus the resulting frequency resolution was 128 Hz (Nyquist Frequency)/256 = 0.5 Hz. From the results of the STFT, we calculated the average magnitude of spectral power of five bands of frequencies: 0.16 - 4 Hz (Delta); 8 - 13 Hz (Alpha); 14 - 21 Hz (Lower Beta); 21 - 30 Hz (Upper Beta); 30 - 40 Hz (Gamma).

For each participant and each epoch, we subtracted the power of the baseline (the open-eye session) from the power averaged over that epoch, thus removing the baseline effect. As a result, we obtained five band-power values for each epoch and each electrode. As this analysis was conducted for all 32 electrodes, each epoch was finally represented by $5 \times 32 = 160$ features. For classification of emotions, each epoch was tagged with a label of the target emotion of the video that played during the epoch was generated. We decided to use such labeling because the evaluation of participants' assessment [28] indicated that target stimuli were properly assessed.

2.7.3. Feature Reduction and Classification

Linear Discriminant Analysis [LDA] is a well known method for reducing dimensionality of the features space. For a classification task, LDA computes a linear combination of features which characterizes or separates two or more classes of objects. We used LDA for features reduction because it is known to perform better than PCA or factor analysis in classification tasks where there are pre-assigned labels to data.

With LDA, for each epoch we obtained three features which are the linear combination of 160 original features representing one epoch. Beside these three features, LDA computed also the LDA weight matrix which was used later to classify test data (as described below). After having a reduced the number of features, we applied several clustering and supervised learning algorithms to classify the EEG epochs into their emotional categories. We applied the LDA feature dimensionality reduction followed by classification analyses to data individually measured. Further, as it is usually done in machine learning computational procedures, we divided data into two subsets (training and test data set). The procedure was to apply the LDA classification learned with the training subsample data to the test subset and then to evaluate the test data classification accuracy rate using the classification functions resulting from clustering and supervised learning algorithms computed on training data (**Figure 4**).

We randomly selected epochs from the original data set to populate these 2 subsamples. We individually computed the accuracy rate resulting from four participants' data (**Figure 5**) in order to identify how to split the original data between the two samples. For each classification computed, we considered the percentage of data used in both subsamples: from 20% to 80% in the training subsample and the remaining data in the test one. Considering the outcomes of this basic analysis, we decided to equally split the original data (50% in each subset).

The common procedure to cross validate data of participants is to evaluate all participants in the same data set. On top of that, we cross-validated our data sets using individual training and test subset because our analysis revealed that there was very low correlation (r < 0.1) among LDA weight matrices of participants. We therefore assumed that the weight matrices (and therefore the emotional representation) are unique among participants due to individual differences.

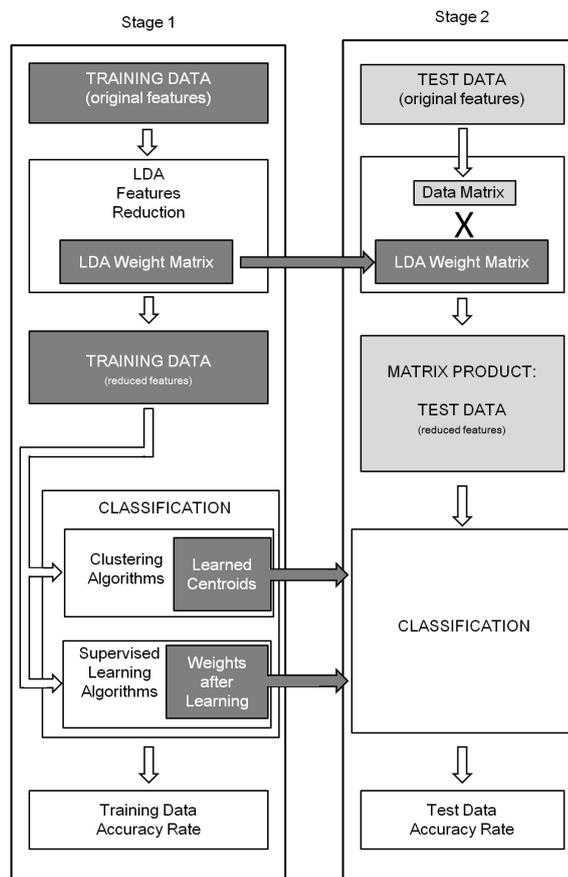To verify this assumption we cross validated our data using the leave-one-out method, leaving out one par-

Figure 4. We displayed how classification accuracy is computed (stage 1 and stage 2) for two subsamples of the original data set (called training and test, respectively). To reduce the features of test data set we used LDA weight matrix evaluated on training data. To evaluate accuracy rates computed on test data set we used the classification functions evaluated on training data by means of cluster and supervised learning algorithms.

ticipant at a time. We made this choice because, using LDA method in case of strong differences among individual data, there is a high risk to evaluate just small portion of information that is probably related to common noise across the participants instead to salient emotional information. Classification accuracy evaluated by leave-one-out method is extremely poor: on average 33.51% among all participants. This indicates that participants are very different (no common pattern emerges among participants) and ultimately supports the use of within-subjects analysis.

