AutoTutor Detects and Responds to Learners Affective and Cognitive States

Sidney D'Mello[1], Tanner Jackson[1], Scotty Craig[1], Brent Morgan[1], Patrick Chipman[1], Holly White[1], Natalie Person[1], Barry Kort[2], Rana el Kaliouby[2], Rosalind Picard[2], and Art Graesser[1]

[1] Institute for Intelligent Systems, University of Memphis, Memphis, TN 38152 USA
sdmello@memphis.edu

[2] MIT Media Laboratory, Massachusetts Institute of Technology, USA
picard@media.mit.edu

Abstract. This paper provides a synthesis of our research towards the development of an affect-sensitive Intelligent Tutoring System called AutoTutor. The affect-sensitive AutoTutor detects the emotions (boredom, flow/engagement, confusion, frustration) of a learner by monitoring conversational cues, gross body language, and facial features. It is also mindful of the learners' affective and cognitive states in selecting its pedagogical and motivational dialogue moves. Finally, the AutoTutor embodied pedagogical agent synthesizes affective responses through animated facial expressions and modulated speech. The paper provides an overview of our theoretical framework, methodology, implementation details, and results.

1. Introduction

Attempts to acquire a deep level understanding of conceptual information through effortful cognitive activities such as a systematic exploration of the problem space, generating self-explanations, making bridging inferences, asking diagnostic questions, causal reasoning, and critical thinking often lead to episodes of failure and the learner experiences a host of affective responses (Mandler, 1984, 1999; Stein & Levine, 1991). Negative emotions are experienced when expectations are not met, failure is imminent, and important goals are blocked. For example, confusion occurs when learners face obstacles to goals, contradictions, incongruities, anomalies, uncertainty, and salient contrasts (Festinger, 1957; Graesser & Olde, 2003; Piaget, 1952). Unresolved confusion can lead to irritation, frustration, anger, and sometimes even rage. On the other hand, a learner may experience a host of positive emotions when misconceptions are confronted, challenges are uncovered, insights unveiled, and complex concepts are mastered. Students' that are actively engaged in the learning session may have a flow like experience, when they are so engrossed in the material that time and fatigue disappears (Csíkszentmihályi, 1990). They may also experience other positive emotions such as delight, excitement, and even one of those rare eureka (i.e. "a ha") moments. Simply put, emotions are systematically affected by the knowledge and goals of the learner, as well as vice versa (Dweck, 2002; Mandler, 1984; 1999; Stein & Levine, 1991). Cognitive activities such as causal reasoning, deliberation, goal appraisal, and planning processes operate continually throughout the experience of emotion. It is important to emphasize that the theories of emotion and learning proposed by Mandler, Stein, Dweck (and others) assume that cognition and emotion are inextricably bound. This scientific literature is quite different from the popular folklore that emotions are detached from cognition (e.g., "right brain being different from left brain").

Given this inextricable link between emotions and learning, it is reasonable to hypothesize that an Intelligent Tutoring System (ITS) that is sensitive to the affective and cognitive states of a learner would positively influence learning, particularly if deep learning is accompanied by confusion, frustration, anxiety, boredom, delight, flow, surprise and other affective experiences (D'Mello, Picard, & Graesser, 2007; Graesser et al., 2007; Lepper & Woolverton, 2002; Picard, 1997). An affect-sensitive ITS would incorporate assessments of the students' cognitive, affective, and motivational states into its pedagogical strategies to keep students engaged, boost self-confidence, heighten interest, and presumably maximize learning. For example, if the learner is frustrated, the tutor would need to generate hints to advance the learner in constructing knowledge, and make supportive empathetic comments to enhance motivation. If the learner is bored, the tutor would need to present more engaging or challenging problems for the learner to work on.

However, a number of technological challenges need to be overcome before the benefits of an affect-sensitive ITS can be fully realized. An affect-sensitive ITS needs to be fortified with sensors and signal processing algorithms to robustly detect the affective states of a learner within real time constraints. The tutor also needs to select pedagogical and motivational moves that maximize learning while positively influencing the learner's affect.

