A Novel Method to Build and Validate an Affective State Prediction Model from Touch-Typing

Subrata Tikadar, Samit Bhattacharya

A Novel Method to Build and Validate an Affective State Prediction Model from Touch-Typing

Subrata Tikadar and Samit Bhattacharya

Indian Institute of Technology Guwahati, Guwahati, India
subratatikadar@gmail.com, samit3k@gmail.com

Abstract. Affective systems are supposed to improve user satisfaction and hence usability by identifying and complementing the affective state of a user at the time of interaction. The first and most important challenge for building such systems is to identify the affective state in a systematic way. This is generally done based on computational models. Building such models requires affective data. In spite of the extensive growth in this research area, there are a number of challenges in affect induction method for collecting the affective data as well as for building models for real-time prediction of affective states. In this article, we have reported a novel method for inducing particular affective states to unobtrusively collect the affective data as well as a minimalist model to predict the affective states of a user from her/his typing pattern on a touchscreen of a smartphone. The prediction accuracy for our model was 86.60%. The method for inducing the specific affective states and the model to predict these states are validated through empirical studies comprising EEG signals of twenty two participants.

Keywords: Affective State, Emotion Induction Method, Mobile-HCI, Smartphone Interaction, Touchscreen Keyboard, Typing Pattern, Unobtrusive Approach, Valence and Arousal.

1 Introduction

A system that is able to identify and/or complement the affective states of its users at the time of interaction can be termed as affective system [29]. The first and the most challenging step to build such systems is to identify the affective states in real-time. The challenges are more if the target systems are small and mobile, compared to that of large static systems. This is because small mobile systems may not always allow additional expensive hardware and/or probes for supporting services. Otherwise, the mobility and affordability of the devices (probably the two of the most important reasons for them being popular) might be affected. Extensive research has been going on since last decade for addressing the challenges [30]. Nevertheless, issues still persist. The issues include erroneous and fake data in the existing methods of inducing affect and emotion together with the way of collecting the affective data, and complex model to identify the states in real-time. Additionally, in most of the cases neither the induction methods nor the model to identify the states are systematically validated.

Researchers depend on users' feedback for this purposes, which is not always reliable [31].

Our contribution in this research area is twofold. The first contribution is proposing a novel method for inducing specific affective states to unobtrusively collect the affective data. We have proposed a gaming approach for this. We have developed an android game having multiple stimulants for affect and emotion. The game is able to induce specific affective states, as well as to collect and label the affective data. The data collection and labeling are done without making the users aware of it. This helps to reduce the inaccuracy and imitation in affective data. Our second contribution is proposing a minimalist model to identify the affective states of the users from their typing pattern on touchscreen keyboard of smartphones. Taking the lowest number of input features from users' typing pattern, the model is able to classify a user into any of the four states with high accuracy: *positive-high, positive-low, negative-high, and negative-low*. The states have been chosen based on the circumplex model of emotion [32, 34], where positive and negative specify the level of valence whereas high and low indicate the level of arousal.

We have ascertained the efficiency of our proposed method and model through controlled empirical studies. We have used EEG (electroencephalogram) data of twenty two participants for these purposes. Additionally, we have also tested whether our model works for other applications (social network and instant messaging apps) where typing data is available, not restricted to the special purpose game application developed by us.

The rest of the article is organized as follows. In Section 2, we have presented the literature review for identifying related works of existing methods for inducing affects and emotions, as well as of existing models to identify users' affective states from users' typing pattern. We have also highlighted the importance and novelty of our proposed method and model in this section. Section 3 presents the proposed approach where the method for inducing specific affective states, and the model for identifying the states are described in detail. In Section 4, we discussed about the variation of the model, and strength and limitations of our approach with future scope. Section 5 concludes the article.

2 Literature Review

2.1 Literature Review for Existing Works to Induce Affects and Emotions

Affective data are required to build and train models for automatic identification of the affective states in real-time. Collection of the affective data from 'natural source' is cumbersome and most of the time is infeasible [8, 9]. Affective data from natural source means the corresponding affective states are stimulated naturally, not induced in laboratory setting. For example, someone has got a mail with an extremely bad/good news and s/he is typing for replying the mail. In this case, the typing data can be considered as affective data from natural source. Moreover, the way of labeling these naturalistic data may involve inaccuracy. The data are labeled by either observation or feedback based method. Inaccuracy in observation based methods occurs due to the fact that feeling and perceiving the affective states are not the same [19]. At the same time, the user-feedback method for labeling the data may sometimes be unreliable because human beings may not always be able to report their own exact affective state of a particular moment [19]. Furthermore, limited types of affective data (e.g., speech, image) can be collected from natural sources, which are copyrighted most of the time [10]. As a result, particular affects and emotions are induced in laboratory setting to collect and label the data [14, 28, 31]. It can be noted that although the terms "affect", "emotion", "mood", and "personality traits" are sometimes used interchangeably by the researchers, they are not the same. They are distinguished based on the duration, intensity, cause, and other aspects [2]. In this work, we have worked with the Circumplex model [32, 34] of emotional states. The states in this model can generally be termed as 'affective states'.

Existing methods for inducing affective states include affective music and film, hypnosis, making unnecessary delay, stressful interview, sudden attack, surprise exam, threat of painful electric shock, and unsolvable puzzle and game [14, 28, 31, 39].

The methods are not free from issues like involvement of fake and inaccurate data. In case of inducing emotion through affective film and music, the participants may be aware of the fact that they are being induced. As a result, they may not behave naturally [31]. There is a chance of imitation in emotion and thereby resulting in fake data collection. If a decision is taken based on the fake affective state, the decision may be wrong. Moreover, when the participants are aware of the fact that they are being induced, some additional affects and emotions (anxiety, fear, hesitation, hostility, and nervousness) are induced with the intended specific affect and emotion to be induced [31]. The additional affects and emotion may be incorporated in the attack and threat methods as well. Furthermore, most of the existing methods for inducing emotion together with the collection of affective data consider participants' feedback for labeling the data. For the data labeling, the participants are first induced with some specific affective states. After that, they are given some tasks to be completed, through which the affective data can be collected. The participants are asked to report their states for labeling the collected data afterwards (sometimes beforehand the task-execution), as it is not possible for them to execute the tasks and give feedback at exactly the same time. The affective states at the time of induction, task execution, and taking feedback may not be the same. This is because of the fact that affects and emotions stays for very short duration (e.g., emotion stays in facial expression up to five seconds) [2, 28]. As a result, data may be wrongly labeled, and consequently inaccuracy in the affective data occurs.

