Extraction of Facial Features as Indicators of Stress and Anxiety

M. Pediaditis, G. Giannakakis, F. Chiarugi, D. Manousos, A. Pampouchidou, E. Christinaki, G. Iatraki, E. Kazantzaki, P. G. Simos, K. Marias and M. Tsiknakis

Abstract— **Stress and anxiety heavily affect the human wellbeing and health. Under chronic stress, the human body and mind suffers by constantly mobilizing all of its resources for defense. Such a stress response can also be caused by anxiety. Moreover, excessive worrying and high anxiety can lead to depression and even suicidal thoughts. The typical tools for assessing these psycho-somatic states are questionnaires, but due to their shortcomings, by being subjective and prone to bias, new more robust methods based on facial expression analysis have emerged. Going beyond the typical detection of 6 basic emotions, this study aims to elaborate a set of facial features for the detection of stress and/or anxiety. It employs multiple methods that target each facial region individually. The features are selected and the classification performance is measured based on a dataset consisting 23 subjects. The results showed that with feature sets of 9 and 10 features an overall accuracy of 73% is reached.**

I. INTRODUCTION

Stress and anxiety are psycho-somatic states being present as a side effect of modern, accelerated life rhythms. Stressors are perceived by the human body as threats, mobilizing all of its resources for defense. Stress is regarded as a factor associated with the risk of acquiring congestive heart failure [1]. Anxiety is considered as a psychological state where someone experiences the unpleasant feelings of worrying, fear and uneasiness. Under certain circumstances, anxiety can be conceived as a mental disorder called generalized anxiety disorder characterized by uncontrollable and irrational worry in common life activities that is disproportionate to the actual extent of worry [2]. The detection of stress and anxiety in its early stages turns to be of great significance, especially if achieved without the excessive use of sensors or other monitoring equipment, which may cause extra stress to the individual.

Initially, psychological stress was considered purely from a psychiatric viewpoint. Various questionnaires were used and are still used for the psychological assessment of an individual. These questionnaires are designed based on various factors like daily hassles, happiness scale, perception of the present and the future by an individual, personality traits, depressive life events etc. [3] Although these questionnaires, with certain scoring methods, can assess the degree of stress involved, there are several drawbacks, such as the subjective bias of an individual towards the assessment of his/her psychological state and the tendency of an individual to maintain secrecy regarding personal matters while avoiding any disclosure even when confidentiality is maintained [4]. These factors pose an upper limit to questionnaire based evaluation of an individual for stress and raise serious queries regarding validation of the estimation. Thus, for a more comprehensive and robust evaluation of a psychological state –mainly the emotions– automatic, non-invasive methods based on the analysis of facial expressions have been reported. Several studies have been published, using facial expressions in the recognition of 6 basic emotions [5], but only few studies report approaches of stress and anxiety recognition. Symptoms of stress can be linked with fluctuations either in physiological (e.g. heart rate, blood pressure, galvanic skin response) or in physical measures, as well as with facial features. In fact, gaze spatial distribution, saccadic eye movement, pupil dilation and blink rate carry information that can distinguish stress levels [6]. Anxiety affects the total psycho-emotional human state triggering both psychological and physical symptoms. It is considered a composite feeling that is affected mainly by fear [7]. Therefore, when individuals are experiencing anxiety, facial signs of fear would be expected. In addition, it is argued that anxiety and fear separation is not reachable by using just psychometric means [8]. Although facial signs of anxiety can be ambiguous and literature is not consistent yet, main effects of anxiety on human face are considered in alteration in eye behavior (blinks, eye opening and eyebrow movements), reddening and lip deformations. Further facial symptoms include strained face, facial pallor, dilated pupils, and eyelid twitching [9].

This paper describes an approach of stress and anxiety assessment using facial signs consisting of the mouth activity, head motion, heart rate, blink rate and eye movements. Section II describes the methods used for the extraction of the features related to these signs, as well as the

*Research supported by the SEMEOTICONS project (FP7-ICT-2013-10-611516) and partially by the PredictES project of the COOPERATION 2011 framework under the NSRF 2007-2013 program of the Greek Ministry of Education.

M. Pediaditis, G. Giannakakis, F. Chiarugi, D. Manousos, A. Pampouchidou, E. Christinaki, G. Iatraki, E. Kazantzaki, P. G. Simos, K. Marias, and M. Tsiknakis are with the Foundation for Research and Technology – Hellas, Institute of Computer Science, Computational BioMedicine Laboratory, Vassilika Vouton, 71110, Heraklion, Crete, Greece (phone: +30 2810 391340; fax: +30 2810 391428; e-mail: mped; ggian; chiarugi; mandim; pampouch; echrist; giatraki; elenikaz; pgsimos; kmarias; tsiknaki@ics.forth.gr).

