The Impact of System Feedback on Learners' Affective and Physiological States

Conference Paper · June 2010

DOI: 10.1007/978-3-642-13388-6_31 · Source: DBLP

Omar AlZoubi
Jordan University of Science and Technology

Rafael A Calvo
The University of Sydney

The Impact of System Feedback on Learners' Affective and Physiological States

Payam Aghaei Pour[1], M. S. Hussain[1], Omar AlZoubi[1], Sidney D'Mello[2], and Rafael A. Calvo[1]

[1] School of Electrical and Information Engineering, University of Sydney, Australia
[2] Institute for Intelligent Systems, University of Memphis, Memphis, USA
{payama, sazzzad, oalzoubi, rafa}@ee.usyd.edu.au
sdmello@memphis.edu

Abstract. We investigate how positive, neutral and negative feedback responses from an Intelligent Tutoring System (ITS) influences learners' affect and physiology. AutoTutor, an ITS with conversational dialogues, was used by learners (n=16) while their physiological signals (heart signal, facial muscle signal and skin conductivity) were recorded. Learners were asked to self-report the cognitive-affective states they experienced during their interactions with AutoTutor via a retrospective judgment protocol immediately after the tutorial session. Statistical analysis (Chi-square) indicated that tutor feedback and learner affect were related. The results revealed that after receiving positive feedback from AutoTutor, learners mostly experienced 'delight' while surprise was experienced after negative feedback. We also classified physiological signals based on the tutor's feedback (Negative vs. Non-Negative) with a support vector machine (SVM) classifier. The classification accuracy, ranged from 42% to 84%, and was above the baseline for 10 learners.

Keywords: feedback, emotion, affective computing, multimodal interfaces, AutoTutor

1 Introduction

The connection between emotions and deep learning has recently received increased attention in the interdisciplinary arena that spans psychology, education [1, 2], neuroscience, and computer science [3-5]. Although several important questions pertaining to the affect-cognitive relationship during the learning process still remain unanswered [6], there is some evidence that the more typically studied six basic emotions [7] (happiness, sadness, surprise, disgust, anger and fear) are not the emotions most pertinent to learning, at least not for short learning sessions that last under 2 hours. Instead, affective states such as confusion, boredom, flow, curiosity, interest, surprise, delight, and frustration have emerged as highly relevant and influential to the learning experience; many of these states are frequently experienced during tutorial sessions with both Intelligent Tutoring Systems as well as human tutors [4, 8-11].

Inspired by the inextricable link between affect and learning, some researchers have worked to endow ITS with the ability to detect learners' affective states (e.g. confusion, frustration, etc.), respond to these states, and generate appropriate emotional expressions by embodied pedagogical agents. These affect-sensitive ITSs aspire to narrow the interaction bandwidth between computer tutors and human tutors with the hope that this will lead to an improved user experience and enhanced learning gains [12-14].

Accurate affect-detection is clearly an essential challenge to be overcome before functional affect-sensitive ITSs can become a reality. Although a number of studies have attempted to recognize learners' affect from facial expressions and speech [15-17], studies using physiological signals especially in educational contexts are relatively scarce. This is because, physiological sensors are often considered invasive and not suitable for learning environments as the sensors might interfere with the primary task of learning or problem solving. Fortunately, concerns pertaining to intrusiveness of the sensors, are somewhat less problematic with the recent advent of wearable physiological sensors [18]. Therefore, a re-investigation of the possibility of inferring a learner's affective state by monitoring physiological sensors is warranted. It should also be noted that although physiological responses to affective events has been a century long endeavor, most of the investigations traditionally studied the basic emotions, and little is known about the physiological manifestations of the learning-centered cognitive-affective states such as confusion and frustration.

Toward more real-world environment for physiological data collection, and the imperative to better understand the role of emotion in learning, we describe a study that collected physiological data to investigate the viability of detecting learning-centered affective states from physiology. We begin with a discussion of relevant affective computing research using physiological signals.

