Proceedings of the Federated Conference on Computer Science and Information Systems pp. 75–80
DOI: 10.15439/2015F264
ACSIS, Vol. 5

Emotion Monitor - Concept, Construction and Lessons Learned

Agnieszka Landowska
Gdansk University of Technology, Narutowicza St. 11/12, 80-233, Gdansk, Poland
E-mail: nailie@eti.pg.gda.pl

***Abstract*—This paper concerns the design and physical construction of an emotion monitor stand for tracking human emotions in Human-Computer Interaction using multi-modal approach. The concept of the stand using cameras, behavioral analysis tools and a set of physiological sensors such as galvanic skin response, blood-volume pulse, temperature, breath and electromyography is presented and followed by details of Emotion Monitor construction at Gdansk University of Technology. Some experiments are reported that were already held at the stand, providing observations on reliability, accuracy and value the stand might provide in human-systems interaction evaluation. The lessons learned at this particular stand might be interesting for the other researchers aiming at emotion monitoring in human-systems interaction.**

I. Introduction

THIS paper concerns challenges in automatic multimodal affect recognition. Although it seems that the domain is well established and there are many off-the-shelf solutions, the reliability, accuracy and granularity of emotion recognition is still a challenge. In 2013 a project was started at Gdansk University of Technology (GUT) to build an emotion monitor stand that uses existing technologies in order to extend human-systems interaction with emotion recognition and affective intervention. The concept of the stand assumed combining multiple modalities used in emotion recognition in order to improve the accuracy of affect classification. The considered input channels included the ones that are most frequently used in the emotion recognition: physiological signals (skin conductance, respiration, electromyography, EEG, heart rate, peripheral temperature) [1], video input for facial expression analysis [2], keyboard and mouse usage patterns [3] as well as textual inputs for sentiment analysis [4]. The hardware layer of the stand was constructed in 2013, and the software layer in 2014 and 2015, however, the latter still requires extension, including improvement of classification algorithms. The paper describes the concept and construction details of the Emotion Monitor stand at GUT. Selected experiments held at the stand are described that provide an insight into practical aspects of automatic emotion recognition. The experiments are diverse, from the ones that aimed at establishment of reliable measurement procedures of physiological signals, ones that aimed at classification algorithms training and others that were practical applications of the stand in systems design. Although not every experiment was successful, all of them contributed to the knowledge on practical aspects of multimodal emotion monitoring. Lessons learned on the way are the main theme of this paper and the research questions of the article might be formulated as follows: *how to monitor emotional states in Human-Computer Interaction with acceptable reliability, accuracy and granularity and what are the main challenges in automatic multimodal affect recognition?* The main purpose of the paper is to evaluate the usability of the stand as well as to express the main limitations of emotion recognition in human-computer interaction.

This work was supported by Polish-Norwegian Financial Mechanism Small Grant Scheme under the contract no Pol-Nor/209260/108/2015 and by DS Funds of ETI Faculty, Gdansk University of Technology.

II. Related work

Works that are mostly related to this research fall into two categories: research on applicability of emotion monitoring in the context of human-system interaction and studies on emotion recognition based on different input channels.

First group of related papers provides rationale for emotion recognition application in human-systems interaction. Software usability testing can be extended with observation of human emotions [5][6] and it is also possible to measure and optimize software development processes [7][8][9]. Based on the methods for usability evaluation, it would be possible to evaluate educational software and resources designed for e-learning [10][11][12]. Physiological parameters can be also used for optimization of other emotion recognition algorithms and in affect-aware games and other intelligent personalized systems [13].

There are numerous emotion recognition algorithms that differ on input information channels, output labels, affect models and classification methods. As literature on affective computing tools is broad and has already been summarized several times (eg. [14]), only example papers are referenced. The most frequently used emotion recognition methods that might be considered for emotion monitor stand include:

- facial expression analysis (requires video as an input channel, however expressions might be partially controlled by people, especially when they know they are being observed or recorded) [2][14][16];
- audio (voice) signal analysis in terms of modulation (this method is seldom used in human-computer interaction as the voice communication channel is rarely used) [2][16];

- textual input analysis (sentiment analysis requires conversational system interface) [17];
- physiological signals - although very precise and cannot be controlled by most of the people, require specialized equipment [1][15];
- behavioral patterns analysis (keystroke dynamics and mouse usage patterns) combined with other modalities can improve the accuracy of affect recognition; moreover those are the most natural input channels in HCI [3].

