Modeling Student Affect and Behavior using Biometric Readings, Log Files and Low Fidelity Playbacks

Vincent Raymond C. Castillo[1]
uowp@yahoo.com

Kathrina Blanca V. Villaflor[1]
kathrinavillaflor@yahoo.com

Ramon L. Rodriguez[1]
ramrod_aul@yahoo.com

Ma. Mercedes T. Rodrigo[1]
mrodrigo@ateneo.edu

ABSTRACT
Affective computing is computing that relates to user emotion, feelings, moods, temperament and motivation. One of its core problems that it tries to address is the automatic detection of user affect. In this paper, attempts were made to develop models of affective and behavioral states that users exhibit and experience while using Aplusix, an intelligent tutoring system for Algebra. To this end, we gathered both user interaction log data and biometrics data from first year Information Technology students at the Mapua Institute of Technology. We synchronized both logs, cut them into time frames, and labeled them following rules that we formulated for identifying the specific states of interest. We then used two supervised learning algorithms, J48 decision tree and logistic regression, to model student affect and behavior based on log files. We focused on modeling the affective states of boredom, flow and confusion, and on-task and off-task behavior. Given our data set, logistic regression resulted as the more accurate model due to better correlation as compared to J48.

Keywords
Biometrics, log file analysis, affect, behavior, data mining, intelligent tutoring systems, Aplusix.

1. INTRODUCTION
Affective computing is computing that relates to user emotion, feelings, moods, temperament and motivation [12]. One of its core problems that it tries to address is the automatic detection of user affect.

In recent years, researchers have attempted to model user affect using indicators such as biometrics [6], log files [5], human observations [13] and combinations thereof. In our study, we attempted to model the student boredom, flow, and confusion and on-task and off-task behaviors based on biometrics and log files we gathered as the students used Aplusix, an intelligent tutoring system for algebra.

The paper is organized as follows: Section 2 discusses the related literature related to biometric systems, log file analysis, and software used. Section 3, discusses the methodology. The analysis and presentation of data are described in Section 4. Section 5 includes the discussion.

2. RELATED LITERATURE
This review of related literature gives an overview of the two types of data that we use in this study: biometrics and log files.

2.1 Biometrics
Biometrics refers to the measurable or biological characteristics that identify a person [11]. They are used for identity verification, as well as the recognition of emotion or behavior. The latter process is achieved by combining human judgments with readings from hardware sensors that gather data from a person's physical characteristics or body movement [14]. The data can then be used to arrive at patterns that can be associated with different types of emotions.

A number of affective computing researchers have made use of biometrics to identify user affect while using learning software. Conati et al [6] used muscle movement and skin conductance as indicators of children's affective states while using the math game PrimeClub. Based on these findings, the researchers were able to determine frustration and surprise. Frustration was conveyed when the person frowns and surprise when eyebrows were raised. The study also found that internal arousal weakens when users voice out their opinions and feelings towards others.

Asteriadis et al [1] tracked facial muscle movements and hand gestures in an personalized reading environment to determine six general user states which are Frustration, Not paying attention, Attentive, Tired/sleepy, Distracted and Full of interest. The study made use of a web camera to capture images of the learner and used facial feature detection to identify the state based from the position of the feature points such as the mouth, eyebrows and eyes. This data, along with gaze vector and inter-ocular distance was analyzed by a Sugeno-type system to generate a model that estimates user attention. Results showed that the performance of the model for both attentive and non-attentive states were 87.7% accurate.

2.2 Log files and log file analysis
User interactions with a computer-based system can be recorded for analysis in the form of log files [9]. Log files can be analyzed while the person is using the software (real-time analysis) or afterwards (post-hoc analysis). Intelligent Tutoring Systems (ITS) use real-time analysis to model student knowledge as the student interacts with the system. The model is then used to determine what to present to the user next and how best to do so.

Several studies made use of post-hoc analysis. Cocea and Weizbelzahl [5] were able to determine if a learner was engaged or not by analyzing log files produced by HTML Tutor, an interactive learning environment that teaches web publishing. The study wanted to provide evidence that log files can provide sufficient information about a learner's motivational level from interpreting commonly-logged data. With the use of C4.5 algorithm, they were able to determine that average learners who spent forty-five minutes on the tutor are moderately motivated and continue to read and take tests. Once they realize that they have enough knowledge less motivation is exerted in learning. They were unable to determine a learner's motivation within the first forty-five minutes or determine the learner's goal orientation before using the tutor due to the limited set of log samples.

