RECOGNITION AND GENERATION OF EMOTIONS IN AFFECTIVE e-LEARNING

Efthymios Alepis, Maria Virvou
Department of Informatics, University of Piraeus, 80 Karaoli & Dimitriou St., 18534, Piraeus, Greece

Katerina Kabassi
Department of Ecology and the Environment, Technological Educational Institute of the Ionian Islands
2 Kalvou Sq., 29100 Zakynthos, Greece

Keywords: e-Learning, Affective Interaction, Bi-modal interaction, Generation of emotions, OCC theory, Decision making theories, Educational agents.

Abstract: This paper presents an educational system that incorporates two theories namely SAW and OCC in order to provide an improved affective e-learning environment. Simple additive Weighting (SAW) is used for the recognition of possible emotional states of the users, while the cognitive theory of emotions (OCC) is used for the generation of emotional states by educational agents. The system bases its inferences about users' emotions on user input evidence from the keyboard and the microphone, as two commonly used modalities of human-computer interaction. The actual combination of evidence from these two modes of interaction has been performed based on a sophisticated inference mechanism for emotions and a multi-attribute decision making theory. At the same time, user action evidence from the two modes of interaction also activates the cognitive mechanisms of the underlying OCC model that proposes emotional behavioural tactics for educational agents who act for pedagogic purposes. The presented educational system provides the important facility to authors to develop tutoring systems that incorporate emotional agents who can be parameterized so as to reflect their vision of teaching behaviour.

1 INTRODUCTION

One of the major scientific challenges is the exploration of how humans interact with their environment and with each other. Perceiving, learning and adapting to the world around us are commonly labelled as intelligent behaviour (Pantic & Rothkrantz, 2003). In many situations human-computer interaction may be improved by multimodal emotional interaction in real time (Jascanu, 2008), (Bernhardt, 2008). Affective computing has recently become a very important field of research because it focuses on recognizing and reproducing human feelings within human computer interaction. Human feelings are considered very important but only recently have started being taken into account in software user interfaces. Thus, the area of affective computing is not yet well understood and needs a lot more research to reach maturity.

In the last decade, education has benefited a lot from the advances of Web-based technology. Indeed, there have been many research efforts to transfer the technology of ITSs and authoring tools over the Internet. Past reviews (Lane, 2006, Brusilovsky, 1999) have shown that all well-known technologies from the areas of ITS have already been re-implemented for the Web. Some important assets include platform-independence and the practical facility that is offered to instructors of authoring e-learning courses at any time and any place. However, this independence from real instructors and classrooms may cause emotional problems to students who may feel deprived of the benefits of human-human interaction. This may affect the educational process in a negative way. A remedy for these problems may lie in rendering human-computer interaction more human-like and affective in educational software. To this end, the incorporation of speaking, animated educational agents in the user interface of the educational application can be very important.

Alepis E., Virvou M. and Kabassi K. (2009).
RECOGNITION AND GENERATION OF EMOTIONS IN AFFECTIVE e-LEARNING.
In *Proceedings of the 4th International Conference on Software and Data Technologies*, pages 273-280
DOI: 10.5220/0002242802730280

Indeed, the presence of animated, speaking educational agents has been considered beneficial for educational software (Johnson et. al., 2000, Lester et. al., 1997). Instructors that may use educational authoring tools should not necessarily be computer experts and should be helped to develop sophisticated educational applications in an easy and cost-effective way (Virvou & Alepis, 2005).

Affective computing may be incorporated into sophisticated educational applications by providing adaptive interaction based on the user's emotional state. Regardless of the various emotional paradigms, neurologists/psychologists have made progress in demonstrating that emotion is at least as and perhaps even more important than reason in the process of decision making and action deciding (Leon et al., 2007). Moreover, the way people feel may play an important role in their cognitive processes as well (Goleman, 1995).

Indeed, Picard points out that one of the major challenges in affective computing is to try to improve the accuracy of recognizing people's emotions (Picard, 2003). Ideally, evidence from many modes of interaction should be combined by a computer system so that it can generate as valid hypotheses as possible about users' emotions. It is hoped that the multimodal approach may provide not only better performance, but also more robustness (Pantic & Rothkrantz, 2003).

In previous work, the authors of this paper have implemented and evaluated with quite satisfactory results from the users' perspective, other educational systems with emotion recognition capabilities (Alepis et al. 2007). As a next step we have extended our affective educational system by employing fully programmed educational agents that are able to express a variety of emotions.