Finally, to consider individual data as unique is consistent with the idea that emotion representations is individually developed based on the link between the sensory representation of a stimulus and its value in the affective working memory [29]. Affective working memory is shaped by changes in visceromotor response (e.g., autonomic, chemical, and behavioral). Visceromotor responses affect our body state, contributing to creation of pleasant and unpleasant feelings (e.g., a vasodilatation is perceived as pleasant while a vasoconstriction is perceived as negative).

3. Results

3.1. Classification

Individual test data was analyzed using the following algorithms for classification of epochs which are already
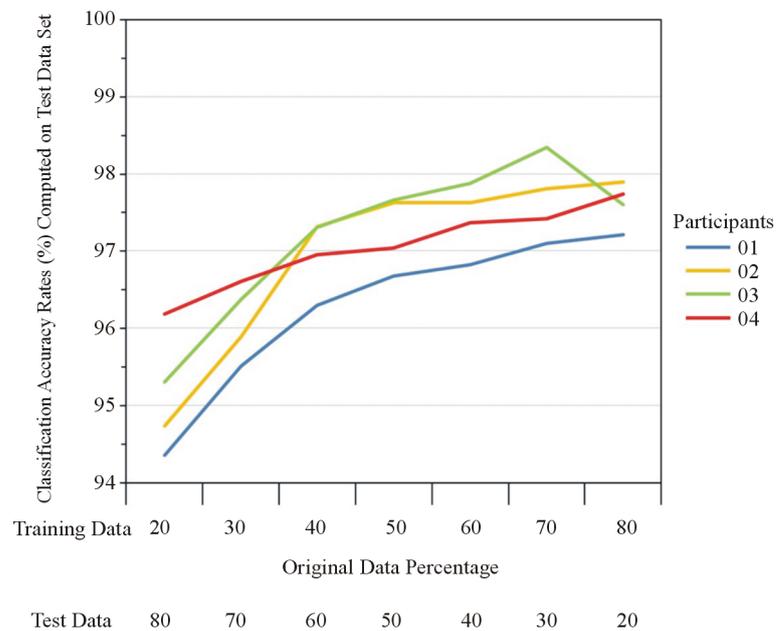
Figure 5. Classification accuracy rates as function of dimension of training and test data set. Individual results of 4 participants are displayed. Note that, for all evaluated participants, accuracy rates were barely improved by increasing the dimension of training data over 50%. 100% of original data were always split in such two subsamples.

tagged with target emotions. Basically, we used two types of algorithms: 1) Supervised learning algorithms and 2) Clustering algorithms.

1) Supervised learning algorithms:

a) Error Back-propagation [BP] (Number of hidden neuron = 5, Output neurons = 4, Min MSE = 1E-5, in delta MSE = 1E-8, Min Step length = 1E-8, Iterations: 1000, Percentage of Training data: 80, test = 20);

b) Learning Vector Quantization [LVQ] (number of neurons = 4, Tolerance = 1E-5, Iterations: 1000, Learning rate = 0.1);

c) Support Vector Machine [SVM] (All-against-one, soft margin = 2, gamma = 2, kernel = linear, percentage of training data = 50, test data = 50).

2) Clustering algorithms:

a) Vector Quantization [VQ] (Learning rate = 0.1, Max Iterations: 50, Tolerance: 0.00001);

b) Fuzzy C-Means Clustering [FCM] (Fuzzifier m = 2, Max iterations: 500, Tolerance = 0.00001);

c) K-means (L2, Max Iterations = 50, Tolerance: 0.0001);

d) K-medians (L1, Max Iterations = 50, Tolerance: 0.0001).

Classification accuracy results are reported in **Table 1**. Using SVM we obtained the best results in terms of classification accuracy: on average 97.2% among all participants evaluated on test data (within interval 96.7% - 98.3%).

Table 1. Classification accuracy rate has been individually computed using several supervised learning and clustering algorithms. The highest accuracy in terms of average is obtained with SVM algorithm.

3.2. EEG Power Spectrum

Several studies support argument that among all cortical areas, it is just the frontal lobes that may significantly express neuronal correlates of emotional responses [30]. The hemispherical activity asymmetries found by us are consistent with the previous observation about the frontal EEG alpha asymmetry and withdrawal model [5]. Regarding the frontal activity asymmetry in the alpha band, Coan found that a response to negative emotions is related to larger frontal right activity [31] (we shall refer as frontal right asymmetry); other studies showed that a larger frontal left activity is related to a response to positive emotions [32] [33].