2 Recognizing affective states through physiology

Some important questions need to be addressed before functional physiological-based affect detectors for learning environments can be developed. Perhaps the most vital is whether distinct physiological patterns can be associated with particular emotions. Although the common answer is an enthusiastic "yes", the scientific research is much more controversial [19]. What is clear, however, is that some physiological correlations of the "basic" emotions can be identified more reliably than others. For example, fear has been related to an increase of heart rate, skin conductance level and systolic blood pressure [20], while anger has been related to an increase of heart rate, and both systolic and diastolic blood pressure [21]. In contrast, the physiology of sadness has proven to be difficult to pin-point physiologically, and has been associated with both an increase [22] and a decrease [20] in heart rate. Whether the more learning-centered emotions would prove to be as physiologically elusive as sadness or more consistent like anger awaits further research.

In addition to these challenges, the situation is more complicated because several of the studies investigating the physiology of emotion have adopted experimental protocols that have little relevance for ITSs, which have to operate in the real world. Pioneering work studying physiological states during ITS interactions by Conati was not conclusive on how these studies could be best performed [23]. More recently, recordings with one physiological sensor in naturalistic scenarios have been reported, so important progress is being made [24].

The current paper reports on a study that investigated the physiological embodiment in response to tutor feedback during learning sessions with AutoTutor, an ITS that provides conversational dialogues (described below). In addition to investigating the physiological correlates of the affective states we also focus on the feedback that the tutor provides to the learners. Focusing on feedback is critical because, in addition to being *directive* (i.e. tells students what needs to be fixed) and *facilitative* (i.e. helps students conceptualize information), feedback can also influence learners' affective and motivational processes [25].

The present study monitored learners' cognitive-affective states, their physiological signals, and the tutor's feedback during a 45 minute session with AutoTutor on topics in Computer Literacy [26]. AutoTutor is a validated ITS that helps students learn topics in Newtonian physics, computer literacy, and critical thinking via a mixed-initiative natural language dialogue. AutoTutor's dialogues are organized around difficult questions and problems (called main questions) that require reasoning and explanations in the answers. When presented with these questions, students typically respond with answers that are only one word to two sentences in length. In order to guide students in their construction of an improved answer, AutoTutor actively monitors learners' knowledge states and engages them in a turn-based dialogue. AutoTutor adaptively manages the tutorial dialogue by providing feedback on their answers (e.g. "good job", "not quite"), probing the learner for more information (e.g. "What else"), giving hints (e.g. "What about X"), prompts (e.g. "X is a type of what "),correcting misconceptions, answering questions, and summarizing topics.

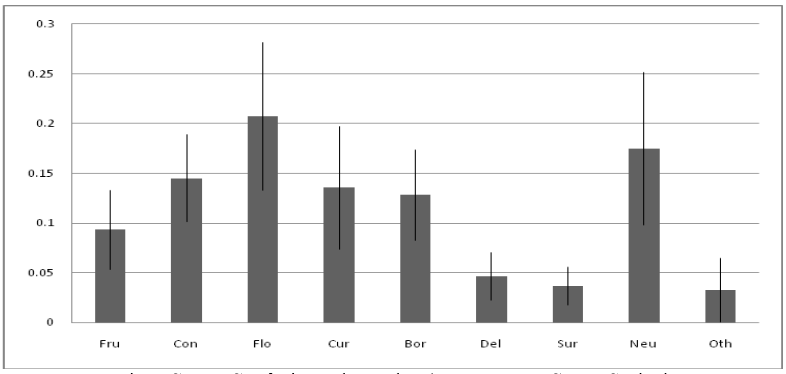
Although affect and learning connections have been explored in previous studies with AutoTutor [27], the current study focuses specifically on the physiological states of the learners. This topic has not been explored in previous studies with AutoTutor. In particular, the specific research questions that motivated the current study include: (a) What is the relationship between the tutor's feedback and the learner's self-reported affective states?, (b) What are the physiological correlates of these cognitive-affective states, and (c) How does the tutor's feedback influence the learners' physiological signals?

