Extending Log-Based Affect Detection to a Multi-User Virtual Environment for Science

Ryan S. Baker[1], Jaclyn Ocumpaugh[1], Sujith M. Gowda[2], Amy M. Kamarainen[3], Shari J. Metcalf[4]

[1] Teachers College Columbia University, New York, New York
[2] Arizona State Unversity, Tempe, Arizona
[3] NY Hall of Science, New York, New York
[4] Harvard University, Massachusetts

baker2@exchange.tc.columbia.edu, jo2424@columbia.edu, mgsujith@gmail.com, akamarainen@nyscience.org, shari_metcalf@gse.harvard.edu

Abstract. The application of educational data mining (EDM) techniques to interactive learning software is increasingly being used to broaden the range of constructs typically incorporated in student models, moving from traditional assessment of student knowledge to the assessment of engagement, affect, strategy, and metacognition. Researchers are also broadening the range of environments within which these constructs are assessed. In this study, we develop sensor-free affect detection for EcoMUVE, an immersive multi-user virtual environment that teaches middle-school students about casualty in ecosystems. In this study, models were constructed for five different educationally-relevant affective states (boredom, confusion, delight, engaged concentration, and frustration). Such models allow us to examine the behaviors most closely associated with particular affective states, paving the way for the design of adaptive personalization to improve engagement and learning.

Keywords: student modeling, educational data mining, intelligent tutoring system, science inquiry, MUVEs, affect detection

1 Introduction

Researchers are increasingly interested in automated affect detection within educational software [cf. 8], which can be used both to drive automated intervention [3,16] and to conduct basic research on affect and learning [13,18]. One popular approach in affect detection within educational software is to leverage physical sensors of various sorts, including visual images obtained through webcams posture sensors, and electroencephalograms [1, 21, 29, 30]. Detectors built using physical sensors have typically been successful at identifying student affect in laboratory settings [12], but in classrooms as well [2]. One limitation of physical sensors in education research, however, is that they can be both costly and fragile. Combined with bandwidth restrictions, these issues can reduce the practicality of real-time sensor-based detection, especially in school environments.

Consequently, many researchers are now working towards sensor-free affect detection [see 6, 11, 17, 22, 26]. The quality of detectors developed in this fashion has now reached a point where detector agreement with expert field coders is about half as good as inter-rater agreement between human experts [cf. 6, 22]. Furthermore, sensor-free detectors of student affect have been able to predict standardized measures of student learning [22] and even which students will choose to attend college several years later [27]. Many of the sensor-free affect detectors developed have been developed for intelligent tutoring systems [6, 12, 17, 22].

In this paper, we build detectors that can infer a range of student affective states within the context of a multi-user virtual environments (MUVE), a computer-based learning program where each student controls an avatar who moves through a virtual world in a more autonomous fashion, interacting with non-player characters and objects in order to solve puzzles and learn educational content [15]. Affect detection has been developed previously for one MUVE, Crystal Island [26]. Their model was developed from a combination of baseline data obtained from a series of questionnaires and data from student interactions with the MUVE. The present study builds on this pioneering work, developing a sensor-free affect detector for EcoMUVE without using baseline questionnaires.

2 Data

The data analyzed were collected from 153 students studying with two teachers at a suburban middle school in the Northeastern United States. Students were predominantly White and Asian-American, with small numbers of Latino and African-American students. Only 1% was eligible for the free/reduced-price lunch program, considerably below the national average.

Students in this study were using EcoMUVE, a computer-based curriculum designed to teach about ecosystems. This 3D virtual world simulates real-life ecological environments, allowing students to develop an understanding of the complex interrelationships characteristic of ecosystems by maneuvering avatars throughout pond (module 1) and forest (module 2) ecosystems like that shown in Figure 1 [19]. Each 2-week module allows students to explore the simulated ecosystem over a number of virtual days, providing opportunities to observe interactions among ecological components (e.g., water, algae, fish) and the impact of human development on these multifaceted relationships. One module, for example, introduces a pond environment negatively affected by the nearby development of human infrastructure. Through investigation of the ecosystem (e.g., measuring bacterial composition of the pond, interviewing residents), students uncover causes of observed changes—in this instance, fertilizer runoff from an adjacent golf course and housing development is producing an algal bloom [19].