To investigate this, we calculated the spatial distribution of alpha band power by averaging epochs from all participants that are labeled with each emotion, obtaining four grand-averaged head maps for the four emotion categories, as showed in panel A of **Figure 6**. Considering that the relatively higher brain activity is associated to a relatively lower alpha band activation [34], it is possible to see that the rate of a frontal right asymmetry is larger in categories of negative emotions (sad and disgust).

This frontal right asymmetry became weaker for relatively more positive emotion (neutral) and weakest for a positive emotion (amusement). Ranking of investigated emotions obtained using the ratio of left over right frontal power (**Figure 6(b)**) is consistent with the valence dimension indicated in the circumplex model of affect [35].

Finally, as it is possible to see in panel A of **Figure 6**, we found other EEG power asymmetric differences among other areas (temporal, parietal and occipital lobes). To best of our knowledge, in previous studies such asymmetries were not specifically associated to processes related to emotions.

Table 1. Classification accuracy rates.

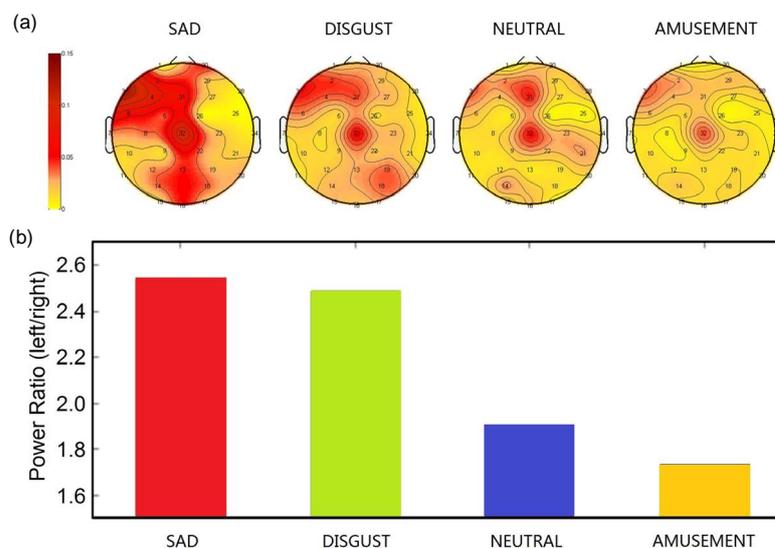| Participants	Supervised Learning Algorithms				Clustering Algorithms		
	SVM	BP	LVQ	FCM	K-Means	K-Medians	VQ
1	96.8	94.8	95	95.2	94.7	91.8	91.7
2	96.9	95.8	95.6	94.9	89.1	87.1	86.8
3	97.8	96.1	95.6	95.5	89	93.7	92.6
4	97.7	96	95.9	95.9	93.7	91.3	95.8
5	97.2	96.1	95.8	95.5	84.3	91.8	84.7
6	96.7	95.7	95.4	95	93.8	87.7	88.9
7	98.3	96.8	96.1	96.1	89.6	89.4	91.7
8	96.9	96	94.8	92.6	91.9	91.8	92.5
9	96.8	95.1	94.6	94	91.9	92.1	91.5
AVG	97.2	95.8	95.4	95	90.9	90.7	90.7

Figure 6. (a) Average of participants' EEG power spectrum among stimuli categories (categories correspond to the four emotions) measured approximately on EEG alpha band (from 8 - 13 Hz). We subtracted the individual EEG power recorded during the baseline (eye-open) before to evaluate participants' average. In frontal areas clear EEG power spectrum differences are evident among all conditions. (b) Ratio of average of participants' EEG alpha power spectrum measured in left frontal electrodes (2, 3 and 4 in panel A corresponding to AF3, F7 and F3) over right frontal electrodes (27, 28 and 29 in panel A corresponding to F4, F8 and AF4) during the four categories of emotional stimuli used in the study.

3.3. Electrodes Reduction

Electrode set reduction is commonly investigated mainly to facilitate a possible online data evaluation even without powerful processors [36]. Many methods were proposed to reduce the electrodes number [37] [38]. Sometimes functional hypotheses may lead to the optimal electrodes placement but ultimately is always the accuracy rate that defines the best pool of electrodes.

Reviewing the literature we found that several emotional studies with promising results employed similar pools of electrodes to classify human emotions [38]-[40]. Several of these studies shared exactly the same pool of electrodes because they relied on EEG data taken with commercial device (such those produced by Emotiv© that uses electrodes located in AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8 and AF4 according to the American Electroencephalographic Society Standard).

Six frontal electrodes indicated in these previously adopted pools of electrodes (AF3, AF4, F3, F4, F7 and F8) were the same to the one suggested from our EEG power results (**Figure 6(a)**). In our study, as previous studies suggested [5] [31] the (asymmetrical) activity in frontal lobes seems to be meaningful in the emotional processes. According to psychological constructionist approaches, among other areas, ventromedial prefrontal cortex (VMPFC) and medial temporal lobe are involved in the emotion representation (probably the representation of pleasantness-unpleasantness), indeed.

