Limitations of Emotion Recognition from Facial Expressions in e-Learning Context

Agnieszka Landowska, Grzegorz Brodny and Michal R. Wrobel
Department of Software Engineering, Gdansk University of Technology, Narutowicza Str. 11/12, 80-233 Gdansk, Poland

Keywords: e-Learning, Emotion Recognition, Facial Expression Analysis, Intelligent Tutoring Systems.

Abstract: The paper concerns technology of automatic emotion recognition applied in e-learning environment. During a study of e-learning process the authors applied facial expressions observation via multiple video cameras. Preliminary analysis of the facial expressions using automatic emotion recognition tools revealed several unexpected results, including unavailability of recognition due to face coverage and significant inconsistency between the results obtained from two cameras. The paper presents the experiment on e-learning process and summarizes the observations that constitute limitations of emotion recognition from facial expressions applied in e-learning context. The paper might be of interest to researchers and practitioners who consider automatic emotion recognition as an option in monitoring e-learning processes.

1 INTRODUCTION

There are numerous emotion recognition algorithms that differ on input information channels, output labels or affect representation model and classification method. From the perspective of e-learning applications, the most important classification is based on input channel, as not all channels are available in the target environment. Proposed in the field of Affective Computing algorithms differ on information sources they use (Landowska, 2015b). Therefore some of them have limited availability in e-learning context. Assuming that a learner works in a home environment, more specialized equipment is not available, eliminating e.g. physiological measurements as an observation channel. However it can be expected that a home e-learning environment will be equipped with a mouse, a keyboard, a microphone and a low to medium quality camera. Voice channel is an option for synchronous classes and videoconferences. In asynchronous e-learning observation channels include: monitoring standard input devices usage, facial expression analysis using cameras and scanning of textual inputs for sentiment (for free-text only). Authors of the paper are aware of the synchronous and blended model of e-learning, however this study focuses on asynchronous learning process in home environment.

Authors of the paper designed and conducted an experiment that aimed at monitoring e-learning process using automatic emotion recognition. Facial expression was among the observation channels and we have expected to reveal information on a learner affect from automatic analysis. However, the analysis of the channel led to unexpected results, including unavailability of recognition due to face coverage and significant discrepancy between the results obtained from two cameras. This paper aims at reporting the limitations of emotion recognition from facial expressions applied in e-learning context.

The main research question of the paper is given as follows: *What availability and reliability of emotion recognition might be obtained from facial expression analysis in e-learning home environment?* The criteria for analysis will include availability and reliability of emotion recognition. The quasi-experiment of e-learning process monitoring was performed to spot realistic challenges in automatic emotion recognition. As a result, a number of concerns were identified for affect acquisition applied in e-learning context.

The paper is organized as follows. Section 2 provides previous research we based our study on. Section 3 includes operationalisation of variables and experiment design, while Section 4 – study execution details and results. Section 5 provides

Landowska, A., Brodny, G. and Wrobel, M.
Limitations of Emotion Recognition from Facial Expressions in e-Learning Context.
DOI: 10.5220/0006357903830389
In *Proceedings of the 9th International Conference on Computer Supported Education (CSEDU 2017) - Volume 2*, pages 383-389
ISBN: 978-989-758-240-0

summary of results and some discussion, followed by concluding remarks (Section 6).

2 RELATED WORK

Works that are mostly related to this research are studies on emotion recognition from facial expression analysis.

The most frequently used emotion recognition methods that might be considered in monitoring e-learning include facial expression analysis (Szwoch and Pieniazek, 2015), audio (voice) signal analysis in terms of modulation and textual input analysis (Kolakowska, 2015).