The study by Stoica, et al [13] was able to interpret user activities of an image game and a text game such as time completion, on a mobile device. The study made use of the coIAT tool [2] which interrelates log files generated by the software with video captures and observation notes. Although the main goal of this study was to create a prototype for group collaborations and does identify affective states, the tool was able to prove that a logged action viewed using various media such as audio and video provides a clearer interpretation of user's behavior compared to the use of log files alone.

Baker et. al [3] used text replays of logs to classify whether or not a student was gaming the system. Gaming the system is defined as exploitation of system regularities in order to progress through the curriculum without learning. Examples of gaming the system include systematic guessing and hint abuse. Text replays are text-only playbacks of user interactions with a system. Baker found out that labeling data using text replays is up to 40 times faster that using field observation or high-fidelity replays, eg. replays using video and other media. Training of data mining algorithms with labels generated using text replays led to better classifiers than those trained with quantitative field observation data.

For our study, we made of biometrics data and log files as inputs for our analysis.

3. DESCRIPTION OF SOFTWARE USED

Aplusix [4] is educational software that teaches basic algebra. Its interface as seen in Figure 1 comprises of a mathematical equation at the start of the program where the user is free to manipulate the question until the solution has been derived. As a student derives a new solution, the equation moves to a new line situated at the bottom of the original equation. The vertical parallel lines indicate that the user is leading to the correct solution of the expression. When the lines are black, that means that the steps are equivalent. When the lines are red and have an X, the steps are not equivalent.

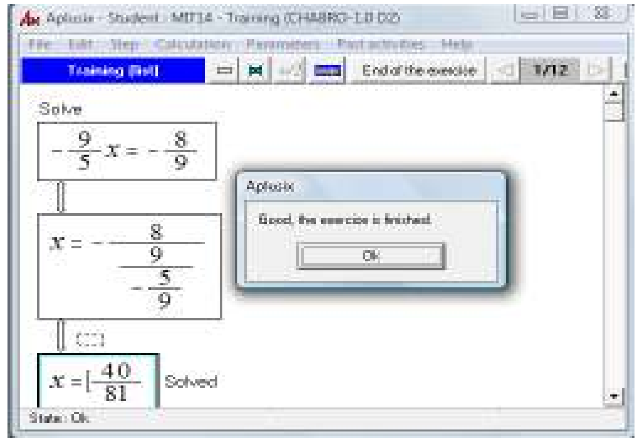


Figure 1. The Aplusix environment.

Brainfingers [10] (Figure 2) is a software / hardware combination that measures biometrics signals. The package comprises of a headband, an interface box and its software. The headband is worn on a person's forehead and serves as a sensor that detects electronic signals from the skin. These signals are divided into three – electrooculography (EOG), which is used to record eye movements, electroencephalography (EEG), which responds to electric activity within the brain, and electromyography (EMG), which detects facial muscle activity. Brainfingers has been used as an alternative input device for game control as well as a computer interface for people with disabilities. Ours is the first study that uses Brainfiners to detect user affect and behavior.