The best recognition results are obtained when fusing information from diverse input channels and early and late fusion can be distinguished [18]. Early fusion methods combine features derived from separate input channels to create a common feature vector for classification [19]. Late fusion combines the classification results provided by separate classifiers for every input channel; however, this requires some mapping between emotion representation models used as classifier outputs [20]. The highest accuracies are obtained mainly for two-class classifiers and multimodal input channels (including physiological measurements).

III. Emotion Monitor Concept

The emotion monitor stand objective is to conduct experiments on computer users affective states retrieval and analysis. The stand is equipped with computers, cameras and a set of biosensors, which allow to monitor user activities and record multiple user observation channels at the same time. The data are then processed further to extract features and classify emotional states from single or multimodal input channels. Schema of the emotion monitor stand is provided in Figure 1.

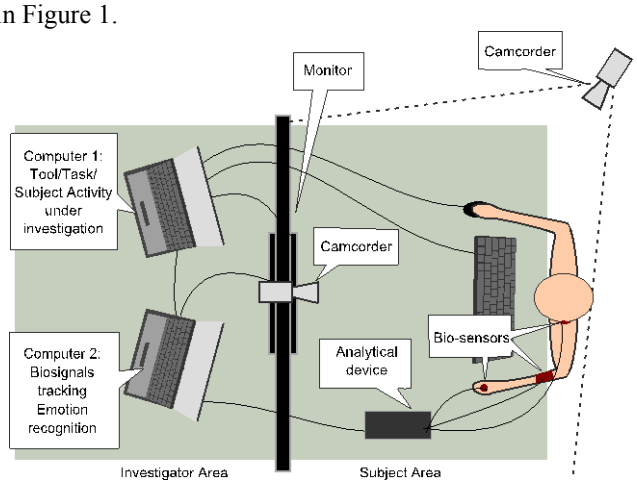

Fig. 1 The concept of emotion monitor stand hardware configuration

The emotion monitor stand is divided into two separated areas: a subject area, where a user performs tasks that are under observation and an investigator area, where a researcher is able to monitor user's activities and track biometric parameters. During investigation, when different emotional states would be evoked on purpose, it would disturb experiments, when a subject would have a possibility to turn to the investigator personally, therefore the investigator and subject areas are separated.

The emotion monitor stand is equipped with the following devices: (1)biometric sensors set, including skin conductance, blood-volume pulse, respiration, temperature, electromyography and EEG sensors, (2) analytical device that allows to simultaneously sample multiple channels with high frequency, (3)front camera that records face and upper part of the subject's body, (4)side camera that records experiment execution, (5)computer 1, which allows the subject to perform tasks under investigation, (6)computer 2, which allows the investigator to monitor user activities and parameters.

Software layer of emotion monitor includes an application to store and track biometric data, tools for observation and recording of video images, keyboard and mouse usage tracker and user activity logger. Apart from the applications recording input channels, the main emotion monitor's application is the one that combines input channels and multiple classifiers in order to provide an affective state estimate. The result might be displayed with diverse visualization tools (general or dedicated for emotion representations).

IV. Emotion Monitor Hardware Layer Construction

In 2013 an emotion monitor stand was constructed at Gdansk University of Technology as a dedicated stand in research laboratory room. The investigator's area and the subject's area are separated with a part-wall made out of furniture, which allows for visual, and partly acoustic separation. Photos of the subject's and the investigator's area are provided in Fig. 2 (left and right respectively).