We are in the process of implementing this two-phase strategy (affect detection and response) into an existing ITS, AutoTutor. AutoTutor is an intelligent tutoring system that helps learners construct explanations by interacting with them in natural language and helping them use simulation environments (Graesser, Jackson, & McDaniel, 2007). AutoTutor helps students learn Newtonian physics, computer literacy, and critical thinking skills by presenting challenging problems (or questions) from a curriculum script and engaging in a mixed-initiative dialog while the learner constructs an answer. AutoTutor provides *feedback* to the student on what the student types, *pumps* the student for more information, *prompts* the student to fill in missing words, gives *hints*, fills in missing information with *assertions*, identifies and corrects *misconceptions* and erroneous ideas, *answers* the student's questions, and *summarizes* topics. While the current version of AutoTutor adapts to the cognitive states of the learner, the affect-sensitive AutoTutor would be responsive to both the cognitive and affective states of learners (D'Mello, Picard, & Graesser, 2007).

This paper provides a synthesis of our research aimed at developing an affect-sensitive AutoTutor. We first describe our classification algorithms that attempt to detect the affective states of the learner followed by a description of how the tutor can be responsive to the learners' emotions. We focus on a model of learners' emotions that include boredom, engagement/flow, confusion, frustration, delight, and surprise.

These emotions were selected on the basis of four empirical studies that used multiple methodologies (i.e. observational, emote-aloud, retrospective judgments by multiple judges) to monitor the emotions that learners' experienced during tutoring sessions with AutoTutor (Craig et al., 2004; D'Mello et al., 2006; Graesser et al., 2006; 2007).

2. Detecting learners' affective states

Robust emotion measurement is a critical component for any ITS that aspires to respond to a learners' emotions. Simply put, an ITS that cannot detect a learner's emotions cannot adequately respond to a learner's emotions. Although, the last decade has witnessed a burst of innovative and exciting research aimed at developing computational systems to monitor the emotions of a person (see Pantic & Rothkrantz (2003) for a comprehensive review and the proceedings of ACII 2007 edited by Paiva, Prada, & Picard (2007) for recent advances), most of the affect detection systems focus on recognizing Ekman and Friesen's (1978) "basic" emotions. However, there is some evidence that these "basic" emotions (i.e. anger, fear, sadness, happiness, disgust, and surprise), though ubiquitous to everyday experience, are not particularly relevant to learning (D'Mello et al., 2006; Kort, Reilly, & Picard, 2001; Lehman et al., 2008). Consequently, we developed computational systems to detect the presence of boredom, engagement, confusion, and frustration (delight and surprise were excluded because they are quite rare).

Our affect detection system monitors conversational cues, gross body language, and facial features. They use supervised learning methods for affect classification. These classifiers were trained on data obtained in a study that involved synchronization and data recording of the sensors while 28 learners interacted with AutoTutor (see Figure 1) (Graesser et al., 2006). Manually annotated affect labels, which are required for the supervised learning systems, were obtained by multiple human judges including the learner (self judgments), an untrained peer, and two trained judges.

2.1 Sensors used for affect detection

Conversational cues (dialogue features). A one-on-one tutoring session with AutoTutor yields a rich trace of contextual information, characteristics of the learner, episodes during the coverage of the topic, and social dynamics between the tutor and learner. These conversational cues cover a broad and deep feature set that includes assessments of deep meaning, world knowledge, and pragmatic aspects of communication. Therefore, several conversational features and discourse markers (collectively called dialogue features) were extracted from AutoTutor's log files and were utilized to infer the learner's affect. The dialogue features were computed for each student-tutor turn (i.e. student submits response, tutor provides feedback, tutor presents next question). They included *temporal* features (e.g. time on problem, response time), assessments of *response verbosity* (e.g. number of characters, speech act), assessments of the *conceptual quality* of the student's response obtained by

Latent Semantic Analysis (LSA) (Landauer & Dumais, 1997), *conversational directness* (i.e. how much information the tutor is explicitly providing to the student), and *tutor feedback* (negative, neutral, positive).