Video input is most commonly used channel for emotion recognition, as it is universal and not disturbing method of user monitoring. Algorithms analyze face muscle movements in order to assess user emotional state based on Facial Action Coding System (FACS) (Sayette et al., 2001). There are many algorithms that differ significantly on the number of features and methods of data extraction, feature selection and classification process. Classifiers are usually build on one of the known artificial intelligence tools and algorithms, including decision trees, neural networks, Bayesian networks, linear discriminate analysis, linear logistic regression, Support Vector Machine, Hidden Markov Models (Kołakowska et al., 2013). Depending on the classification method, input channels and selected features, accuracy of affect recognition differs significantly, rarely achieving more than 90 percent. It is important to emphasis that highest accuracies are obtained mainly for two-class classifiers. As literature on affective computing tools is very broad and has already been summarized several times, for a more extensive bibliography on affective computing methods, one may refer to Zeng et al. (Zeng et al., 2009) or to Gunes and Schuller (2013).

The emotion recognition techniques provide results in diverse models of emotion representation. Facial expression analysis usually provide the results using Ekman’s six basic emotions model extended with neutral state – usually a vector of seven values is provided, each value indicating an intensiveness of: anger, joy, fear, surprise, disgust, sadness, neutral state (Kołakowska et al., 2015).

Emotion recognition from facial expressions is susceptible to illumination conditions and occlusions of the face parts (Landowska, 2015b).

Facial expression analysis has a major drawback – mimics could be to some extent controlled by humans and therefore the recognition results might be intentionally or unintentionally falsified (Landowska and Miler, 2016).

Self-report on emotions, although subjective, is frequently used as a “ground truth” and this approach will be applied in this study. The second approach from the literature is multi-channel observation and consistency check (Bailenson et al., 2008). Another approach is manual tagging by qualified observers or physiological observations, but this approach was not used in this study.

The abovementioned results influenced decisions on the design of this study, especially use of more than one observation channel and improving illumination conditions. Detailed study design is reported in Section 3.

3 RESEARCH METHODOLOGY

In order to verify applicability of emotion recognition in e-learning context a quasi-experiment was conducted. It was based on a typical on-line tutorial in using a software tool extended with monitoring user emotion recognition channels. The concept was to engage observation channels that are available in typical home environment, although the experiment was held at lab setting.

3.1 Experiment Design

The aim of the experiment was to investigate emotional states while learning using video tutorials. Video tutorials, such as published on Youtube, are popular, especially among the younger generation form of gaining knowledge on how to use specific tools, perform construction tasks, and even play games.

The experiment was held at Emotion Monitor stand at Gdansk University of Technology. The stand is a configurable setting allowing to multi-channel observation of a computer user (Landowska, 2015a). The experiment hardware setting consisted of three computers, specialized lighting set and two cameras. Software component included:

- Inkscape as a tool to learn by a participant,
- web browser as a main tool leading a participant (with a dedicated website developed to set tasks and collect questionnaire data),
- Morae Recorder and Observer to record user's actions,
- video recording software that might record two cameras consecutively.

A participant of the study had one computer with one monitor and standard input devices at disposal, the other equipment were used for observation purpose. There were two cameras fronting user face, one located above and one below the monitor, both at monitor center. The cameras were intentionally a standard computer equipment, as usually is available at home desk and medium quality Logitech webcams were used. There was one factor uncommon for home environment: specialized lighting set that allowed to maintain stable and adequate illumination conditions. The set on is a prerequisite of Noldus FaceReader, an emotion recognition tool, to work properly, as defined by the software producer. Recognition rates decease with uneven and inadequate lighting and this condition was explored before, therefore we have designed an experiment rather to observe camera location condition. The experimental setting is visualized in Figure 1.

