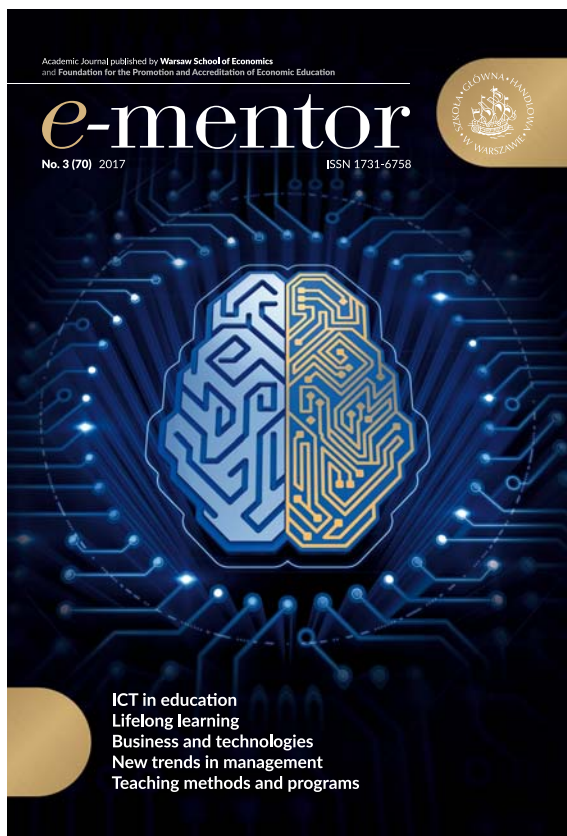


e-mentor

DWUMIESIĘCZNIK SZKOŁY GŁÓWNEJ HANDLOWEJ W WARSZAWIE
WSPÓŁWYDAWCA: FUNDACJA PROMOCJI I AKREDYTACJ KIERUNKÓW EKONOMICZNYCH

2017, nr 3 (70)



Agnieszka Landowska, Grzegorz Brodny, *Investigation of educational processes with affective computing methods*, „e-mentor” 2017, nr 3(70), s. 15–24, <http://dx.doi.org/10.15219/em70.1304>.

Investigation of educational processes with affective computing methods



Agnieszka
Landowska



Grzegorz
Brodny

This paper concerns the monitoring of educational processes with the use of new technologies for the recognition of human emotions. This paper summarizes results from three experiments, aimed at the validation of applying emotion recognition to e-learning. An analysis of the experiments' executions provides an evaluation of the emotion elicitation methods used to monitor learners. The comparison of affect recognition algorithms was based on the criteria of availability, accuracy, robustness to disturbance, and interference with the e-learning process. The lessons learned in these experiments might be of interest to teachers and e-learning tutors, as well as to those researchers who want to use affective computing methods in monitoring educational processes.

Introduction

Humans are emotional, and that characteristic influences all aspects of individual functioning. Although computers are mechanical entities, interaction with them is influenced by the user's affective states. In some computer application domains, the emotionality of users may influence the effectiveness of performed tasks.

One of the domains where emotions might play a crucial role is e-learning. Research on educational processes provided evidence that some emotional states support learning processes, while other suppress them (Hudlicka, 2003, pp. 1–32; Paiva, Dias, Sobral, & Woods, 2004; Picard, 2003, pp. 55–64; Sheng, Zhu-ying, & Wan-xin, 2010, pp. 269–272). In a traditional classroom, teachers address affective issues, such as fatigue, a lack of concentration, low motivation or boredom (Landowska, 2013, pp. 16–31). Human mentors devote at least as much time and attention to emotional goals as they do to the achievement of cognitive goals (Elliott, Lester, & Rickel, 1999, pp. 195–211).

The domain that deals with recognizing user emotional states in human-computer interaction (HCI) is affective computing (Picard, 2003, pp. 55–64). Affective computing has developed methods for recognizing human affect, which all differ on their input information channels, output labels or representation models, and classification methods.

Although a number of emotion recognition algorithms are known, the question of their applicability in monitoring educational processes remains open. This paper aims to investigate this issue and evaluate the applicability of emotion recognition techniques in e-learning process analysis. During the last 2 years, a series of experiments was undertaken at Gdansk University of Technology, aiming to monitor emotions in educational contexts. These experiments involved the observation of users performing e-learning tasks in a laboratory setting. Multiple observation input channels were recorded and analyzed using emotion recognition techniques. We noticed that the availability and value brought by specific observation channels differ a lot, depending on the educational task or even the learner. The methods differ also on compatibility, robustness to noise and subjectivity. In the literature, we found no clues for choosing a particular observation channel for monitoring affect in education. This paper presents a cross-analysis of three experiments, performed from the perspective of the availability of observation channels and the value provided for monitoring affect in education. The paper summarizes the lessons learned. The research question of the paper is formulated as follows: *Which observation channels used in emotion recognition are available, robust and valuable, in the educational context?*

This paper is organized as follows: the *background* section reports on previous research on affective learning and affective methods. The *research methods* section includes the operationalization of variables and a description of the experiments. The *study results* section presents an analysis of the affect recognition methods investigated in these experiments and a comparison of those methods based on defined criteria. The *summary of lessons learned and discussion* section provides a summary of the results and some discussion, followed by the conclusion.

Background

Works that are related to this research fall into three categories:

- (1) studies on emotional states in learning processes,



- (II) research on affective computing methods, including emotion recognition, representation and processing by intelligent software,
- (III) studies on affective learning supported with emotion recognition technologies.