Specifically the VMPFC belongs to a functional circuit responsible of a neural representation that modulates the visceromotor control (e.g., autonomic and behavioral responses) as part of value-based representation of an object [41]. Further, VMPFC may provide information used in affective judgments [29]. The medial temporal lobe, associated with VMPFC and orbitofrontal cortex, is part of another functional circuit that entails the threat or reward value of a stimulus [29] by connecting the sensory information about stimulus with the perceived state induced by somatovisceral changes [42].

Assuming that in proximity of the above defined medial temporal lobe we could derive some information related to the emotional representation, we extended the set of six electrodes to eight by adding two temporal electrodes (T7 and T8) as also suggested in previous studies [38] [39]. Combining these findings we obtained a pool of eight electrodes: AF3, AF4, F3, F4, F7, F8, T7 and T8. To evaluate what classification accuracy we can achieve with these eight electrodes instead of the original 32, we replicated our initial classification procedure on the reduced electrode set. Results are reported in **Table 2**. Using 8 electrodes out of 32, the best rate of individual classification is 92.5% and the average classification rate is 87.5%.

Table 2. For each participant we computed the individual classification accuracy rates by means of SVM algorithm using a selected pool of electrodes. The electrodes used were: AF3, AF4, F3, F4, F7, F8, T7 and T8.

3.4. Fantasy Scale and Classification Accuracy

Assuming that the classification accuracy computed individually reflects how well participants' emotional representations were consistent with target emotions, we tested whether there was a correlation between their empathic disposition toward fictional scenes (see 2.6.2.2. for details) and the emotional representation. We

Table 2. Classification accuracy rates for selected pool of electrodes.

Participants	Supervised Learning Algorithms
	SVM
1	90.5
2	89.8
3	81.4
4	90.6
5	86.0
6	84.1
7	84.5
8	92.5
9	88.1
AVG	87.5

expected that a relatively high reactivity to fictional stimuli was associated with a better elicitation of target emotion, resulting in relatively better individual emotional representation and higher classification accuracy. However, correlation analysis between fantasy scale scores and accuracy rates revealed no significant correlation (r = 0.31, p = 0.2).

4. Discussion

In this paper, we reported effective classification of four emotional states using EEG features extracted by means of STFT. We support the idea that EEG can effectively classify conscious emotions (or components of emotion representation). We applied several well-known machine learning algorithms for feature classification. The classification accuracy of 97% that we obtained is quite high and promising for real life application. As we expected, supervised learning algorithms like SVM and BP provides higher accuracy, though at the cost of longer training times. The FCM clustering algorithm provides the best accuracy on average among the clustering methods applied.

Considering the immense signal differences among participants, we took individually all the epochs for each participant and extracted LDA features, resulting in computation of a unique LDA weight matrix. As our analysis was performed offline, extension to online classification of emotions in real time requires new data to be classified correctly and for each new participant it is necessary: first to record some EEG data with emotional stimuli, second to compute STFT and LDA features, and as third step to use the LDA weight matrix for classification of new data from the same participant. We consider this process of extracting LDA weight matrix as a form of "calibration", before the system is ready to classify specific emotions online. We tested this hypothesis splitting available data into training and test data sets, where training data were used as calibration for new data (test data). The resultant high classification accuracy indicates that this method is very promising for online detection of emotions.

At present, the proposed method is not designed to work online and the offline calibration is unavoidable for classification of new data. Nevertheless, such offline solution should be not considered a limit of this study because even online BCI systems require an individual calibration/learning stage; single user has to learn how to produce different patterns of signal (biofeedback) and at the same time the machine (the algorithm) has to learn how to recognize these patterns. This process usually gradually improves by means of trial repetition. In this regard our method is faster and easier to implement, since it does not require the user to actively learn how to perform a specific mind task.

Further, we could obtain such excellent classification results of test data just because we collected EEG data of high quality[1]. In other words, good results can be obtained if data within specific (emotion) category are consistent. In our study, considering that we proposed several controls to assure that the recorded EEG signals were in response to specific emotions, high accuracy results suggest that our data were representative of investigated emotions. Finally, the effectiveness of the protocol used is further established by considering that, in contrast to most of studies that evaluated many complex and sometimes parameters proposed ad hoc, computing just simple power spectrum data we obtained the best accuracy rates so far.

Nevertheless, we want to emphasize that it is nearly impossible to make a comparison among classification studies. Comparison is difficult because, especially in the data acquisition protocol, studies widely differ from each other. One key difference in this study is that while most other studies used relaxation task (such as breathing task) or presented relaxing video/audio between stimuli we used a distractor. In our view, using a distractor task is more effective in neutralizing the emotional state of participants during the experiment. Beside the novelties proposed in the current study, to the best of our knowledge these set of emotions and the actual video-clips have never been used before in any similar extensive EEG study.