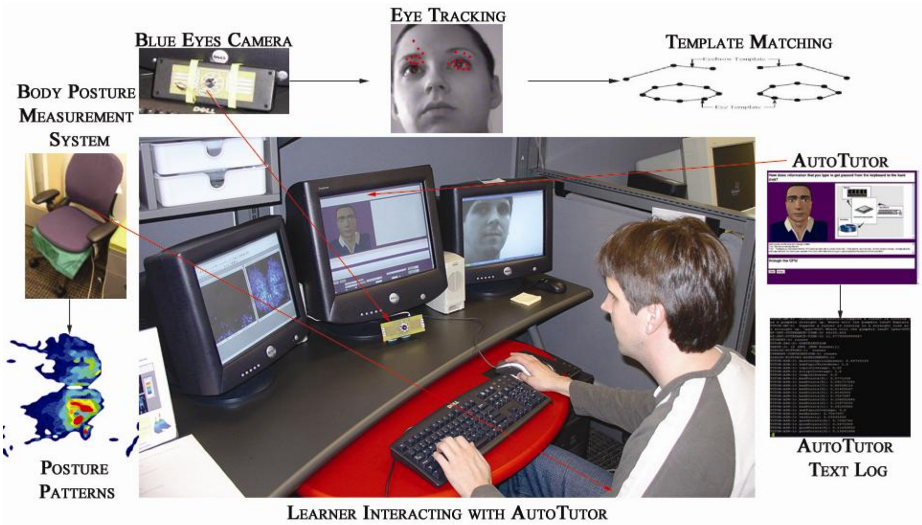

Fig. 1. Sensors used for affect detection as learner interacts with AutoTutor

Gross body language (posture features). The Body Posture Measurement System (BPMS), developed by Tekscan™, was used to monitor the gross body language of a student during a session with AutoTutor (see left panel of Figure 1). The BPMS consists of a thin-film pressure pad (or mat) that can be mounted on a variety of surfaces. The output of the BPMS system consisted of two 38x41 matrices (for the back and seat) with each cell in the matrix corresponding to the amount of pressure exerted on the corresponding element in the sensor grid.

We relied on an *attentive-arousal* framework (Bull, 1987) to interpret relationships between the posture features and the affective states of the learner. One can think of heightened pressure in the seat as resonating with a tendency to position one's body towards the source of stimulation (i.e., high attentiveness since the learner is positioning his or her body towards the AutoTutor interface, or a short distance between the nose and the screen). On the other hand, an increase in pressure on the back of the chair suggests that the learner is leaning back and detaching one's self from the stimulus (low attentiveness). Arousal was operationally defined by the rate of change of pressure exerted on the back and the seat of the pressure sensitive chair and is similar to the degree of activation.

Facial feature tracking. We used the IBM BlueEyes system developed by Picard and colleagues (Kapoor & Picard, 2005) to monitor facial expressions for affect recognition (see top panel of Figure 1). Recognition of emotions is a four stage process. First, the system locates and tracks the pupils of the eye (step 1). The system achieves real-time, highly accurate tracking (less than one pixel RMS error per control point), with no calibration required. Second, the system fits templates to the upper facial features (eyes and brows), in real time (step 2). Third, the system labels facial action units (muscle movements in the face, Ekman & Friesen, 1978) with an accuracy of 68% on six upper facial action units (step 3); it should be noted that a person has to score 75% to qualify as a human expert. The fourth stage involves linking patterns of action unit activations to the different emotions (step 4). This is achieved by identifying the action units that accompany each emotion. For example, confusion is expressed by a lowered brow and the tightening of the eye lids (Craig et al., in press; McDaniel et al., 2007). Steps 1,2, and 3 are currently fully automated and we are in the process of implementing Step4. This would yield a fully automated facial feature based affect recognition system.

2.2. Classification accuracy

We are currently exploring some of the technical challenges associated with the automated detection of the facial expressions. However, experimental simulations indicate that conversational cues and gross body language are viable channels for affect detection. The Waikato Environment for Knowledge Analysis (WEKA)(Witten & Frank, 2005) was used to evaluate the performance of various standard classification techniques in an attempt to detect learners' affect.. The classification algorithms were compared in their ability to detect boredom, confusion, flow, and frustration from neutral. Classification reliability was evaluated on 17 standard classification algorithms using k-fold cross-validation (k = 10). Conversational cues alone yielded accuracies of 69%, 68%, 71%, and 78%, in individually detecting boredom, confusion, flow, and frustration from neutral (chance=50%) (D'Mello et al., 2008). Classification accuracies obtained from gross body language were 70%, 65%, 74%, and 72% in detecting boredom, confusion, flow, and frustration versus the neutral baseline (baserate = 50%) (D'Mello, Picard, & Graesser, 2007). Taken together, classification accuracies are 73% when each affective state is aligned with the optimal sensory channel.

2.3. Combining channels for multimodal affect detection

We are currently exploring a number of strategies to combine the information from the different sensory channels into one emotion classifier that can be used in AutoTutor. One option is to use a sensory-level fusion technique, which involves grouping features from the various sensors before attempting to classify emotions. Alternatively, in decision level fusion, the affective states would first be classified from each sensor and then integrated to obtain a global view across the various sensors. We are currently exploring both of these alternatives. We will test the hypothesis that classification performance from multiple channels will exhibit *super-additivity*, i.e., performance superior to an additive combination of individual channels. An alternative hypothesis is that there will be *redundancy* across the channels, i.e., adding additional channels yields negligible incremental gains.

