Making a Case for Machine Perception of Trainee Affect to Aid Learning and Performance in Embedded Virtual Simulations

Robert Sottilare, Ph.D.
U.S. Army Research Development and Engineering Command
Simulation & Training Technology Center
12423 Research Parkway, Orlando, FL, USA 32826

robert.sottilare@us.army.mil

ABSTRACT

For our purposes, machine perception is defined as the ability of a computer-based training system to sense the behavior and affective state (e.g. mood or emotions) of trainees and interpret whether they are engaged, bored, frustrated, confused or even hostile during the training process. This paper puts forward the notion that the maturation of machine perception of trainee affect is critically important to optimizing learning for individuals and teams in embedded virtual simulations and other isolated training environments. Embedded training applications within operational platforms (e.g. tanks, aircraft, ships and individual Warfighting systems) continue to be explored today in many NATO countries (e.g. United States, Germany and the Netherlands). The lack of human tutors within operational platforms limits the understanding of each trainee's affective state and the completeness of the trainee model, the representation of the trainee's state within intelligent tutoring systems. Tutor technology is currently not sufficiently mature to provide accurate, portable, affordable, passive and effective sensing and interpretation of the trainee's affective state and limits the adaptability and effectiveness of the instruction in today's embedded training systems. This paper rationalizes the need for machine perception of affect in future embedded virtual simulations.

1.0 INTRODUCTION

Warfighters are exposed to artificial intelligence (AI) through computer-based tutors, virtual characters in games and other simulations, and expert decision support tools in their training environments today. This paper puts forward the notion that machine perception of affect will be a vitally important AI capability to support isolated, distributed training within operational systems also known as embedded training. Machine perception of affect in an intelligent tutoring context is the ability of computers to sense and interpret images, sounds and behaviors to determine which actions to take or strategies to employ to optimize the learning and performance of trainees.

As noted in Figure 1, instruction can be provided via human or intelligent tutoring systems (computer or machine-based systems). Intelligent tutoring systems support one-to-one training experiences, have limited capability to support one-to-many (collective) training and will be critical in the future to support embedded training applications within operational platforms (e.g. tanks, aircraft, ships and individual Warfighting systems) in NATO countries.

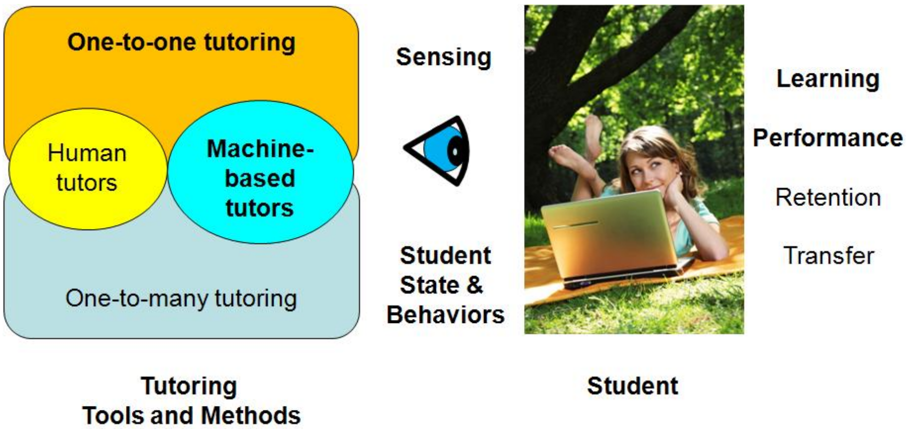

Figure 1: Conceptual Tutoring Model

The value of one-on-one tutoring vice group tutoring (i.e. traditional classroom teaching) has been documented among students who work one-to-one with expert human tutors and often score 2.0 standard deviations higher than students in a conventional classroom (Bloom, 1984).

A model of intelligent tutoring system interactions is shown in Figure 2. For the purposes of this paper, we will concentrate on machine perception of affect and its relationship to the trainee model (also referred to as the student or user model) and the pedagogical module.