Previous studies mostly investigated classifications of basic emotions. From a categorical approach perspective [43] every basic emotion belongs to complete distinct family of emotions. Nevertheless, it is still not clear which emotional framework is most suitable to define emotions: categorical or dimensional approach [44]. Since we chose emotions belonging to different families, we evaluated possible differences in the emotional spaces (dimensional approaches) in order to select emotions defined as well distinct by both emotional frameworks. We investigated different emotional spaces because the number of dimensions that composes the emotional space is relative to the chosen model. We evaluated different dimensional models [27] [34] and [45] to identify the most

---

[1]EEG data collected in this experiment are available for download at: http://bsi-ni.brain.riken.jp/index.html.en

different emotions in terms of spatial distance. It is therefore possible to argue that our results are related to a classification of very different emotional states. Intuitively it is easier to discriminate between two very different emotional states rather than between similar ones.

Display of relative EEG alpha spectral power is consistent with the frontal EEG alpha asymmetry approach and withdrawal model. Classification accuracy of a pool of reduced electrodes, besides indicating the importance of the frontal lobes in the emotional processes, suggests that those electrodes provide the most meaningful information (at least in terms of emotions classification). Consistent with psychological constructionist approaches, location of electrodes selected in the pool suggests that the information provided is relative to some components of emotion representations that are connected with the valence dimension (pleasant-unpleasant). Alpha power asymmetries analysis indicates four distinct degrees of withdrawal (and therefore valence) and, at least from a construct perspective (Russell's Circumplex model of affect), the four emotional states investigated are characterized by different levels of valence (and arousal). Classification accuracies obtained using 32 and 8 electrodes are very encouraging in future employment and development of EEG-based methods for online emotion classification.

Finally, assuming that we measured some components of emotional representation, high accuracy rates may indirectly suggest that participants well perceived and recognized the affective stimuli used. It appears that emotional stimuli were also well assessed as it is supported by results obtained from analysis of self assessment scores (self-assessment classification accuracy rate averagely >90%). These results may support our initial idea that some participants are more suitable for emotional studies. Considering that our participants were not low in emotional intelligence and with high variance in their measures of empathy toward fictional scenes (see 2.6.2.2.), we concluded that for them different empathetic skills were not relevant to define actual differences in emotional processes. In other words in this study the mere participants screening based on emotional intelligence would have been sufficient to grant consistency between emotional states and physiological data.

5. Conclusion

We presented an approach which allows us to considerably improve classification accuracy of human emotions. Results suggest that it is possible to classify some selective human emotions using EEG in a reliable way and that it will be feasible to perform online emotion classification. Our spectral power analysis is consistent with the frontal EEG alpha asymmetry effect and withdrawal model. Electrodes reduction localized the most valuable information to classify emotions in ventromedial prefrontal, frontal cortex and medial temporal lobe. Previous studies indicate such areas are involved in representation of emotional valence (pleasant-unpleasant). This suggests that the four emotion categories we used are distinctly separable at least in the valence dimension. Further investigation regarding classification of emotions with similar perceived valence is needed to test the effectiveness of alpha asymmetry as a potential detector of distinct emotions. We also believe our improved experimental protocol played a key role in achieving very high classification accuracy. Finally, we argued that, since there is very high individual difference in representation of emotions, participants should be analyzed individually. We conclude that future experiments with a large number of participants (e.g., N = 100) are needed to further understand the degree of individual difference, separability of emotion-classes and to finally understand the emotion specific EEG patterns that underlie the mental states. Such study will probably enable us to propose a generic classification model of emotion.

Acknowledgements

References

[1] Rottenberg, J., Gray, R.D. and Gross, J.J. (2007) Emotion Elicitation Using Films. Oxford University Press, London, 9-28.

[2] Janssen, J.H., IJsselsteijn, W.A. and Westerink, J.H. (2013) How Affective Technologies Can Influence Intimate Interactions and Improve Social Connectedness. *International Journal of Human-Computer Studies*, **72**, 33-43.

[3]   Pell, M.D., Monetta, L., Paulmann, S. and Kotz, S.A. (2009) Recognizing Emotions in a Foreign Language. *Journal of Nonverbal Behavior*, **33**, 107-120. http://dx.doi.org/10.1007/s10919-008-0065-7

[4]   Chanel, G., Ansari-Asl, K. and Pun, T. (2007) Valence-Arousal Evaluation Using Physiological Signals in an Emotion Recall Paradigm. *IEEE International Conference on Systems*, *Man and Cybernetics*, Montreal, 7-10 October 2007, 2662-2667.