Educational agents may be parameterized in many aspects, the way they speak, the pitch, speed and volume of their voice, their body-language, their facial expressions and the content of their messages. Educational agents are able to express specific pedagogical emotional states by the incorporation of the OCC (Ortony et. al., 1990) model. The resulting educational system incorporates an affective authoring module that relies on the OCC theory. The system uses the OCC cognitive theory of emotions for modelling possible emotional states of users-students and proposing tactics to the instructors for improving the interaction between the educational agent and the student while using the educational application. Through the incorporation of the OCC model, the system may suggest that the tutoring educational agent should express a specific emotional state to the student for the purpose of motivating her/him while s/he learns. Consequently, the educational agent may become a more effective instructor, reflecting the instructors' vision of teaching behaviour.

However, as yet there are no authoring tools that provide parameterization in user interface components such as speech-driven, animated educational agents. The present educational system provides the facility to authors to develop tutoring systems that incorporate speaking, animated emotional agents who can be parameterized by the authors-instructors in a way that reflects their vision of teaching behaviour in the user interface of the resulting applications.

2 OVERVIEW OF THE SYSTEM

The educational application is installed either on a public computer where both students and instructors have access, or alternatively each student may have a copy on his/her own personal computer. The underlying reasoning of the system is based on the student modelling process of the educational application. The system monitors and records all students' actions while they use the educational application and tries to diagnose possible problems, recognise goals, record permanent habits and errors that are made repeatedly. Help is provided through the tutoring agents that not only support the students' educational process, but also interact affectively with the students by expressing emotional states. The incorporated model that controls the tutoring agents' behaviour is described in section 4. The inferences made by the system concerning the students' characteristics are recorded in their student model. Hence, the system offers advice adapted to the needs of individual students. The system's database is used to hold all the necessary information that is needed for the application to run and additionally to keep analytical records of the performance of all the students that use the educational application.

While using the educational application from a desktop computer, students are able to retrieve information about a particular course. In the example of Figure 1 a student is using the e-learning system for a medical course about anatomy. The information is given in text-form while at the same time an animated agent reads it using a speech engine. Students may choose specific parts of the theory and the available information is retrieved from the system's database.

Figure 1 illustrates the main form of the educational application on a desktop computer.

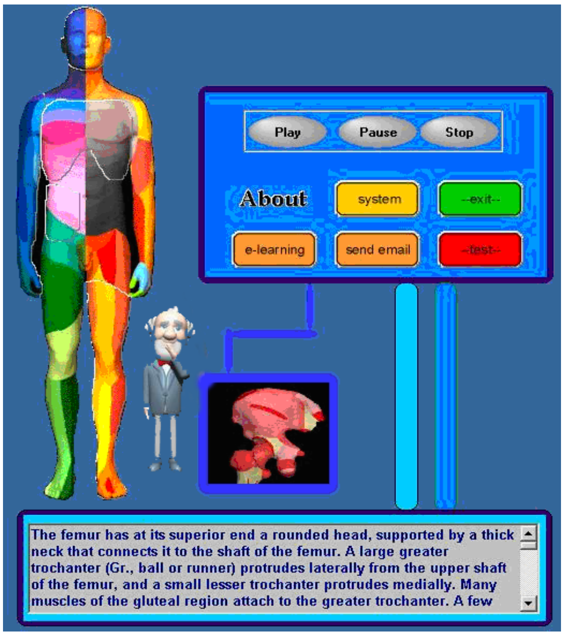

Figure 1: The main form of the application with the presence of the tutoring agent.

Similarly, students are able to take tests that include questions, answers, multiple-choice, etc, concerning specific parts of the theory. The tutoring agent is also present in these modes, in order to make the interaction more human-like and to assist the student by providing pedagogical assistance when it is needed.

3 OVERVIEW OF THE EDUCATIONAL AGENTS

The educational applications that result from the authoring process described in this paper incorporate a tutoring agent that is a cartoon-doctor. The cartoon-doctor is a fully programmable agent who can move around the tutoring text and can show parts of the theory in real time (Figure 2). It has also incorporated features of human body-language. The educational agent may show patience while the student reads the theory, boredom if the student is not responding to the system, wonder if the student makes an unexpected move, etc. The cartoon-doctor's behaviour is programmatically controlled by an underlying mechanism that relies on the OCC theory, described in the next section.