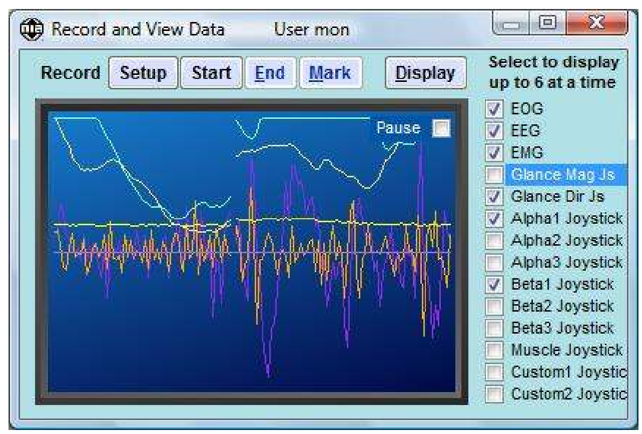
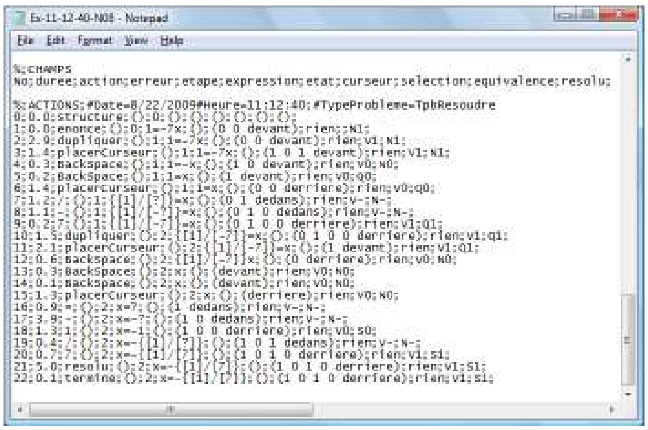


Figure 2. The Brainfingers environment.

4. METHODOLOGY

The experiment was conducted at the MAPUA Institute of Technology during the first term of the quarter semester of school year 2009 - 2010. Twelve freshman students from Information Technology volunteered to be study participants. They were composed of six females and six males and their ages ranged from 15 to 18 years old. All students already have basic knowledge on algebra. Out of the twelve students, biometrics data from only 10 students was usable. Biometrics data from the remaining two students was not recorded for undetermined reasons.

Because we only had two sets of Brainfingers, we could only record data from two participants at a time. Each pair of students was given a brief introduction covering the purpose of the research before starting. A pre-test followed which contained linear equations from basic algebra. Each was given ten (10) minutes to solve the questions on the pre-test. When the students were finished, we asked them to wear the Brainfingers headbands. After the headbands were calibrated, we began recording the students' biometrics. We gave the students one Aplusix manual each. We then instructed them to use Aplusix for 40 minutes. During the interaction, Aplusix recorded student actions into logs. After the interaction with the software, the student took a 10-minute post-test. Figure 3 shows an excerpt of a log file recorded by Aplusix. The data fields are presented at the middle of the file under the **%;CHAMPS** header and separated by commas while their corresponding data based on the order of the fields are displayed under the **%;ACTIONS** header. Not all the data was of interest to us at this point. Table 1 enumerates the different data fields that we filtered from the data and their corresponding descriptions.


```
Ex-11-12-40-N08 - Notepad
File Edit Format View Help

%;CHAMPS
No;duree;action;erreur;etape;expression;etat;curseur;selection;equivalence;resolu;

%;ACTIONS;#Date=8/22/2009#Heure=11:12:40;#TypeProbleme=TpbResoudre
0;0.0;structure;();0;();();();();();();
1;0.0;enonce;();0;1=-7x;();(0 0 devant);rien;;N1;
2;2.9;dupliquer;();1;1=-7x;();(0 0 devant);rien;v1;N1;
3;1.4;placerCurseur;();1;1=-7x;();(1 0 1 devant);rien;V1;N1;
4;0.3;BackSpace;();1;1=-x;();(1 0 devant);rien;V0;N0;
5;0.2;BackSpace;();1;1=x;();(1 devant);rien;V0;Q0;
6;1.4;placerCurseur;();1;1=x;();(0 0 derriere);rien;v0;q0;
7;1.2;/;();1;{[1]/[?]}=x;();(0 1 dedans);rien;v-;N-;
8;1.1;-;();1;{[1]/[-?]}=x;();(0 1 0 dedans);rien;v-;N-;
9;0.2;7;();1;{[1]/[-7]}=x;();(0 1 0 0 derriere);rien;V1;Q1;
10;1.5;dupliquer;();2;{[1]/[-7]}=x;();(0 1 0 0 derriere);rien;v1;q1;
11;2.1;placerCurseur;();2;{[1]/[-7]}=x;();(1 devant);rien;V1;Q1;
12;0.6;BackSpace;();2;{[1]/[-7]}x;();(0 derriere);rien;V0;N0;
13;0.3;BackSpace;();2;x;();(devant);rien;V0;N0;
14;0.1;BackSpace;();2;x;();(devant);rien;V0;N0;
15;1.3;placerCurseur;();2;x;();(derriere);rien;V0;N0;
16;0.9;=;();2;x=?;();(1 dedans);rien;v-;N-;
17;3.9;-;();2;x=-?;();(1 0 dedans);rien;v-;N-;
18;1.3;1;();2;x=-1;();(1 0 0 derriere);rien;V0;S0;
19;0.4;/;();2;x=-{[1]/[?]};();(1 0 1 dedans);rien;v-;N-;
20;0.7;7;();2;x=-{[1]/[7]};();(1 0 1 0 derriere);rien;v1;S1;
21;5.0;resolu;();2;x=-{[1]/[7]};();(1 0 1 0 derriere);rien;v1;S1;
22;0.1;termine;();2;x=-{[1]/[7]};();(1 0 1 0 derriere);rien;v1;S1;
```