We therefore propose a novel method, following a gaming approach, to induce specific affective states. Our method neither directly/indirectly informs the participants that they are getting induced nor considers participants' feedback for collecting and labeling the data. It is expected that the proposed method addresses the challenges in existing methods of inducing affects and emotion for collecting the affective data.

2.2 Literature Review for Existing Works to Identify Affects and Emotions

Literature contains many works to identify humans' affective states. The states can be identified from facial expression [12, 15, 22], gestures and posture [3, 4, 20], speech and voice [26, 33], physiological signals (like EBR, EEG, EMG, GSR, HR) [17, 18, 36] and multimodal inputs [18, 22] (where more than one input modalities are considered). Although the works are substantial, these are not free from issues. For instance, identifying affective states from facial expression involves expensive computation like computer vision and image processing, fixed camera position to capture the facial image, and proper lighting conditions. These constraints may not be satisfied everywhere, particularly in the context of mobile environment. Moreover, facial expression may not always be real (e.g., sometimes we smile for the sake of formalities, although we are not actually happy at that moment). The issues may also exist in case of identifying affective states from humans' gestures and postures. There may be some complications in the method of identifying the affective states form speech and voice as well. One of the major complications is surrounding noise, particularly in the mobile environment. Moreover, the accuracy of identifying affects and emotions depends on the corpus used for training the models developed in this method. Furthermore, the languages and geographical locations are two additional obstacles for the lack of accuracy in real-time emotion identification through this method. This is because of the fact that a particular word may be pronounced differently based on the ethnicities of the speakers. Although the physiological signals and multimodal based methods identify users' affective states with high accuracy, there are some difficulties to apply these for real-time identification of the states. These methods requires involvement of additional and highly expensive sensors and/or probes to sense the signals, which is not acceptable and affordable in mobile environment, particularly in the context of smartphones. Moreover, no user is expected to be comfortable and willing to wear these additional sensors and probes at the time of interaction with the mobile device. Researchers are therefore looking for alternative input for identifying affect and emotion so that the concept of affective computing can be practically applied. The use of interaction behavior can be a good alternative.

Interaction behavior may include haptic behavior, device handling patterns, touch patterns, and typing patterns. The advantages of considering the behavioral inputs include inexpensive computation and hardware setting. However, choosing the appropriate behavioral input is difficult. The inputs for identifying affect and emotion may vary based on the environments and applications. For instance, identifying emotion from mouse clicking pattern [43] may not be applicable for the context of smartphone interaction. Existing models for identifying emotion from users' typing pattern on physical keyboard [13, 21, 40] also may not be suitable in case of smartphone interaction as typing pattern on physical keyboard and smartphone's keyboard are not the same (we use almost all the fingers of our two hands while typing on physical keyboard whereas generally only two thumbs for typing on touchscreen keyboard of the smartphone). Literature contains few works where users' affect and emotion have been identified from touch patterns on touchscreen (e.g., [5, 35, 38]). Although these are suitable for smartphone interaction, we cannot always assume that all the interactions are performed through touches. Considering typing pattern on touchscreen of smartphone may be one of the useful inputs in this context. Although the typing on smartphone keyboard is performed through touch, the touch pattern while typing on the keyboard and the pattern while touching on any place of the screen for touch interactions may not be the same. For example, differences in frequency of touches may be observed in these two different scenario.

Lee et al [23, 24] used virtual keystroke data to identify the affective state of the users. Although their approach is suitable in the context smartphone interaction, we identified few limitations in their work including data collection and feature selection. The participants were emotionally induced by “International Affective Picture System (IAPS)” before the actual data collection. The affective state may change within the period of emotion-induction and data collection because emotions remains for very short duration [2, 28]. Secondly, users were asked to indicate their affective states after finishing the typing by posting ‘smilies’/self-assessment manikin (SAM). This alone can change the affective state of the users. The approach, hence, is not unobtrusive in true sense. Moreover, some of the input features used in their model (‘special symbol frequency’, ‘erased text length’, and ‘backspace key press frequency’) are unnecessary and/or inappropriate for the current context. We therefore propose a model with minimum number of appropriate features.

3 Proposed Approach

Taking users’ typing pattern on touchscreen keyboard of a smartphone as input, our proposed model classifies the affective state of a user into any of the following four categories:

- Positive-High,
- Positive-Low,
- Negative-High, and
- Negative-Low.

Here, positive and negative specify the valence level whereas high and low indicate the level of arousal. We chose these affective states based on the Circumplex model of affective states [32, 34], where continuous states are discretized based on the level of arousal and valence. The reason behind choosing the Circumplex model of emotion is its vast application areas [11, 39]. Moreover, identifying affective states based on the Circumplex model is most suitable for some specific applications. For instance, identifying the affective states of a student in the classroom should be in the form of valence and arousal for its better utilization [42].

The novel game for inducing these specific affective states, data collection through the game, building the minimalist model with these data, and the study for validating the induction method as well as the model is described in detail in this section.

3.1 Method for Inducing the Specific Affective States

We designed a special purpose android typing game for inducing the specific affective states without the knowledge of the users, as well as for labelling and collecting data without any user feedback. The approach was expected to induce the states minimizing the imitation in emotion (as it was done without users' knowledge), and result in labeled data minimizing the inaccuracy (as the method was unobtrusive).