Fig. 1. Screenshots of the EcoMUVE virtual environment and tools.

3 Method

3.1 Obtaining Ground Truth Labels through BROMP

Student affect and behavior was observed *in situ* during EcoMUVE use. Both were coded simultaneously by an expert field observer, the second author, using the Baker-Rodrigo Observation Method Protocol (BROMP), [20]. BROMP has been used for several years to study behavior and affect in educational settings [5, 7, 25, 29] and has been used as the basis for successful automated detectors of affect [6, 22]. At present, 59 individuals have been certified, achieving inter-rater reliability (Cohen's Kappa > 0.6) with one or more other BROMP-certified coders. The coder in this study has successfully certified six other BROMP coders and has experience coding for a variety of different educational systems in populations that represent a wide range of regional, ethnic, and socioeconomic differences.

BROMP is implemented using the HART field observation synchronization software [6] developed for the Android platform. This software enforces a pre-determined order that prevents observers from being biased towards coding only the most interesting classroom events. HART also synchronizes observations to internet time so that they can be matched with the corresponding log files of the educational software being observed. In line with previous research on optimizing affect identification [23], BROMP-trained coders make holistic judgments based on contextualized observations of the student's actions, utterances, facial expressions, posture, and interactions with teachers or peers [7, 20]. Coders use side-glances to minimize observer effects, ignoring the affect and behavior of any students other than the one currently being observed. They record the first affective state they observe but have up to 20 seconds to make their observation.

In this study, seven affective states were recorded: boredom, engaged concentration (the affective state associated with flow – cf. [7]), confusion, delight, disgust, frustration, and sorrow. The categories were selected based on several criteria, including evidence about prevalent categories from previous learning research [7, 10], evidence about the prevalence of delight in games [24], qualitative reports from teachers and EcoMUVE developers, and discoveries made during a pilot study. Affective states which did not fit these categories and observations which occurred when a student could not otherwise be coded (e.g. if the student left the room or the teacher paused EcoMUVE activity for lecture), were coded with a "?".

Students were observed over the course of up to three class days (one class period per day). Observations for which a "?" was recorded or that occurred when the student was logged out of EcoMUVE were excluded, resulting in 2187 observations across all students and an average of 14.29 observations per student (SD = 5.35).

Within the field observations, the most common affective state was engaged concentration (67%). The remaining affective states were far less frequent. Delight was observed 7.1% of the time (much higher than typically seen in intelligent tutors), and confusion was recorded in 3.1% of the observations. Frustration (0.9%), boredom (0.5%), disgust (0.4%) and sorrow (0.1%) accounted for less than 2% of the data combined. The remaining 20.9% of the observations were labeled with the "?" that BROMP coders use when another affective state is being presented, when a student's affective state is ambiguous, or when the student otherwise cannot be observed. In this study, many of these cases involved the teachers pausing EcoMUVE activity for lecture or asking students to get out of their seats for group activities.

3.2. Creation of Affect Models

Log file data was synchronized with BROMP data, so that each 20-second period preceding the entry of an observation (termed a clip) was tagged with the corresponding affect and behavior labels. Models were constructed at the clip level for the five most common affective states (boredom, confusion, delight, engaged concentration, and frustration).

Features were distilled from available information within EcoMUVE's log files. As in previous research of affect and other educationally relevant constructs, features included specific descriptions of individual actions (e.g. picking up a particular object), classification of different actions by types (e.g. picking up similar objects), information about whether or not an action was novel or repetitive, and temporal information. As with previous investigations of virtual environments, they also included information about EcoMUVE's virtual locations (e.g. whether an action was completed in the submarine or near the pond) and interactions between students.

Attempts were made to fit each detector using six common classification algorithms (i.e., K* JRip, J48, REPTree, Bayesian Logistic Regression, and Linear Regression), which are representative of a variety of different patterns but are less susceptible to over-fitting than many other algorithms.

Features for machine learning algorithms were chosen using forward selection, an iterative process in which features are added individually. At each iteration, the feature that most improves model goodness is added; this process continues until model performance no longer improves. In this study, cross-validated Cohen’s (1960) Kappa, which scales from -1 to 1, was used as the goodness metric during feature selection. Features that performed below chance in single-feature models (Kappa ≤ 0) were excluded prior to this process in order to reduce the chance of over-fitting.