Once we have isolated the individual channel or combination of channels that maximizes the discriminability of each of the four affective states from neutral, we would require a super classifier to integrate the outputs of the individual affect-neutral classifiers. We envision a collection of affect-neutral classifiers that would first determine whether the incoming dialogue pattern resonated with any one or more of the emotions (versus a neutral state). If there is resonance with only one emotion, then that emotion would be declared as being experienced by the learner. If there is resonance with two or more emotions, then a conflict resolution module would be launched to decide between the alternatives. Perhaps this would be a second level affect classifier. We have indeed been encouraged by our preliminary experiments that calibrate the accuracy of such a multi-layered emotion classifier.

3. Responding to learners' affective and cognitive states

Classification of learner emotions is an essential step in building a tutoring system that is sensitive to the learner's emotions. The other essential component is to build mechanisms that empower AutoTutor to intelligently respond to these emotions, as well as to their states of cognition, motivation, social sensitivity, and so on. In essence, how can an affect-sensitive AutoTutor respond to the learner in a fashion that optimizes learning and engagement? Therefore, the next phase of our research focused on fortifying AutoTutor with the necessary pedagogical and motivational strategies to address the cognitive and the affective states of the learner.

Our previous research has revealed that boredom, flow/engagement, confusion, and frustration were the most prominent states that learners' experienced while interacting with AutoTutor (Craig et al., 2004; D'Mello et al., 2006; Graesser et al., 2006; 2007). Boredom, confusion, and frustration are negative emotions, and are states that, if addressed appropriately, can have a positive impact on engagement and learning outcomes. Flow, on the other hand, is a highly desirable positive affective state that is beneficial to learning. Although, most tutoring environments would want to promote and prolong the state of flow, any intervention on the part of the tutor runs the risk of adversely interfering with the flow experience. Therefore, the current version of the affect-sensitive AutoTutor does not respond to episodes of flow. Instead, we focus on addressing the affective states of boredom, frustration, and confusion.

At this point in science, there are no empirically proven strategies to address the presence of boredom, frustration, and confusion. Therefore, possible tutor reactions to student emotions were derived from two sources: theoretical foundations of pedagogy/affect and recommendations made by pedagogical experts.

3.1. Theoretical perspective and recommendations by pedagogical experts

An examination of the literature provided some guidance on how best to respond to the states of boredom, confusion, and frustration. We focused on two major theoretical perspectives that address the presence of these negative emotions. These included attribution theory (Batson, Turk, Shaw, & Klein, 1995; Heider, 1958; Weiner, 1986) and cognitive disequilibrium during learning (Piaget, 1952; Craig et al. 2004; Festinger, 1957).

Attribution theory to address boredom and frustration. Attribution theory is based on the *explanations* people make to explain their success or failure. According to this theory, the cause of the success or failure can be based on three dichotomous factors: *internal* or *external; stable* or *unstable; controllable* or *uncontrollable.* A basic principle of attribution theory is that a person's attributions for success or failure determine the amount of effort the person will expend on that activity in the future and that people tend to make attributions that allow them to maintain positive views of themselves. So, success will be attributed to stable, internal, and controllable factors and major failures will be attributed to external, uncontrollable factors. However, it is important to get learners to change this failure attribution so that their *failures* are attributed to internal, unstable factors over which they have control (e.g., effort) (Heider, 1958; Weiner, 1986). In order to change this attribution, learners must be encouraged to focus on learning goals. People who emphasize learning goals are likely to seek challenges, if they believe the challenges will lead to greater competence; and they tend to respond to failure by increasing their effort (Dweck, 2002).

Empathy has been indicated as an important emotional response for attributions. In this case empathy serves two functions. First, displaying empathy portrays an awareness of blocked goals and a willingness to help. When displays of empathy are observed, the learner is more likely to anticipate the goals of the other displaying empathy (Batson, et al., 1995). So an example from a tutoring context would be the tutor displaying empathy for the student will cause the student to understand the tutor is attempting to help and will make the student more likely adopt the learning goals put forth by the tutor. Therefore, both boredom and frustration can be handled in similar ways using empathetic responses by the tutor.