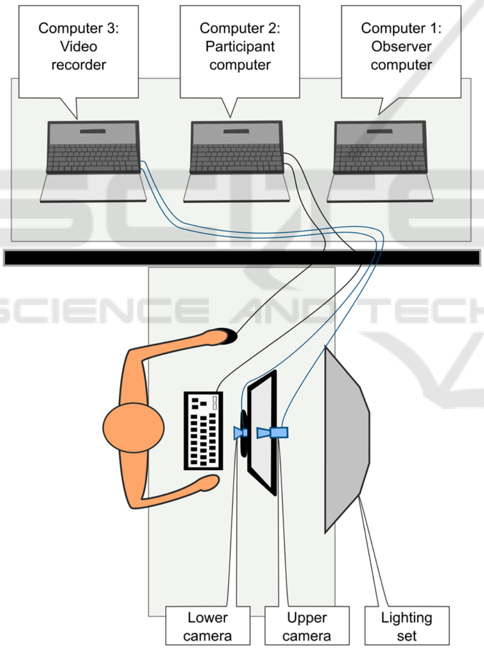

Figure 1: Experimental setting design.

During the study, data were collected from independent channels, which allow to make assumptions on emotional state of user: video, key stroke dynamics, mouse movements and self-report.

The experiment procedure started with an informed consent and followed scenario implemented as consecutive web pages:

(1) Experiment instruction (contained information on experiment procedure and also the Self-Assessment Manikin (SAM) emotional scale description, as was used in the following questionnaires).
(2) Preliminary survey to fill-in, which included questions about age, gender, level of familiarity with the graphical software including Inkscape and assessment of current emotional state (SAM scale).
(3) Tutorial #1.
(4) Post-task questionnaire (SAM scale and descriptive opinions).
(5) Tutorial #2.
(6) Post-task questionnaire (SAM scale and descriptive opinions).
(7) Tutorial #3.
(8) Post-task questionnaire (SAM scale and descriptive opinions).
(9) Final questionnaire summarizing the completed course.

In this manner there were presented and evaluated three consecutive tutorials – the intention was to capture reactions to tasks of diverse difficulty and duration. The first tutorial presented a relatively simple operation in Inkscape (putting a text on a circle path) and lasted for 3 minutes. The second one was the most complicated (a text formatting that imitates carving in a wood) – it was 6:42 minutes long, however users often had to stop and rewind the video in order to perform the task properly. The last tutorial was moderately a complicated (drawing a paper folded in a shape of a plane) and it lasted for 6:32 minutes. While watching a tutorial, the user was meant to perform operations shown in the film. It was not required to achieve the final result in the Inkscape, the user could move to the next stage, when the tutorial video has finished.

3.2 Operationalisation of Variables

The main research question of the paper: *What availability and reliability of emotion recognition might be obtained from facial expression analysis in e-learning home environment?* was decomposed to more detailed metrics that might be retrieved based on experiment results.

Availability factor characterizes, to what extent video observation channel is available throughout time. There are several conditions of unavailability: a face might be not well visible due to partial or total occlusion, relocation of face position due to body movements (camera position usually is set and face might be partially visible, if a learner moves intensively), a face angle towards camera might be too high for an recognition algorithm to work properly. Following metrics were proposed: (1)

percentage of time, when a face was not recognizable at video recording (both overall and per user, denoted UN1); (2) percentage of time, when face was visible, but no emotion is recognizable at video recording (both overall and per user, denoted UN2); (3) percentage of time-based availability of emotion recognition recordings from video (both overall and per user, denoted AV). We assumed that if overall and per-user availability is greater than 90% of time, the conditions for analysis are good, while we expect at least 70% availability (minimum level) per user in order to make any conclusions based on the emotion observations.

Reliability factor indicates, how trustworthy are recognized emotional states – to what extent we might assume, they are the actual emotions of a learner during the process. As there is no way to know the ground truth regarding emotional state, in the experiment we have employed an approach of multi-channel observation and consistency measures to validate the reliability. There were two cameras and the video recordings were analyzed independently (after synchronization). The following metric is proposed: (1) percentage of time when emotion recognition results from the two cameras are consistent – the same dominant emotion is recognized (both overall and per user, denoted REL1); (2) direct difference between recognized states in valence-arousal representation model (both overall and per user, denoted REL2).