Detectors were evaluated at the student level using 5-fold cross-validation (e.g. detectors were trained on data from four student groups and tested on data from a fifth). In addition, students were stratified into fold assignments based on their training labels, guaranteeing a representative number of majority and minority class observations in each fold. After the creation of each fold, an alternate version of each training fold was created through resampling so that an equal number of examples where the construct was present or absent occurred in each fold. In this process, clips that contain the construct being detected were duplicated in order to artificially increase that construct’s frequency within the training set. However, in order to ensure validity, model performance was always tested on data that had not been resampled.

In addition to Cohen’s Kappa, which was also applied during the forward selection process, A' was used to assess final detector performance and select the optimum algorithm for each detector. A' scales from 0 to 1 (chance = 0.5) and assesses the probability that the detector will correctly identify whether a specific affective state is present or absent in a specific clip. A' is equivalent to W, the Wilcoxon statistic, and closely approximates the area under the Receiver-Operating Curve [14]. Because current implementations of AUC ROC available in data mining and statistics packages over-estimate goodness for the special case where multiple data points have the same confidence, A’ was calculated using software available at http://www.columbia.edu/~rsb2162/computeAPrime.zip.

4 Results

Each of the five detectors constructed for EcoMUVE performed better than chance under cross-validation. In particular, Kappa values for these models were generally comparable to values seen for sensor-free affect detection in recent papers [e.g. 6, 22, 26], but A’ was somewhat lower than the values seen in [22].

The best detector of Boredom used JRip, and achieved a Kappa of 0.31 and an A’ of 0.65. It relied upon five features: (1) a normalized metric of student speed based on a calculation called TimeSDtype (see below), (2) the largest amount of time between two actions in a clip, (3) the number of zoom changes within the submarine, (4) the number of times the student has viewed the data he or she has collected, and (5) and the number of player actions in the clip.

Table 1: Features used in boredom detector

Boredom	
B1	Across the last 5 actions, the sum of TimeSDType. For each student action, TimeSDType is the degree to which the current student action is faster or slower than the average action by all students involving the same type of action (e.g. air temperature measurements or entering the submarine), in standard deviations faster (-) or slower (+) than the average.
B2	The largest amount of time between two actions during the clip (not considering whether or not the student logged out between actions).
B3	The total number of *zoom* changes within the submarine so far.
B4	The total number of times the student used the *view data* application so far.
B5	The number of student actions in the clip.

The best detector of Confusion used J48 and achieved a Kappa of 0.23 and an A' of 0.60. It relied upon five features: (1) the ratio between the number of times the student has viewed his or her data and taken measurements, (2) the number of times the student has repeated the same measurement in the current zone, (3) the total number of measures that the student has taken so far, (4) the average number of characters in each text chat, and (5) the number of player actions in the clip.

Table 2: Features used in confusion detector

Confusion	
C1	The ratio between the number of times the student *viewed data* up until the current action, and the number of times the student took measurements up until the current action.
C2	The number of times the student has repeated the same measurement in the current zone so far.
C3	The total number of measures that the student has taken so far.
C4	The average number of characters in each text chat the player engaged in during the clip. (Actions other than chats are counted as 0 characters).
C5	The number of student actions in the clip.

The best detector of Delight used Bayesian Logistic Regression, achieving a Kappa of 0.19 and an A' of 0.62. It relied upon six features: (1) the number of student actions involving plants, (2) the percentage of photos not followed by accessing the relevant species page, (3) a largest number of measures that the student has taken per trip to each zone, (4) the ratio between the amount of time spent in the submarine and the number of measures taken in it, and (5) the largest value of the second feature in this model, and (5) the number of student actions in the clip.

Table 3: Features used in delight detector

Delight	
D1	The total number of student actions involving plants so far.
D2	The percentage of photographs a student takes without immediately accessing the relevant species page.
D3	The number of measurements taken per student trip to each zones so far. Then the largest value of this feature at any point in the clip is taken.
D4	The ratio between the amount of time spent in the submarine and the number of measurements taken in the submarine
D5	The percentage of photographs a student takes without immediately accessing the relevant species page. Then the largest value of this feature at any point in the clip is taken.
D6	The number of student actions in the clip.