Cognitive disequilibrium theory to address confusion. Cognitive disequilibrium is believed to play an important role for in comprehension and learning processes (Graesser & Olde, 2003; Piaget, 1952). Deep comprehension occurs when learners confront contradictions, anomalous events, obstacles to goals, salient contrasts, perturbations, surprises, equivalent alternatives and other stimuli or experiences that fail to match expectations (Mandler, 1984, 1999; Schank, 1986). Cognitive disequilibrium has a high likelihood of activating conscious, effortful cognitive deliberation, questions and inquiry that aim to restore cognitive equilibrium.

When a learner enters a state of confusion due to the content they are learning this is equivalent to entering cognitive disequilibrium. The tutor’s first step should be to encourage the tutee to continue working so they can reach a state of equilibrium again and by doing so reach the full benefit of the state of disequilibrium. However, if the learner persists in a state of cognitive disequilibrium for too long the tutor should display empathy with the learner’s attempts thereby acknowledging their attempts to reach their goals and direct them out of the state of confusion before they give up.

3.2. Recommendations by pedagogical experts

In addition to theoretical considerations, the assistance of experts was enlisted to help create the set of tutor responses. Two experts in pedagogy, with approximately a decade of related experience each, were provided with excerpts from real AutoTutor dialogues (including both the tutor and student dialogue content, screen capture of the learning environment, and video of the student's face as illustrated in Figure 1). There were approximately 200 excerpts averaging around 20 seconds in length, each of which included an affective response by the student. The experts were instructed to view each of the excerpts and provide an appropriate follow-up response by the tutor. These example responses were placed into similar groups that loosely resembled production rules. For example, if a student is frustrated then the tutor should provide encouragement to continue and establish a small sub-goal. The tutor might also provide motivational and empathetic statements to alleviate frustration because this approach has been shown to be quite effective in reducing frustration (Klein, Moon, & Picard, 2002).

3.3. Production rules to respond to learners' affective and cognitive states

We derived a set of production rules that addressed the presence of boredom, confusion, and frustration by amalgamating perspectives from attribution theory and cognitive disequilibrium theory with the recommendations made by the experts. Although, the rules created by the pedagogical experts allowed for any possible action on the part of the tutor, AutoTutor can only implement a portion of those actions. For example, one possibility to alleviate boredom would be to launch an engaging simulation or a seductive serious game. However, the current version of the tutor does not support simulations or gaming, so such a strategy is not immediately realizable. Consequently, we selected production rules that could be implemented by AutoTutor's primitive actions which include feedback delivery (positive, negative, neutral), a host of dialogue moves (hints, pumps, prompts, assertions, and summaries), and facial expressions and speech modulation by AutoTutor's embodied pedagogical agent (EPA).

The production rules were designed to map dynamic assessments of the students' cognitive and affective states with tutor actions to address the presence of the negative emotions (See Figure 2). There were five parameters in the student model and 5 parameters in the tutor model. The parameters in the student model included, (a) the current emotion detected, (b) the confidence level of that emotion classification, (c) the previous emotion detected, (d) a global measure of student ability (dynamically updated throughout the session), (e) the conceptual quality of the student's immediate response. AutoTutor incorporates this 5 dimensional assessment of the student and responds with: (a) feedback for the current answer, (b) an empathetic and motivational statement, (c) the next dialogue move, (d) an emotional display on the face of the AutoTutor embodied pedagogical agent, and (e) emotionally modulating the voice produced by AutoTutor's text to speech engine. As indicated above, the mappings between the student model and the tutors' actions were informed by theories on emotion and learning as well as recommendations made by the pedagogical experts.

As a complete example, consider a student has been performing well overall (high global ability), but the most recent contribution wasn't very good (low current contribution quality). If the current emotion was classified as boredom, with a high probability, and the previous emotion was frustration then AutoTutor might say the following, "Maybe this topic is getting old. I'll help you finish so we can try something new". This is a randomly chosen phrase from a list that was designed to indirectly address the student's boredom and to try and shift the topic a bit before the student becomes disengaged from the learning experience. This rule fires on several different occasions, and, each time it is activated, AutoTutor will select a dialogue move from a list of associated moves. In this fashion, the rules are context sensitive and are dynamically adaptive to each individual learner.