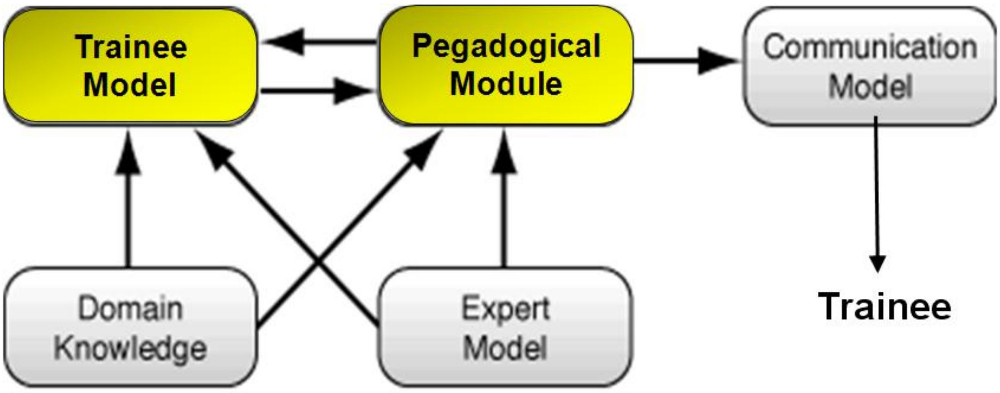

Figure 2: Intelligent Tutoring System Model (Beck, Stern and Haugsjaa, 1996)

The trainee model has generally been a record of the trainee's knowledge and performance history. It stores information specific to each individual learner including a history of performance and other pertinent data. This could include affective state information (i.e. personality, mood or emotions). The trainee model also records observable actions and may infer non-observable states (i.e confusion, boredom or other emotions). "Since the purpose of the student model is to provide data for the pedagogical module of the system, all of the information gathered should be able to be used by the tutor [pedagogical module]." (Beck, Stern and Haugsjaa, 1996)

The pedagogical module provides a model of the instruction process and contains logic for making decisions about when to review information, when to present new topics or concepts, and what instructional strategies to use. The sequencing of topics is controlled by the pedagogical module. Once the topic has been selected, a problem must be generated for the trainee to solve and then feedback is provided on the trainee's performance. As noted above, the trainee model is used as input to this component, so the pedagogical decisions should reflect the differing needs of each trainee (Beck, 1996).

Today, intelligent tutors focus primarily on trainee performance and their progress relative to training objectives. They tend to ignore other factors which might influence learning and performance outcomes. In general, computer-based training provides every trainee the same instruction regardless of their experience, competency level or state (e.g. cognitive, physical, affective state). The few embedded training environments that exist lack the human tutoring support that is typically part of instruction at military training centers.

Future embedded training systems should evolve to include intelligent tutors that will develop and maintain a complex trainee model that includes the real-time assessment of the trainee's state and other underlying influencers of performance. Each trainee will receive a tailored experience from an adaptive, computer-based intelligent tutor. The maturity of adaptive intelligent tutoring technology will be critical in providing training experiences that are comparable to human tutoring.

With improved machine perception, future intelligent tutoring systems within embedded training systems will be able to assess the trainee's affective state and adapt instruction to maintain the trainee's focus or overcome barriers unique to each individual Warfighter. This paper rationalizes the need for additional research in machine perception of affect for future embedded virtual simulations.

2.0 RATIONALE FOR ENHANCED MACHINE PERCEPTION OF AFFECT

Both the U.S. Army and Air Force have recently focused on requirements for "adaptive/tailored training, innovative learning models, strategies or tools" and "precision learning". "Precision learning delivers the appropriate education, training, or experience at the right time and place, in the right format, to generate the right effect. Precision learning relies on customized learning, mass collaboration, push and pull learning systems, distributed learning opportunities, increased use of simulated and virtual technology, and enhanced use of visualization technologies. It focuses learning on the learner" (Air Education & Training Command, 2008).

However, intelligent tutors within embedded training systems today have very limited or no ability to:

- evaluate the learner's cognitive/social/physical needs and adapt instruction to meet those needs and mitigate learning risks (e.g. boredom or confusion) (Sottilare & Proctor, 2009)
- perceive/predict the learner's affective state (e.g. mood or emotional) (Picard, 2006)
- adapt strategies and feedback to build rapport (trust) between learners and virtual tutors (Kang, Gratch & Wang, 2008)

One of the major limiting factors of intelligent tutoring systems is that they have yet to match the perceptiveness of human tutors. Intelligent tutors will eventually have the potential to post similar results to human tutors when they have the same capabilities to perceive the behaviors and state (e.g. cognitive, physical, emotional state) of their trainees to determine whether they are engaged, bored, frustrated, confused or even hostile during the training process. Why is this important?