The best detector of Engaged Concentration used J48, achieving a Kappa of 0.24 and an A' of 0.56. It relied upon seven features: (1) a normalized measure of how fast or slow student actions are based on their peers' typical response time for the same kind of action, (2) the amount of time the student has spent using EcoMUVE during the real-world day, (3) the amount of time the student has spent interacting with NPCs, (4) the amount of time the student has interacted with plants, (5) another normalized measure of how fast or slow student actions are, (6) the number of photos taken in the real-world day, and (7) the number of student actions per clip.

Table 4: Features used in engaged concentration detector

Engaged Concentration	
E1	Across the last 3 actions, the sum of TimeSDType. (See definition of TimeSDType in boredom detector features, Table 1.)
E2	The total time the student spent using EcoMUVE so far in the real world day, as calculated from the beginning of the clip.
E3	The total time spent by the player interacting with NPCs so far.
E4	The total amount of time spent by the player interacting with plants.

E5	Across the last 5 actions, the sum of TimeSDObjectType. For each student action, TimeSDObjectType is the degree to which the current player action is faster or slower than the average action by allstudent involving the same type of action, but only for actions involving interaction with an object/animal/plant (e.g. taking a photo of an animal), in standard deviations faster (-) or slower (+) than the average. Then, the largest value of this feature at any point in the clip is taken.
E6	The total photographs taken in the current, real-world day so far.
E7	The number of student actions in the clip.

The best detector of Frustration used K*, achieving a Kappa of 0.27 and an A' of 0.65. It relied upon seven features: (1) the average amount of time the student has taken to read a species page, (2) the average time per measure in the submarine, (3) the smallest values of the second feature, (4) the ratio between the time spent reading (for the first time) and rereading a species page, (5) the number of player actions in the clip, (6) the time spent in the submarine divided by the number of measurements taken during that time, and (7) the time per field guide access divided by the number of times a student has accessed a species page for the first time.

Table 5: Features used in frustration detector

Frustration	
F1	The average amount of time a student takes to read a species page.
F2	The average time per measurement taken in the submarine.
F3	The average time per measurement taken in the submarine. Then the smallest value of this feature at any point in the clip is taken.
F4	The ratio between the total time a student spent reading species pages and the number of times he or she re-reads a species page for the second or subsequent time.
F5	The number of student actions in the clip.
F6	Time spent in submarine (*from enter submarine* to next *enter zone*), divided by number of measurements taken from within submarine.
F7	Time per field guide access (from opened to closed), divided by total number of species page accesses—but *only* for cases where student is accessing a species page for the first time.

As can be seen, there were some key commonalities between the features utilized by the different models. Most notably, the number of actions that a student made within the 20-second clip being examined was included in the detector of every single construct. This feature gives some information about the speed with which students are working within the system. Temporal features were generally important, forming part of the model even beyond this feature for engaged concentration, frustration, and boredom. Student measurements played a prominent role for confusion and frustration – and also, somewhat surprisingly, for delight. In particular, repeating the same measurement was an indicator of confusion. Students who access the data without regard to the measurements they are taking are more likely to be bored, while those who are doing so frequently, though with more purpose, are more likely to be confused. Similarly, actions within the submarine were indicators of both positive and negative emotions (delight, frustration, and boredom). Students who took many photos, but who did not follow them up by reading the species page, were more likely to be delighted. Text chat with other students was an indicator of confusion, while interaction with the game's non-player characters was an indicator of engaged concentration.

5 Discussion and Conclusions

In this paper, we present five sensor-free models of educationally-relevant affective states for the virtual environment, EcoMUVE, a multi-user virtual environment (MUVE) for learning about ecosystems. In recent years, it has been demonstrated that affective models can be developed for a range of online learning environments. To our knowledge, this is the first paper demonstrating that sensor-free affect detectors can be developed for a MUVE without changing the student experience in any way. In the prior work on affect detection in MUVEs [26], questionnaire data was incorporated into the models, and the student experience was changed in order to develop models, with students completing pop-up surveys on their affect. We extend this pioneering work by developing models using non-intrusive BROMP field observations, and develop models that can make decisions using no data other than the unmodified interactions between the student and the learning system.