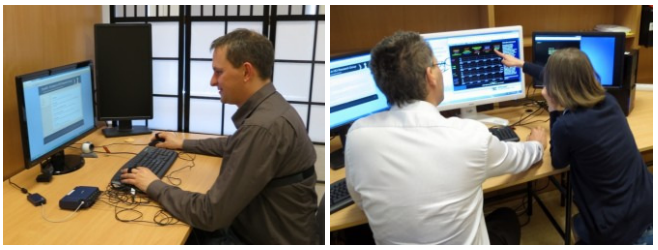

Fig. 2 Emotion monitor stand at GUT (left - the participant's area, right - the investigator's area)

The equipment of the stand was chosen based on capabilities and availability. Coder FlexComp Infiniti by Thought Technology, Canada was chosen as biosensors analytical device. The coder is a ten-channel multimodal device dedicated for real-time biofeedback, psychophysiology training and monitoring. It is connected with computer database via TT-USB device and allows to simultaneously record up to ten channels with sampling rate 2048 samples per second. The coder removes noise from all input channels and performs signal amplification and preliminary filtration that is adjusted to sensor type. EEG

sensors, which are compatible with the coder, allow to perform impedance measurement, which allows to provide EEG signal of high quality. Other available devices had less input channels, lower sampling rates or did not allow to perform impedance measurement.

The biometric sensor set for the emotion monitor stand was designed to measure physiological parameters that are commonly used in affect recognition. and the choice was justified by literature review. Sensors are compatible with the FlexComp Infiniti coder and other coders produced by Thought Technology. Some of the sensors in the predefined set were doubled in order to try out multiple locations at the same time. Detailed list of sensor types includes: skin conductance, electromyography, respiration, temperature, electroencephalography and blood-volume pulse sensors.

Additionally three standard computer sets with two monitors each were provided for the stand. One computer is dedicated for biometric recording and emotion recognition and two monitors allow to display more parameters at the same time. The second computer is provided for a subject to perform tasks under investigation, additional monitor allows an investigator to track user's progress with the tasks. The third computer was added for the investigator-subject communication (investigator displays commands for a subject) – see a black monitor on the left photo in Figure 2.

The stand is normally equipped with three cameras: two in front of the subject and one side camera. The front cameras include one standard RBG camera of medium quality and additionally RGB-D camera with infrared depth sensor that allows for posture analysis independent of illumination for special applications.

V. Emotion monitor software layer structure

There is a number of applications installed at the stand, however not all of them are used simultaneously. The applications might be divided into the following categories: (1) tools for input channels recoding and pre-processing; (2) applications for the data processing into feature vectors and classifiers' training; (3) software for classification and validation of the results (might be the same tools as above); (4) tools for emotional state visualization and interpretation.

The conceptual and actual diagram regarding software layer of the emotion monitor stand is provided in Fig. 3.

A. Data acquisition tools

Thought Technology BioGraph Infiniti application was installed for gathering biometric data from the coder. The system was chosen due to compatibility with the coder, but also due to signal quality optimization features including verifying signal quality and adjusting sensor placement, integrated electrode impedance measurement as well as artifact rejection feature, both automatic and manual. Additionally, an optional Physiology Suite, which is specifically designed for monitoring and assessing physiological functions: recording biomeasurement sessions, reviewing recorded data for the purpose of artifact rejection, generating session reports and demonstrating the results, was installed.

Emotion monitor is also equipped with a set of applications for computer user behavioral observations: Mobii eye tracker, keystroke tracker and mouse tracker, Morae Recorder and Observer, and Logitech Video Capture.

As in some experiments there is a need for simultaneous recording of up to 6 camera images, the change of the software recording video is considered, as it allows to capture one camera image only (at one computer) and in the experiments with multiple cameras is not sufficient.