(I) There are multiple studies on how emotional states influence the learning process and many of them precede automatic affect recognition technology. The findings from this literature constitute the background for our analysis and can be formulated as follows:

- emotional states of very high or very low arousal (both positive and negative valence) disturb learning processes (Elliott et al., 1999, pp. 195–211),
- educational processes are supported by states of engagement, concentration and flow (Klein & Picard, 2002, pp. 141–169; Baker, 2007, pp. 1059),
- different emotional states support different learning tasks (Kapoor, Mota, & Picard, 2001, pp. 2–4),
- slightly negative states are better than positive ones (surprisingly, negative states foster critical and analytical thinking) (Baker, 2007, p. 1059; Ben Ammar, Neji, Alimi, & Gouardères, 2010, pp. 3013–3023),
- emotional states with a higher dominance factor support the learning process (moderate anger is better than fear in an educational environment) (Ben Ammar et al., pp. 3013–3023; Hone, 2006, pp. 227–245).

(II) The literature on affective computing tools is very broad and has already been summarized several times, for example by Zeng et al. (Zeng, Pantic, Roisman, & Huang, 2007, p. 126) or by Gunes et al. (Gunes & Schuller, 2013, pp. 120–136). Here the most relevant literature findings related to emotion recognition are summarized below:

- Emotion recognition methods might be based on diverse input channels (and not all of them are available in the target e-learning environment). One can distinguish algorithms based on visual information processing (Zeng et al., 2007, p. 126; Binali, Potdar, & Wu, 2009, p. 259–264; Bailenson et al., 2008, pp. 303–317), algorithms based on body movement analysis (Zeng et al., 2007, p. 126; Boehner, Depaula, Dourish, & Sengers, 2007, pp. 275–291), algorithms based on text analysis (Maria & Zitar, pp. 695–716; Neviarouskaya, Prendinger, & Ishizuka, 2009, pp. 278–281; Binali et al., 2010, pp. 172–177; Ling, Bali, & Salam, 2006; Li & Ren, 2008), algorithms based on voice signal processing (Elliott et al., 1999, pp. 195–211), algorithms based on standard input device usage pattern processing (Kołakowska, 2013), and algorithms based on physiological measurement interpretation (Bailenson et al., 2008, pp. 303–317; Picard & Daily, 2005; Wioleta, 2013, pp. 556–561).

- The emotion recognition techniques provide results in diverse models of emotion representation: a valence-arousal dimensional model (Yik, Russell, & Barrett, 1999, pp. 600–619), the PAD (pleasantness-arousal-dominance) model (Mehrabian, 1996, pp. 261–292) or Ekman's six basic emotions (Scherer & Ekman, 1984) model, extended by neutral state. Some emotion elicitation algorithms provide results as a single label pair (e.g., stress-no stress).
- Emotion recognition algorithms differ significantly in emotion recognition accuracies and granularity. The most accurate results are usually obtained for two-class recognition. Another approach to increase accuracy is using a combination of the observation channels and fusing the results.
- All input channels are susceptible to disturbances and the applicability of emotion recognition might be significantly influenced by that fact. For example, emotion recognition via facial expressions is susceptible to illumination conditions and occlusions of parts of the face (Landowska, 2015, pp. 1–9). Using a combination of observation channels is also an option for addressing this issue.
- All emotion recognition techniques are based on observable symptoms of affective states. The symptoms (e.g., facial expressions) might have a major drawback – they could be to some extent controlled by humans and therefore the recognition results might be intentionally or unintentionally falsified (Landowska & Miler, 2016, pp. 1631–1640).
- Self-reports on emotions, although subjective, are frequently used as a “ground truth”. The second approach from the literature is multi-channel observation and consistency checks (Bailenson et al., 2008, pp. 303–317). A third approach is manual tagging by qualified observers or physiological observations.

These findings influenced the decisions on the design of the experiments described in this study, especially the use of more than one observation channel.

(III) The combination of affective learning and affective computing is not new, as affective computing researchers often apply their methods in the learning process analysis. The most spectacular application scenarios include examples of affective learning companions and affective tutors; however, the affect-awareness mechanisms might be much simpler. Moreover, lots of methodological and didactic findings are achieved by means of technology-enhanced learning monitoring (Landowska, 2013, pp. 16–31). The unquestionable achievement of affective computing is research on the states of frustration and flow and on emotional states in different types of educational tasks (Maria & Zitar, 2007, pp. 695–716). However, the research usually focuses on one perspective or



one emotional state only, e.g., only on frustration or only on boredom (Bessi re, Newhagen, Robinson, & Shneiderman, 2006, pp. 941–961; Scheirer, Fernandez, Klein, & Picard, 2002, pp. 93–118; Ang, Dhillon, Krupski, Shriberg, & Stolcke, 2002, pp. 2037–2040). The concept of concentration or flow states (effective states) as an oscillation between boredom and frustration (ineffective states) is the most popular approach in studying e-learning processes. For example Woolf, Burleson, and Arroyo (2009) proposed a set of useful cognitive-affective terms scales for emotion labeling dedicated to learning processes. The states were additionally assigned with a numeric representation of desirability in educational processes (for example concentration was rated 2 – highly desirable, while boredom was rated 0 – ‘not desirable’) (Woolf et al., 2009, pp. 129–164).

As far as we are aware, there is no systematic review of emotion recognition methods summarizing applicability in monitoring e-learning processes. Moreover, no criteria have been defined to choose among competitive solutions for emotion recognition.

Research method

Within the study of emotion recognition in e-learning processes monitoring, three quasi-experiments were undertaken. The experiments were held at Emotion Monitor stand at Gdansk University of Technology. The stand is a configurable setting allowing for the multi-channel observation of a computer user (Landowska, 2015, pp. 75–80). Based on the experiments’ results, the criteria for evaluation of applicability were selected and analyzed.