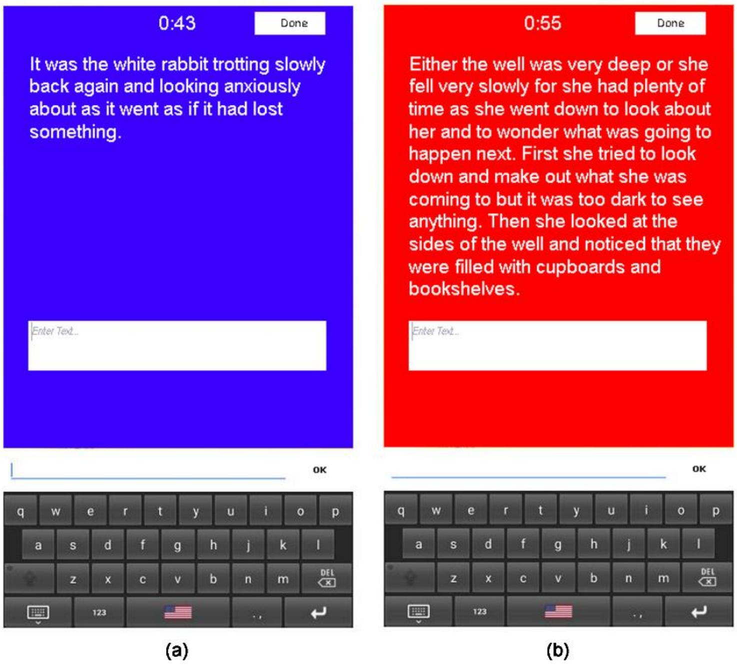


Fig. 1. Snapshots of two modes of the game interface; (a) fascinating mode, and (b) dull mode

Game Description. The game had two modes: *fascinating* mode and *dull* mode. While playing the game, a player had to play two fascinating modes and two dull modes. In the fascinating modes s/he had to type a small paragraph (of 110-120 characters – possible to type for a normal person) whereas in the dull mode s/he had to type a long paragraph (of approximately 350-400 characters – might be impossible to type for a normal person) within the time limit of sixty seconds. Lengths of the paragraphs were decided based on the work of [21], [25], and [27]. Based on the correctness and speed, the score was displayed on the screen in the fascinating modes (players were informed that rewards will be given as per the score). For the correctness of typing, we used a string matching algorithm. The algorithm compares each of the character of the typed text with each of the character of the given paragraph. The typing was considered to be accurate when both the texts match exactly. Erased characters and backspace key were not considered for string matching. The score was displayed in the fascinating modes only. In the dull modes, the players were informed that they would be rewarded if and only if they could complete the typing task within the time limit. Two affective colors were kept as background colors in the two modes of the game to aid the induction method. Background color for the fascinating mode was kept blue whereas that of the dull mode was kept red [1, 37]. Figure 1 shows the snapshots of the two modes of the game: (a) when a user was playing fascinating mode, and (b) when a user was playing dull mode.

It was expected that the multiple stimulants of the fascinating mode (ease and possibility of typing a short paragraph, keenness of earning a maximum reward, and the blue color in background) were able to induce positive emotions to the players. The emotions here were assumed to be of positive valence. On the other hand, stimulants used in dull mode (asking for completing an impossible typing job, practically no reward, and the red color in background) were expected to induce negative emotions to the players. The emotions here were assumed to be of negative valence.

All along during the gameplay, background music were played as additional stimulants for the induction method. Music capable of inducing positive valence with high arousal (Tchaikovsky's "Mazurka from Swan Lake Ballet") was kept in one fascinating mode and music capable of inducing positive valence with low arousal (Gluck's "Orpheus and Eurydice") was kept in another fascinating mode. Similarly, music capable of negative valance with high arousal (Mussorgsky's "Night on Bald Mountain") was kept in one dull mode and music capable of negative valence with low arousal (Marcello's "Adagio from Oboe Concerto in D minor") was kept in another dull mode. In between the fascinating and dull mode, we kept a neutral music (Kraftwerk's "Pocket Calculator") to neutralize the induced emotion. We chose affective music based on [6, 39]. For example, Västfjäll [39] has mentioned that a music having major-mode, fast-tempo, medium-pitch, firm-rhythm, dissonance-harmony, and high-loudness has the capability to make one excited. They have also mentioned a list of music pieces responsible for positive, neutral, and negative states specifying high and low arousal.

The modes were placed randomly to avoid bias. In between the two modes score and reward details were shown to neutralize the affects. Music capable of inducing neutral affect were also kept for this. Transitions of the music in between different modes were done smoothly (the volume of the music was reduced) to avoid the feeling of jerk. The players were aware of neither the strategy for the induction mechanism nor the modes of the game.

Validating the Purpose of the Game. We assumed that our special purpose typing game was able to induce the four specific affective states (positive-high, positive low, negative-high, and negative-low) to the users while playing the game, as we kept such stimulants. However, it was required to validate whether the game was really able to induce these states. For this, we used EEG signal of the players. We captured the EEG signals by Emotiv EPOC+[1] (a 14-channel wireless EEG headset: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4; the device collects the signals with the sample rate of 2048 per second – filtered and downsampled to 128 SPS / 256 SPS, user configured; the response frequency (bandwidth) of the device is 0.16 to

[1] https://www.emotiv.com/epoc/

43 Hz.). EEG signals were processed by EmotivPRO[2] software, which was able to analyze the raw EEG and derive six affective states (‘stress’, ‘engagement’, ‘interest (valence)’, ‘excitement (arousal)’, ‘focus’, and ‘relaxation (meditation)’) from the raw signals in real-time. However, for this particular study, we were interested in only the two states (valence and arousal) among the six affective states. Although an earlier study [7] verified the reliability of a previous model of the device for capturing EEG (EPOC), we conducted a pilot study to establish the ground truth for the affective states before validating our induction method through these.

Fig. 2. Pilot Study to establish the ground truth of affective states

Pilot Study for Establishing the Ground Truth of Emotion. Ten students (five males and five females, mean age of 24.7 with SD=3.37) took part in the study. To establish the ground truth for the identified affective states by the EmotivPRO, we asked each of the ten participants to wear the EPOC+ headset (Fig. 2) which was connected with the EmotivPRO software running on a laptop computer. After recording the baseline emotion, the participants were shown two videos having the elements of inducing arousal and valence emotion (one video to check two levels of arousal i.e., high and low arousal, and another to check two levels of valence i.e., positive and negative valence). The works including [39] and [6] helped us to select the videos having emotion inducing element. We downloaded the videos from YouTube. Following are the lists of audio-visual used for this experiment (Table 1).