B. Training, classification and validation tools

This part of the emotion monitor stand is still under

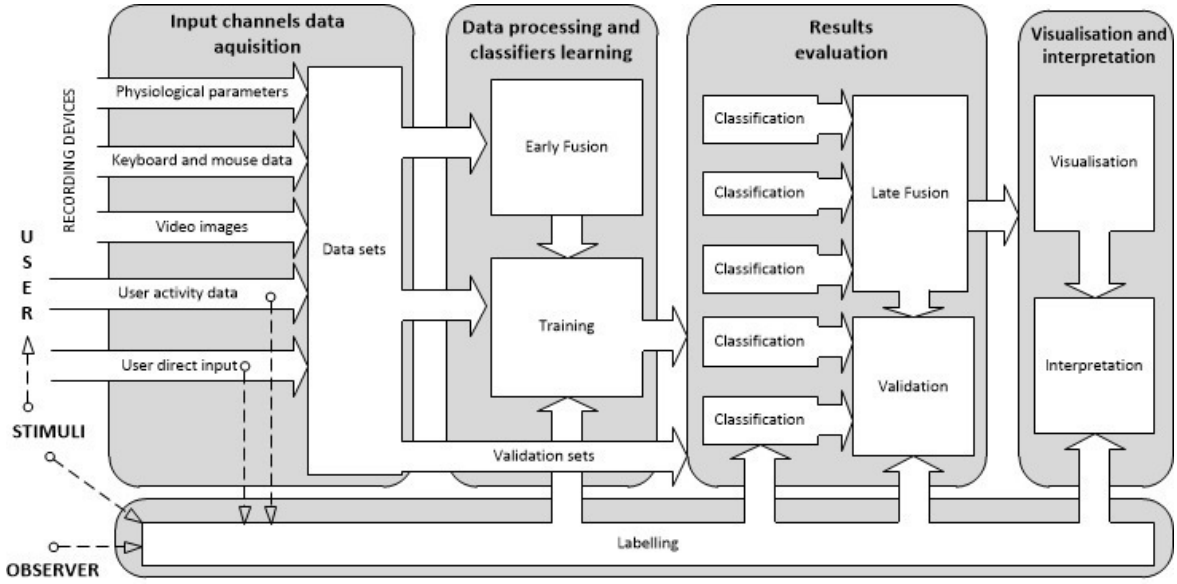

Fig. 3 Emotion monitor stand at GUT - the software layer

development, as there are several challenges in the automatic classification of computer users emotional states. Although the concept was to train classifiers off-line and then use them in real time, this would require more research, than was assumed. Therefore currently, during experiments only data acquisition is performed and the analysis and interpretation is performed afterwards. The analytical and training tools we use include the following: Morae Manager, Knime, Statistica, MatLab, SAS, Origin.

C. Visualization tools

There is an emotional state visualizer that was prepared for the stand at GUT, however now it is not used, as the integration tool for real-time analysis and visualization is under development. Therefore apart from the dedicated tools we use a number of external tools for the data visualization, eg. Knime, Origin.

D. Early and late fusion

The main concept of the stand was to perform the integration of emotional activation information from multiple input channels. One might consider early fusion, in which the data is combined to create common feature vectors. This approach has at least two important drawbacks: timing synchronization and temporary unavailability of input channels.

The first challenge we have encountered is timing synchronization. It seems simple with the timestamps provided by computers, however in practical acquisition of the affective activations the issue lies in the delay of the emotion expression for different channels, eg. heart rate change and skin conductance change last up to one/two seconds, facial expressions are delayed in comparison with the physiological signals. This is the result of the sympathetic and parasympathetic system activation and this is how it works. As a result it is hard to assign the same label to exactly the same period of time for different channels.

Another challenge is temporary unavailability of the input channels. All of the channels used are subject to temporal unavailability: for biosignals movement artifacts must be removed, as they interfere with informative peaks, face recognition is dependent on: face position and illumination conditions eg. if a user moves head a little bit, face recognition tool must follow the face (find it again), moreover keyboard and mouse are usually used interchangeably, with pauses. The common feature vectors are full of blank values, asynchronously for the input channels.

Late fusion that is based on the integration of the recognized emotional states from different classifiers suffers from diverse emotion representation models and lack of mapping between them. Facial expression analysis algorithms for emotion recognition use Ekman's Facial Coding System and provide six basic emotions as a result. The biosignals are best in recognition of the arousal dimension of emotional state, and not the valence (positive and negative experience might cause the same activation of the nervous system). There is a constant challenge of labeling, which will be depicted with the experiment in section VI.

VI. Experiments in emotion monitoring

There were 5 experiments already held at the stand: (1) a study on reliability of physiological signals measurements in the HCI context, (2) experiment with picture stimuli, two experiments (3) and (4) with sound that provided the knowledge required for the optoelectronic system for autistic children, and finally (5) game experience monitoring.