The empathetic and motivational statement represents AutoTutor's attempt to address the presence of negative emotional states. The AutoTutor agent also expresses different emotional states which convey an emotion to the user. These states include surprise, delight, disappointment, compassion, and skepticism. For example, surprise is displayed when the tutor detects a degree of novelty such as a low domain knowledge student providing an exceptionally good answer. Expressions of delight parallel positive feedback statements while disappointment accompanies expressions of negative feedback. The tutor displays compassion when a learner has been making a serious attempt but is simply not grasping the material.

Student Model		**Tutor Action**
Current Emotion		**Feedback**
boredom, confusion, frustration		positive, neutral, negative
Classification Confidence		**Empathetic and motivational statement**
high or low		
Previous Emotion	$\longrightarrow$	**Next Dialogue Move**
boredom, confusion, frustration		hint, pump, prompt, splice, assertion
Global Student Ability		**Facial Expression**
high or low		surprise, delight, compassion, skeptical
Quality of Current Answer		**Speech intonation**
high or low		pitch, intensity, speech rate, etc

Fig. 2. Production rules to respond to learners' affective and cognitive States

Examples of preliminary affective displays are illustrated in Figure 3. These displays were created with the Haptek People Putty Software Development Kit. The facial expressions in each display were informed by Ekman's work on the facial correlates of emotion expression. Surprise is characterized by having both eyebrows raised, eyes wide open, and mouth agape (action units 1+2, 5, 26; Ekman & Friesen, 1978). This is followed almost immediately by delight, consisting of a cheek raise and a wide smile (action units 6 + 12; Ekman & Friesen, 1978). Displeasure, a less intense version of disgust, is characterized by a pursing of the lips and a shaking of the head (action units 24, 57; Ekman & Friesen, 1978). Compassion is a sense of understanding displayed to the user. This is manifested by an inner eye-brow raise, eyes open, and lips slightly pulled down at the edges (action units 1, 5, 15; Chovil, 1991). Finally, skepticism is a combination of confusion and curiosity, characterized by a furrowing of the brow, an eye squint, and one outer eyebrow is raised (action units 2, 4, 7).

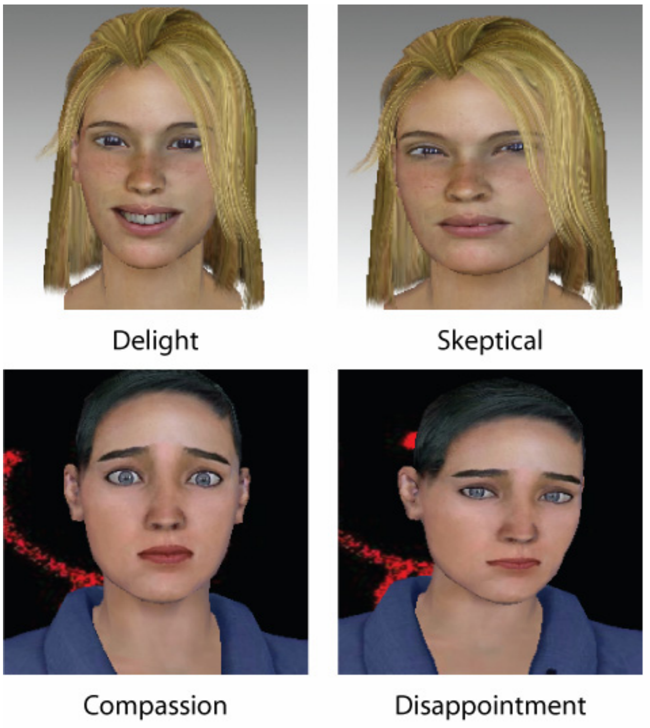

Fig. 3. Affect synthesis by embodied pedagogical agents

The facial expressions of emotion displayed by AutoTutor will be augmented with emotionally expressive speech synthesized by the agent. A monotonous, metered voice would be replaced with one that is emotionally expressive by variations in pitch, speech rate and other prosodic features. Previous research (Abelin & Allwood, 2003; Griffiths, 2006; Paeschke & Sendlmeier, 2000; Schmidt, 2005) has led us to conceptualize AutoTutor's affective speech on the following indices of pitch range, pitch level, and speech rate. The agent's speech was modulated using the Speech Synthesis Markup Language (SSML) in accordance to the parameters specified in Table 1.