Table 1. List of audio-visual used in the pilot study

Name of the Audio-Visual	Link (URL)	Used to Examine

[2] https://www.emotiv.com/emotivpro/

High Arousal Music Playlist	https://www.youtube.com/watch?v=WHUo4pZLAAA	Arousal levels
The Best Of Eagle Attacks 2018 - Most Amazing Moments Of Wild Animal Fights! Wild Discovery Animals	https://www.youtube.com/watch?v=RB4RCOe-ZEw	Arousal levels
World most Shocking Video that made the whole world cry!	https://www.youtube.com/watch?v=QJxwE7mdGns	Arousal levels
14 Strange Ways of Life the Ancient Egyptians Practiced	https://www.youtube.com/watch?v=GiMbVa6XzTw	Valence levels
What Will Happen to Humans Before 2050?	https://www.youtube.com/watch?v=Cip3LmqQ7Y0	Valence levels
25 True Facts That Will Shock You	https://www.youtube.com/watch?v=HChCEGR_0lg	Valence levels

We showed the videos partially (for a maximum two minutes) on a large projector screen with amplified sound. Short videos were shown because we have observed that participants were unable to report their affective states if these were asked few minutes later. This proves that the obtrusive approaches to collect emotional data may have incorrect information. The affective states (arousal and valence) were recorded by EmotivPRO while the participants were watching the videos. After the completion of each video, we told them about their affective states (valence and arousal i.e., interest and excitement) at different points of the video (after every ten seconds) and asked them to report on their agreement with the output by EMotivPRO. We replayed the videos for them to recall their affective state at the time of watching the particular instances of the videos. We compared the affective states reported by the participants and the affective state identified by EmotivPRO, and found the similarity of 92.08%. Thus, we considered the afffective state identified by the EmotivPRO as the ground truth of the users' affective state.

Validating the Induction Method. Considering the affective states identified by the EmotivPRO as ground truth, we conducted a validation study with ten different volunteers (five male and five female, mean age of 25.3 years with SD=3.94). In the study, the participants were asked to wear the EPOC+ headset which was connected to the EmotivPRO software installed on a laptop computer. We then asked the participants to play the special purpose game we developed. We observed that the players' interest (valence) was higher in the fascinating mode and lower in dull mode. As the EmotivPRO does not show the valence in negative scale (it shows both the arousal and valence in a 100 point scale ranging from 0 to 100 with a scale division of 10), we considered the higher reading (51-100) of valence as positive valence and lower reading (0-50) of valence as negative valence. We observed that negative valence was shown while the players played the dull mode, and positive valence was shown at the time of playing fascinating mode. These indicate that the game is capable of inducing two required levels of valence. We also observed that the arousal was high in one fascinating mode and the same was low in another fascinating mode. As expected, similar observations for arousal were made in the dull modes as well (arousal was observed high in one dull mode, whereas the same was observed in another dull mode). This was because we kept such stimulants in the particular modes (the music capable of eliciting high arousal in one fascinating and dull mode, and the music capable of eliciting low arousal in another fascinating and dull mode). Figure 3 contains

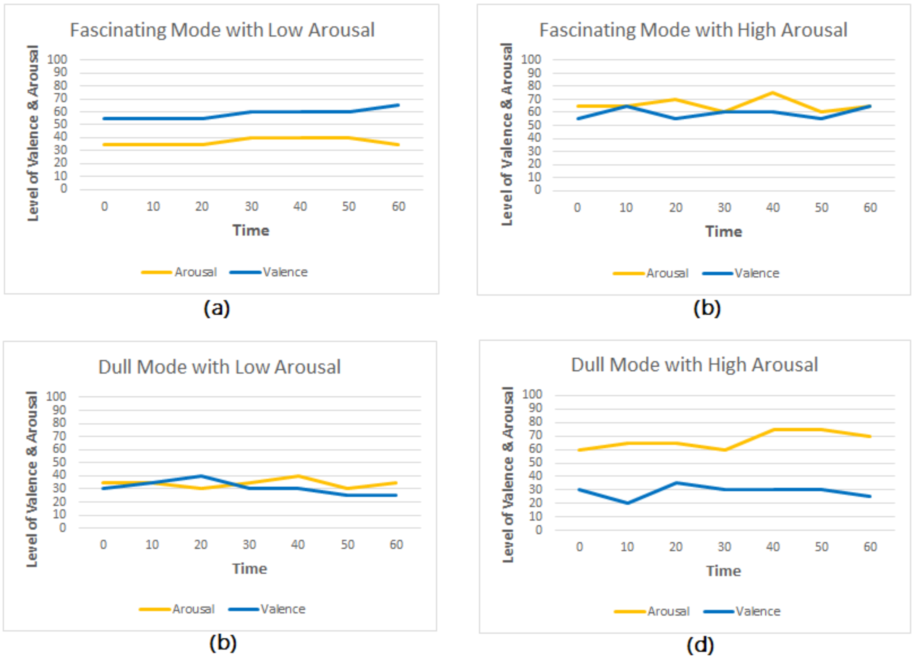


Fig. 3. Four specific affective states induced to a particular participant by our proposed method (identified from EEG signals by EmotivPRO)

the level of arousal and valence of fascinating and dull mode of a particular participant (identified by EmotivPRO): 3(a) when the participant was playing one of the fascinating modes where music capable of eliciting low arousal was there; 3(b) when the participant was playing another fascinating mode of the game where music capable of eliciting high arousal was there; 3(c) when the participant was playing one of the dull modes where music capable of eliciting low arousal was there; and 3(d) when the player was playing another dull mode where music capable of eliciting high arousal was there. If we observe closely, we can see that fascinating mode of the game is able to induce positive valence whereas dull mode is able to induce negative valence. We have considered valence ≤ 50 (shown by EmotivPRO) as negative, as discussed earlier. Keeping the valance level fixed, it is possible to induce two different level of arousal by changing the additional stimulant (affective music). This means, through our game it is possible to induce four specific affective states to a participant: positive-high, positive-low, negative-high, and negative-low. Similar observations were made for other nine participants as well. The observations validate the fact that the gaming approach with multiple stimulants, proposed by us, is able to induce the four specific affective states.

3.2 Building the Minimalist Model

Data Collection through the Game. The special purpose game application we developed is able to induce specific affective states. At the same time, it can store the typing data (all the key-press details) after labelling these as per the induced affective states. However, it was required to know what exact typing data should be collected for building a model for identifying affective states. Initially, we considered the following seven as input features for our proposed model as they seem to have some relations with affect and emotion [23, 24].