Figure 3. Excerpt of a log file from Aplusix.

Table 1. Data fields used from the Aplusix log file.

Data field	Description
Duree	Duration between two actions that occurred.
Action	Registered keyboard presses, mouse clicks and problem initialization.
Expression	Current state of the equation/expression.
Equivalence	Classifies if the expression is visible or invisible.
Resolu	Classifies if the expression is solved, quasi-solved or not solved.

Figure 4 displays an excerpt of a log file recorded by Brainfingers. All field names are arranged on the first line under the **[Data]** header followed by the data on the succeeding lines. These labels were chosen from the software's settings before any data has been recorded, however much focus is on EEG, EOG and EMG. Other fields that we might find useful were included such as joystick values, glance directions and muscle movement.

Figure 4. Excerpt of a log file from Brainfingers.

At the end of the data gathering session, all log files from Aplusix and the biometrics data from Brainfingers were extracted and collated into a database using Microsoft Access. All data were arranged in a table and separated in 20-second segments. For each segment, the Aplusix logs were used to determine the affective and behavioral states of the user.

There were only two possible behaviors: on-task or off-task. On-task behavior refers to productive interaction with the software. Off-task behavior refers to the opposite.

For affective states, a segment could have one of three possible tags:

- Flow - user interacts with the system, showing an understanding of the task at hand while advancing its difficulty.
- Confused - user interacts with the system but shows difficulty understanding the current task.
- Bored - user does not show any interest in the task

Although it is possible for a student to experience more than one of these states in a 20-second time period, we only assigned one tag per clip for tractability.

Through a series of discussions, we arrived at the following rules to help us identify whether a student was demonstrating a particular affective state or behavior (see Tables 2 and 3).

Table 2. Rules in determining affect types.

Affect	Rules
Flow	An equation was solved, which was implied when a row's Action column was tagged as **resolu** followed by **termine**. The Resolu column was also tagged as **S1.**
	The Expression column continuously changed towards a possible solution
Confused	The Action column comprised mostly of entries that showed cursor movements such as **droite** (right), **gauche** (left), **haut** (up) and **bas** (down) without movement in the Expression column.
	The same equation occurred several times in the frame due to undoing and redoing of one part.
	The user almost reached a solved expression but the expression was not in the the proper form.

Bored	There were long intervals between actions which generated a value in the Duree column that was more than or equal to twenty (in seconds).
	The user moved on to the next problem without solving the current equation. The Action column was tagged as **termine** without the **resolu** entry prior to it.
	The Resolu column was tagged with either of the following: **N-**, **N0** or **N1**
	Continuous **backspace** entries in the Action column were present until the Expression column returned to the original equation.

There were several rules used to tag behavior that were somewhat associated with affect. Focus was given on the duration of actions because it was more likely that the people are off-task if they are doing activities not related to the software.

Table 3. Rules in determining behavior type.

On-task	Off-task
There were at least two recorded user actions in the segment.	There were less than two records of data or there was no data recorded.
	The next row recorded occurred more than 20 seconds later.
	Unrelated data or text was entered in the Expression column.

There were some special considerations used in tagging the affective and behavioural states. We observed during the experiment that the participants read the manual while using the software at the same time, hence generating large numbers in the log's Duree column. Instead of tagging these as bored and off-task, these were tagged as confused and on-task.