P. G. Simos is also with University of Crete, Department of Neuro-psychology (e-mail: akis.simos@gmail.com).

M. Tsiknakis is also with the Department of Informatics Engineering, Technological Educational Institute of Crete (e-mail: tsiknaki@ep.teicrete.gr).

selection and acquisition of the dataset, with which the methods were tested. The feature selection and classification results are presented in Section III, followed by a conclusion in Section IV.

II. METHODS

A. Dataset acquisition and video selection

Twenty-three volunteer subjects (sixteen men and seven women) were recoded while watching three clips which aimed to elicit the feelings of anxiety, stress and relaxation. Participants were asked if they have any negative tendency towards heights or closed and small spaces, so that the most effective video could be selected for watching. The selected scenes were not extreme by means of inducing heavy stress, anxiety or even fear. This was done for a) satisfying the guidelines of the study's ethical committee, and b) to follow the aim of the study, which is on early, moderate manifestations of stress and/or anxiety. After watching each video, each participant filled a self-report questionnaire reporting the experienced feeling during the video. This was done by using a rating from 1 to 5, where 1 stands for "Relaxed" and 5 for "Stress or Anxiety". The latter represented a single class since, according to the psychologist's opinion the correct self-assessment of stress and anxiety cannot be taken for granted. Additionally, two psychologists reviewed independently, and in a blind manner, the videos that were collected. For the conflicting cases, a third independent expert psychologist compared the data of the other two annotators.

The selection of the video sequences for the study at hand was performed by accumulating the labels for each class as given from the two experts in conjunction with the subjective rating given by each participant. The selection resulted in 10 videos for the "Relaxed" class and 12 for the "Stress or Anxiety" class. The videos had an average duration of 1 minute at a resolution of 526x696 pixels at 50 fps.

B. ROI detection

The ROI detection is based on the Viola-Jones detection algorithm [10], as implemented in OpenCV 2.4.8 [11], which uses a rejection cascade of boosted classifiers, in this case, decision trees, working with an extended Haar-like feature set. The detection method assumes that only one face exists in the video sequence and that it is in frontal position with only slight out-of-plane rotations. The method subsequently takes the largest among multiple positives to be the actually detected ROI. Within the detected area of a face, a second scan is performed for the detection of the mouth. Assuming again that the largest positive is the correct region of interest, with the additional constraint that the mouth must be located within the lower third of the detected face. If a face is not detected, then a scan for the mouth is performed on the whole image. Outliers in the ROI detection are corrected manually.

C. Head motion estimation

Head motion is measured in terms of horizontal and vertical deviation from a specific reference point. The detected face ROI is used for selecting a sub-region of interest, which is the region between the eyes and mouth and has less facial expressions, thus being the most appropriate for measuring head motion (discarding movements that are related to facial expressions, such as mouth movements, eye blinks etc.). The next step selects specific feature points that are located at the 4 edges of the new sub-ROI and then applies a Kanade-Tomasi-Lucas (KLT) tracker [12] . In addition, a motion filter that discards erratic trajectories and unstable points is applied. For that purpose, the maximum distance traveled by each point between consecutive frames is calculated and points with a distance exceeding the mode of the distribution are discarded. From the filtered point trajectories, the following features are gained: *the mean, median and standard deviation of xDistance; yDistance; xyNorm; xVelocity; yVelocity; Velocity magnitude.*

D. Facial color for heart rate estimation

Heart rate (HR) estimation from video streaming via spatial and temporal analysis was performed implementing a method similar to Poh et al. [13]. This approach exploits the photoplethysmography principle with the use of ambient light in order to detect subtle color changes caused from the amount of reflected light due to change of volume in the facial blood vessels during the cardiac cycle. The method initially detects and tracks the face ROI. The intensity values in this region, for each color channel, in each frame, are spatially averaged to form three raw traces while extreme values are excluded from the mean of the sample data. These traces, after de-trending and normalization, are decomposed into three independent signal components using Joint Approximate Diagonalization of Eigenmatrices (JADE) as elaborated in [14] for the specific task. Afterwards, the spectral flatness is measured and the component with the lowest measure is selected as the most appropriate. Finally, the fast Fourier transform is applied on the selected source signal. The pulse frequency is designated as the frequency that corresponded to the highest power of the spectrum within the operational range to [0.75, 4] Hz (corresponding to [45, 240] bpm). This method obviously returns the following feature: *Heart rate*.