1. *Typing speed*: Number of characters typed per second.
2. *Backspace key press frequency*: Number of backspace key presses per second.
3. *Special symbol frequency*: Number of special symbols (for instance, emojis, smileys, astrological symbol) typed per second.
4. *Max text length*: Maximum number of characters typed without pressing the 'del' key for a second.
5. *Erased text length*: Number of characters erased in a second.
6. *Touch count*: Total number of key presses (including 'del' key) per second.
7. *Device shake frequency*: Number of times the device's position in space changed above an acceptable level of displacement (above a threshold value) per second.

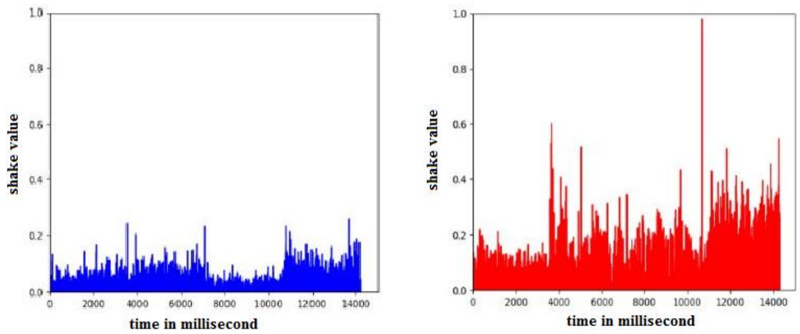
However, 'Special symbol frequency' was discarded as it may not be available and/or not important in case of some applications. For instance, unlike chat application, it may not be required to exchange emotion with friends while taking class note. Moreover, if we take special symbol, for instance 'emojis', the approach will not remain unobtrusive as it seems to be one kind of user feedback (users themselves report their states through the emojis). One can hide her/his actual emotion while exchanging emotion using 'emojis'. There is a chance of collecting fake affective data, in this case. The 'Erased text length' has also been discarded because in case of android keyboard, multiple characters can be deleted with a single backspace key press.

We, therefore collected the typing data for the remaining five features: typing speed, backspace key press frequency, max text length, touch count, and device shake frequency. Among these five features, we required a threshold value for the 'device shake frequency'. Otherwise, it was difficult to determine how much 'displacement' of the device could be considered as a shake. The absolute shake value is obtained when a trigger[3] is detected. The trigger refers to automatic detection of the displacement of the device. The android version we worked on (Marshmallow 6.0.1), checks for a trigger every 0.02 seconds. The absolute shake value is calculated as the scalar displacement in terms of the x, y and z coordinates of the accelerometer sensor following the equation (1).

$$displacement = \sqrt{x^2 + y^2 + z^2} \tag{1}$$

[3] https://developer.android.com/reference/android/hardware/TriggerEvent.html

For determining the shake threshold, we conducted an experiment with another set of ten participants (five male and five female, mean age of 25.1 years with SD=3.14). All the participants were asked to play the same game application twice per level. Once they were asked to type the target text, keeping the device as still as possible. A Second time, the users were allowed to handle the device as they usually do it while typing. Figure 4 represents the plots obtained corresponding to low and high shake.



(a) Low shake value **(b) High shake value**

Fig. 4. Low and high shake value

It clearly indicates that the shake values in the usual typing occasion are much higher than the same in the first occasion where the device was kept still as much as possible while typing. This indicates that the data can be naturally clustered into two separate groups. Therefore, we calculated the shake threshold as the mean of the two cluster centers. Cluster centers were decided by taking the means of the shake values in the individual clusters. The threshold value was found to be 0.04. We did not require any threshold value to record the data for the other four features.

We collected the data from thirty three participants for further analysis to build the minimalist model. Data was collected by using a smartphone (having 5.5 inches screen, 'Android Marshmallow' OS, 2.5GHz quad-core processor, and 3GB RAM) where the game we developed was installed.

Selection Strategy and Details of the Participants for Data Collection. We wrote a formal mail to all of the students in our institute for voluntarily participating in the data collection experiment. On arrival, we asked a set of questions to the students to know their current health condition as well as the history of their health for selecting them as participants. We did not collect data from the students who were in severe medical condition (e.g., serious headache, shortness of breath), who consumed any intoxicating substance in last three hours, or did not sleep well in last night for at least six hours [7]. Some of the volunteers could not participate due to their finger sizes (were relatively large compared to the size of the keys). This was because we wanted to avoid the 'fat finger' problem. We followed the same strategy to select the participants for every study mentioned in this paper. Once selected, each participant was asked to sign a 'participant consent form' which was signed by one of the authors (as experimenter) as well. We selected thirty three participants for the data collection. Among the thirty three participant seventeen were male and sixteen were female. The mean age of the participants was 24.19 with SD=2.91. Although the participation was voluntary, we offered some packaged food at the end of the study to thank the volunteers.

Data Analysis. We analyzed the collected data (a) to minimize and finalize the number of features, (b) to decide the appropriate time slice for which the final set of selected features should be taken together before applying the classifier and (c) to decide the classifiers to be used. The detail analyses are described below.

(a) Feature Reduction. We used the Principle Component Analysis (PCA) to identify the number of features which could be rejected. We plotted a graph for cumulative proportion of variance against the number of features (Fig. 5).

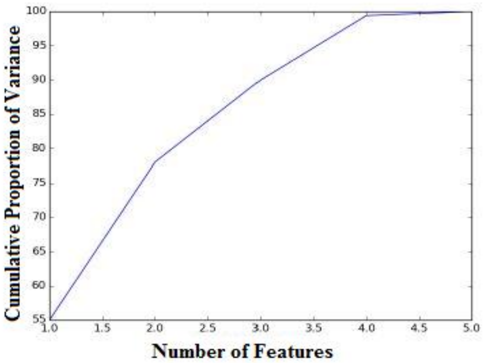


Fig. 5. Result after applying PCA

It was observed that the curve became almost parallel to the x axis after four features. Hence, we decided that one among the five features could be rejected. Therefore, our final set of features could be of size four. We applied the 'f-regression technique' to identify the feature which could be discarded among the five features. The f-regression technique calculates the correlation between two variables. It selects the features one by one, which has a high correlation with the output variable (valence and arousal level, in our case), and at the same time very less dependency with the already selected features. We observed that except for 'backspace key press frequency', all the other four features: 'typing speed', 'touch count', 'max text length' and 'device shake frequency' had high correlation with valence and arousal level. Thus backspace key press frequency was rejected from the final set of features.