Instructors may choose from 27 available speech engines that the system incorporates. These speech engines are synthesisers that produce different voices. The system also offers the facility of parameterising these voices by changing the pitch, speed and volume, as illustrated in figure 3. Thus, the resulting tutoring system may use the voices differently in different contexts to show enthusiasm, when the student is doing particularly well, to imitate whisper, when it judges that the student needs help, or even to show anger when the student is consistently careless and does not pay any attention to the educational system.

Figure 2: The cartoon-doctor.

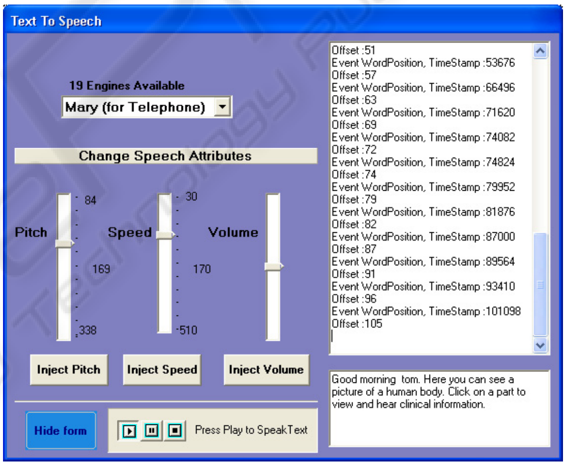

Figure 3: Setting parameters for the voice of the tutoring agent.

In order to produce an "angry" tone of speaking for the animated agent, as an example, instructors may increase the pitch the speed and the volume of the speech engine. This may also be achieved by selecting an appropriate speech engine from the ones that are available. Additionally, the instructor may use the form of Figure 4 that provides more specific and detailed controls. In this form instructors have also the ability to set the exact pronunciation of a word by using phonemes.

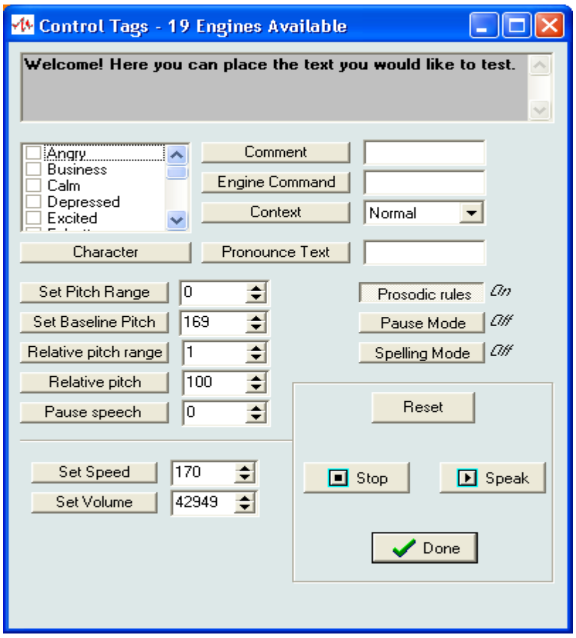

Figure 4: Detailed controls for the voice of the tutoring character.

The system incorporates built in tools, to which only the instructors have access. These tools help the instructors modify the behaviour of the characters further, with the agents' emotion generation facility as the final objective. Not only can the instructor command the assistant to say something under certain circumstances, but s/he can also add commands in the text that will be spoken, in a way that the agent may seem to express a specific emotional state. These commands are understood by the system and are interpreted into changing speech attributes, body movements, facial expressions, etc.

4 EMOTION RECOGNITION AND EMOTION GENERATION

4.1 Recognizing Emotional States

A user monitoring component has been used to capture all user input data during the interaction with the educational application. The monitoring component is illustrated in figure 5. Input data consist of audio information that has been collected through the keyboard, as well as audio information that has been collected through the microphone.