Emotion observation channels evaluation criteria

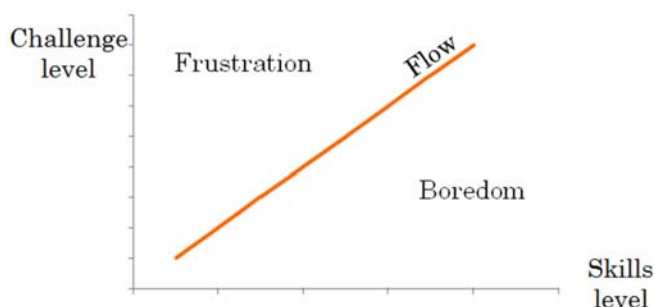
The observation channels used in each experiment were evaluated. The general research question on the availability and value of emotion observation channels was distilled into five criteria, provided below.

1. **Availability** – this represents the degree to which an emotion recognition technique based on a particular observation channel is available in the application context. E-learning can occur in various locations (e.g., at school/university, at home, or outdoors) and with diverse models

(e.g., synchronous, asynchronous). Both the location and model influence the availability of observation channels. Moreover, the availability of the channel might also be task-dependent (that is, in one task a channel might be available and in another it might not, e.g., textual input) or user-dependent (e.g., a user might not have a camera or might switch it off). The availability metric in this study is assigned the following values: AV – channel available, NA – unavailable, TD – task-dependent or UD – user dependent.

2. **Robustness** – this represents the degree to which an observation channel can handle disturbances in the context of e-learning. All channels are susceptible to disturbance; however, some might be more or less likely to be disturbed in an e-learning environment. The robustness metric is rated on a scale of low–medium–high, where a higher score is better.
3. **Compatibility** – this represents a factor of match between the emotional states under investigation and the emotional states provided by the emotion recognition tools. According to related work on affective learning (summarized in background section), the most frequent subject of interest in e-learning is the state of flow as an oscillation between boredom and frustration (see Figure 1). The emotion recognition techniques provide outputs with diverse emotion representation models, which are more or less compatible with the point of interest in e-learning. The compatibility factor is rated on a scale of low–medium–high, where a higher score is better.
4. **Independence on human will** is an important factor, if one wants to use the emotional states as an additional source of knowledge on the subjective internal cognition of systems. The emotion recognition techniques differ significantly in independence (e.g., self-report is the most dependent on human will and is also susceptible to intentional and unintentional falsification). The independence factor is rated on a scale of low–medium–high, where a higher score is better.

Figure 1. Representation of flow as a state between frustration and boredom



Source: Landowska (2016).

5. The **convenience** factor measures the influence of an emotion recognition application on e-learning processes. Environmental changes introduced by applying emotion recognition should not disturb or interfere with learning itself. Some emotion recognition techniques blend naturally into an e-learning environment, while others can significantly disturb a user. The scale for convenience used in this study is low–medium–high, where a higher score is better.

Experiment 1. Learning via playing educational game

The aim of this experiment was to investigate emotional states during learning using an educational game (Landowska & Miler, 2016, pp. 1631–1640). A game on IT project management, called GraPM, was chosen mainly due to the availability of the target group. The target group of the game is computer science students and study participants were recruited from this group. Full randomization of the participant selection was not possible (only volunteers took part in this study), and therefore the study was a quasi-experiment. The participants were asked to play a game several times and both their emotional state and educational outputs were measured. From the perspective of this study, the most important aspect is the type of emotion recognition channel used: facial expressions, self-report, and physiological signals. Details for the channel input data analysis for all three experiments are provided in section *Data processing and analysis methodology*.

Ten participants took part in a study and the recording sessions lasted for 40-90 minutes. The following conclusions regarding the applicability of emotion recognition from the channels to monitoring educational activities were found.

1. The observation channels were susceptible to disturbances. Facial expressions were susceptible to occlusions of the face (e.g., moustache, glasses, hand near the face, or atypical haircuts) and camera location. Physiological signals were disturbed by keyboard typing and other bodily movements.
2. The availability of observation channels for emotion recognition might be user-dependent – some of the participants did not want to have their facial expressions recorded. Moreover, some people draw their face closer to the screen when reading details (possibly due to sight problems), making a face only partially visible in the camera above the monitor.
3. The emotion recognition methods used provided estimates of emotional states, with diverse representation models and diverse accuracy. We decided to use multiple observation channels hoping to fuse them afterwards to achieve more accurate results. However, fusion was not possible due to the diversity in emotion representation models used by the solutions. Facial expression analysis software provided results that almost exclusively fall into Ekman's six

emotions model; the self-report used the most popular questionnaire, the Self-Assessment-Manikin, which provides a Pleasure-Arousal-Dominance model; and physiological signals provided information on arousal only.

4. The emotion elicitation techniques provided contradictory results. We compared the valence from facial expressions (mapped from Ekman's six basic emotions) to the self-report and compared the arousal obtained via physiological parameters with the self-report and found low correlation between all of these. Based on the experimental results, it was hard to determine which estimates were actually correct. Another issue is timing: the self-report is captured sporadically (between tasks), while other measurements are almost continuous (captured several to hundreds of times per second, independently of the task duration).

Experiment 2. Learning with a Moodle course

The aim of this experiment was to investigate emotional states while using a Moodle course with diverse activities. Three Moodle activity types were employed: watching a lecture (slide show with narration), completing a quiz, and adding a forum entry on a pre-defined subject. There were 10 participants recruited from computer science students; the course subject was "IT strategies".