[5]   Davidson, R.J., Ekman, P., Saron, C.D., Senulis, J.A. and Friesen, W.V. (1990) Approach-Withdrawal and Cerebral Asymmetry: Emotional Expression and Brain Physiology: I. *Journal of Personality and Social Psychology*, **58**, 330-341. http://dx.doi.org/10.1037/0022-3514.58.2.330

[6]   Lindquist, K.A., *et al.* (2012) The Brain Basis of Emotion: A Meta-Analytic Review. *Behavioral and Brain Sciences*, **35**, 121-143. http://dx.doi.org/10.1017/S0140525X11000446

[7]   Hatada, T., Sakata, H. and Kusaka, H. (1980) Psychophysical Analysis of the "Sensation of Reality" Induced by a Visual Wide-Field Display. *SMPTE Journal*, **89**, 560-569. http://dx.doi.org/10.5594/J01582

[8]   Zhao, J. (2006) The Effects of Induced Positive and Negative Emotions on Risky Decision Making. Talk Presented at the 28th Annual Psychological Society of Ireland Student Congress, Maynooth, May 2006, 22 p.

[9]   Estrada, C.A., Isen, A.M. and Young, M.J. (1997) Positive Affect Facilitates Integration of Information and Decreases Anchoring in Reasoning Among Physicians. *Organizational Behavior and Human Decision Processes*, **72**, 117-135. http://dx.doi.org/10.1006/obhd.1997.2734

[10]   Lerner, J.S., Tetlock, P.E. (1999) Accounting for the Effects of Accountability. *Psychological Bulletin*, **125**, 255-275. http://dx.doi.org/10.1037/0033-2909.125.2.255

[11]   Gross, J.J., Sutton, S.K. and Ketelaar, T. (1998) Affective-Reactivity Views. *Personality and Social Psychology Bulletin*, **24**, 279-288. http://dx.doi.org/10.1177/0146167298243005

[12]   Schutte N.S., Malouff, J.M. and Bhullar, N. (2009) The Assessing Emotions Scale. Assessing Emotional Intelligence. Springer, Berlin, 119-134.

[13]   Knez, I. (1995) Effects of Indoor Lighting on Mood and Cognition. *Journal of Environmental Psychology*, **15**, 39-51. http://dx.doi.org/10.1016/0272-4944(95)90013-6

[14]   Anderson, C.A., Deuser, W.E. and De Neve, K.M. (1995) Hot Temperatures, Hostile Affect, Hostile Cognition, and Arousal: Tests of a General Model of Affective Aggression. *Personality and Social Psychology Bulletin*, **21**, 434-448. http://dx.doi.org/10.1177/0146167295215002

[15]   Flynn, J.E. (1977) The Effect of Light Source Color on User Impression and Satisfacti*on. Journal of the Illuminating Engineering Society*, **6**, 167-179.

[16]   Burroughs, H.E., Hansen, I. and Shirley, J. (2011) Managing Indoor Air Quality. Fairmont Press, Lilburn, 149-151.

[17]   Jakobs, E., Manstead, A.S. and Fischer, A.H. (2001) Social Context Effects on Facial Activity in a Negative Emotional Setting. *Emotion*, **1**, 51-69. http://dx.doi.org/10.1037/1528-3542.1.1.51

[18]   Fridlund, A.J., Kenworthy, K.G. and Jaffey, A.K. (1992) Audience Effects in Affective Imagery: Replication and Extension to Dysphoric Imagery. *Journal of Nonverbal Behavior*, **16**, 191-212. http://dx.doi.org/10.1007/BF00988034

[19]   Mehrabian, A. (1996) Pleasure-Arousal-Dominance: A General Framework for Describing and Measuring Individual Differences in Temperament. *Current Psychology*, **14**, 261-292. http://dx.doi.org/10.1007/BF02686918

[20]   Levenson, R.W. (1988) Emotion and the Autonomic Nervous System: A Prospectus for Research on Autonomic Specificity. *Social Psychophysiology and Emotion*: *Theory and Clinical Applications*, Wiley, London, 17-42.

[21]   Philippot, P. (1993) Inducing and Assessing Differentiated Emotion-Feeling States in the Laboratory. *Cognition & Emotion*, **7**, 171-193. http://dx.doi.org/10.1080/02699939308409183

[22]   Gross, J.J. and Levenson, R.W. (1995) Emotion Elicitation Using Films. *Cognition & Emotion*, **9**, 87-108. http://dx.doi.org/10.1080/02699939508408966

[23]   Schaefer, A., Nils, F., Sanchez, X. and Philippot, P. (2010) Assessing the Effectiveness of a Large Database of Emotion-Eliciting Films: A New Tool for Emotion Researchers. *Cognition and Emotion*, **24**, 1153-1172. http://dx.doi.org/10.1080/02699930903274322

[24]   Davis, M.H. (1983) The Effects of Dispositional Empathy on Emotional Reactions and Helping: A Multidimensional Approach. *Journal of Personality*, **51**, 167-184. http://dx.doi.org/10.1111/j.1467-6494.1983.tb00860.x

[25]   Davis, M.H. (1980) A Multidimensional Approach to Individual Differences in Empathy. *JSAS Catalog of Selected Documents in Psychology*, **10**, 85.