We then generated four data sets, one per state of interest: On-task/Off-task, flow/not in flow, bored/not-bored, confused/not confused. All rows that matched the state of interest were tagged with the word "yes" while the other states are tagged with the word "no".

5. ANALYSIS AND PRESENTATION OF DATA

The Waikato Environment for Knowledge Analysis (WEKA) [8] was used to analyze each data set using a tree-based and a function-based algorithm. We chose J48 decision tree algorithm and logistic regression for comparison because the two algorithms work with a nominal class and a combined set of nominal and ordinal features. Tree-based algorithms have characteristics that are possible to determine a state based from the given rules. Meanwhile, function-based algorithms can determine the probability of an affect or behavior to occur based from the data at hand.

Data that was analyzed by WEKA using both the J48 algorithm and logistic regression generated models for each affective and behavioral state. A sample of a J48 model can be seen in Figure 5. The model produced branches based on how the state matches the rules that have been identified in Tables 2 and 3. Figure 6 meanwhile shows a sample of a logistic regression for the same state where each variable displays the corresponding coefficients.

Both algorithms also produced a classified summary of cross-validation data that can be seen in Table 4. A detailed summary of the said data can be seen in Table 5.

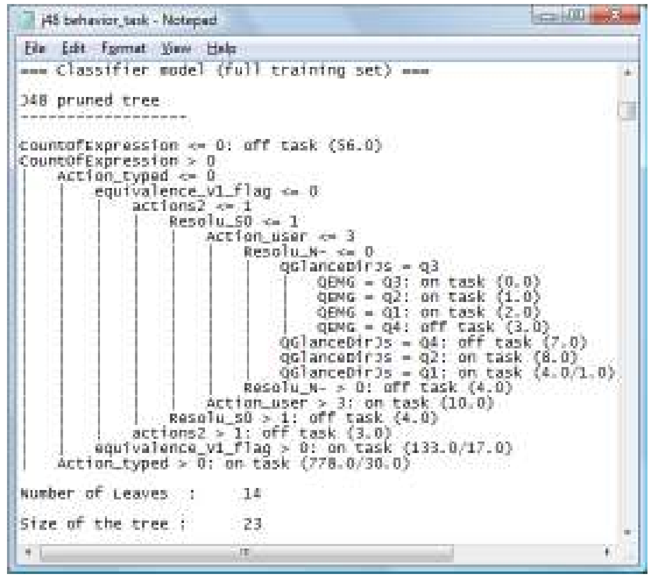


Figure 5. The J48 pruned tree model for Behavior.

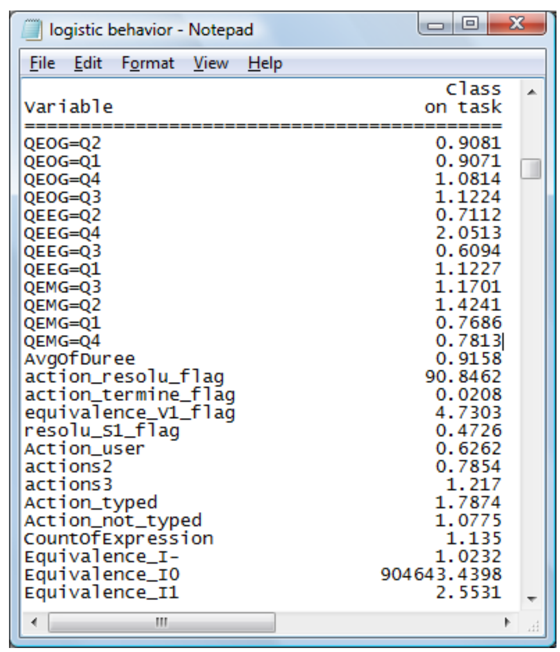


Figure 6. An excerpt of the Logistic Regression model for behavior.

The models generated by J48 and logistic regression used both features from the Brainfingers and Aplusix logs. In figure 5, the main branch feature used by the tree is the count of expression and its inner leaves use the quartiled EMG and Glance Direction joystick values. Each leaf of the tree is assigned with a class. For the behavior model, either "on task" or "off task" was used while for the affective models "yes" or "no" was used. The numbers in the parenthesis indicate the number of predictions assigned to the leaf, followed by the number of incorrectly classified instances. When the model was tested via cross-validation, the last leaf "Action_typed > 0" was used 778 times, but 30 of the predictions were incorrectly classified.