The experiment design was modified according to the observations from the first experiment. One of the main changes was the simultaneous use of four cameras for recording facial expressions. The four cameras were used in order to overcome the challenge of a face moving out of the camera frame. Sentiment analysis of textual input was added as a new observation channel for this study. Self-report (between-task) and physiological measurements were also employed. The following conclusions regarding the applicability of emotion recognition from the channels to monitoring educational activities were found:

1. The availability of some emotion observation channels is task-dependent. For example, sentiment analysis depends on whether or not a task is writing-based. Moreover, the occurrence of facial expressions differs for passive versus active tasks. More mimicking activity is observed in passive tasks (e.g., watching a lecture) than for taking a quiz. According to our observations, the more a person is focused on a task, the fewer facial actions are observed.
2. A sentiment analysis of textual inputs should be based on free-text or opinion-like texts only (the textual inputs on a given subject in this experiment were usually neutral).
3. Physiological signals only provide information on the fluctuation of arousal and not on the valence of an emotional state. The differentiation of boredom and frustration based on physiology is weak – we might observe a participant's arousal rise due to positive emotions.

4. Self-report is the most dependent on human will and should be confirmed via another observation channel. Moreover, our observation is that there is no good timing for self-report: when disrupting the task, it influences the process we are monitoring. When self-report is used between tasks only, we might only get temporal information, which might not reflect the perception of the entire task.
5. The location of the camera is a crucial factor influencing recognized emotional states. There were observable discrepancies between the emotions recognized based on the data from each of the four cameras. As camera location in a learning environment is user-dependent, the observations made on the basis of facial expressions analysis must take that into account.

Experiment 3. Learning with online tutorials

The aim of this final experiment was to investigate emotional states while learning via watching video tutorials. Video tutorials, such as those published on YouTube, are a popular form of gaining knowledge on how to use specific tools, perform certain tasks, or even play games, especially among the younger generation (Landowska, Brodny, & Wróbel, 2017, pp. 383–389). We chose the tool Inkscape for the participants to learn, and three tutorials of varying difficulty were selected for the experiment. There were 23 participants and each session lasted for 40–90 minutes.

During the study, data were collected from the following observation channels: facial expression recordings (two cameras located below and above the monitor), sentiment analysis of opinion-like text, and self-report. Physiological signal measurement was not used. Another new observation channel was included: behavioral characteristics derived from key-stroke dynamics and mouse movement patterns.

The main observations from the study regarding the observation channels include following:

1. The availability of camera recordings in an e-learning environment is acceptable; further, the upper camera (above monitor) availability is better than for the location below the monitor. Moreover, when one camera recording is unavailable, recording from the second one is usually available.
2. When using two cameras, the inconsistency of emotion recognition is relatively high. The lower camera tends to overestimate surprise, while the upper one tends to overestimate anger.
3. Opinion-like textual inputs are appropriate for sentiment analysis; however, opinions are usually close to neutral in experimental settings.
4. The peripheral (mouse, keyboard) usage patterns reveal information on affect of low granularity and accuracy, and should be combined with other observation channels.

Data processing and analysis methodology

The processing and analytical techniques depend mainly on the input channel used for emotion recognition. The data gathered from the experiments included the following:

- facial expressions (videos),
- textual inputs by users,
- self-reported emotional states (before/after task completion),
- behavioral data from peripheral usage,
- physiological data (skin conductance, blood-volume pulse, respiration).

The data were processed by algorithms that automatically recognize emotions based on the input channels. For facial expression analysis, we used three algorithms. Two of the algorithms are off-the-shelf solutions: Noldus FaceReader and QuantumLab Ellen. Another algorithm is a homemade solution based on neural networks and the Luxand library. The solutions provided a time-series of estimated emotional states represented by a vector of basic emotions, where each emotion was assigned a number between 0 and 1, indicating the recognized intensity of the emotion. Some algorithms provide additional data, like the angle of the face towards a camera, or recognized occlusions, while none provides an indication of the recognition confidence.

Textual inputs are analyzed using sentiment analysis algorithms. There are two solutions we analyzed the texts with:

1. a rule-based algorithm, that was designed for opinion-mining on learning activities and
2. a more general algorithm for sentiment analysis that is based on lexical analysis and a dictionary of affect-annotated words.

The first solution only works for opinions about learning activities, while the other solution might be used for mining sentiment in diverse textual inputs, but is also less accurate. The rule-based algorithm provides a set of phrases that were represented in the text, e.g., “interesting subject” or “helpful teacher”. The more general algorithm provides a sentiment value of the sentence or a phrase based on words found in an affect-annotated dictionary; the result is provided in a representation model that the dictionary is annotated with. As textual inputs were provided in Polish, three Polish lexicons were used: NAWL, SentiD and ANPW.

As there is no “ground truth” for a person’s emotional states, self-reported emotional states might be used as labels, and that approach was used in the analysis of the data from these experiments.

Peripheral usage was stored as a series of key-strokes and mouse movements and raw data had to be processed to more general metrics. The metric for using keyboard included the ones based on frequency of keystrokes and time-based metrics (e.g., flight time, delays). Mouse analysis requires a three-step analysis. First, the detailed data on separate movements and key presses is combined to form complex movements like “drag’n’drop”. Then, the metrics are calculated for

discrete time periods, such as velocity or acceleration. Finally, the temporal metrics are analyzed using statistical descriptive metrics (min/max values, averages, standard deviations, etc.).

Raw physiological data recorded by Thought Technology coder were exported as time series files with a pre-defined frequency. The data were gathered using diverse frequency rates (16Hz-2kHz); while exporting them, the highest frequency rates were used (2kHz). The data were re-calculated using Biograph Infinity tool to reveal the heart rate from the blood-volume pulse and respiration rate via skin conductivity. The heart rate, respiration rate and skin conductance were analyzed, using a person-specific baseline in order to find changes in the signals that indicate a change in arousal. In most of the cases, changes in the signal reflects a stimuli, but some peaks are hard to interpret.