E. Eye-related feature extraction

Features from eyes movements' dynamics used in this study are eye blinks and eye opening. Face region was fitted using Active Appearance Models (AAM) [15] resulting in 68 landmarks placed in predefined facial locations. Points related to the eyes mark out the eyeball perimeter with specific landmarks (6 landmark points used for each eye). Then, the average distance between upper and lower eye points was calculated. This measure determines the eye opening. For the detection of eye blinks, a threshold is established after visual inspection of data to be an appropriate distance below the average distance of points. A parameter of consecutive 5 samples (which corresponds to time interval 100msec) should be below the threshold in order an eye blink to be detected. Summarizing, this method gives the following features: *Blink rate; Eye opening.*

F. Mouth-related feature extraction

The area of the mouth is being analyzed with two different approaches. The first uses dense optical flow for mouth motion estimation and the second is based on tracking Eigen features for mouth opening detection.

The first method uses optical flow, which is a velocity field that transforms one image to the next image in a

sequence. In this work, the velocity vector for each pixel is calculated within each ROI by using *dense* optical flow as described by Farnebäck [16]. It is used to extract the maximum velocity, or magnitude, in two ROIs, the upper lip upper and the lower lip. The upper lip area has a height of 35% of the total mouth ROI height. The remaining 65% is for the lower lip area, while the width is the same for all ROIs. Optical flow is applied only on the Q channel of the YIQ transformed image, since the lips appear brighter in the Q channel [17]. Finally, for each signal 23 features are extracted by using a sliding window of 60 sec in duration. These features are: *Variance of time intervals between adjacent peaks (VTI); ENR*[1]*; Minimum; Maximum; Maximum minus Minimum; Median; Variance; Standard deviation; Root mean square; Interquartile range; Skewness; Kurtosis; Zero crossing rate; Mean crossing rate; Normalized energy; Entropy in energy bins; 25% Spectral power frequency*[2]*; Power (0-3Hz); Power (3-6Hz); Dominant frequency; Entropy in spectral bins; Spectral roll off, Spectral centroid*.

The second method, for detecting mouth openness, is based on Eigen-feature detection and tracking. Eigen-feature points [18] are computed within the mouth ROI and the strongest 100 points are selected for further processing. Based on that set, the following two points are computed:

$$Q_1 = [mean(\boldsymbol{X}), \min(\boldsymbol{Y})] \quad (1)$$

$$Q_2 = [mean(\boldsymbol{X}), \max(\boldsymbol{Y})]$$

Where **X** and **Y** correspond to the coordinate vectors of the points within the set. The baseline mouth openness (MO_0) is computed as the difference of between Q_1 and Q_2 on a user-specific baseline image where the mouth is known to be closed. The total set of points is being tracked with the KLT tracker [12]. For each frame (Q1, Q2) are recomputed resulting to a new mouth openness (MO_n). The mouth status is set to OPEN when the following condition is satisfied:

$$MO_n > MO_0 * OC \quad (2)$$

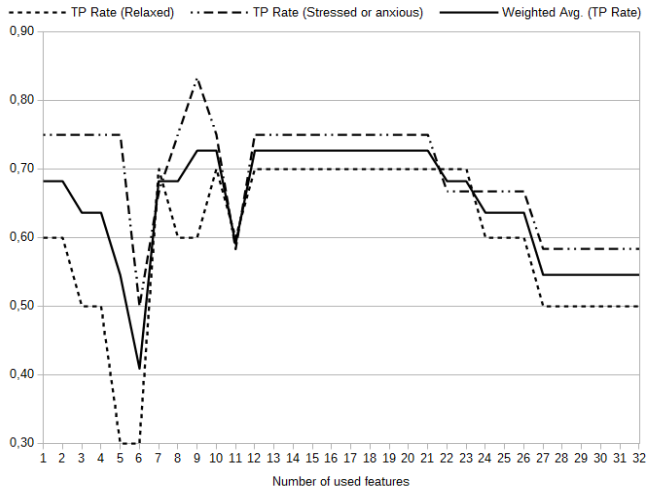
Where OC is the opening coefficient as a percentage of the expansion of the baseline mouth. This method returns the following three features: *Normalized mouth openings; Average mouth openness duration; Average MO intensity*.