The analysis of the data collected by the monitoring component, revealed some statistical results that associated user input actions through the

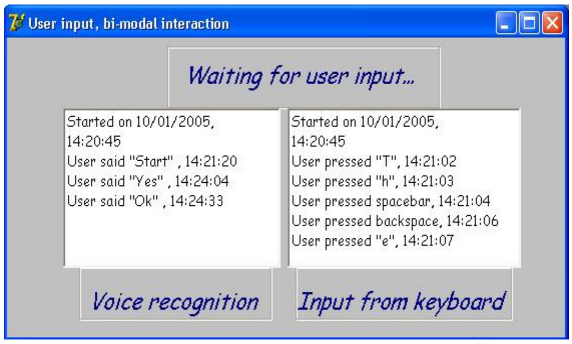

Figure 5: Snapshot of operation of the user modelling component.

computer's keyboard and microphone with possible emotional states of the users. More specifically, considering the keyboard we have the following categories of user actions: a) user types normally b) user types quickly (speed higher than the usual speed of the particular user) c) user types slowly (speed lower than the usual speed of the particular user) d) user uses the "delete" key of the keyboard often e) user presses unrelated keys on the keyboard f) user does not use the keyboard.

Considering the users' basic input actions through the computer's microphone we have 7 cases: a) user speaks using strong language b) users uses exclamations c) user speaks with a high voice volume (higher than the average recorded level) d) user speaks with a low voice volume (low than the average recorded level) e) user speaks in a normal voice volume f) user speaks words from a specific list of words showing an emotion g) user does not say anything.

Therefore, whenever an input action is detected the system records a vector of input actions through the keyboard (k1, k2, k3, k4, k5, k6) and a vector of input actions through the microphone (m1, m2, m3, m4, m5, m6, m7).

All the above mentioned attributes are used as Boolean variables. In each moment the system takes data from the bi-modal interface and translates them in terms of keyboard and microphone actions. If an action has occurred the corresponding attribute takes the value 1, otherwise its value is set to 0. Therefore, for a user that speaks with a high voice volume and types quickly the two vectors that are recorded by the system are: k= (0, 1, 0, 0, 0, 0) and m= (0, 0, 1, 0, 0, 0, 0). These data are further processed by the decision making model for determining the emotion of the user.

A previous empirical study revealed the attributes that are taken into account when evaluating different emotions (Alepis & Virvou, 2006). However, these attributes were not equally important for evaluating different emotions. In this study human experts resulted that one input action does not have the same weight while evaluating different emotions. Therefore, the weights of the attributes (input actions) were calculated in order to be used by the decision making model.

For the evaluation of each alternative emotion the system uses SAW (Fishburn, 1967, Hwang & Yoon, 1981) for a particular category of users. According to SAW, the multi-attribute utility function for each emotion in each mode is estimated as a linear combination of the values of the attributes that correspond to that mode.

The SAW approach consists of translating a decision problem into the optimisation of some multi-attribute utility function U defined on A. The decision maker estimates the value of function $U(X_j)$ for every alternative X_j and selects the one with the highest value. The multi-attribute utility function U can be calculated in the SAW method as a linear combination of the values of the n attributes:

$$U(X_j) = \sum_{i=1}^{n} w_i x_{ij} \tag{1}$$

where X_j is one alternative and x_{ij} is the value of the i attribute for the X_j alternative.

In view of the above, for the evaluation of each emotion taking into account the information provided by the keyboard is done using formula 2.

$$\boxed{em_{ke_1} = w_{e_1k1}k_1 + w_{e_1k2}k_2 + w_{e_1k3}k_3 + w_{e_1k4}k_4} \\ \boxed{+ w_{e_1k5}k_5 + w_{e_1k6}k_6} \tag{2}$$

Similarly, for the evaluation of each emotion taking into account the information provided by the second mode (microphone) is done using formula 3.

$$\boxed{em_{me_1} = w_{e_1m1}m_1 + w_{e_1m2}m_2 + w_{e_1m3}m_3 + w_{e_1m4}m_4} \\ \boxed{+ w_{e_1m5}m_5 + w_{e_1m6}m_6 + w_{e_1m7}m_7} \tag{3}$$

em_{ke_1} is the probability that an emotion has occurred based on the keyboard actions and em_{me_1} is the probability that refers to an emotional state using the users' input from the microphone em_{ke_1} and em_{me_1} take their values in [0,1].

In formula 1 the *k*'s from *k1* to *k6* refer to the six attributes that correspond to the keyboard. In formula 2 the *m*'s from *m1* to *m7* refer to the seven attributes that correspond to the microphone. The *w*'s represent the weights of the attributes. These weights correspond to a specific emotion and to a specific input action and were calculated in the pre mentioned empirical study.