All time series derived from diverse input channels were then analyzed using statistical approaches (descriptive statistics, correlation measures, and recognition methods based on machine learning, providing that labels are assignable).

Study results

The emotion recognition input channels were analyzed from the perspective of applicability in user experience evaluations, using the criteria defined in section *Emotion observation channels evaluation criteria*.

Availability

As availability differs depending on the context of the e-learning process, the following contexts were considered:

1. a user learning at home, in an asynchronous mode,
2. a learner taking part in a virtual classroom meeting,
3. learning in a computer lab at school/at university,
4. learning outside, via a mobile device;
5. monitoring e-learning processes in an experimental setting.

The experimental setting is the easiest way to monitor e-learning processes with emotion recognition channels, but, even in this context, the availability of some channels is dependent on the performed task. The availability of the channels in each context is summarized in Table 1. The last column contains the availability in an experimental setting, while the other columns contain an anticipated availability in diverse environments where e-learning might occur: at a home desk (asynchronous or synchronous learning), in a computer lab, or outside, using mobile devices (a ubiquitous learning model).

Self-reporting is always available, as a questionnaire on affect can be added anywhere within the learning path. The technique has drawbacks, which are discussed below. In any context, some of the techniques might be temporarily unavailable as they depend on which task is being currently performed. Most asynchronous learning activities happen in silence; therefore, speech analysis might be a difficult option to use. When performing passive tasks like listening or watching, sentiment analysis and peripherals usage patterns are also not available. There are also input channels, for which availability is user-dependent, i.e., facial expression analysis or speech prosody. A camera or microphone might be not available at the learner's home desk, their network throughput might be too limited to carry video signal, and the user has the right and ability to switch the camera/microphone off. Tracking of the peripheral devices should be, for ethical reasons, performed with informed consent, which implies unavailability in the case when the user does not give consent. There are studies on the perceived disturbance of observation channels and the level of disturbance depends significantly on the context of use, with higher disturbance levels found in home environments (Landowska & Brodny, 2017, pp. 26–41).

Physiological measurements are usually available in the controlled contexts of experimental settings. However, technological progress might solve this problem, as smart devices measuring physiological signals like heart rate or skin conductance are currently being developed and introduced into the market.

Table 1. Availability of emotion monitoring input channels in e-learning contexts

Emotion elicitation channel	Context				
	Home desk, asynchronous model	Home desk, virtual classroom	Computer laboratory	Mobile learning	Experimental setting
Self-report	AV	AV	AV	AV	AV
Facial expression analysis	UD	AV	AV	UD	AV
Peripherals usage patterns	TD, UD	TD, UD	TD	TD, UD	TD
Prosody of speech*	TD, UD	AV	TD	TD	TD
Sentiment analysis	TD	TD	TD	UN	TD
Physiological measurements	UN	UN	UN**	UN**	AV

(AV – available, TD – task-dependent availability, UD – user-dependent availability, UN – unavailable, * – channel not used in the experimental setting, ** – availability might change with upcoming technologies)

Source: own study.

In the experiments performed and reported in this study, the following emotion elicitation techniques were used: self-report, facial expression analysis, peripheral usage patterns, sentiment analysis and physiological measurements. Speech prosody was not analyzed, as there was no task where voice communication was employed.

While considering emotion elicitation techniques for a specific e-learning context, the availability of the observation channel is the first factor to consider. For example, in a design of the online educational game GraPM (Landowska & Miler, 2016, pp. 1631–1640), the only available emotion observation channel was the facial expression analysis. If more channels are available, other criteria may be considered.

Robustness, compatibility, independence and convenience

All emotion elicitation channels are susceptible to some disturbance that limits the practical applicability in an e-learning context, even though a channel is available. Applicability was distilled into robustness, compatibility, independence and convenience criteria, as defined in section *Emotion observation channels evaluation criteria*. A summary of the techniques comparison is provided in Table 2, while justification for the scores, based on observations from the experiments, follows.

Robustness to disturbance is high for self-report only – there are few sociological phenomena that might influence self-reported emotional states; however, for most e-learning contexts, the self-report is robust against disturbance. For the video channel, robustness is relatively low, as the channel suffers from insufficient or uneven illumination conditions and face occlusions (facial hair, atypical haircuts, glasses, or even a hand near the face), which are typical conditions for a user learning at home. In the first experiment, emotion recognition from facial expressions was available on average 77% of the time (measured by the number of frames for which emotion recognition was available against the total number of frames), with variations among users from 42% to 100% availability. In the second experiment, the average availability reached 89% of the time, with variation among participants from 59% to 99% of the time. In

the third experiment, the availability was higher with an average of 84% of the time and variability of 43–98% (depending on the user). The availability was enhanced with the use of two cameras in the third experiment and four cameras in the second experiment.

Peripheral usage patterns, which were used as an emotion elicitation technique in the third experiment, were significantly task-dependent. They might be used for comparing emotional states within the same task; however, comparison among tasks is difficult. As both mouse movements and keystrokes are analyzed, the recognition results suffer from the temporal unavailability of one of the input types. In the first experiment, which was based on a game, keyboard data were unavailable, as users mainly used the mouse for controlling the game. In the second experiment, which was based on a Moodle course, the passive activities (watching and listening) did not provide any keyboard/mouse entries, whereas the active parts of the course mainly provided mouse patterns. Due to the missing keyboard entries, the analysis of peripheral usage patterns was limited. Moreover, another complication might be imposed, while using this type of analysis at a learner's home desk – the patterns are dependent on keyboard and mouse types, and therefore comparability among participants might be limited.