(b) Identification of Appropriate Time Slice. We also identified the appropriate time slice for which all the features should be taken together before applying the classifier. We empirically found out the appropriate time slice by statistically testing the following hypothesis for each time slice from one second to ten seconds.

The null hypothesis corresponding to each feature was:

"For feature X; the mean of feature values of X for low arousal is same as the mean of feature values X for high arousal".

Hence, the alternative hypothesis for each feature was:

"For feature X; the mean of feature values of X for low arousal is different from the mean of feature values X for high arousal"

Table 2. Hypothesis testing for different time slices

Time Slice	Minimum significance level to reject all null hypothesis using t-test	
	for valence	for arousal
1 second	0.09	0.08
2 seconds	0.08	0.09
3 seconds	0.08	0.08
4 seconds	0.14	0.13
5 seconds	0.13	0.15
6 seconds	0.16	0.14
7 seconds	0.04	0.04
8 seconds	0.16	0.17
9 seconds	0.08	0.12
10 seconds	0.18	0.20

We applied the t-test for each feature for all the time slices to find out the appropriate time slice for which the hypothesis for all remaining features were rejected. In the similar way, the statistical test was done for valence as well. The results are presented as follows (Table 2).

The result above showed that the window size of 7 seconds had the minimum required significance level ($P<0.05$) to reject all null-hypothesis. Moreover, it was the only window size for which null hypothesis for all the features got rejected. We therefor chose 7 seconds as an interval, i.e., in each seven second our model should predict the affective stated based on the four features. We used the t-test also to verify whether rejection of the backspace key frequency feature was justified or not. The null hypothesis for backspace key frequency could not be rejected even for a significance level as high as 0.80. This implies that our decision to remove the particular feature was justified.

(c) Choice of the Classifiers. As our intension was to classify the users' states based on two levels of binary classifications (first for valence and then for arousal), the state-of-art binary classifier Support Vector Machine (SVM) with linear kernel could be chosen as one of the classifiers. We chose multilevel SVM for the classification. We assumed that all the four affective states identified by our predictive model were almost equally probable. This was because we trained our model with such type of data where data about all the four emotional sates were equally distributed. At the same time, the size of the data set was small. Therefore, Naïve Bayes classifier was also a good choice for the classification. It also tends to avoid the problem of overfitting. We, therefore explored both the SVM and Naïve Bayes classifiers.

Results. The model was trained by approximately 80% (data of twenty six participants) of total data (data of thirty three participants). Remaining data (data of seven participants) were used for testing. We did not do this in a simple manner because in that case the model could not learn anything from the set of data used only for testing. Hence, we have applied cross validation technique (CV). We followed Leave-One-Subject-Out cross-validation (LOSOCV) technique for training and testing the model. We chose this particular CV because Hammerla and Plötz [16] have argued that this CV is best applicable for the classifiers we chose. They have also argued that LOSOCV is best for the type of experiment we conducted. In LOSOCV, instead of random division, dataset is divided participant wise for the cross validation. The classification accuracy found after applying the two classifiers are shown in Table 3. We observed that the Naïve Bayes gave the highest average classification accuracy of 86.60%. The phrase ‘average classification accuracy’ denotes the summation of accuracy in each trial divided by the number of trials (five in our case – as we made five trials to cover the data of all the participants). It may be noted that although the Naïve Bayes gave higher average accuracy, the SVM also gave high accuracy. Moreover, the classification accuracy in case of the SVM is comparatively more stable. Therefore, we can claim that both the classifiers are equally likely to become candidates for our model.

Table 3. Affective states prediction accuracy

Trials	Percentage of Accuracy	
	Naïve Bayes	SVM
1	92.02	86.50
2	88.27	83.89
3	78.54	85.28
4	89.96	75.72
5	84.23	84.68
Average percentage of accuracy	**86.60**	**83.21**

Validating the Model. Other than testing the model with collected data, we conducted a separate validation study. For this we used EEG data of a different group of twelve participants (six male and six female, mean age of 24.67 years with SD=2.53).

The participants were asked to play the same game application we developed. Data was recorded as before. However, this time they wore the EPOC+ headset at the time of playing the game. The EPOC+ was connected with EmotivPRO software installed on a laptop computer through a dedicated Bluetooth dongle. The clocks of the laptop and the smartphone where the game was installed were synchronized before conducting the experiment. We recorded the valence and arousal identified by the EmotivPRO when the participants were playing the game.

After the experiment, using the collected feature value, we predicted the affective states using our model and compared these with the affective states identified by the EmotivPRO. We found 88.05% similarities for this. The high similarity validates the fact that our model is able to predict users’ affective state into any of the specific four circumplex state with high accuracy.

Testing the Generalizability of the Model. We also wanted to test whether our model works for other applications where typing data is available as the input for interactions, not just restricted to the game application we have developed. Thus, we could validate the generalizability of the model. We conducted an additional study with a different set of ten participants (five male and five female, mean age of 24.9 years with SD=1.85) for this. The participants were asked to wear the EPOC+ and access their favorite instant messaging service (IMS, e.g., WhatsApp, Facebook Messenger) and social networking (SNS e.g., Facebook, Twitter) apps for ten minutes each. In case of IMS, they were instructed to chat with their friends/relatives whereas for SNS, they were instructed to write some comments on some shared image/video by their friends/relatives. All the participants chose Facebook as SNS and WhatsApp as IMS as they have accounts in these. The EmotivPRO recorded the valence and arousal while the participants were performing the tasks. The typing data were collected by a background app written by us. We compared the four specific affective states predicted by our model (with Naïve Bayes classifier) and the same identified by the EmotivPRO. States predicted by our model and predicted by the EmotivPRO at each instant (after every 7 secs) is compared. In 87.90% cases, the predictions were the same. This indicate that our model can be applied in other application areas where interactions involve touch-typing data.

4 Discussion

4.1 Variation in the Model

We further extended our research making slight variation in the model based on particular research argument and target application area.

Variation with Meaningless Text. Epp et al [13] argued that the classifier works best in this type of model when we use simple English words with simple structure of sentence avoiding complex linguistic features. Complex linguistic features in sentences introduce memory biases. In our work, we tested the model using text from children's novel "Alice in Wonderland" (Fig. 1) which was also used by Epp et al [13]. To determine the correlation between the feature used and arousal level after completely removing the linguistic features from the text we made a variation in the model. We used text with meaningless words generated randomly. The following is an example of randomly generated meaningless text.