In cases where both modals (keyboard and microphone) indicate the same emotion then the probability that this emotion has occurred increases significantly. Otherwise, the mean of the values that have occurred by the evaluation of each emotion using formulae 1 and 2 is calculated.

The system compares the values from all the different emotions and selects the one with the highest value of the multi-attribute utility function. The emotion that maximises this function is selected as the user's emotion.

4.2 Agents that Act Emotionally

Through the incorporation of the OCC theory, the system may suggest that the educational agent should express a specific emotional state to the student for the purpose of motivating her/him while s/he learns. Accordingly, the agent becomes a more effective instructor.

In OCC theory, emotional states arise from cognitive models that measure positive and negative reactions of users to situations consisting of events, agents and objects. Correspondingly, events match user goals that are key elements in the OCC theory.

Table 1: Variables for calculating the intensity of events for the OCC theory.

Event variables
• a mistake (the user may receive an error message by the application or navigate wrongly)
• many consecutive mistakes
• absence of user action for a period of time
• action unrelated to the main application
• correct interaction
• many consecutive correct answers (related to a specific test)
• many consecutive wrong answers (related to a specific test)
• user aborts an exercise
• user aborts reading the whole theory
• user requests help from the agent
• user takes a difficult test
• user takes an easy test
• user takes a test concerning a new part of the theory
• user takes a test from a well known part of the theory

Table 2: Variables for user actions through the microphone and the keyboard.

Variables of user actions through keyboard and microphone
• user types normally • user types quickly (speed higher than the usual speed of the particular user) • user types slowly (speed lower than the usual speed of the particular user) • user uses the backspace key often • user hits unrelated keys on the keyboard • user does not use the keyboard • user speaks words from a specific list of words showing an emotion • user does not say anything • user speaks with a low voice volume (lower than the average recorded level) • user speaks in a normal voice volume • user speaks with a high voice volume (higher than the average recorded level)

Tables 1 and 2 illustrate representative subsets of intensity variables concerning user input actions and application events that are used by the system's adapted OCC emotion model in order to propose an emotional state as a educational tactic for the animated agent. The variables illustrated in tables 1, 2 have been specified in our own implementation and adaptation of OCC in our educational application. The application's user interface is multi-modal, thus it is possible for the system to monitor and record user actions such as speed of typing through the keyboard as well as low voice volume through the microphone etc. The proposed authoring system integrates the OCC model by comprising a subset of five basic emotional states, namely happiness, sadness, anger, fear and surprise. Each one of the above mentioned five emotional states can be synthesized by the animated agent, as it is illustrated in figure 6.

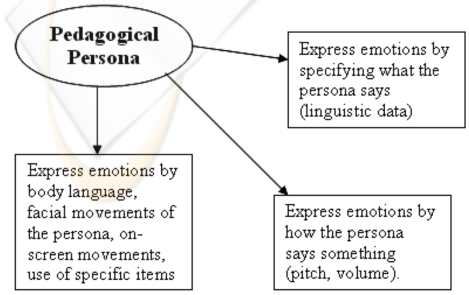

Figure 6: Events-Actions of the agent for the synthesis of an emotional state.

As an example we describe a situation where a student is taking a multiple choice test after having read the corresponding theory of that lesson. The "default" goal for each user is to succeed in answering correctly the questions of each test. In our example we assume that the difficulty level of the test is high and the student has already answered a couple of questions correctly. At this point in accordance with the system's incorporated OCC model the student is pleased that s/he has answered correctly the previous questions and is also experiencing hope that s/he will continue answering correctly. The corresponding intensity variables for this event are illustrated in table 1 as "many consecutive correct answers (related to a specific test)" and "user takes a test concerning a new part of the theory". The second variable indicates that succeeding in such a test is difficult, thus can invoke admiration by the educational agent. The user's hope of continuing to answer correctly may then be encouraged by the educational agent by expressing admiration for the student's success and encouraging her/him to continue answering successfully. In this case the student has a "goal" for answering correctly. If the student continues her/his successful course the educational agent will express happiness, by saying something in a "happy voice" and/or by smiling or doing a positive gesture. This behaviour by the agent results by the analysis of the event variables of the interaction as well as by the goals both the student and the agent have set. The incorporation of the OCC theory provides the reasoning mechanism in deciding which emotional state is more appropriate for the agent in each sequence of events and user actions. Finally each one of the possible OCC states are translated by means of the five basic emotions the agent can express (for example confirmed hope, joy and admiration are OCC states that the agent expresses as happiness). Intensity variables as well as user and agents goals are illustrated in our simplified OCC model in figure 7.