Sentiment analysis is available for text-based tasks. However, only free-text inputs should be analyzed, as topical answers rarely reveal sentiment. In the third experiment, the user was periodically asked to input an opinion into a text field of a certain length and the results were inconclusive. Spontaneous entries, for example in online forums, might reveal more information on sentiment.

Physiological measurements, which were used in the first and second experiments, are easily disturbed, as most of the sensors are located on the fingers, sometimes even on the finger tips. In an e-learning context, the learner is using a computer or another device, which operated by hand; therefore, typing and hand gestures introduce significant disturbance to measuring physiological signals.

Compatibility represents a factor of match between the emotional states under investigation and the emotional states usually provided by emotion recognition tools. If state of flow is to be detected, no emotion elici-

Table 2. Emotion monitoring input channels in e-learning contexts – comparison (high–medium–low scale was used with the higher value indicating better applicability in an e-learning context)

Emotion exploration technique	Robustness to disturbances	Compatibility with e-learning	Independence on human will	Convenience in e-learning
Self-report	High	Medium	Low	Medium
Facial expression analysis	Low	Low	Medium	High
Peripherals usage patterns	Medium	Medium	High	High
Speech prosody*	Low	High	High	Medium
Sentiment analysis	Medium	Medium	Low	High
Physiological measurements	Low	High	High	Low

Source: own study.

tation technique can directly provide such an output. Some of the input channels might reveal information about arousal (physiology, prosody), which might be a good differentiator between boredom and frustration. Some information on arousal might be derived from peripheral usage pattern and sentiment analysis. The boredom and frustration states might be also asked directly in a self-report. Well-established questionnaires for emotion retrieval (e.g., the Self-Assessment Manikin [Bailenson et al., 2008, pp. 303–317]) or adjective lists (Strapparava & Valitutti, 2004, pp. 1083–1086) are only partially transferrable to a boredom-frustration model. Emotion recognition from facial expression analysis is usually based on the FACS model and provides output as a combination of six basic emotions (joy, anger, disgust, sadness, surprise, and fear) and a neutral state. The emotion representation model is incompatible with the one required in e-learning, as it is hard to represent boredom with Ekman's six basic emotions. In the experiments, facial expression analysis was used, but a mapping to valence-arousal model had to be applied, adding an additional factor of mapping accuracy to overall uncertainty about the recognized emotional state.

Independence on human will is another factor that should be taken into account when choosing an emotion elicitation technique. People behave differently when they know they are being observed, and that fact might influence the emotion elicitation results. The channels that are highly controllable by humans include self-reported states and textual inputs. Facial expressions are to some extent controlled and might be intentionally falsified. The channels that are hard to control by an individual include physiological measurements, prosody, and peripheral usage patterns. If an emotion elicitation channel is susceptible to intentional or unintentional falsification, an advisable approach, which was applied in the experiments, is to use more than one observation channel, including the ones that are less dependent on human will.

Last but not least, the **convenience** factor measures the influence of emotion recognition application on e-learning processes. Some emotion recognition techniques blend naturally into an e-learning environment, especially when using a computer setting: peripheral usage patterns, facial expression analysis and sentiment analysis are among the most convenient approaches. A word of warning must be given about the self-report, which is only partially natural and might interrupt a state of flow, if timing is not carefully considered. Imagine any system asking for an emotional state every minute – the focus and concentration on an educational task would be constantly interrupted. On the other hand, if an emotional state is reported in large time intervals (e.g., once every few hours), it might not reflect the emotions occurring during the entire period, just last (best remembered) time. In our experiments, emotions were reported after each task, and the interval depended on task duration as well as user efficiency. As a result, the time interval between consecutive self-reports varied from 5 minutes to over an hour.

This limitation is hard to overcome in an e-learning context and priority should be given to educational task efficiency over emotion elicitation precision.

For virtual classrooms (synchronous learning using an audio channel), prosody analysis might be convenient. Physiological measurements are usually incompatible with the e-learning process of using a computer, as sensors might disturb the user while they operate a keyboard, mouse or touch screen.

Summary of lessons learned and discussion

The results of the experiment analysis are summarized with the following statements:

- Multiple observation channels can be used in monitoring e-learning processes: self-reported states, facial expression analysis, peripheral usage patterns, prosody, sentiment analysis, and physiological measurements; the main criteria remains availability in a specific setting.
- There is no best method for emotion elicitation in an e-learning context. The channels that are available have other drawbacks (e.g., limited robustness to disturbance, or low compatibility of the output model of emotions).
- The availability of input channels that can be used in emotion recognition is often limited and also task- and user-dependent.
- The choice of observation channels for monitoring e-learning processes should be based on following criteria: tasks to be performed, setting for observation (available equipment), required additional consent for affect recognition, and robustness to disturbance.
- The self-report, although always available, is the most subjective channel and should be supported with another observation channel for credibility.
- Priority should be given to the efficiency of the e-learning process and not to monitoring activities. Therefore, the observation should be as unobtrusive as possible, making convenience an important matter to consider.

Based on our experience, it is possible to monitor educational processes with affective computing methods. This study revealed the advantages and disadvantages of emotion observation channels from the perspective of their usefulness in monitoring educational processes. A practical implication of the study is that the selection of observation channels should be tailored to what is needed (what processes are monitored, what emotions are fostering or suppressive), what channels are available (whether the availability depends on the user and task) and whether observation will affect the observed process (this factor is often overlooked in planning). An analysis of the literature shows that a single channel of observation is usually used, without much reflection on its limitations. Having a tool for analyzing emotions from facial expressions does not mean that the data received from



it will be valuable for the purpose of the study. While applying automatic emotion recognition in e-learning, one must be aware of the limitations of the methods that are currently available.