Linnenbrink and Pintrich (2002) found that many trainees experience some confusion when confronted with information that does not fit their current knowledge base, but those in a generally positive affective state will adapt their known concepts to assimilate the new information. Trainees in a generally negative affective state will usually reject this new information. This infers the need for tutors (human or otherwise) to be able to perceive and address the affective state of the trainee and adapt instruction to optimize the assimilation of new information and enhance performance.

Improving machine perception will provide additional information for the trainee model within intelligent tutoring systems and enhanced trainee models will increase the probability of the intelligent tutor making more appropriate instructional strategy decisions for each individual trainee. To highlight current capabilities, a comparison of selected machine perception for tutoring research is explored in the next section.

3.0 COMPARING CURRENT METHODS OF MACHINE PERCEPTION FOR TUTORING

Given the large scale on which embedded training may be applied in military systems, machine perception methods should be accurate, portable, affordable, passive and effective:

- accurate: correctly senses images, distances, movements and other behaviors and precisely interprets their meaning
- portable: easily integrates into the operational equipment with minimal impact on weight and power
- affordable: inexpensive to acquire, integrate, use and maintain on a large scale
- passive: does not interfere or detract from the learning process; unobtrusive
- effective: enhances trainee models and selects appropriate instructional strategies that optimize learning and skill development

Below are three selected examples of recent machine perception methods, and their strengths and limitations relative to the established criteria.

Yun, Shastri, Pavlidis and Deng (2009) demonstrated passive sensing and interpretation of thermal images to estimate user stress. They altered the difficulty levels of game play for users based on singular input from StressCam, which monitors heat dissipation through a thermal imaging-based camera and analysis system. Since stress levels are related with increased blood flow in the forehead and higher blood flow equates to more heat, StressCam can passively and continuously sense and interpret thermal images as stress and frustration level.

One limitation of this technology is the narrow focus on frustration while other emotions are not interpreted. Another is that the cause of increased stress/frustration is unknown using this method. Yet another is the portability of this system, which is questionable at best as part of a man-worn system based on its weight and power requirements, but may be satisfactory for vehicle-based systems. System cost may be an issue now, but expect the cost of thermal imaging camera to decrease over the next five years. Finally, the focus of the StressCam experiment was to manipulate the difficulty level of the game to increase/reduce stress and no evidence was provided regarding the use of this technology as a tool for determining instructional strategies (e.g. challenging, supporting or pumping).

Neji and Ben Ammar (2007) investigated an intelligent tutoring system that included an embodied conversational agent. In addition to the two-way conversational input and output, the agent behavior was informed about the emotional state of the trainee through a machine vision system. The machine vision system perceived changes in facial expressions of the trainee and based on distances between facial landmarks classified the expression as one of six universal emotional states (joy, sadness, anger, fear, disgust and surprise) or a neutral expression. Emotional state was then used in the ITS to determine which tutoring strategy (e.g. sympathizing or non-sympathizing feedback, motivation, explanation, steering). The internal state of the agent is based on the PECS (Physical conditions, Emotional state, Cognitive capabilities and Social status) architecture proposed by Schmidt (2000).

A significant drawback to Neji and Ben Ammar's (2007) "Affective e-Learning Framework" is the cost of the vision system which limits its deployability to operational systems for embedded training. While their approach provides key components (emotional sensing and perception, selection of instructional strategies and interactions based on learner emotional state and the PECS architecture) for an adaptable tutoring system, it does not assess: whether the intelligent tutor's perception of the affective state of the learner aids the intelligent tutor in selecting appropriate instructional strategies that result in enhanced learning outcomes or performance; the influence of other affective variables (e.g. mood components like pleasure and arousal) have on learner outcomes or how these affective variables might influence each other.

Finally, D'Mello, Craig , Sullins and Graesser (2006) and D'Mello and Graesser (2007) used frequent conversation patterns to predict affective states when trainees emote aloud. Frequent conversation patterns significantly predicted trainees' affective states (i.e. confusion, eureka, frustration) and provided feedback, pumps, hints and assertions to influence the trainee's progress. The primary drawback to this approach was the requirement for trainees to "emote aloud" which has some of same drawbacks as other self-report methods and may be incompatible with trainees with lower openness scores in personality assessments like the Big Five Personality Test. Another drawback was the low participant throughput for the experiment based on the labor intensive nature of the data collection and analysis. Given the variability among trainees and the associated time to baseline each trainee, this method is unsuitable as is for embedded training applications.