As some of the experiments were already reported [21] [22][23][24], this section would summarize results of experiments (1), (3), (4) and (5) as well as provide more detailed description of experiment (2), which was not reported before. The descriptions would focus on revealing observations on usefulness of the emotion monitor stand.

A. Experiment 1. The challenge of biosignals acquisition in human-computer interaction

After the stand was constructed at GUT in 2013, the first challenge was encountered in the sensitiveness of biometric sensors readings to movements. The typical locations of the sensors are finger tips or finger bases, which is inconvenient while using mouse and/or keyboard in human-computer interaction. Therefore, an experiment aiming at eliminating sensors from hands and finding alternative locations that are as good as typical finger placement was held. As the experiment was already reported in detail [21], only the major findings are summarized. The experiment allowed to draw some conclusions on human-computer interaction monitoring based on bio-measurements of muscle electric activity, respiration, temperature, pulse or skin conductance:

1. Emotion recognition in human-computer interaction should not use EMG measurements placed on trapesius muscle nor sensors located at finger tips (temperature, BVP sensor). Alternative locations of temperature and BVP sensor on earlobes could be accepted as a solution for human-computer interaction monitoring.
2. Skin conductance sensor location on forearm is perceived as less disturbing than location on fingers. Location on forearm is also less sensitive to mouse movements.
3. Respiration sensor are perceived as low disturbing and insensitive to movements, except from body movements.
4. For the signal recorded by all sensors artifacts connected with large movements (body movements) should be removed independently of their location on body.

In emotion recognition algorithms, which are based on biomeasurements, relative values should be provided rather than absolute values, as there are significant differences between individuals. Normalization or standardization procedure should be performed with personal average, instead of overall average calculated for all subjects.

Therefore baseline recording is important for experimental settings, as well as natural environments [21]. The experiment allowed to find a method for reliable acquisition of physiological signals in human-computer interaction.

B. Experiment 2. The challenge of labelling

The concept of the experiment assumed registration of biometric parameters when evoking emotions on the basis of pictures. GAPED (Geneva Affective Picture Database) set was chosen, as the pictures in the set are labeled with the emotional activations in PAD (Pleasure Arousal Dominance) model. The experiment aimed at: acquisition of data for learning algorithms that cause emotions as well as determination whether the readings in alternative locations of sensors vary with emotions.

The pictures were grouped into 6 groups of similar emotional activations for: fear, sadness, anger, disgust and joy. The sixth group represented photos considered neutral for the reference. In each group 5 pictures were chosen and displayed one after another. Picture groups were separated with rest, when no picture was shown. Moreover, before the slide show of photos, a baseline was recorded. Necessity of the baseline recording results from the diversity of individual biometric readings.

The experiment succeeded in determining whether the readings in alternative locations of sensors vary with emotions, however failed in training classifiers for emotion recognition. The resulting recognition rate for the six emotional states never reached more than 30% independently of the classifier, training method and its parameters. Detailed analysis of the results for specific people revealed that pictures were not efficient stimuli to evoke emotions for part of them. There were a number of computer science students involved in the experiment, for whom even the drastic pictures from wars and hospitals were not enough to evoke reactions, as they are used to such views by intensively playing shooter games. For this group of people, the slide show of pictures that lasted for 20 minutes was simply boring and it was visible in physiological signals that headed towards relaxation state, independently of how drastic picture was shown.

Another interesting group of participants reacted to each and every picture independently of what was presented, even a photo of a blue mug or a box caused the reaction (visible with skin conductance fluctuation). This group of participants, which might be described as highly-reactive, also sometimes exhibited skin conductance raise before the picture was actually shown.

The minority of the participants exhibited the expected reactions: high for the drastic photos and low for more neutral ones.

The experiment revealed that choosing a stimuli is an important matter (obviously), but also that stimuli's label is not enough for labeling the data for emotion recognition. In the picture set there was one picture repeated twice in different contexts (in different picture groups) and the reaction to it depended mostly on the context, and not the actual photo (if any reaction was there). As a result we have tried alternative labeling methods, however they were found insufficient in this experiment.

However, the experiment brought to deeper understanding of the challenge of labeling with emotional states and those are the lessons learned:

1) There is a difference between the expression of emotion and the actual emotion eg. one might smile, although feeling embarrassed by the picture.