3 Data and Methods

The participants in this study consisted of 16 paid volunteers, most of whom were engineering students, from The University of Sydney. All participants signed an informed consent form approved by a Human Research Ethics Committee. The study typically lasted 2.5 hours for each participant.

Participants completed a 45 minute AutoTutor training session on one out of three randomly assigned topics in computer literacy (hardware, Internet, or operating systems). During this interactive session, a video of the learner's face and a video of his or her computer screen were recorded. In addition, three physiological sensors measured heart activity (electrocardiogram - ECG), face muscle activity (electromyogram - EMG), and galvanic skin response recorded from the tip of the fingers (GSR). The physiological signals were acquired using a BIOPAC MP150 system with AcqKnowledge software at a sampling rate of 1,000 Hz for all channels. ECG was collected with two electrodes from the wrists and a ground on the left ankle. Only one channel of EMG was recorded from a corrugator muscle with two electrodes each 2 cm apart. GSR was recorded from the index and middle finger (left hand).

Learners retrospectively provided self-report affect judgments immediately after their AutoTutor session. The collection of these self-reported judgments proceeded in the following way. A learner would view a video of himself/herself during the interaction simultaneous with a video capture of his/her screen interaction with AutoTutor. The videos automatically paused every 20 seconds at which point learners were asked to select one o r more affective states from a list of eight states, in addition to an "other" category. These states were: confusion, frustration, boredom, flow/engagement, curiosity, surprise, delight, and frustration. Fig. 1 reflects the proportional values and 95% confidence interval for the affective states reported by all learners. Where more than one state was selected, the learner was then asked to indicate the most pronounced state. The current analysis only considers the more pronounced affective state for each 20-second block.



Notes. Fru = Frustration, Con = Confusion, Flo = Flow/engagement, Cur = Curiosity, Bor = Boredom, Del = Delight, Sur = Surprise, Neu = Neutral, Oth = Other

Fig. 1. Proportional values with 95% confidence interval for affective states reported by all learners.

4 Results and Discussion

AutoTutor's short feedback (positive, neutral, negative) is manifested in its verbal content, intonation, and a host of other non-verbal conversational cues. . Positive feedback accompanies correct answers, negative feedback incorrect answers, while the tutor provides neutral feedback when the student's answer lies between these two extremes. Although the feedback selection mechanism is more sophisticated that this simple description suggests, of greatest relevance is the fact that the feedback evokes strong emotional responses from participants [6, 8]; this emotional-elicitation quality of the tutor's feedback is very relevant for the present paper.

Our analyses are organized around three questions: (a) how tutor feedback impacts the affective state of the student (based on the self-report)?, (b) how tutor feedback influences the physiological state of the students?, and (c) how the self-reported affective states correlate with physiology. These three questions were addressed by constructing supervised classifiers. The assumption is that if a classification model with accuracy higher than a baseline can be built, then the affective response to a stimulus is not random.

Self-report annotations were synchronized with AutoTutor's feedback type (mined from AutoTutor's logs) and with corresponding physiological signals. Then, 20-second blocks for the affect annotations and 10-second blocks after the feedback were extracted from the physiological signals.

The Waikato Environment for Knowledge Analysis (Weka), a data mining package [28], was used for the classification of the three pairs of data described here, all based on a 10-fold cross validation. The default parameter values were used for classification in this study. The ZeroR classifier represents the baseline for classification accuracy; differences in baseline accuracy are based on differences in the class distributions of individual learners. Support vector machine (SVM) classifier with a linear kernel was utilized for training classification models which is based on John Platt's sequential minimal optimization algorithm for training a support vector machines classifier [29]. A feature selection algorithm was used to reduce the dimensionality of the physiological data. A chi-square (X^2) feature selection as implemented in Weka was used for selecting the ten most relevant features. The X^2 feature selection evaluates attributes by computing the value of the chi-squared statistic with respect to the class label either feedback or self-report emotion [30].