For consistency analysis, the un-recognized face and emotion condition frames are excluded. We expect overall and per-user consistency to be greater than 70%, while 50% is the minimal consistency per user in order to make any conclusions based on the emotion observations.

3.3 Data Analysis Methods and Tools

Video recordings were analyzed using Noldus FaceReader software, that recognizes facial expressions based on FACS. The facial expressions are then interpreted as emotional state intensity. The tool provides detailed results as intensiveness vector, containing values (0-1) for: joy, anger, fear, disgust, surprise, sadness and neutral state, or, alternatively it might provide the values of valence and arousal. FaceReader might also provide discrete results – each frame is assigned a dominant emotion as a label. Both result types were analyzed. From the perspective of the emotion recognition from facial expression analysis, the following events would be disturbing: looking around and covering part of the face with a hand. In order to apply automatic face analysis, face position should be frontal to the camera.

If a face is not found on a frame, FIND_FAILED label is returned. If a face was found, but a program was unable to recognize an emotional state a FIT_FAILED label is returned. The error labels are used in this study in calculating availability rates.

Data pre-processing and analysis was performed using Knime analytical platform. Significance tests were performed, whenever necessary – the results are provided in the following sections.

4 EXPERIMENT EXECUTION AND RESULTS

The experiment was held in 2016 and 17 people took part in it. Videos were recorded with 1280x720 resolution and 30 fps frequency. Two video recordings were broken, therefore in this paper we report results based on 15 participants. Among those, 13 were male and 2 female, aged 20 to 21.

From the study execution the following observations should be declared. Participants differed in task execution duration – the shortest study lasted 55 and the longest 103 minutes. Some subjects did not achieve the final result in one or multiple tasks. The participants were not advised on this – the decision of proceeding to another task before previous one was accomplished was up to them.

4.1 Availability

In order to evaluate the quantitative distribution of the availability over time, analysis of data exported from FaceReader emotions recognition software has been performed. Availability metrics UN1, UN2 and AV (for definitions see Section 3.2) were calculated for upper and lower camera independently and for both. The results are provided in Table 1. All means are statistically significant, except for UN1 for upper camera, which was denoted with an asterix. Significance was confirmed by single sample t-test – 95% confidence interval was assumed.

Upper camera was characterized by average 89,7% availability, which is close to threshold defined as good analytical conditions. There were only two participants that had availability below 70% of the recording time.

Table 1: Availability metrics (all means are statistically significant except for one marked with *).

Participant	Upper cam.			Lower cam.			Both cameras		
	UN1	UN2	AV	UN1	UN2	AV	UN1	UN2	AV
P01	0,1	0,4	99,5	0,9	4,1	95,0	0,5	2,3	97,3
P03	0,3	1,8	97,9	1,0	6,1	92,9	0,6	4,0	95,4
P04	1,7	13,8	84,5	2,6	9,5	87,9	2,1	11,7	86,2
P05	2,5	2,7	94,7	26,4	43,0	30,6	14,5	22,9	62,7
P06	4,9	1,4	93,7	0,0	2,3	97,7	2,5	1,8	95,7
P07	0,8	2,0	97,1	1,1	28,6	70,3	1,0	15,3	83,7
P08	0,2	8,6	91,2	0,7	5,0	94,3	0,4	6,8	92,8
P09	30,0	11,3	58,7	0,2	4,7	95,1	15,1	8,0	76,9
P10	0,9	3,8	95,2	2,4	59,9	37,6	1,7	31,8	66,5
P11	0,0	0,0	99,9	0,0	2,1	97,9	0,0	1,0	98,9
P12	0,3	1,8	97,8	0,0	0,0	100,0	0,2	0,9	98,9
P14	19,2	14,6	66,2	38,0	42,3	19,8	28,6	28,4	43,0
P15	0,3	0,8	98,9	6,1	3,9	90,0	3,2	2,4	94,4
P16	0,5	2,1	97,4	21,7	18,1	60,2	11,1	10,1	78,8
P17	1,4	6,4	92,3	0,9	7,5	91,5	1,2	7,0	91,9
Mean (SD)	**5,1 (8,6)***	**5,2 (4,9)**	**89,7 (12,3)**	**7,2 (11,9)**	**14,6 (18,7)**	**78,2 (27,3)**	**6,2 (8,2)**	**9,9 (10,1)**	**83,9 (16,2)**