2) People exhibit more facial expressions when talking with others than in front of the computer screen (eg. typical surprise reaction of "jaw dropping" was never encountered). Micro-expressions must be recognized and interpreted instead.

3) Expectation of a stimuli results in pre-condition reaction (difficult to synchronize it with label).

4) There is no way of actually determine, what the emotional state of a person is (emotion has external expressions, however it is an internal phenomena, there is a two-factor theory of emotional reaction: both stimuli and the interpretation is required for the emotion to appear).

5) Some people do not exhibit facial expressions, for the others, the actual expression varies significantly.

6) When labeling pictures, sounds, videos or other stimuli with questionnaires one might obtain anticipated emotional state instead of the actual one.

7) People differ significantly in the ability to recognize and express their own emotional states.

8) Emotional reactions might be caused by some internal thoughts (some participants exhibited skin conductance fluctuations during baseline and rest recordings).

Although finished with no success, it was a valuable experiment, which revealed lots of research issues to work.

C. Experiments 3 and 4. Practical applications

The other three experiments were practical applications of the stand. Experiments (3) and (4) were conducted to construct an optoelectronic system supporting behavioral therapy of autistic children [22][23]. The first experiment aimed at selection of physiological parameters which are closely correlated with person's emotional state. Measuring changes of those parameters and adequate data processing can show the emotion of investigated person. The experiment used a stimuli of 1 minute sound that started with 1kHz constant sound that was gradually silenced and finished with a shot sound. Only few parameters gave very strong change in measured signal after the shot sound: skin conductance, respiration and electromyography, however the individual reaction varied for subjects. The respiration rate and skin conductance changes were chosen to be monitored in the elaborated system. Another experiment (4) was conducted in order to evaluate the prototype of the

optoelectronic system supporting behavioral therapy of autistic children. Recently, this system has been tested in one kindergarten for children with disabilities. Such support will be very useful and can significantly improve psychological treatment of those children [23].

VII. Results and discussion

Apart from typical challenges in classification: feature selection, choosing classifier and its structure, proper training methods and validation, there are more challenges in automatic emotion recognition. The first one ins reliable data acquisition, as all input channels are subject to some noise and temporal unavailability. The main challenge seem still the labeling with emotional states. One might consider labeling with user reports, expert observations, user activity or stimuli labels, however all of the techniques could be questioned. Perhaps a combination of the two or three different labeling methods is a way, however the problem of blending in the case of constrictions remains. Another challenge is the fusion (early or late) of the emotional expressions estimate from different input channels and this field has gathered some researcher attention so far [19], however still much is to be revealed.

The author is aware of the fact that this study is not free of some limitations. First of all, the report on the experiment is subjective one, as the paper aimed at sharing lessons learned rather than reporting the actual experiments.

The question formulated at the beginning of the paper *how to monitor emotional states in Human-Computer Interaction with acceptable reliability, accuracy and granularity?* remains open and requires further research. However, the second question *what are the main challenges in automatic multimodal affect recognition?* , which was the main purpose of this study, was provided with some answers.

VIII. Conclusion

Although there are some off-the-shelf solution for recognizing human affect (produced by Affectiva or Empathica) as well as many "smart" watches that track physiology, determining the actual emotional state of a human being is still a challenge, even for qualified psychologists. With all the complicated equipment and algorithms the only thing we could track is emotion's external symptoms. Moreover the reliability and the accuracy of the provided estimate depends on many conditions: availability of the input channels, air conditions (temperature, humidity) and the context a human is in. Perhaps the internal phenomena of the emotion is what makes us really unpredictable, i.e. humans.

References

[1] Szwoch W. (2013) Using physiological signals for emotion recognition, In Proc of HSI, Gdańsk, Poland, 556-561.

[2] Zeng Z, Pantic M, Roisman G, Huang T.S (2009) A survey of affect recognition methods: Audio, visual, and spontaneous expressions. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 31(1), 39-58

[3] Kołakowska A. (2013) A review of emotion recognition methods based on keystroke dynamics and mouse movements, In Proc of HSI, Gdańsk, Poland, 548–555.