Table 1. Acoustic-prosodic correlates of tutors emotional expressions

Affective State	Pitch Range	Pitch Level	Speech Rate
Surprise/Delight	Wide	Very High	Fast
Empathy	Narrow	Low	Slow
Skeptical	Narrow	High	Slow
Disappointment	Narrow	Low	Slow

4. Conclusions

We have described a new version of AutoTutor that aspires to be responsive to learners' affective and cognitive states. The affect-sensitive AutoTutor aspires to keep students engaged, boost self-confidence, and presumably maximize learning by narrowing the communicative gap between the highly emotional human and the emotionally challenged computer. In order to test whether an affect-sensitive cognitive tutor is effective, we will compare two different versions of AutoTutor: one that is sensitive to learner emotions and one that is not. The original AutoTutor has a conventional set of fuzzy production rules that are sensitive to cognitive states of the learner, but not to the emotional states of the learner. Our improved AutoTutor is sensitive to these affective states. The obvious prediction is that learning gains and the learner's impressions should be superior for the affect-sensitive AutoTutor.

The affect-sensitive AutoTutor represents one out of a handful of related efforts made by a number of researchers who have a similar vision (e.g. Conati, 2002; Kort, Reilly, & Picard, 2001; Litman & Forbes-Riley, 2004; McQuiggan & Lester 2007; Woolf, Burelson, & Arroyo, 2007). Our unified vision is to advance education, intelligent learning environments, and human-computer interfaces by optimally coordinating cognition and emotions. Whether the affect-sensitive AutoTutor positively influences learning and engagement awaits further development and empirical testing.

Acknowledgments

This research was supported by the National Science Foundation (REC 0106965 and ITR 0325428). Any opinions, findings and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of NSF.

References

1. Abelin, A. and Allwood, J. (2003) Cross linguistic interpretation of expressions of emotions. In Proceedings of VIII Simposio Internactional de Communicacion Social, pages 387–393.
2. Batson, C., Turk, C., Shaw, L., & Klein, T. (1995). Information function of empathic emotion: Learning that we value the other's welfare. *Journal of Personality and Social Psychology, 68*, 300-313.
3. Bull, E. P. (1987). *Posture and Gesture*. Pergamon Press.
4. Chovil. N. (1991). Discourse-oriented facial displays in conversation. *Research on Language and Social Interaction,* 25:163-194.
5. Conati C. (2002). Probabilistic Assessment Of User's Emotions In Educational Games. *Journal Of Applied Artificial Intelligence, 16*, 555-575.
6. Craig, S.D., Graesser, A. C.,Sullins, J., & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning. *Journal of Educational Media, 29,* 241-250.

7. Csikszentmihalyi, M. (1990). Flow: The Psychology of Optimal Experience. Harper-Row: NY.
8. D'Mello, S. K., Craig, S. D., Sullins, J., & Graesser, A. C. (2006). Predicting Affective States Through An Emote-Aloud Procedure From AutoTutor's Mixed-Initiative Dialogue. *International Journal Of Artificial Intelligence In Education, 16*, 3-28.
9. D'Mello, S. K., Craig, S. D., Witherspoon, A. W., McDaniel, B. T., & Graesser, A. C. (2008). Automatic Detection of Learner's Affect from Conversational Cues. *User Modeling and User-Adapted Interaction*, 18(1-2), 45-80.
10. D'Mello, S. K., Picard, R., & Graesser, A. C. (2007). Towards an Affect Sensitive AutoTutor, *IEEE Intelligent Systems, 22*(4), 53-61.
11. Dweck, C. S. (2002). Messages that motivate: How praise molds students' beliefs, motivation, and performance (in surprising ways). In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 61-87). Orlando, FL: Academic Press.
12. Ekman, P, & Friesen, W. V. (1978). The facial action coding system: A technique for the measurement of facial movement. Palo Alto: Consulting Psychologists Press.
13. Festinger, L. (1957) *A theory of cognitive dissonance*, Stanford, CA: Stanford University Press
14. Graesser, A. C. Chipman, P., King, B., McDaniel, B., and D'Mello, S (2007). Emotions and Learning with AutoTutor. *13th International Conference on Artificial Intelligence in Education (AIED 2007)*. R. Luckin et al. (Eds), (pp 569-571). IOS Press.
15. Graesser, A.C., Jackson, G.T., & McDaniel, B. (2007). AutoTutor holds conversations with learners that are responsive to their cognitive and emotional states. *Educational Technology, 47*, 19-22.
16. Graesser, A.C., McDaniel, B., Chipman, P., Witherspoon, A., D'Mello, S., & Gholson, B. (2006). Detection Of Emotions During Learning With AutoTutor. *Proceedings of the 28th Annual Conference of the Cognitive Science Society,* Mahwah, NJ: Erlbaum, 285-290.
17. Graesser, A. C. & Olde, B. (2003) How does one know whether a person understands a device? The quality of the questions the person asks when the device breaks down, *Journal of Educational Psychology*, 95, 524–536.
18. Griffiths, S. (2006). *Expressive Speech and Speech Output*. Unpublished Masters thesis. Bielefeld, NRW, Germany: University Bielefeld.
19. Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York: John Wiley & Sons
20. Kapoor, A. and Picard, R. E. (2005), Multimodal Affect Recognition in Learning Environments, ACM MM'05, November 6-11, 2005, Singapore.
21. Klein, J., Moon, Y. and Picard, R. (2002). This Computer Responds to User Frustration. *Interacting with Computers*, 14(2), 119-140.
22. Kort, B., Reilly, R., & Picard, R. (2001). An Affective Model Of Interplay Between Emotions And Learning: Reengineering Educational Pedagogy—Building A Learning Companion. *Proceedings IEEE International Conference on Advanced Learning Technology: Issues, Achievements and Challenges,* Madison, Wisconsin: IEEE Computer Society, 43-48.
23. Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211-240.
24. Lehman, B. A., Matthews, M., D'Mello, S. K., and Person, N. (in press). Understanding Students' Affective States During Learning. *Ninth International Conference on Intelligent Tutoring Systems (ITS'08).*
25. Litman, D. J., & Forbes-Riley, K. (2004). Predicting student emotions in computer-human tutoring dialogues. In *Proceedings of the 42nd annual meeting of the association for com-*