Lower camera was characterized by average 78,2% availability, which is below the defined threshold, however might be acceptable, as exceeds 70% of time. For the camera, 4 participants had low (under minimal) availability, meaning that in practice they should be excluded from analysis. For two participants availability of emotion recognition through video channel was as low as 20-30 % of time.

In most of the cases, when one camera was highly unavailable, the data from the other one were available, which is an argument for using two. Although there was a difference between average availability of the lower and upper camera, the differences for metrics UN1 and AV are not statistically significant (only difference for UN2 metric is statistically significant), which was confirmed with paired t-test, assuming confidence interval of 95%.

A more detailed analysis of the cases with the lowest availability rates was performed. In the vast majority of cases disturbance was caused by leaning the chin on the hand. For example participant P14 held a hand near the face for more than half of the recording time. Such position is typical for high level of concentration or state of deep thoughts. In art, for example, it is used to represent characters of thinkers and philosophers. Figure 2 shows one of the experiment participant among two most famous sculptures of thinkers, Rodin's Le Penseur, and Michelangelo's Il Penseroso. However, this position may also be associated with fatigue and boredom.

4.2 Reliability

Reliability metrics results are provided in Table 2. Metric REL01 refers to consistency based on labels of dominant emotions and for almost all participants is below a threshold of 50%. For 4 participants the emotion labels are different for more than 90% of time. Such huge discrepancy was the first our observation while analyzing results. More detailed analysis indicate that upper camera tends to overestimate anger (as eyebrows are recorded from upper perspective, they seem more lowered than in zero angle position). The lower camera seems to overestimate surprise, as eyebrows are recorded from lower perspective, they seem more up than in zero angle position). Confusion matrixes based on recognized labels show that also neutral state from one camera is paired with another emotion from the second camera. As label-based consistency was very low, we have decided to analyze consistency of the emotion recognition results in valence-arousal model of emotions. Metric REL02 was calculated for both dimensions and the results are provided in Table 2.

Figure 2: Hand by the face posture while thinking.

The consistency for arousal is high – in 13 out of 15 participants exceeds 90%, only 2 have the consistency above 80%. Valence inconsistency is significantly higher – 90% threshold is exceeded only in one case, while another two are above 80%. For majority of participants the consistency of valence recognition from the two camera location is lower than 50%, and even for one is reported as 0. Difference of valence is statistically significant, which was confirmed by paired t-test with 95% confidence interval.

5 SUMMARY OF RESULTS

The presented study revealed the following results:

- availability of camera recordings in e-learning environment is acceptable;
- upper camera availability is higher than for the location below the monitor;
- when one camera recording is unavailable, recording from the second one is usually available, making an advantage of using two;
- when using two cameras the inconsistency of emotion recognition is relatively high and for majority of the participants below the acceptable threshold;
- lower camera tends to overestimate surprise, while upper one – anger.

All automatic emotion recognition algorithms are susceptible to some disturbances and facial expression analysis is not an exception – suffers from face oval partial cover, location of the camera, insufficient or uneven illumination. When compared to a questionnaire (self-report), all automatic emotion recognition methods are more independent on human will and therefore might be perceived as a more reliable source of information on affective state of a user, however inconsistency rate is alarming.

The study results permit to draw a conclusion that automatic emotion recognition from facial expressions should be applied in e-learning processes tests with caution, perhaps being confirmed by another observation channel.

The authors acknowledge that this study and analysis has some limitations. The main limitations of the study include: limited number of participants and arbitrarily chosen metrics and thresholds. More case studies as well as additional experiments that practically would validate the findings are planned in the future research.