The logistic regression model meanwhile, generated a positive or negative coefficient for each feature in the dataset. For ordinal features, a coefficient was assigned directly. But for nominal features, each possible answer was given a coefficient. Figure 6 shows the logistic regression model for the Behavior state. Its QEOG variable has 4 possible values: Q1, Q2, Q3, and Q4. Each of these values has a different coefficient value from the others.

Table 4. WEKA stratified cross-validation summary

Class	Algorithm	Correctly Classified Instances	Incorrectly Classified Instances	Correctly Classified Instances %	Incorrectly Classified Instances %	Kappa statistic
Behavior	J48	935	78	92.30%	7.70%	0.5938
	LogReg	931	82	91.91%	8.09%	0.6041
Bored	J48	918	95	90.62%	9.38%	0.5038
	LogReg	924	89	91.21%	8.79%	0.5657
Confused	J48	752	261	74.23%	25.77%	0.32
	LogReg	773	240	76.31%	23.69%	0.3683
Flow	J48	782	231	77.20%	22.80%	0.5233
	LogReg	786	227	77.59%	22.41%	0.5305

Overall, logistic regression yielded better results versus J48. In terms of Behavior, the kappa value of 0.6041 for the model created by logistic regression is slightly higher than 0.5938 by J48. For Boredom, the model by logistic regression is higher, from 0.5657 compared to 0.5038 of the J48 version. In terms of Flow, logistic regression performed better with a kappa value of 0.5305 compared to a kappa value of 0.5233 for J48. For Confused, the model produced with logistic regression had a kappa value of 0.3683 which was better than the kappa value of 0.32 for J48.

Both algorithms were also able to classify more data correctly for all affective and behavioral states, although the incorrectly classified instances for the both the Confused and Flow states are slightly higher than the rest. These could be a result of the set fuzziness in the application of our labeling rules. Although we had rules that guided the labeling process, it was possible for two affective states to overlap because the user might have been experiencing either of two different affective states at the same time or successively within the same time period.

Table 5. Detailed Accuracy by Class

Class	Algorithm	Detailed Accuracy By Class	TP Rate	FP Rate	Precision	Class
Behavior	J48		0.976	0.456	0.938	on task
			0.544	0.024	0.764	off task
		Weighted Avg.	0.923	0.403	0.917	
	LogReg		0.963	0.392	0.946	on task
			0.608	0.037	0.697	off task
		Weighted Avg.	0.919	0.348	0.915	
Bored	J48		0.465	0.03	0.686	yes
			0.97	0.535	0.927	no
		Weighted Avg.	0.906	0.472	0.896	
	LogReg		0.559	0.037	0.683	yes
			0.963	0.441	0.938	no
		Weighted Avg.	0.912	0.39	0.906	
Confused	J48		0.439	0.139	0.553	yes
			0.861	0.561	0.797	no
		Weighted Avg.	0.742	0.442	0.728	
	LogReg		0.46	0.118	0.604	yes
			0.882	0.54	0.807	no
		Weighted Avg.	0.763	0.422	0.749	
Flow	J48		0.827	0.308	0.796	yes
			0.692	0.173	0.733	no
		Weighted Avg.	0.772	0.253	0.771	
	LogReg		0.835	0.311	0.797	yes
			0.689	0.165	0.742	no
		Weighted Avg.	0.776	0.251	0.774	

The Detailed Accuracy by Class presents the performance of the algorithm per possible answer. The number of "yes" answers greatly differs from the number of "no" answers, thus one answer could dominate the total number of correctly classified instances. True positives (TP Rate) are the correctly classified instances, while false positives (FP Rate) are the incorrectly classified instances.

The overall correctly classified instances for the Confused state are more than 75% for both algorithms. However, the number of TP Rate for the "yes" class is only 46%. This suggests that the "no" class dominated the classification.

6. CONCLUSION

This study concludes that it is possible to use interaction logs in conjunction with biometrics to create models for detecting human affect or behavior.