putational linguistics (pp. 352-359). East Stroudsburg, PA: Association for Computational Linguistics.

26. Lepper, M. R., & Woolverton, M. (2002). The wisdom of practice: Lessons learned from the study of highly effective tutors. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 135-158). Orlando, FL: Academic Press.
27. Mandler, G. (1984). *Mind and body: Psychology of emotion and stress*. New York: Norton.
28. Mandler, G. (1999). Emotion. In B. M. Bly & D. E. Rumelhart (Eds.), *Cognitive science. Handbook of perception and cognition* (2nd ed., pp. 367-384). San Diego, CA: Academic Press.
29. McQuiggan, S. W., and Lester, J. C. (2007). Leveraging Affect for Narrative-Centered Guided Discovery Learning Environments. Supplementary Proceedings of the *13th International Conference on Artificial Intelligence in Education (AIED 2007)*, (pp 67-74)
30. Paeschke, A., and Sendlmeier, W. (2000): Prosodic characteristics of emotional speech: Measurements of fundamental frequency movements, In *SpeechEmotion-2000*, 75-80.
31. Paiva, A., Prada, R., and Picard, R.W (Eds.) (2007). *Affective Computing and Intelligent Interaction*. Springer.
32. Pantic, M. & Rothkrantz, L. J. M. (2003). Towards an Affect-sensitive Multimodal Human-Computer Interaction. *Proceedings of the IEEE, Special Issue on Multimodal Human- Computer Interaction (HCI*), *91* (9), 1370-1390.
33. Piaget, J. (1952). *The origins of intelligence*. New York: International University Press.
34. Picard, R.W. (1997). *Affective Computing*. Cambridge, Mass.: MIT Press 1997.
35. Schank, R.C. (1986). Explanation Patterns: Understanding Mechanically and Creatively. Hillsdale, NJ: Erlbaum.
36. Schmidt, S. (2005). *Pers önlichkeitsaspekte in stimmen von anthropomorphen produkten.* Unpublished Master's thesis, Universit ät des Saarlandes.
37. Stein, N. L., & Levine, L. J. (1991). Making sense out of emotion. In W. Kessen, A. Ortony, & F, Kraik (Eds.), *Memories, thoughts, and emotions: Essays in honor of George Mandler* (pp. 295-322). Hillsdale, NJ: Erlbaum.
38. Woolf, B., Burelson, W., and Arroyo, I. (2007). Emotional Intelligence for Computer Tutors. Supplementary Proceedings of the *13th International Conference on Artificial Intelligence in Education (AIED 2007)*, (pp 6-15)
39. Weiner, B. (1986). An *attributional theory of motivation and emotion*. New York: Springer-Verlag.
40. Witten, I. H. & Frank E. (2005). Data Mining: Practical machine learning tools and techniques (2nd ed.). San Francisco, CA: Morgan Kaufmann.