[26]   Schutte, N.S., Malouff, J.M., Hall, L.E., Haggerty, D.J., Cooper, J.T., Golden, C.J. and Dornheim, L. (1998) Development and Validation of a Measure of Emotional Intelligence. *Personality and Individual Differences*, **25**, 167-177. http://dx.doi.org/10.1016/S0191-8869(98)00001-4

[27] Fontaine, J.R., Scherer, K.R., Roesch, E.B. and Ellsworth, P.C. (2007) The World of Emotions Is Not Two-Dimensional. *Psychological Science*, **18**, 1050-1057. http://dx.doi.org/10.1111/j.1467-9280.2007.02024.x

[28] Valenzi, S., Islam, T. and Cichocki, A. (2013) Mixed Approach to Emotion Assessment. *Poster Presented at Annual Meeting of the Society for Neuroscience*, San Diego, 12 November 2013, 263. http://www.sfn.org/~/media/SfN/Documents/Annual%20Meeting/FinalProgram/FullAbstractPDFs/AbstractsPDF_Poster_Tues_PM.ashx

[29] Barrett, L.F., Mesquita, B., Ochsner, K.N. and Gross, J. (2007) The Experience of Emotion. *Annual Review of Psychology*, **58**, 373-403. http://dx.doi.org/10.1146/annurev.psych.58.110405.085709

[30] Phillips, M.L., Drevets, W.C., Rauch, S.L. and Lane, R. (2003) Neurobiology of Emotion Perception I: The Neural Basis of Normal Emotion Perception. *Biological Psychiatry*, **54**, 504-514. http://dx.doi.org/10.1016/S0006-3223(03)00168-9

[31] Coan, J.A., Allen, J.J. and Harmon Jones, E. (2001) Voluntary Facial Expression and Hemispheric Asymmetry over the Frontal Cortex. *Psychophysiology*, **38**, 912-925. http://dx.doi.org/10.1111/1469-8986.3860912

[32] Tomarken, A.J., Davidson, R.J. and Henriques, J.B. (1990) Resting Frontal Brain Asymmetry Predicts Affective Responses to Films. *Journal of Personality and Social Psychology*, **59**, 791-801. http://dx.doi.org/10.1037/0022-3514.59.4.791

[33] Wheeler, R.E., Davidson, R.J. and Tomarken, A.J. (1993) Frontal Brain Asymmetry and Emotional Reactivity: A Biological Substrate of Affective Style. *Psychophysiology*, **30**, 82-89. http://dx.doi.org/10.1111/j.1469-8986.1993.tb03207.x

[34] Lindsley, D.B. and Wicke, J. (1974) The Electroencephalogram: Autonomous Electrical Activity in Man and Animals. *Bioelectric Recording Techniques*, **1**, 3-83. http://dx.doi.org/10.1016/B978-0-12-689402-8.50008-0

[35] Russell, J.A. (1980) A Circumplex Model of Affect. *Journal of Personality and Social Psychology*, **39**, 1161-1178. http://dx.doi.org/10.1037/h0077714

[36] Ansari-Asl, K., Chanel, G. and Pun, T. (2007) A Channel Selection Method for EEG Classification in Emotion Assessment Based on Synchronization Likelihood. *EUSIPCO 2007*, 15*th European Signal Processing Conference*, Poznan, 3-7 September 2007, 1241-1245.

[37] Rizon, M., Murugappan, M., Nagarajan, R. and Yaacob, S. (2008) Asymmetric Ratio and FCM Based Salient Channel Selection for Human Emotion Detection Using EEG. *WSEAS Transactions on Signal Processing*, **4**, 596-603.

[38] Khosrowabadi, R., Quek, H.C., Wahab, A. and Ang, K.K. (2010) EEG-Based Emotion Recognition Using Self-Organizing Map for Boundary Detection. 2010 20*th International Conference on Pattern Recognition* (*ICPR*), Istanbul, 23-26 August 2010, 4242-4245.

[39] Khalili, Z. and Moradi, M.H. (2009) Emotion Recognition System Using Brain and Peripheral Signals: Using Correlation Dimension to Improve the Results of EEG. *IJCNN* 2009, *IEEE International Joint Conference on Neural Networks*, Atlanta, 14-19 June 2009, 1571-1575.

[40] Liu, Y., Sourina, O. and Nguyen, M.K. (2010) Real-Time EEG-Based Human Emotion Recognition and Visualization. 2010 *International Conference on Cyberworlds* (*CW*), Singapore, 20-22 October 2010, 262-269.