III. Results

All features were collated to form a single feature vector consisting of 70 features for 22 instances. The resulting dataset was initially analyzed statistically to extract the most prominent features that can discriminate between the two states. T-tests were performed to compare means between the two groups to see their univariate behavior. In addition, one way ANOVA was used for features extracted from the same facial signs (e.g. upper-lower lip, left-right eye) or from different approaches (mean, median of feature time series) to ascertain if some features do not provide additional information regarding the dataset. The procedure allowed to eliminate features with redundant information, i.e. information included in other features.

In a subsequent step the data mining software Weka v3.7.12 was used for classification experiments. The feature selection was performed through evaluating the worth of a feature by measuring the Pearson's correlation between the feature itself and the class. The features were ranked by their individual evaluations and the all features with a value above 0.25 (32 features) were selected for further classification tests using leave-one-out cross-validation and a multilayer Perceptron artificial neural network (ANN) classifier. Leave-one-out cross-validation was chosen, because it presents the most reliable evaluation metrics in cases with a small number of instances such as the one at hand. Tests involving multiple classifiers showed that the ANN returns the best results in terms of balance between the two classes, while other classifiers (Naïve Bayes, Bayes network, SVM, Decision tree) showed a tendency to high true positive (TP) rates for only one of the two classes. The selected classifier uses backpropagation for training and sigmoid nodes. The number of hidden layers results from the number of features plus the number of classes. The aforementioned setup was repeated multiple times while removing the feature with the lowest rank each time, until only one feature was left. This enables to identify the smallest feature set the represents both classes the best under the given circumstances. The result from this procedure is shown in Figure 1. The overall accuracy (weighted average) reaches its maximum for feature sets 9, 10 and 12–21. The highest TP rate for the "Stressed or anxious" class is achieved with the set of 9 highest ranked features, while for the "Relaxed" class feature sets 7, 10 and 12–23 return the best TP rate. From these cases, the feature sets with 9 and 10 features are the most interesting since they are the smallest sets that return the best overall and well balanced results. The detailed classification results for the two feature sets are shown in Tables I and II.

Figure 1. Classification results with reducing number of features according to their rank.



The first set includes the following features (in descending order by their rank): *Mean crossing rate (max.*

[1] ENR is the energy ratio of the last 75% to the first 25% of the autocorrelation sequence, as a measure for motion manifested as quasi-periodic spikes (randomness).

[2] It is the 25% spectral power frequency defined as the upper bound of the frequency band starting at 0 Hz that contains 25% of the total spectral power. Motion patterns containing isolated sharp spikes generate a broader band while patterns with many (near periodic) spikes produce a narrow.

vel. upper lip), Zero crossing rate (max. vel. upper lip), Mean crossing rate (max. vel. lower lip), Zero crossing rate (max. vel. lower lip), Maximum velocity (upper lip), Median x axis speed (head), Mean eye opening, Spectral centroid (max. vel. lower lip) and Spectral roll off (max. vel. lower lip). The second set additionally includes the *25% Spectral power frequency (max. vel. lower lip)*. It is immediately noticeable that the heart rate is not included in the selected feature set as already revealed by the previous statistical analysis. Blink rate and mouth opening rate are also not included.

TABLE I DETAILED ACCURACY BY CLASS - 9 FEATURES WITH THE HIGHEST RANK

Class	TP Rate	FP Rate	Precision	F-Measure
Relaxed	0.60	0.17	0.75	0.67
Stressed or anxious	0.83	0.40	0.71	0.77
Weighted avg.	0.73	0.29	0.73	0.72

TABLE II DETAILED ACCURACY BY CLASS -10 FEATURES WITH THE HIGHEST RANK

Class	TP Rate	FP Rate	Precision	F-Measure
Relaxed	0.70	0.25	0.70	0.70
Stressed or anxious	0.75	0.30	0.75	0.75
Weighted avg.	0.73	0.28	0.73	0.73