In particular, we believe that our model of delight makes an important contribution to the nascent area of sensor-free, automated affect detection in online learning systems, since, to the best of our knowledge, this is the first time that a cross-validated sensor-free model of this construct has performed above chance for the original data distribution. While delight is less common in the intelligent tutoring systems within which much of the work on sensor-free affect has taken place, it is prominent within game-like environments, as shown by its relatively high frequency in the EcoMUVE data. As this area of research advances and suites of sensor-free affect detectors become available for more systems, delight will likely prove to be an important indicator of engagement.

The resultant models presented here achieve an average performance of Kappa = 0.25 and A' = 0.61, an average Kappa higher than that seen in [26] and not far below the values seen for sensor-free affect detectors developed for intelligent tutoring systems [cf. 6, 22]. While considerable room for improvement remains, it is worth noting that detectors of comparable goodness were recently developed for ASSISTments using BROMP observations and similar data mining techniques [CITE]. ASSISTments' affect detectors have since been successful at predicting long-term learning outcomes for middle-school students, including success on state standardized exams [22] and even which students will attend college several years after using a learning system [27].

Even without further refinement, these detectors should be sufficient for the development of fail-soft interventions that can be implemented without interrupting the learning process and for discovery with models research, such as the approach used to make long-term predictions in ASSISTments. Further study of EcoMUVE log files using these detectors is also likely to provide insight into what aspects of the learning system are most boring, frustrating, or confusing, supporting the development of design changes to make EcoMUVE more engaging and effective.

Acknowledgements. Portions of this material are based upon work supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A080514 to Christopher J. Dede and Tina A. Grotzer. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not represent the views of the Institute or the U.S. Department of Education. This research was also supported by grant (#OPP1048577) from the Bill & Melinda Gates Foundation. Thanks also to Belinda Yew for help in editing and to the teachers and students who made this study possible.

6 References

1. Alzoubi, Omar, Rafael A. Calvo, and Ronald H. Stevens. "Classification of eeg for affect recognition: an adaptive approach." *AI 2009: Advances in Artificial Intelligence*. Springer Berlin Heidelberg, 2009. 52-61.
2. Arroyo, I., Cooper, D. G., Burleson, W., Woolf, B. P., Muldner, K., & Christopherson, R. "Emotion Sensors Go To School." *AIED*. Vol. 200. 2009.
3. Arroyo, Ivon, Beverly Park Woolf, David G. Cooper, Winslow Burleson, and Kasia Muldner. "The impact of animated pedagogical agents on girls' and boys' emotions, attitudes, behaviors and learning." In *Advanced Learning Technologies (ICALT), 2011 11th IEEE International Conference on*, pp. 506-510. IEEE, 2011.
4. Baker, Ryan SJD, Sujith M. Gowda, Albert T. Corbett, and Jaclyn Ocumpaugh. "Towards automatically detecting whether student learning is shallow." In *Intelligent Tutoring Systems*, pp. 444-453. Springer Berlin Heidelberg, 2012.
5. Baker, R.S.J.d., Mitrovic, A., Mathews, M.: Detecting Gaming the System in Constraint-Based Tutors. Proceedings of the 18th Annual Conference on User Modeling, Adaptation, and Personalization, (2010), 267-278.