There are also issues that were not addressed and evaluated within this study, i.e. consistency with other emotion recognition channels and perhaps self-report. Those factors require a much deeper experimental project.

6 CONCLUSIONS

There is a lot of evidence that human emotions influence interactions with computers and software products. No doubt that educational processes supported with technologies are under that influence

Table 2: Reliability metrics.

Participant	REL01	Valence						Arousal					
		Upper Cam.		Lower Cam.		Diff	REL02	Upper Cam.		Lower Cam.		Diff	REL02
		Mean (SD)		Mean (SD)				Mean (SD)		Mean (SD)			
P01	**43,5**	0,00	(0,02)	-0,02	(0,03)	0,02	**100,00**	0,25	(0,05)	0,23	(0,06)	0,02	**100,00**
P03	**36,6**	-0,46	(0,21)	-0,18	(0,14)	0,28	**38,46**	0,34	(0,05)	0,30	(0,04)	0,04	**100,00**
P04	**19,9**	-0,33	(0,20)	-0,78	(0,17)	0,45	**11,22**	0,33	(0,08)	0,32	(0,08)	0,01	**100,00**
P05	**17,5**	-0,50	(0,18)	-0,19	(0,15)	0,31	**30,43**	0,27	(0,07)	0,34	(0,05)	0,07	**94,20**
P06	**50,2**	-0,10	(0,14)	-0,13	(0,12)	0,03	**89,47**	0,30	(0,03)	0,23	(0,08)	0,07	**91,23**
P07	**20,9**	-0,86	(0,07)	-0,29	(0,10)	0,57	**1,69**	0,28	(0,05)	0,30	(0,04)	0,02	**98,31**
P08	**7,4**	-0,70	(0,19)	-0,11	(0,16)	0,59	**8,33**	0,36	(0,06)	0,35	(0,07)	0,01	**100,00**
P09	**26,9**	-0,20	(0,14)	-0,53	(0,24)	0,34	**23,81**	0,35	(0,06)	0,36	(0,08)	0,01	**80,95**
P10	**9,6**	-0,52	(0,25)	-0,20	(0,18)	0,31	**28,85**	0,30	(0,09)	0,32	(0,05)	0,01	**92,31**
P11	**5,8**	-0,75	(0,10)	-0,01	(0,01)	0,75	**0,00**	0,29	(0,03)	0,30	(0,03)	0,01	**100,00**
P12	**8,4**	-0,02	(0,03)	-0,08	(0,13)	0,05	**89,47**	0,28	(0,03)	0,24	(0,05)	0,04	**100,00**
P14	**12,4**	-0,56	(0,21)	-0,27	(0,19)	0,29	**35,82**	0,41	(0,0)	0,33	(0,08)	0,08	**85,07**
P15	**17,5**	-0,83	(0,14)	-0,17	(0,12)	0,66	**2,74**	0,28	(0,04)	0,33	(0,05)	0,04	**100,00**
P16	**44,3**	-0,92	(0,09)	-0,50	(0,16)	0,42	**5,95**	0,29	(0,06)	0,34	(0,05)	0,04	**97,62**
P17	**37,0**	-0,65	(0,14)	-0,41	(0,13)	0,24	**33,33**	0,35	(0,04)	0,36	(0,05)	0,01	**100,00**

too. Therefore investigating emotions induced by educational resources and tools is an object of interest of designers, producers, teachers and learners, as well.

This study contributes to identifying practical concerns that should be taken into account when designing e-learning processes monitoring and when interpreting the results of automatic emotion recognition.

ACKNOWLEDGEMENTS

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

REFERENCES

Bailenson, J.N., Pontikakis, E.D., Mauss, I.B., Gross, J.J., Jabon, M.E., Hutcherson, C.A., Nass, C. and John, O., 2008. Real-time classification of evoked emotions using facial feature tracking and physiological responses. *International Journal of Human-Computer Studies*, 66(5), pp.303–317.