4.1 Feedback and Affect

Affective states of all learners are significantly dependent with AutoTutor feedback. A 5 x 9 Chi-square (X^2) analysis revealed this dependency. A chi square value of approximately 165.0 with 32 degrees of freedom ($\alpha = 0.05$) was obtained. Table 1 shows the contingency table of all feedback types and self-reported affective states for all learners.

Table 1. A contingency table of 16 learners for all feedback types and affective states (self-report)

	Affect								
Feedback	Fru	Con	Flow	Cur	Bor	Del	Sur	Neu	Oth
Positive	7	7	24	12	19	39	2	27	5
Neutral Positive	2	12	5	7	2	1	1	7	1
Neutral	12	22	29	25	30	22	3	34	1
Neutral Negative	8	5	7	3	6	0	4	3	0
Negative	57	65	60	41	45	4	29	50	16

Notes. Fru = Frustration, Con = Confusion, Flow = Flow/engagement, Cur = Curiosity, Bor = Boredom, Del = Delight, Sur = Surprise, Neu = Neutral, Oth = Other

Using the data for all learners, the five different feedback types produced by AutoTutor were used to classify the self-reported affective states. The frequency of each cognitive-affective state varied: Boredom (102), confusion (111), Delight (66), Frustration (86), Neutral (121), Curiosity (88), Frustration (86), Flow (125), Surprise (39). Table 2 gives the classification results for affective states that had accuracy above the ZeroR baseline. As a result the study looked at pairs of cognitive-affective states that can best be separated, or in other words, determining the effect of system feedback on the cognitive-affective state of the learner using only pairs of cognitive-affective states. The results showed that the pair Delight and Surprise could best be separated with a recognition rate of 86.67% and kappa 0.7 where, Delight was related to a positive feedback and Surprise to a negative feedback. Delight in general was best separated from other affective states, whereas, the Delight and Frustration pair had a high separation of 82.89 %; frustration is related to a negative feedback as well. Flow, Neutral, and Curiosity could not be separated effectively from other classes. Boredom, Frustration, and Confusion showed weak separation among others.

Table 2. Classification results for discriminating affect-pairs from feedback

Affect Pair	**ZeroR (baseline) (% Correct))**	**SVM (%Correct)**
Boredom -Confusion	52.1	56.8
Boredom -Delight	60.7	69.0
Boredom-Frustration	54.2	57.98
Confusion-Delight	62.71	73.45
Confusion-Neutral	52.16	58.62
Curiosity-Delight	57.14	70.13
Curiosity-Frustration	50.57	60.34
Delight - Flow	65.45	69.63
Delight-Frustration	56.58	82.89
Delight-Neutral	64.71	67.91
Delight-Surprise	62.86	86.67
Flow-Neutral	50.81	53.25
Frustration -Neutral	58.45	56.04

4.2 Feedback and Physiology

The results of SVM classification from the physiological data based on feedback for all features and selected features are given in Table 3. For some learners, there were data sparseness problems with tutor feedback, which made automatic classification unfeasible. In order to alleviate this problem, we grouped AutoTutor's feedback types into two classes; the majority feedback (negative) in a 'Negative' class and the rest in the 'Other' class.

Table 3. Physiological data classification based on Feedback, Negative Class Vs. Other Class

Learner	Negative Count	Other Count	ZeroR (Baseline) (% Correct)	SVM (All features) (% Correct)	SVM (10 features)) (% Correct)
1	24	23	42.55	72.34	63.83
2	18	18	44.44	47.22	63.89
3	32	17	65.31	69.39	67.35
4	26	20	56.52	41.30	50
5	31	26	54.39	47.37	50.88
6	19	27	58.70	39.13	56.52
7	28	35	55.56	57.14	71.43
8	18	24	57.14	78.57	76.19
9	20	18	52.63	73.68	84.21
10	29	35	54.69	50	42.19
11	23	30	56.60	60.38	66.04
12	20	21	51.22	60.98	58.54
13	22	22	45.45	65.91	72.73
14	18	27	60	53.33	60
15	25	27	51.92	59.62	61.54
16	14	24	63.16	57.89	65.79
Total	367	394	51.77	49.67	52.56