6. Baker, R.S.J.d., Gowda, S.M., Wixon, M., Kalka, J., Wagner, A.Z., Salvi, A., Aleven, V., Kusbit, G., Ocumpaugh, J., Rossi, L. Towards Sensor-free Affect Detection in Cognitive Tutor Algebra. *Proceedings of the 5th International Conference on Educational Data Mining*, (2012), 126-133.
7. Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. "Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments." *International Journal of Human-Computer Studies* 68.4 (2010): 223-241.
8. Calvo, R. A., & D'Mello, S. "Affect detection: An interdisciplinary review of models, methods, and their applications." *Affective Computing, IEEE Transactions on* 1.1 (2010), 18-37.
9. Cohen, J.: A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), (1960), 37-46.
10. Craig, S., Graesser, A., Sullins, J., & Gholson, B. "Affect and learning: an exploratory look into the role of affect in learning with AutoTutor." *Journal of Educational Media* 29.3 (2004), 241-250.
11. Conati, C., & Maclaren, H. "Empirically building and evaluating a probabilistic model of user affect." *User Modeling and User-Adapted Interaction* 19.3 (2009), 267-303.
12. D'Mello, S. K., Craig, S.D., Witherspoon, A. W., McDaniel, B. T., and Graesser, A. C. Automatic Detection of Learner's Affect from Conversational Cues. *User Modeling and User- Adapted Interaction*, 18(1-2), (2008), 45-80.
13. D'Mello, Sidney, and Art Graesser. "The half-life of cognitive-affective states during complex learning." *Cognition & Emotion* 25.7 (2011), 1299-1308.
14. Hanley, J., McNeil, B.: The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. Radiology, 143 (1982), 29-36.
15. Ketelhut, Diane Jass, et al. "A multi□user virtual environment for building and assessing higher order inquiry skills in science." *British Journal of Educational Technology* 41.1 (2010), 56-68.
16. Lehman, B. A., D'Mello, S. K., Strain, A., Millis, C., Gross, M., Dobbins, A., Wallace, P., Millis, K., & Graesser, A. C. Inducing and tracking confusion with contradictions during complex learning, International Journal of Artificial Intelligence in Education, 22(2), (2013), 85-105.
17. Litman, Diane J., and Kate Forbes-Riley. "Recognizing student emotions and attitudes on the basis of utterances in spoken tutoring dialogues with both human and computer tutors." *Speech communication* 48.5 (2006), 559-590.
18. Liu, Z., Pataranutaporn, V., Ocumpaugh, J., & Baker, R. S. Sequences of Frustration and Confusion, and Learning.*Proceedings of the 6th International Conference on Educational Data Mining*, (2013), 114-120.

19. Metcalf, Shari, Amy Kamarainen, M. Shane Tutwiler, Tina Grotzer, and Chris Dede. "Ecosystem science learning via multi-user virtual environments." *International Journal of Gaming and Computer-Mediated Simulations* 3, no. 1 (2011), 86.
20. Ocumpaugh, J., Baker, R.S.J.d. and Rodrigo, M.A. "Quantitative Field Observation (QFOs) Baker-Rodrigo Observation Method Protocol (BROMP) 1.0 Training Manual version 1.0 Oct. 17, 2012." (2012).
21. Pantic, Maja, and Leon JM Rothkrantz. "Toward an affect-sensitive multimodal human-computer interaction." *Proceedings of the IEEE* 91.9 (2003): 1370-1390.
22. Pardos, Z. A., Baker, R. S., San Pedro, M. O., Gowda, S. M., & Gowda, S. M. "Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes." In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, pp. 117-124. ACM, (2013).
23. Planalp, S., DeFrancisco, V.L., Rutherford, D. Varieties of Cues to Emotion in Naturally Occurring Settings. Cognition and Emotion, 10 (2), (1996), 137-153.
24. Rodrigo, M.M.T., Baker, R.S.J.d. Comparing Learners' Affect While Using an Intelligent Tutor and an Educational Game. *Research and Practice in Technology Enhanced Learning*, 6 (1), (2011), 43-66.
25. Rodrigo, M.M.T., Baker, R.S.J.d. Coarse-Grained Detection of Student Frustration in an Introductory Programming Course. *Proceedings of ICER 2009: the International Computing Education Workshop*. (2009).
26. Sabourin, J., Mott, B., & Lester, J. C. Modeling learner affect with theoretically grounded dynamic bayesian networks. In *Affective Computing and Intelligent Interaction* Springer Berlin Heidelberg. (2011), 286-295.
27. San Pedro, M.O.Z., Baker, R.S.J.d., Bowers, A.J., Heffernan, N.T. Predicting College Enrollment from Student Interaction with an Intelligent Tutoring System in Middle School. *Proceedings of the 6th International Conference on Educational Data Mining*, (2013), 177-184.
28. San Pedro, M.O.Z., Baker, R.S.J.d., Gowda, S.M., Heffernan, N.T. Towards an Understanding of Affect and Knowledge from Student Interaction with an Intelligent Tutoring System. *Proceedings of the 16th International Conference on Artificial Intelligence and Education*, (2013), 41-50.
29. Sebe, Nicu, Ira Cohen, Theo Gevers, and Thomas S. Huang. "Multimodal approaches for emotion recognition: a survey." In *Electronic Imaging 2005*. International Society for Optics and Photonics, (2005), 56-67.
30. Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *31*(1), (2009). 39-58.