Other methods explored (Zhou and Conati, 2003; Zimmermann, Guttormsen, Danuser and Gomez, 2003; Burleson and Picard, 2004) have similar limitations relative to the criteria noted above.

4.0 TOWARD MACHINE PERCEPTION OF AFFECT IN EMBEDDED TRAINING

The methods of machine perception selected for evaluation in this paper were based on single mode, passive physiological measures. The primary advantages of passive physiological methods are that the technology is not invasive and therefore does not detract from the learning process; and the attention of the trainee is not drawn to what the intelligent tutoring system is trying to measure. The disadvantages of passive physiological methods are the time and difficulty to setup sensors for measurement; and the analysis and interpretation of physiological data is often difficult given the variability of individual trainees.

McIntyre and Göcke (2007) have advocated a multimodal approach to reduce the uncertainty associated with physiological measurements. They have also identified two major problems to be overcome in developing a multimodal computer system capable of sensing the affective state of a trainee:

- lack of an ontology to describe affective states
- failure to incorporate individual nuances resulting in an oversimplified picture of affective expression

A multimodal approach could provide tailored training, but there are drawbacks per Bjork (2006) in having the tutor (human or computer-based) take too active a role in guiding training. Bjork (2006) suggests that consideration should be given to balance near-term objectives (performance and learning) and longer term objectives (retention and transfer) by offering sufficient challenges so that training experiences will have lasting effect.

Given the strengths and limitations of current machine perception technology, and considerations for learning, performance and retention, questions will need to be resolved to realize a fully autonomous embedded training capability. What barriers need to be overcome and what capabilities are required to realize machine perception of affect in future embedded training systems? Below are some suggested capabilities to realize machine perception of affect for embedded training systems in the future:

- Ontology: an ontology specific to machine perception of affect which defines affective states and describes distinguishing factors between each
- Personalized Sensors: unobtrusive, multimodal sensors tailored to each individual with sufficient historical data to accurately classify (with > 99% probability) the trainee’s affective state
- Personalized Strategies: instructional strategies tailored to each individual with sufficient historical data to accurately classify the appropriate intervention to mitigate the risk of negative predictions of performance or to maintain positive predictions of performance; strategies should be scenario independent and balance near (performance and learning) and long term objectives (retention and transfer)
- Independent Measures: establish the weights of independent sensing modes
- Real-time or Near Real-time Strategies: instructional strategies fed by continuous updates to the affective trainee model to support learning objectives
- Embedded Capabilities: weight, power and size tailored to support individual and platform-based embedded training; most challenging design for individual Soldiers/Marines

5.0 CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

Many of the limitations noted for intelligent tutoring systems and machine perception are not unique to embedded training, but are exacerbated by the challenging environments in which embedded training takes place: in vehicles, ships, aircraft and on individual Soldiers/Marines. The current state-of-the-art in machine perception has limitations, but also significant potential on which to build future capabilities.

Machine perception will be key in expanding the dimensions of the trainee model to include affect will provide additional information from which to make decisions on instructional content, coaching strategies and feedback to the trainee as shown in Figure 3.

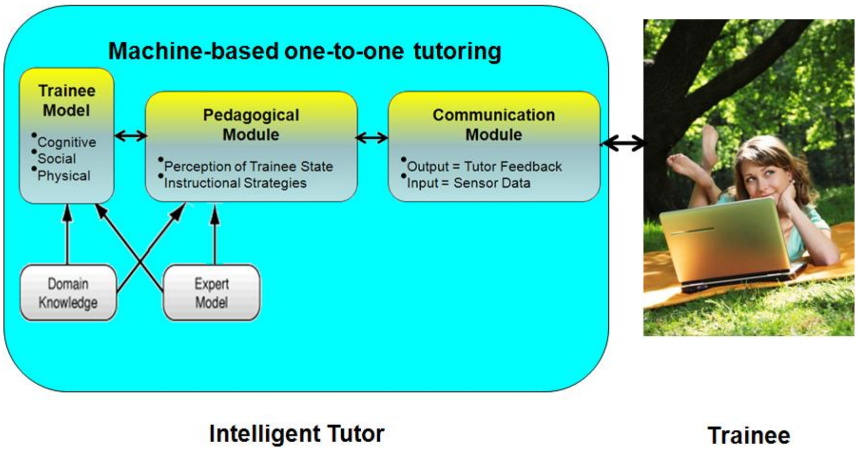

Figure 3: Machine-based one-to-one tutoring

Three areas of machine perception research are recommended to expand embedded training capabilities:

- Sensors: research to improve the reliability and accuracy of multimodal, passive sensor suites for individual and collective embedded training
- Predictive Models: research to improve the accuracy of predictive models across populations, training missions and scenarios to select instructional strategies that are optimized for learning, retention, transfer and performance for individuals and teams in embedded training environments
- Building Trust: research to use data from the affective trainee model to shape the behavior of virtual human tutors in embedded training environments to build/maintain trust/rapport between the tutor and the trainee

REFERENCES

[1] Air Force Education & Training Command (2009). On Learning: The Future of Air Force Education and Training.

[2] Beck, J., Stern, M., and Haugsjaa, E. (1996). Applications of AI in Education, ACM Crossroads.

[3] Bjork, R. (2006). How We Learn Versus How We Think We Learn: Implications for the Design and Evaluation of Instruction. Reinvention Center Conference, Washington, D.C., November 2006.

[4] Bloom, Benjamin S. (1984) The 2-sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher 13: 4-16.

[5] Burleson, W. and R. Picard (2004). Affective agents: Sustaining motivation to learn through failure and a state of stuck. In Social and Emotional Intelligence in Learning Environments Workshop In conjunction with the 7th International Conference on Intelligent Tutoring Systems.

[6] D'Mello, S.K., Craig , S.D, Sullins J. and Graesser, A.C. (2006). Predicting Affective States expressed through an Emote-Aloud Procedure from AutoTutor's Mixed-Initiative Dialogue, International Journal of Artificial Intelligence in Education, v.16 n.1, p.3-28, January 2006

[7] D'Mello, S. and Graesser, A. (2007). Mind and Body: Dialogue and Posture for Affect Detection in Learning Environments. In Proceedings of the 13th International Conference on Artificial Intelligence in Education (AIED), Marina del Rey, CA.

[8] Kang, S., Gratch, J., & Wang, N. Agreeable people like agreeable virtual humans. Proceedings of Intelligent Virtual Agent 2008, Tokyo, Japan, H. Prendinger, J. Lester, and M. Ishizuka (Eds.): pp. 253–261, Springer-Verlag, Berlin Heidelberg

[9] Linnenbrink, E. A. and Pintrich, P. R. (2002). Motivation as an enabler for academic success. School Psychology Review, 31, 313-327.

[10] McIntyre, G. and Göcke, R. (2007). Towards Affective Sensing. In: Jacko, Julie A. (ed.) HCI International 2007 - 12th International Conference - Part III 2007. pp. 411-420.

[11] Neji, M. and Ben Ammar, M. (2007). Agent-based Collaborative Affective e-Learning Framework. Electronic Journal of e-Learning Volume 5 Issue 2, pp. 123 – 134.

[12] Picard, R. (2006). Building an Affective Learning Companion. Keynote address at the 8th International Conference on Intelligent Tutoring Systems, Jhongli, Taiwan. Retrieved from http://www.its2006.org/ITS_keynote/ITS2006_01.pdf

[13] Schmidt, B. (2000). The Modelling of Human Behaviour. SCS-Europe BVBA, Ghent.

[14] Sottilare, R. and Proctor, M. (2009). Using student mood and task performance to train classifier algorithms to select effective coaching strategies within Intelligent Tutoring Systems (ITS). Unpublished manuscript submitted to the International Journal of Artificial Intelligence in Education.

[15] Yun, C., Shastri, D., Pavlidis, I. and Deng, Z. (2009). O' Game, Can You Feel My Frustration?: Improving User's Gaming Experience via StressCam. Computer-Human Interaction (CHI) 2009.

[16] Zhou X. and Conati C. (2003). Inferring User Goals from Personality and Behavior in a Causal Model of User Affect. In Proceedings of IUI 2003.

[17] Zimmermann, P., Guttormsen, S., Danuser, B. and Gomez. P. (2003). Affective Computing - A Rationale for Measuring Mood with Mouse and Keyboard. International Journal of Occupational Safety and Ergonomics.