At this point we may also describe a situation with a negative emotional state issue. A student has already made many mistakes (consecutive mistakes event variable) in taking a test concerning parts of the theory that s/he is expected to know well. The educational agent makes a remark about this, advising her/him to be more attentive. More consecutive mistakes will trigger the state of reproach for the educational agent and finally will conclude in the expression of the agent's anger or sadness. In both cases (positive and negative) other event variables may be triggered at the same time.

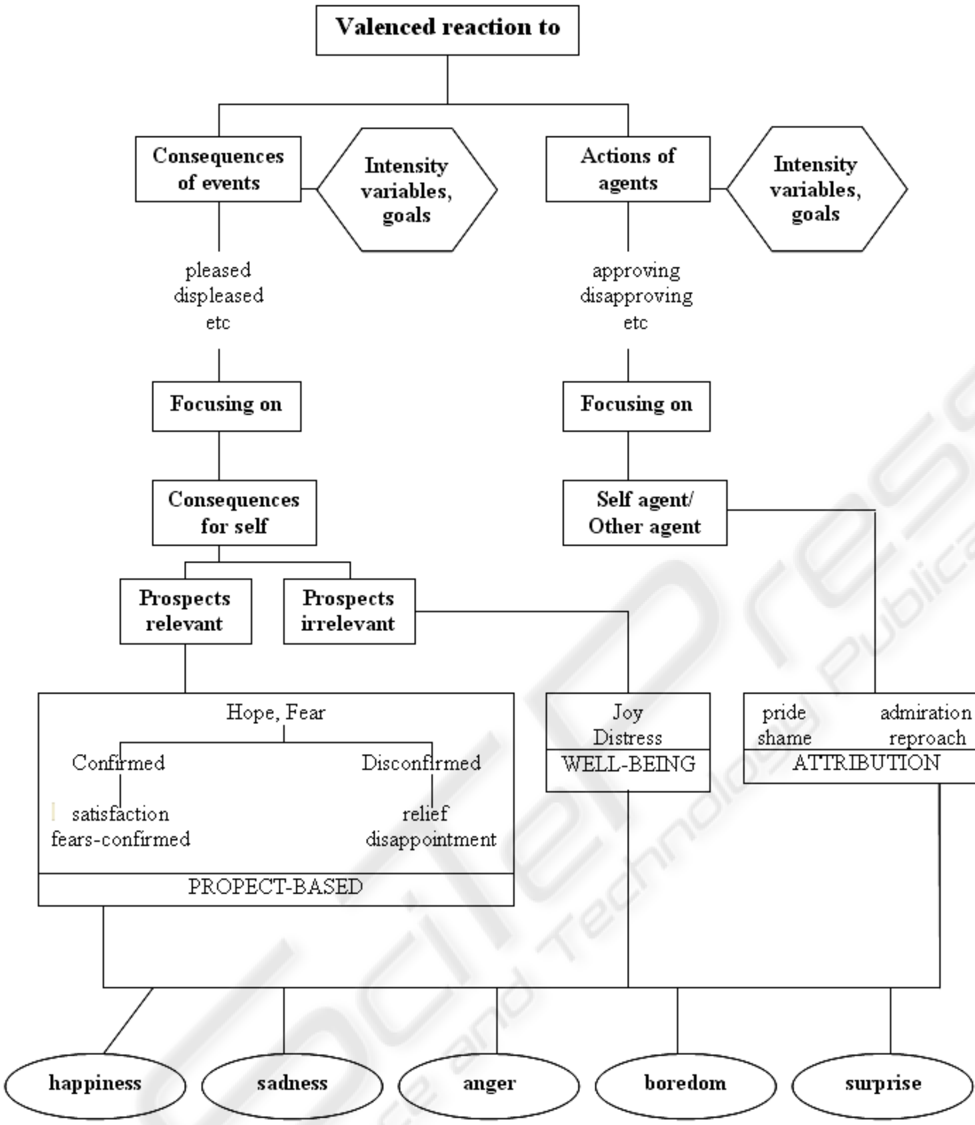

Figure 7: Incorporation of the OCC model for specifying the agent's emotional state.