We are aware of the fact that the presented study has some drawbacks. The experiments are not reported in detail, as two of them were previously reported (Landowska & Miler, 2016; Landowska, Brodny, & Wróbel, 2017), although in different context. This paper summarizes the lessons learned while applying automatic emotion recognition to monitoring e-learning processes, rather than reporting detailed data. The assignment of labels related to certain criteria (apart from robustness) is subjective and based on our experience. Moreover, not all observation channels were applied in all three experiments – e.g., prosody was not used, as there was no speaking activity within the educational process being monitored. The channel has been included in the analysis, as it might be used; however, the criteria for that channel were evaluated by anticipation rather than experimental results.

Conclusion

Teachers know well that emotions are a crucial component of successful education. There are learners who, despite high intelligence and potential, fail in accomplishing their learning activities. Goleman (2006) proposed the term 'emotional intelligence' to name a phenomena of affect influence on human effectiveness in life. Self-sustainability and motivation play a crucial role in accomplishing e-learning activities. The presented experiments allowed us to apply multiple observation channels in monitoring affect during e-learning activities. An affective analysis of a learning process might be used in observing the quality of educational design in systems and resources. Moreover, online analysis, provided to an online tutor might help him or her to address the fluctuation of motivation and concentration in learners. The emotion recognition techniques are applicable to e-learning; however one should be aware of the limitations of these techniques. We are planning more experiments to monitor synchronous activities.

Acknowledgements

The authors thank Dominika Makowiecka and Michał Wróbel, who helped in conducting the experiments. This study is supported in part by the 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.

References

Ang, J., Dhillon, R., Krupski, A., Shriberg, E., & Stolcke, A. (2002). Prosody-based automatic detection of annoyance and frustration in human-computer dialog. *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP 2002)*, 2037–2040.

Bailenson, J.N., Pontikakis, E.D., Mauss, I.B., Gross, J.J., Jabon, M.E., Hutcherson, C.A.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. <http://dx.doi.org/10.1016/j.ijhcs.2007.10.011>

Baker, R.S.J. d. (2007). Modeling and understanding students' off-task behavior in intelligent tutoring systems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1059.

Ben Ammar, M., Neji, M., Alimi, A.M., & Gouardères, G. (2010). The affective tutoring system. *Expert Systems with Applications*, 37(4), 3013–3023. <http://dx.doi.org/10.1016/j.eswa.2009.09.031>

Bessière, K., Newhagen, J.E., Robinson, J.P., & Shneiderman, B. (2006). A model for computer frustration: The role of instrumental and dispositional factors on incident, session, and post-session frustration and mood. *Computers in Human Behavior*, 22(6), 941–961.

Binali, H., Wu, C., & Potdar, V. (2009). A new significant area: Emotion detection in e-learning using opinion mining techniques. *3rd IEEE International Conference on Digital Ecosystems and Technologies*, 259–264. <http://dx.doi.org/10.1109/DEST.2009.5276726>

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

Boehner, K., Depaula, R., Dourish, P., & Sengers, P. (2007). How emotion is made and measured. *International Journal of Human-Computer Studies*, 65(4), 275–291. <http://dx.doi.org/10.1016/j.ijhcs.2006.11.016>

Elliott, C., Rickel, J., & Lester, J.C. (1999). Lifelike pedagogical agents and affective computing: An exploratory synthesis. *Lecture notes on artificial intelligence. Artificial Intelligence Today*, 1600, 195–211.

Goleman, D. (2006). *Emotional intelligence*. USA: Bantam.

Gunes, H., & Schuller, B. (2013). Categorical and dimensional affect analysis in continuous input: Current trends and future directions. *Image and Vision Computing*, 31(2), 120–136. <http://dx.doi.org/10.1016/j.imavis.2012.06.016>

Hone, K. (2006). Empathic agents to reduce user frustration: The effects of varying agent characteristics. *Interacting with Computers* 18(2), 227–245. <http://dx.doi.org/10.1016/j.intcom.2005.05.003>

Hudlicka, E. (2003). To feel or not to feel: The role of affect in human-computer interaction. *The International Journal of Human-Computer Studies*, 59(1–2), 1–32. [http://dx.doi.org/10.1016/S1071-5819\(03\)00047-8](http://dx.doi.org/10.1016/S1071-5819(03)00047-8)

Kapoor, A., Mota, S., & Picard, R.W. (2001). Towards a learning companion that recognizes affect. *Association for Advancement of Artificial Intelligence Fall Symposium*, 543, 2–4.

Kołakowska, A. (2013). A review of emotion recognition methods based on keystroke dynamics and mouse movements. *6th International Conference on Human System Interactions*. <http://dx.doi.org/10.1109/HSI.2013.6577879>

Landowska, A. (2013). Affective computing and affective learning – methods, tools and prospects. *EduAkcja. Magazyn edukacji elektronicznej*, 1(5), 16–31.

Landowska, A. (2015a). Emotion monitor – Concept, construction and lessons learned. *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, FedCSIS 2015*, 75–80.



Landowska, A. (2015b). Towards emotion acquisition in IT usability evaluation context. *Proceedings of the Multimedia, Interaction, Design and Innovation on ZZZ – MIDI*, 1–9.

Landowska, A. (2016). How to design affect-aware educational systems—the AFFINT process approach. *Proceedings on the European Conference of e-Learning 2016*.

Landowska, A., & Brodny, G. (2017). Postrzeżenie inwazyjności automatycznego rozpoznawania emocji w kontekście edukacyjnym. *EduAkcja. Magazyn edukacji elektronicznej*, 1(13), 26–41.