The counts column in Table 3 shows the number of instances for each class. Results emphasize differences among learners; 12 learners had classification accuracy above the baseline. This suggests that there are physiological patterns that can be identified from feedback. It was also noticeable that the learners with higher classification accuracy had a mix of ECG, EMG, and SC as selected features, while those with low accuracies had only ECG features selected. This implies that multimodal features can increase the classification accuracy in automatic affect recognition. The Chi square feature selection improved the classification accuracy in most cases, by selecting the most informative features, and discarding those features that are redundant or irrelevant to the classification task; however for some cases it degraded the classification accuracy due to the loss of some informative features, and this is subject dependent as our results suggest. Further investigations are needed to find more efficient feature selection methods. The classification for the combined data considering all learners was no better than the baseline, which indicates that physiological patterns in response to feedback are different among individuals.

4.3 Self-report and Physiology

The second part of the project is to study how physiological patterns could be mapped into self-reported cognitive-affective states, in the sense that students during interaction with Autotutor system would experience some emotions that affect their physiology. If we were able to build models that can map these physiological changes into affective states, we would be able to adapt the tutoring system to students' current emotional state; we hypothesize that such an adaptation will enhance the learning experience of students in future development of tutoring systems. For this paper we evaluated only standard techniques but results were not significantly above the baseline so they are not discussed in detail. Hence, a more detailed investigation is required. Meanwhile it is interesting to consider the issues that would arise when attempts to identify the cognitive-affective states from the physiological signals.

- The differences between subjects make it unlikely that a classifier trained with data from all subjects would be accurate [31].
- The cognitive-affective state categories have very skewed distributions and training a classifier with highly unbalanced data is more difficult.

5 Conclusion

We investigated the impact of an Intelligent Tutoring System's feedback on learners' self-reported affective states and their physiological states. The results indicated that there was a relationship between tutor feedback and self-reported affective states, as well as between feedback and physiology. Automatic classifiers achieved accuracies above the baseline showing that both affective states and physiology can be predicted from the tutor feedback. These results are significant since different feedback types (negative or positive) from AutoTutor indicate possible influence in learners' emotional and physiological states. This suggests that here is some coherence in the way that learners physiologically respond to tutor feedback.

A preliminary study of possible relationships between the affective states and physiology did not show significant relationships. A more thorough study of these relationships was planned as the second part of this project. The effect of specific stimulus (e.g. a photograph) on subjects' physiology can provide information to create models [32] that can predict learners' affective states in learning scenarios. As for future work, 'normative databases' can help to create such models suitable for learning scenarios.

References

1. Lepper, M., Henderlong, J.: Turning 'play' into 'work' and 'work' into 'play': 25 years of research on intrinsic versus extrinsic motivation. Intrinsic and extrinsic motivation: The search for optimal motivation and performance 257-307 (2000)
2. Linnenbrink, E., Pintrich, P.: The role of motivational beliefs in conceptual change. Practice 115, 135 (2002)