"*fffff sssss eeeee ccccc bbbbb kkkkk lllll pppppp aaaaa zzzzz vvvvv nnnnn rrrrr yyyyy qqqqq mmmmm jjjjj ooooo ttttt uuuuu*".

This was done for removing the linguistic features completely to remove memory biasness. After performing hypothesis testing on this data, we observed that the null hypothesis for all the features other than the 'backspace key press frequency' were rejected for a significance level as low as 0.01. This concludes that the correlation between the features (used in the model) and arousal level increases when linguistic features are removed. We, therefore, suggest to avoid complex linguistic features in this kind of experiment.

Variation without 'Device shake frequency' Feature. In some application, sometimes it is possible that the input devices are stable (e.g., 'smart' classroom with smartphone/tablet for each student fixed on the desk [41]). Moreover, sometimes user may put their device on the table/desk/bed for typing. 'Device shake frequency' is insignificant in these scenarios. Therefore, we wanted to make another variation of the model where device shake frequency was not considered. Through this variation, we examined the performance of the model. Following observation was made (Table 4).

Table 4. Accuracy of prediction with and without considering 'device shake frequency' feature

Classifier used	Average accuracy considering 'device shake frequency' (%)	Average accuracy without considering 'device shake frequency' (%)
Naïve Bayes	86.60	82.46
SVM	83.21	80.25

Although accuracy reduces slightly without considering 'Device shake frequency' feature, still the accuracy is high. However, we advise to consider the device shake frequency feature for additional accuracy.

4.2 Strength, Limitation, and Future Scope

Following our approach, it is possible to build models for predicting users' affect and emotion without taking any user feedback. The data collection method is purely unobtrusive, which helps to reduce the inaccuracy in emotional data. We have built a model following the approach which can predict the specific affective states of the users of small handheld touchscreen devices without their knowledge. This minimizes the imitation in emotion. Moreover, no extra hardware and sensors are required to capture the emotional data in our model. Only typing patterns and dynamics are sufficient to identify the users' affective states. We claim our model to be ubiquitous as the affective states are identified without the knowledge of the user by a device which stays always with the user, and it does not matter where the user is. The number of features to be used in the model has been minimized by excluding the unnecessary and application specific features used by earlier works ([23, 24]). This makes the model simple, generalized, and inexpensive in terms of computation. This in turn increases the adaptability of the model in various applications. Most of the related works (e.g., Lee et al [23, 24]).) were validated based on the users' feedback which may not be always reliable [31]. Our additional validation studies (consist of EEG signals of the participants) demonstrate that our proposed affect induction method is really able to induce the intended specific affective states. These studies also validate that our proposed minimalist model of affection is really able to predict users' affective states with high accuracy.

Despite the novelty and strengths of the proposed approach, there are some limitations and scopes for future work. Our model fits best in some specific application area (e.g., in education [42]), as it is able to detect only the four affective states based on the Circumplex model. However, it may be required to identify the affective states of a user more specifically as per the requirement of various other application areas. For instance, in movie making and promotion industry, it may be required to distinguish 'anger' from 'fear' where both of them are in 'negative-high' state. Our model, in its current state is unable to predict user's state with such distinction. In the future, we intend to refine the model which can identify the users' affect and emotion in a comprehensive way. We conducted a preliminary study for testing the generalizability of the model and observed that the model works (with high accuracy) for two other applications (SNS and IMS) where it is possible to get touch-typing input. The model can therefore be applied in these applications. If applied, the users of these applications will be able to know the affective states (in terms of the level of arousal and valence) of their friends and/or the followers while chatting and/or commenting on some posts. These can be made optional as well as mandatory, as per the requirements. The affective states exchanged in these applications by conventional methods (i.e., through emojis, stickers, GIFs), may not always be true because sometimes the users may intentionally hide their actual states. Nevertheless, an extensive study is required to identify more number of real-time applications where the model can be applied. Further research is also required for developing techniques to complement the identified affective states as well as its effects on users. We want to work on these as well. In future, we also intend to work on applying the model for building affective interactive system like affective classroom.

5 Conclusion

A novel method for inducing specific affective states has been proposed to minimize the inaccuracy and imitation in affective data. We have built a special purpose android game for this. It has been shown that the game is really able to induce the affective states without the knowledge of the users, and collect the affective data without any feedback from them. This in turn reduces the inaccuracy and imitation. We have also built a minimalist model for identifying the affective states of the users of small handheld mobile devices like smartphone from typing pattern on touchscreen. We reduced the unnecessary and application specific features used in earlier works to make the model simple and minimal so the applicability and adaptability increase. It has been shown that the model is really able to predict the specific affective states of the users with high accuracy. We want to work on refining the model, and applying the model to develop affective systems in future.

Acknowledgements. We are thankful to all the volunteers who participated in the empirical studies. We also sincerely acknowledge the contribution of Kshitiz Agarwal, Tarun Sharma, and Yash Mehta in this work. We also thank all the reviewers and the shepherd for their constructive suggestions in improving the quality of the manuscript.