IV. CONCLUSION

This work presents the results of an initial study that aims to find an effective way for detecting early signs of stress and/or anxiety. Several features have been analyzed, which have been calculated by a set of different algorithms, each targeting a specific facial region. The results are promising, showing an overall accuracy or 73%. Some features, such as the eye blink rate or the heart rate, that were expected to play a significant role in the classification process were not employed for the above result. This can be explained by the fact that the study does not take the personal baseline (e.g. heart rate in a relaxed state for each subject individually) into account. It tries to generalize over the dataset population. If the heart rate or the blink rate can be considered as a deviation from a well established personal baseline, additional information for subsequent classification may be gained. A personal baseline could not be calculated due to the limited number of video clips after selection (cf. Section IIAa). Therefore, a larger dataset needs to be investigated that will provide the possibility for deriving a personal baseline, which was not possible with the current dataset, since the overlap of subject IDs in the two classes was very small. A larger dataset will also enable the validation of the current results. In addition, the universality of the features is still questionable since universality has been proven by Ekman [5] only for the six basic emotions. In order to address the above issues, a second campaign has already been planned, with an extended number of participants. It will also allow for an attempt to separately assess stress and anxiety. Finally, the use of a more robust point or ROI tracker is expected to improve the reliability and accuracy of the measured features and consequently of the classifier performance.

ACKNOWLEDGMENT

The authors thank all subjects and project consortium members that participated in the acquisition campaign for the creation of the dataset that was used in this study.

REFERENCES

[1] H. Eriksson, K. Svärdsudd, B. Larsson, L. O. Ohlson, G. Tibblin, L. Welin, L. Wilhelmsen, "Risk factors for heart failure in the general population: The study of men born in 1913," *European Heart Journal,* vol. 10. No. 7, pp. 647–656, 1989.

[2] K. Rowa and M. M. Antony, *Generalized anxiety disorder, Psychopathology: History, Diagnosis, and Empirical Foundations*. p.78, 2008.

[3] S. Joseph, P. A. Linley, J. Harwood, C. A. Lewis and P. McCollam, "Rapid assessment of well-being: The short Depression-Happiness Scale (SDHS)," *Psychology and Psychotherapy-Theory Research and Practice,* vol. 77, pp. 463–478, 2004.

[4] P. Sood, S. Priyadarshini and P. Aich, "Estimation of Psychological Stress in Humans: A Combination of Theory and Practice," *PLoS ONE*, vol. 8, no. 5, e63044, 2013.

[5] P. Ekman, W. V. Friesen and J. C. Hager, *Facial action coding system. The manual on CD-ROM*, 2002.

[6] N. Sharma and T. Gedeon, "Objective measures, sensors and computational techniques for stress recognition and classification: A survey," *Computer Methods and Programs in Biomedicine*, vol. 108, no. 3, pp. 1287─1301, 2012.

[7] J. A. Harrigan and D. M. O'Connell, "How do you look when feeling anxious? Facial displays of anxiety," *Personality and Individual Differences*, vol. 21, no. 2, pp. 205─212, 1996.

[8] A. M. Perkins, S. L. Inchley-Mort, A. D. Pickering, P. J. Corr and A. P. Burgess, "A facial expression for anxiety," *Journal of Personality and Social Psychology*, vol. 102, no. 5, pp. 910─924, 2012.

[9] M. Hamilton, "The assessment of anxiety-states by rating," *British Journal of Medical Psychology*, vol. 32, no. 1, pp. 50─55, 1959.

[10] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001)*, Kauai, HI, USA, 2001.

[11] G. Bradski, "The OpenCV Library | Dr. Dobb's," [Online]. Available: http://www.drdobbs.com/open-source/the-opencv-library/184404319. [Accessed 2015].

[12] C. Tomasi and T. Kanade, *Detection and tracking of point features*. Techincal report, School of Computer Science, Carnegie Mellon Univ. Pittsburgh, 1991.

[13] M. Z. Poh, D.J. McDuff, and R.W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Optics Express*, vol. 18, no. 10, pp. 10762─10774, 2010.

[14] E. Christinaki, G. Giannakakis, F. Chiarugi, M. Pediaditis, G. Iatraki, D. Manousos, K. Marias and M. Tsiknakis, "Comparison of blind source separation algorithms for optical heart rate monitoring," in proc. *EAI 4th International Conference on Wireless Mobile Communication and Healthcare (Mobihealth),* pp. 339─342, 3-5 Nov. 2014.

[15] T. F. Cootes, G. J. Edwards and C. J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681─685, Jun 2001.

[16] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," in *The 13th Scandinavian conference on Image analysis (SCIA'03)*, Göteborg, Sweden, 2003.

[17] N. S. Thejaswi and S. Sengupta, "Lip Localization and Viseme Recognition from Video Sequences," in *National Communications Conference (NCC)*, Mumbai, India, 2008.

[18] J. Shi, J. and C. Tomasi, "Good features to track," *In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'94)*, pp. 593─600, 1994.