For example, if a user spends time without using the educational application, the "absence of user action for a period of time" variable is triggered and the OCC model of the agent may suggest the expression of the agent's boredom. In addition, if the user aborts taking a test this would trigger the "user aborts an exercise" variable and the agent is going to express sadness or anger.

5 CONCLUSIONS AND FUTURE WORK

In this paper we have described how two theories, one from the field of decision making and one from the field of cognitive psychology, have been adapted and incorporated into an affective educational system. More specifically, we describe the implementation of an affective educational application that recognizes students' emotions based on keyboard and microphone input actions and proposes tactics for the behaviour of educational agents based on pedagogical procedures. The resulting educational application employs a bi-modal user interface.

It is among our future plans to evaluate the affective educational system in order to determine the degree of the system's usefulness for the students and also for their instructors. Furthermore, we intend to enrich multi-modal interaction by incorporating a third mode of interaction, visual this time (Stathopoulou & Tsihrintzis, 2005).

ACKNOWLEDGEMENTS

Support for this work was provided by the General Secretariat of Research and Technology, Greece, under the auspices of the PENED-2003 program.

REFERENCES

Alepis, E., Virvou, M., Kabassi, K., 2007. Knowledge Engineering for Affective Bi-modal Human-Computer Interaction, SIGMAP.

Alepis, E. & Virvou, M., 2006. Emotional Intelligence: Constructing user stereotypes for affective bi- modal interaction. In Lecture Notes in Computer Science: “Knowledge-based Intelligent Information and Engineering Systems”, Springer-Verlag Berlin Heidelberg 2006, Volume 4251 LNAI - I, 2006, Pages 435-442

Bernhardt, D., Robinson, P. : Interactive control of music using emotional body expressions, Conference on Human Factors in Computing Systems – Proceedings, pp. 3117-3122 (2008)

Brusilovsky, P., 1999. Adaptive and Intelligent Technologies for Web-based Education. In C. Rollinger and C. Peylo (eds.), Künstliche Intelligenz (4), Special Issue on Intelligent Systems and Teleteaching, 19-25.

Fishburn, P.C., 1967. Additive Utilities with Incomplete Product Set: Applications to Priorities and Assignments, Operations Research.

Goleman, D., 1995. Emotional Intelligence, Bantam Books, New York .

Hwang, C.L., Yoon, K., 1981. Multiple Attribute Decision Making: Methods and Applications. Lecture Notes in Economics and Mathematical Systems 186, Springer, Berlin/Heidelberg/New York.

Jascanu, N., Jascanu, V., Bumbaru, S., 2008. Toward emotional e-commerce: The customer agent, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) Volume 5177 LNAI, Issue PART 1, pp. 202-209

Johnson, W. L, J. Rickel, and Lester, J., 2000. Animated Educational Agents: Face-to-Face Interaction in Interactive Learning Environments. International Journal of Artificial Intelligence in Education, vol. 11, pp. 47-78.

Lane, H.C. 2006. Intelligent Tutoring Systems: Prospects for Guided Practice and Efficient Learning. Whitepaper for the Army's Science of Learning Workshop, Hampton, VA.

Leon, E., Clarke, G., Gallaghan, V., Sepulveda, F., 2007. A user-independent real-time emotion recognition system for software agents in domestic environments. Engineering applications of artificial intelligence, 20 (3): 337-345.

Lester, J., Converse, S., Kahler, S., Barlow, S., Stone, B., and Bhogal, R. 1997. The Persona Effect: affective impact of animated educational agents. In Pemberton S. (Ed.) Human Factors in Computing Systems, CHI’ 97, Conference Proceedings, ACM Press, pp. 359-366.

Pantic, M., Rothkrantz, L.J.M., 2003. Toward an affect-sensitive multimodal human-cumputer interaction. Vol. 91, Proceedings of the IEEE, Institute of Electrical and Electronics Engineers, pp. 1370-1390.

Picard, R.W., 2003. Affective Computing: Challenges. Int. Journal of Human-Computer Studies, Vol. 59, Issues 1-2, pp. 55-64.

Stathopoulou, I.O., Tsihrintzis, G.A., 2005. Detection and Expression Classification System for Face Images (FADECS), IEEE Workshop on Signal Processing Systems, Athens, Greece.

Virvou, M. & Alepis, E. (2005). Mobile educational features in authoring tools for personalised tutoring. Computers & Education, Volume 44, Issue 1, January 2005, Pages 53-68.