Landowska, A., Brodny, G., & Wróbel, M.R. (2017). Limitations of emotion recognition from facial expressions in e-learning context. *9th International Conference on Computer Supported Education*, 383–389. <http://dx.doi.org/10.5220/0006357903830389>

Landowska, A., & Miler, J. (2016). Limitations of emotion recognition in software user experience evaluation context. *Federated Conference on Computer Science and Information Systems*, 1631–1640.

Li, J., & Ren, F. (2008). *Emotion recognition from blog articles*. International Conference on Natural Language Processing and Knowledge Engineering, IEEE NLP-KE.

Ling, H. S., Bali, R., & Salam R.A. (2006). *Emotion detection using keywords spotting and semantic network*. Paper presented at the International Conference on Computing and Informatics ICOCI.

Maria, K.A., & Zitar, R.A. (2007). Emotional agents: A modeling and an application. *Information and Software Technology*, 49(7), 695–716. <http://dx.doi.org/10.1016/j.infsof.2006.08.002>

Mehrabian, A. (1996). Pleasure-arousal. Dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14(4), 261–292.

Neviarouskaya, A., Prendinger, H., & Ishizuka, M. (2009). Compositionality principle in recognition of fine-grained emotions from text. *International Association for Advancement of Artificial Intelligence Conference on Web and Social Media*, 278–281.

Paiva, A., Dias, J., Sobral, D., & Woods, S. (2004). Building empathic lifelike characters: the proximity factor. *International Conference on Autonomous Agents and Multiagent Systems*, 4.

Picard, R.W. (2003). Affective computing: challenges. *The International Journal of Human-Computer Studies*, 59(1–2), 55–64. [http://dx.doi.org/10.1016/S1071-5819\(03\)00052-1](http://dx.doi.org/10.1016/S1071-5819(03)00052-1)

Picard, R.W., & Ahn, H. (2006). Affective cognitive learning and decision making: The role of emotions. *Lecture Notes in Computer Science book series*, 3784, 866–873.

Picard, R.W., & Daily, S. (2005). Evaluating affective interactions: Alternatives to asking what users feel. *Proceedings CHI'05 Workshop on Evaluating Affective Interfaces: Innovative Approaches*.

Picard, R.W., & Klein, J. (2002). Computers that recognise and respond to user emotion: theoretical and practical implications. *Interacting with Computers*, 14(2), 141–169. [http://dx.doi.org/10.1016/S0953-5438\(01\)00055-8](http://dx.doi.org/10.1016/S0953-5438(01)00055-8)

Scheirer, J., Fernandez, R., Klein, J., & Picard, R.W. (2002). Frustrating the user on purpose: a step toward building an affective computer. *Interacting with Computers*, 14(2), 93–118. [http://dx.doi.org/10.1016/S0953-5438\(01\)00059-5](http://dx.doi.org/10.1016/S0953-5438(01)00059-5)

Scherer, K.R., & Ekman, P. (1984). *Approaches to emotion*. Hillsdale, NJ: L. Erlbaum Associates.

Sheng, Z., Zhu-ying, L., & Wan-xin, D. (2010). The model of E-learning based on affective computing. *3rd International Conference on Advanced Computer Theory and Engineering (ICACTE)*, 3, 269–272.

Strapparava, C., & Valitutti, A. (2004). WordNet-Affect: an affective extension of WordNet. *4th International Conference on Language Resources and Evaluation*, 1083–1086.

Wioleta, S. (2013). Using physiological signals for emotion recognition. *The 6th International Conference on Human System Interaction*, 556–561.

Wolf, B., Burleson, W., & Arroyo, I. (2009). Affect-aware tutors: recognising and responding to student affect. *International Journal of Learning Technology*, 4(3/4), 129–164. <http://dx.doi.org/10.1504/IJLT.2009.028804>

Yik, M.S. M., Russell, J.A., & Barrett, L.F. (1999). Structure of self-reported current affect: Integration and beyond. *Journal of Personality and Social Psychology*, 77(3), 600–619. <http://dx.doi.org/10.1037/0022-3514.77.3.600>

Zeng, Z., Pantic, M., Roisman, G.I., & Huang, T. S. (2007). A survey of affect recognition methods. *Proceedings of the 9th International Conference on Multimodal Interfaces*. Paper presented at ICMI 2007 in Nagoya, Japan, November 12–15 (pp. 126–133). <http://dx.doi.org/10.1145/1322192.1322216>

Investigation of educational processes with affective computing methods

This paper concerns the monitoring of educational processes with the use of new technologies for the recognition of human emotions. This paper summarizes results from three experiments, aimed at the validation of applying emotion recognition to e-learning. An analysis of the experiments' executions provides an evaluation of the emotion elicitation methods used to monitor learners. The comparison of affect recognition algorithms was based on the criteria of availability, accuracy, robustness to disturbance, and interference with the e-learning process. The lessons learned in these experiments might be of interest to teachers and e-learning tutors, as well as to those researchers who want to use affective computing methods in monitoring educational processes.

Since 2000, **Agnieszka Landowska** has worked for Gdansk University of Technology, FETI, in the Department of Software Engineering Methods. She is a leader of the Emotions in HCI Research Group and conducts research on software usability, affective computing methods and e-learning systems design. She is a member of the Management Board of Polish Scientific Society on E-Learning (PTNEI) and of the scientific organizations AAAC and EDEN.

Grzegorz Brodny is a Ph.D. Candidate at Gdansk University of Technology (GUT). He received his M.Sc. diploma in Computer Science, with specialization in Software Engineering and Databases in 2015, at GUT. Since 2015 he is a member of the EmoRG science group at GUT.