References

1. Bartram, L., Patra, A., Stone, M.: Affective Color in Visualization. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI 2017), pp. 1364–1374. ACM, Denver, CO, USA (2017).
2. Beedie, C. J., Terry, P.C., Lane, A. M.: Distinctions between Emotion and Mood. Cognition & Emotion 19(6), 847–878 (2005).
3. Bianchi-Berthouze, N., Kleinsmith, A.: A categorical approach to affective gesture recognition. Connection science 16(1), 259–269 (2004).
4. Camurri, A., Lagerlöf, I., Volpe, G.: Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques. International journal of human-computer studies 59(1-2), 213–225 (2003).
5. Ciman, M., Wac, K., Gaggi, O.: iSenseStress: assessing stress through human-smartphone interaction analysis. In Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare, pp. 84-91. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering). , ACM & IEEE, Istanbul, Turkey (2015).
6. Collier, G.L.: Beyond valence and activity in the emotional connotations of music. Psychol-ogy of Music 35(1), 110–131 (2007).
7. Debener, S. et al.: How about taking a low-cost, small, and wireless EEG for a walk?. Psychophysiology 49(11), 1617–1621 (2012).
8. Devillers, L. et al.: Real life emotions in French and English TV video clips: an integrated annotation protocol combining continuous and discrete approaches. In LREC 2006, pp. 1105–1110. Genoa, Italy (2006).
9. Douglas-Cowie, E. et al.: Emotional speech: Towards a new generation of databases. Speech communication 40(1-2), 33–60 (2003).
10. Douglas-Cowie, E. et al.: The HUMAINE database: addressing the collection and annotation of naturalistic and induced emotional data. In International conference on affective computing and intelligent interaction, pp. 488 – 500, Springer, Berlin, Heidelberg (2007).
11. Eerola, T., Vuoskoski, J.K.: A comparison of the discrete and dimensional models of emotion in music. Psychology of Music 39(1), 18-49 (2011).
12. Ekman, P., Sorenson, E.R., Friesen, W.V.: Pan-cultural elements in facial displays of emotion. Science 164(3875), 86–88 (1969).
13. Epp, C., Lippold, M., Mandryk, R.L.: Identifying emotional states using keystroke dynamics. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 715-724. ACM, Vancouver, BC, Canada (2011).
14. Gaffary, Y., et al.: Gestural and Postural Reactions to Stressful Event: Design of a Haptic Stressful Stimulus. In 2015 International Conference on Affective Computing and Intelligent Interaction (ACII), pp. 988 – 992. IEEE, Xi'an, China (2015).
15. Gratch, J., et al.: Felt emotion and social context determine the intensity of smiles in a competitive video game. In 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), pp. 1 – 8. IEEE, Shanghai, China (2013).
16. Hammerla, N.Y., Plötz, T.: Let's (not) stick together: pairwise similarity biases cross-validation in activity recognition. In Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing, pp. 1041 – 1051. ACM, Osaka, Japan (2015).
17. Hazlett, R.L.: Measuring emotional valence during interactive experiences: boys at video game play. In Proceedings of the SIGCHI conference on Human Factors in computing systems (CHI 2006), pp. 1023–1026. ACM, Montréal, Canada (2006).

18. Healey, J., Picard, R.W.: Detecting stress during real-world driving tasks using physiological sensors. IEEE Transactions on intelligent transportation systems 6(2), 156–166 (2005).
19. Gabrielsson, A.: Emotion perceived and emotion felt: Same or different?. Musicae Scientiae 5(1_suppl), 123–147 (2001).
20. Glowinski, D. et al.: Toward a minimal representation of affective gestures. IEEE Transactions on Affective Computing 2(2), 106–118 (2011).
21. Khanna, P., Sasikumar, M.: Recognising emotions from keyboard stroke pattern. International journal of computer applications 11(9), 1-5 (2010).
22. Lang, P.J., et al.: Looking at pictures: Affective, facial, visceral, and behavioral reactions. Psychophysiology 30(3), 261-273 (1993).
23. Lee, H., et al.: Towards unobtrusive emotion recognition for affective social communication. In 2012 IEEE Consumer Communications and Networking Conference (CCNC), pp. 260-264. IEEE, Las Vegas, Nevada, USA (2012).
24. Lee, P.M., Tsui, W.H., Hsiao, T.C.: The influence of emotion on keyboard typing: an experimental study using visual stimuli. Biomedical engineering online 13(1), 81-92 (2014).
25. Lim, Y.M., Ayesh, A., Stacey, M.: The effects of typing demand on emotional stress, mouse and keystroke behaviours. In Intelligent Systems in Science and Information, pp. 209-225. Springer, Cham, London, UK (2014).
26. Murray, I. R., Arnott, J. L.: Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. The Journal of the Acoustical Society of America 93(2), 1097–1108 (1993).
27. Nahin, A.N.H., et al.: Identifying emotion by keystroke dynamics and text pattern analysis. Behaviour & Information Technology 33(9), 987-996 (2014).
28. Philippot, P.: Inducing and assessing differentiated emotion-feeling states in the laboratory. Cognition and emotion 7(2), 171–193 (1993).
29. Picard R. W.: Affective computing. 1st edn. Cambridge: MIT Press (1997).
30. Politou, E., Alepis, E., Patsakis, C.: A survey on mobile affective computing. Computer Science Review 25(1), 79–100 (2017).
31. Polivy, J.: On the induction of emotion in the laboratory: Discrete moods or multiple affect states?. Journal of Personality and Social Psychology 41(4), 803–817 (1981).
32. Posner, J., Russell, J.A., Peterson, B.S.: The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. Development and psychopathology 17(3), 715–734 (2005).
33. Rao, K.S. et al.: Emotion recognition from speech. International Journal of Computer Science and Information Technologies 3(2), 3603–3607 (2012).
34. Russell, J.A.: A circumplex model of affect. Journal of personality and social psychology 39(6), 1161–1178 (1980).
35. Sano, A. and Picard, R.W.: Stress recognition using wearable sensors and mobile phones. In 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII), pp. 671-676. IEEE, Geneva, Switzerland (2013).
36. Schmidt, L.A., Trainor, L.J.: Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. Cognition & Emotion 15(4), 487–500 (2001).
37. Sokolova, M.V., Fernández-Caballero, A.: A review on the role of color and light in affective computing. Applied Sciences 5(3), 275–293 (2015).
38. Tikadar, S., et al.: A Minimalist Approach for Identifying Affective States for Mobile Interaction Design. In 16th IFIP Conference on Human-Computer Interaction (INTERACT), pp. 3-12. Springer, Cham, Mumbai, India (2017).
39. Västfjäll, D.: Emotion induction through music: A review of the musical mood induction procedure. Musicae Scientiae, 5(1_suppl), 173–211 (2001).

40. Vizer, L.M., Zhou, L., Sears, A.: Automated stress detection using keystroke and linguistic features: An exploratory study. International Journal of Human-Computer Studies 67(10), 870-886 (2009).
41. Whitehill, J., et al.: The faces of engagement: Automatic recognition of student engagement from facial expressions. IEEE Transactions on Affective Computing 5(1), 86-98 (2014).
42. Woolf, B., et al.: Affect-aware tutors: recognising and responding to student affect. International Journal of Learning Technology, 4(3-4), 129-164 (2009).
43. Zimmermann, P., et al.: Affective computing—a rationale for measuring mood with mouse and keyboard. International journal of occupational safety and ergonomics 9(4), 539-551 (2003).