In choosing an algorithm for our data set, logistic regression is deemed to be a more effective algorithm against decision trees. Since using this algorithm is not unanimously consistent in producing favorable correlation for all types of states, we acknowledge that not all affective states can be precisely identified by the use of log files.

For future studies, we recommend adding more features to the analysis space. We recommend using fast correlation-based filtering (FCBF) to optimize the feature set. From the optimized feature set, it may be possible to generate more accurate models.

7. ACKNOWLEDGMENTS

We would like to thank Jean-Francois Nicaud of the Laboiratore d' Informatique de Grenoble letting us use the Aplusix. We thank Dr. Andrew Junker for his assistance and support with Brainfingers. We would also like to thank Ryan Baker for suggestions during our analysis. Lastly we would also like to acknowledge the efforts of Abe Esteban for helping us out during our observation and the administration of MAPUA School of IT for letting us use its resources to perform our research. This research undertaking was made possible by the Philippines Department of Science and Technology Engineering Research and Development for Technology Consortium under the project “Multidimensional Analysis of User-Machine Interactions Towards the Development of Models of Affect”.

8. REFERENCES

[1] Asteriadis, S., P. Tzouveli, K. Karpouzis and S. Kollias, “Estimation of behavioral user state based on eye gaze and head pose – application in an e-learning environment,” *Multimedia Tools and Applications*, 41(3), 469-493 (February, 2009).

[2] Avouris, N., G. Fiotakis, G. Kahrimanis, M. Margaritis and V. Komis, "Beyond logging of fingertip actions: analysis of collaborative learning using multiple sources of data," *Journal of Interactive Learning Research*, 18(2), 231-250 (April, 2007).

[3] Baker, R. S.J.d., de Carvalho, A. M. J. A. (2008) Labeling Student Behavior Faster and More Precisely with Text Replays. Proceedings of the 1st International Conference on Educational Data Mining, 38-47.

[4] Chaachoua, H., J. F. Nicaud, A. Bronner and D. Bohineau, “APLUSIX, a learning environment for algebra, actual use and benefits,” Available online, http://www.itd.cnr.it/telma/papers.php.

[5] Cocea, M., and S. Weibelzahl, "*Can log file analysis estimate learner's level of motivation?" Proc. of 14th Workshop on Adaptivity and User Modeling in Interactive Systems*, 32-35 (2006).

[6] Conati, C., R. Chabbal and H. Mclaren, “A Study of Using Biometric Sensors for Monitoring User Emotions in Educational Games.” *Proc. of Workshop on Modeling User Affect and Ac-tions: Why, When and How* (2003).

[7] Froschl, C., “User Modeling and User Profiling in Adaptive E-learning Systems,” Masters Thesis, Graz University of Technology (2005).

[8] Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten, “The WEKA Data Mining Software: An Update”, *SIGKDD Exploration*, 11(1) (2009).

[9] Hulshof, C., “Log File Analysis*.” Encyclopedia of Social Measurement,* Academic Press (2004).

[10] Mateo, J. and M. Faufel, “Computer Access Using Electric Signals from the Forehead: The Cyberlink in Action,” Available online, http://www.wright.edu/lwd/documents/computeraccess_mateo.pdf.

[11] Microsoft Encarta Dictionary, “Biometrics,” Available online, http://uk.encarta.msn.com/encnet/features/dictionary/DictionaryResults.aspx?lextype=3&search=biometrics

[12] Picard, R. W, “Affective Computing,” MIT Press, Cambridge, MA, (1997).

[13] Stoica, A., G. Fiotakis, J. S. Cabrerra, H. M. Frutos, N. Avouris, and Y. Dimitriadis, “Usability evaluation of a mobile device: a case study for a museum application,” *Proc. of PCI 2005* (November, 2005)

[14] Woolf, B., W. Burleson, I. Arroyo, T. Dragon, D. Cooper and R. Picard, “Affect-aware tutors: recognizing and responding to student affect,” *International Journal of Learning Technology*, 4(3/4), 129-163 (2009).

9. AUTHOR’S INSTITUTIONAL AFFILIATIONS

[1] Department of Information Systems and Computer Science, Ateneo de Manila University, Loyola Heights, Quezon City, Philippines, +63 (2) 426-6071