3. Kort, B., Reilly, R., Picard, R.: An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion. 43-48 (2001)
4. 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, (2004)
5. Picard, R.: Affective computing. The MIT Press (1997)
6. D'Mello, S., Graesser, A., Picard, R.: Toward an affect-sensitive AutoTutor. IEEE Intelligent Systems 22, 53-61 (2007)
7. Ekman, P., Levenson, R., Friesen, W.: Autonomic nervous system activity distinguishes among emotions. Science 221, 1208-1210 (1983)
8. D'Mello, S., Craig, S., Witherspoon, A., Mcdaniel, B., Graesser, A.: Automatic detection of learner's affect from conversational cues. User Modeling and User-Adapted Interaction 18, 45-80 (2008)
9. Burleson, W., Picard, R.: Affective agents: Sustaining motivation to learn through failure and a state of stuck. Citeseer(2004)
10. Csikszentmihalyi, M.: Flow: The psychology of optimal experience. New York (1990)
11. Graesser, A., McDaniel, B., Chipman, P., Witherspoon, A., D'Mello, S., Gholson, B.: Detection of emotions during learning with AutoTutor. Proceedings of the 28 th Annual Meetings of the Cognitive Science Society 285-290 (2006)
12. Klein, J., Moon, Y., Picard, R.: This computer responds to user frustration: Theory, design, and results. Interacting with computers 14, 119-140 (2002)
13. Prendinger, H., Ishizuka, M.: The Empathic Companion: A Character-Based Interface That Addresses Users' Affective States. Applied Artificial Intelligence 19, 267-286 (2005)
14. Picard, R.W., Vyzas, E., Healey, J.: Toward machine emotional intelligence: analysis of affective physiological state. IEEE transactions on pattern analysis and machine intelligence 23, 1175-1191 (2001)
15. Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.: Emotion recognition in human-computer interaction. IEEE Signal processing magazine 18, 32-80 (2001)
16. Polzin, T.: Detecting Verbal and Non-verbal cues in the communication of emotion. Unpublished Doctoral Dissertation, School of Computer Science, Carnegie Mellon University (2000)
17. Yacoob, Y., Davis, L.: Recognizing human facial expressions from long image sequences using optical flow. (1996)
18. Picard, R., Vyzas, E., Healey, J.: Toward machine emotional intelligence: Analysis of affective physiological state. IEEE transactions on pattern analysis and machine intelligence 1175-1191 (2001)
19. Barrett, L.: Are Emotions Natural Kinds? , Vol. 1 58 (2006)
20. Christie, I.C.: Multivariate discrimination of emotion-specific autonomic nervous system activity. (2002)
21. Fredrickson, B.L., Mancuso, R.A., Branigan, C., Tugade, M.M.: The undoing effect of positive emotions. Motivation and Emotion 24, 237-258 (2000)
22. Levenson, R.W., Ekman, P., Friesen, W.V.: Voluntary facial action generates emotion-specific autonomic nervous system activity. Psychophysiology 27, 363-384 (1990)
23. Conati, C., Chabbal, R., Maclaren, H.: A study on using biometric sensors for monitoring user emotions in educational games. Workshop on Assessing and Adapting to User Attitudes and Affect: Why, When and How(2003)
24. Arroyo, I., Cooper, D., Burleson, W., Woolf, B., Muldner, K., Christopherson, R.: Emotion Sensors go to School. (2009)
25. Narciss, S.: Motivational Effects of the Informativeness of Feedback. (1999)
26. Graesser, A., Chipman, P., Haynes, B., Olney, A.: AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. IEEE Transactions on Education 48, 612-618 (2005)

27. D'Mello, S., Picard, R., Graesser, A.: Toward an Affect-Sensitive AutoTutor. IEEE Intelligent Systems 22, 53-61 (2007)
28. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems). Morgan Kaufmann (2005)
29. Platt, J.: Machines using Sequential Minimal Optimization. In: Schoelkopf, B., Burges, C., Smola, A. (eds.): Advances in Kernel Methods - Support Vector Learning (1998)
30. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. John Wiley & Sons, New York (2001)
31. Calvo, R.A., Brown, I., Scheding, S.: Effect of Experimental Factors on the Recognition of Affective Mental States Through Physiological Measures. AI 2009: Advances in Artificial Intelligence, Proceedings of 22nd Australasian Joint Conference on Artificial Intelligence Springer-Verlag (2009)
32. Lang, P.J., Greenwald, M., Bradley, M.M., Hamm, A.O.: Looking at pictures: Evaluative, facial, visceral, and behavioral responses. Vol. 30 -274 (1993)