[4] Cambria, E, Schuller, B, Xia, YQ, Havasi, C (2013) New avenues in opinion mining and sentiment analysis. IEEE Intell Syst 28: pp. 15-21

[5] Partala T., Surakka V. (2004) The effect of affective interventions in human-computer interaction, Interacting with Computers, 16, pp. 295-309

[6] Hazlett R., Benedek J. (2007) Measuring emotional valence to understand the user's experience of software, Int. J. Human-Computer Studies, 65, 306-314.

[7] Zimmermann P., Gomez P., Danuser B., Schar S. (2006) Extending usability: putting affect into the user-experience, In Proc. of Nordic Conf. on Human-Computer Interaction, Oslo, pp 27-32.

[8] Kołakowska A, Landowska A, Szwoch M, Szwoch W, Wrobel M R (2013) Emotion Recognition and its Application in Software Engineering, In Proc of HSI, Gdańsk, Poland, 532–539.

[9] Wróbel M.R. (2013) Emotions in the software development process, In Proc of HSI, Gdańsk, Poland, 518-523.

[10] Binali H, Wu C, Potdar V (2009) A new significant area: Emotion detection in e-learning using opinion mining techniques. In: Proc. of 3rd IEEE International Conference on Digital Ecosystems and Technologies, 2009, 259-264

[11] Landowska A (2013) *Affective computing and affective learning – methods, tools and prospects*, EduAction. Electronic education magazine, 1(5), 16-31

[12] Landowska A. (2013) Affective computing and affective learning – methods, tools and prospects, EduAction. Electronic education magazine, 1, 5, pp. 16—31

[13] Chittaro L., Sioni R. (2014) Affective Computing vs. Affective Placebo: Study of a Biofeedback-Controlled Game for Relaxation Training. International Journal of Human-Computer Studies, 72, 8–9, pp. 663–73. doi:10.1016/j.ijhcs.2014.01.007.

[14] Gunes H., Schuller B. (2013) Categorical and dimensional affect analysis in continuous input: Current trends and future directions, Image and Vision Computing, 31, pp. 120-136

[15] Bailenson J.N., Pontikakis E.D., Mauss I.B., Gross J.J., Jabon M.E, Hutcherson C.A.C., Nass C., John O. (2008) Real-time classification of evoked emotions using facial feature tracking and physiological responses, International journal of human-computer studies, 66(5), 303-317

[16] Picard R, Daily S (2005) Evaluating affective interactions: Alternatives to asking what users feel. In CHI Workshop on Evaluating Affective Interfaces: Innovative Approaches

[17] Ling H.S., Bali R, Salam R.A. (2006) Emotion detection using keywords spotting and semantic network, In Computing & Informatics, IEEE, 1-5

[18] Landowska A, Szwoch W, Szwoch M, (2015) Methodology of Affective Intervention Design for Intelligent Systems, Interactions with Computers (unpublished).

[19] Gunes H. and Piccardi M (2005) Affect Recognition from Face and Body: Early Fusion versus Late Fusion, Proc. IEEE International Conference on Systems, Man and Cybernetics, pp. 3437-3443.

[20] Hupont I; Ballano S; Baldassarri S.; Cerezo, E, (2011) Scalable multimodal fusion for continuous affect sensing, IEEE Workshop on Affective Computational Intelligence (WACI), pp.1,8, 11-15

[21] Landowska A.: Emotion monitoring - verification of physiological characteristics measurement procedures, Metrology and Measurement Systems Journal, Vol XXI, No. 4, 2014, pp. 719-732.

[22] Landowska A, Karpienko K, Wróbel M, Jędrzejewska-Szczerska M (2014) Selection of physiological parameters for optoelectronic system supporting behavioral therapy of autistic children, Proc. SPIE Vol. 9290, Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments.

[23] Jędrzejewska-Szczerska M, Karpienko K, Landowska A (2015), System supporting behavioral therapy for children with autism, Journal of Innovative Optical Health Sciences Vol. 8, No. 3, 1541008

[24] Landowska A, Wrobel M (2015) Affective reactions to playing digital games, Int. conf. on Human-Systems Interaction, Warsaw, Poland, pp. 264-270