[41] Koski, L. and Paus, T. (2000) Functional Connectivity of the Anterior Cingulate Cortex within the Human Frontal Lobe: A Brain-Mapping Meta-Analysis. *Executive Control and the Frontal Lobe*: *Current Issues*, Springer, Berlin, 55-65. http://dx.doi.org/10.1007/978-3-642-59794-7_7

[42] Kringelbach, M.L. and Rolls, E.T. (2004) The Functional Neuroanatomy of the Human Orbitofrontal Cortex: Evidence from Neuroimaging and Neuropsychology. *Progress in Neurobiology*, **72**, 341-372. http://dx.doi.org/10.1016/j.pneurobio.2004.03.006

[43] Ekman, P. and Friesen, W.V. (1986) A New Pan-Cultural Facial Expression of Emotion. *Motivation and Emotion*, **10**, 159-168. http://dx.doi.org/10.1007/BF00992253

[44] Barrett, L.F. (1998) Discrete Emotions or Dimensions? The Role of Valence Focus and Arousal Focus. *Cognition & Emotion*, **12**, 579-599. http://dx.doi.org/10.1080/026999398379574

[45] Russell, J.A. and Mehrabian, A. (1977) Evidence for a Three-Factor Theory of Emotions. *Journal of Research in Personality*, **11**, 273-294.

Appendix

We have analyzed a possible relationship between static and dynamic visual features of the video stimuli and their emotional category. The results are summarized in **Figure 3** and **Figure 7**, in which we show that stimuli in all categories cover roughly the same area of the feature space. We show that videos from same emotional category vary in features while in some cases videos from different category have the same characteristics (**Figure 3**). All together we have evaluated 29 features. The first group of features (N = 19) was based on either luminance or color (in two color-spaces RGB and Lab) in different part of the screen. The rest of the features (N = 10) were the mean and standard deviation of inter-frame difference (referred to as "motion energy" here), the mean spatial and temporal frequency, their ratio, and the slopes of amplitude spectra. All luminance measures were highly correlated (**Figure 7**). The **Figure 3** shows the relationship between the mean lightness (L0) and the motion energy. Motion energy was used as the feature with the lowest average correlation with all other features.

Each movie was processed frame by frame in two color spaces, RGB and Lab (L: lightness; a: red-green opponent-dimension; b: yellow-blue opponent dimension). We analyzed the whole frame as well as a square patch with (side of approximately 1deg visual angle) from the center of the screen. In case of spectral features we used the mid-line of the stimulus. Characteristic frequencies $fx0$, $ft0$ were computed as medians of spatial and temporal amplitude spectra. The speed value V was the ratio of the characteristic frequencies $V = ft0/fs0$. The slopes of spectra are computed as the exponent of linear regression to the spectra after transformation into log-log coordinate space.
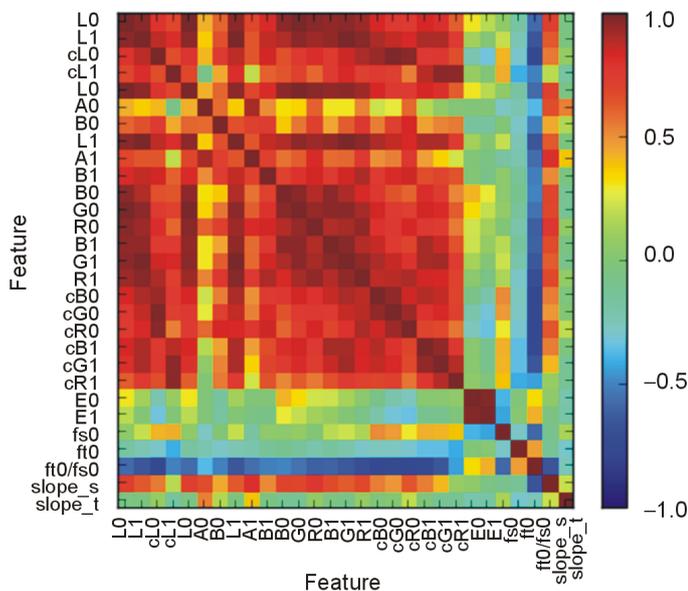

Figure 7. Correlation of features of video stimuli. The figure shows the correlations between different features over time. We chose features that characterize the stimulus' color, lightness, textures (spectral features) and dynamics (temporal spectra and inter-frame difference). In the first group we analyzed color and lightness features: L0, L1, cL0, cL1, L0, A0, B0, L1, A1, B1, B0, G0, R0, B1, G1, R1, cB0, cG0, cR0, cB1, cG1, cR1, the letters LABRGB stand for names of channels of the two color spaces used (Lab and RGB), numbers 0 and 1 stand for mean and standard deviation respectively and presence of letter "c" means that feature was evaluated in the central 1deg patch of the stimulus. E0 and E1 are the mean and standard deviation of inter-frame pixel difference. The last group of measures were computed after frames of video were transformed into the frequency domain, features $fs0$ and $ft0$ are median frequencies of spatial and temporal amplitude spectra, $ft0/fs0$ their ratio and slope_s and slope_t the slopes of linear regression of amplitude spectra in the log-log coordinates.

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or Online Submission Portal.