Binali, H., Wu, C. and Potdar, V., 2010. Computational approaches for emotion detection in text. In *4th IEEE International Conference on Digital Ecosystems and Technologies*. IEEE, pp. 172–177.

Gunes, H. and Piccardi, M., 2005. Affect Recognition from Face and Body: Early Fusion vs. Late Fusion. In *2005 IEEE International Conference on Systems, Man and Cybernetics*. IEEE, pp. 3437–3443.

Gunes, H. and Schuller, B., Categorical and dimensional affect analysis in continuous input: Current trends and future directions. *Image and Vision Computing*, 31(2), pp.120–136.

Wang, J., Yin, L., Wei, X. and Sun, Y., 2006. 3D Facial Expression Recognition Based on Primitive Surface Feature Distribution. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR'06)*. IEEE, pp. 1399–1406.

El Kaliouby, R. and Robinson, P., 2005. Real-Time Inference of Complex Mental States from Facial Expressions and Head Gestures. In *2004 Conference on Computer Vision and Pattern Recognition Workshop*. IEEE, pp. 154–154.

Kolakowska, A., 2015. Recognizing emotions on the basis of keystroke dynamics. In *Human System Interactions (HSI), 2015 8th International Conference on*. pp. 291–297.

Kołakowska, A., Landowska, A., Szwoch, M., Szwoch, W. and Wróbel, M.R., 2013. Emotion Recognition and its Application in Software Engineering. *6th International Conference on Human System Interaction*, pp.532–539.

Kołakowska, A., Landowska, A., Szwoch, M., Szwoch, W. and Wróbel, M.R., 2015. Modeling emotions for affect-aware applications. In *Information Systems Development and Applications*. Faculty of Management, University of Gdańsk, pp. 55--69.

Landowska, A., 2015a. Emotion monitor-concept, construction and lessons learned. In *Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on*. pp. 75–80.

Landowska, A., 2015b. Towards Emotion Acquisition in IT Usability Evaluation Context. In *Proceedings of the Mulitimedia, Interaction, Design and Innnovation on ZZZ - MIDI '15*. New York, New York, USA: ACM Press, pp. 1–9.

Landowska, A. and Miler, J., 2016. Limitations of Emotion Recognition in Software User Experience Evaluation Context. In *Proceedings of the 2016 Federated Conference on Computer Science and Information Systems*. pp. 1631–1640.

Neviarouskaya, A., Prendinger, H. and Ishizuka, M., 2009. Compositionality Principle in Recognition of Fine-Grained Emotions from Text. In *Proceedings of the Third International ICWSM Conference (2009)*. pp. 278–281.

Picard, R.W. and Daily, S.B., 2005. Evaluating affective interactions: Alternatives to asking what users feel. In *CHI Workshop on Evaluating Affective Interfaces: Innovative Approaches*. pp. 2119–2122.

Sayette, M.A., Cohn, J.F., Wertz, J.M., Perrott, M.A. and Parrott, D.J., 2001. A Psychometric Evaluation of the Facial Action Coding System for Assessing Spontaneous Expression. *Journal of Nonverbal Behavior*, 25(3), pp.167–185.

Szwoch, M. and Pieniążek, P., 2015. Facial emotion recognition using depth data. In *2015 8th International Conference on Human System Interaction (HSI)*. IEEE, pp. 271–277.

Vizer, L.M., Zhou, L. and Sears, A., 2009. Automated stress detection using keystroke and linguistic features: An exploratory study. *International Journal of Human-Computer Studies*, 67(10), pp.870–886.

Zeng, Z., Pantic, M., Roisman, G.I. and Huang, T.S., 2009. A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1), pp.39–58.

Zimmermann, P., Gomez, P., Danuser, B. and Schär, S., 2006. Extending usability: putting affect into the user-experience. *Proceedings of NordiCHI'06*, pp.27–32.