Indian Journal of Science and Technology, *Vol 8(24), DOI: 10.17485/ijst/2015/v8i24/80145, September 2015*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Cognitive Intelligent Tutoring System based on Affective State

N. Rajkumar* and V. Ramalingam

Department of Computer Science and Engineering, Annamalai University, Chidambaram - 608002, Tamil Nadu, India; rajimpc@gmail.com, aucsevr@gmail.com

Abstract

This paper presents a novel Intelligent Tutoring System (ITS). The goal of our virtual tutor is to mimic like a human tutor in order to advance the communication and effectiveness of the students learning experience. Towards these goals we designed a Conversational Intelligent Tutoring System (CITS) with multimodal behavior, emotive speech and friendliness. Moreover to improve the student's interaction and to engage their attention, our virtual tutor attempts to ask questions from their thought lesson and assists the students during learning activities. A statistical Friedman analysis was conducted and the results revealed that, our intelligent virtual tutoring system can able to successfully recognize student's behavior and respond according to it. And also participants were asked later to rate the teaching and learning environment of our intelligent system, the review shows that they feel lively.

Keywords: Emotive Speech, Friendliness Behavior, Virtual Agent

1. Introduction

From the last decade intelligent agents are widely in implementing Intelligent Tutoring Systems (ITS). ITS provides individualized content delivery to the students by adding artificial intelligence to improve the effectiveness of a learners experience[1]. Due to the emerging technology in computer vision, speech, and virtual agent behavior provides effective transform interaction with the computer based virtual agents and also Embedding an affect recognition component in an intelligent tutoring system will enhance its ability to provide necessary guidance, and make the tutoring sessions more effective and interactive. ITS can able to provide students with the effectiveness of class room teachers in several ways such as positive, individualized and powerful learning experience for the students. Moreover, research evidence argues that humans interact with computers in such a way similar to the human-human interactions[2,3]. ITS has have been applied to wide range of practical applications such as technical training for military recruits, high school mathematics etc.

The important key factor considered for the successive development of ITS is capable of recognizing students Affective state and reacts accordingly. Despite ITS are effective at supporting student's needs, until now only less researches have been made to promote student's motivation, engagement and interest. Without considering these key factors the ITS will have serious limitation in its efficiency to solve this problem virtual tutor should imitate like human tutor (i.e., be believable and acceptable through appearance, behavior and verbal parameters). The study benefits of ITSs have also been tracked to specific instructional design principles, such as minimizing cognitive load and using immediate feedback[4]. In short, research on ITS has focused on determining how to behave with students as well as how to communicate with students.

Although progress has been made toward the design on more ITSs[5-8] current models are still lacking in terms

**Author for correspondence*

emotive speech and cognitive behavior model. Human communications consist of more than just words. Henry, et al.[9] statically reported that words deliver only 7% of speakers emotional state. Whereas speech signal can deliver 38% of emotional state and other factors such as gesture and facial expressions can express 55% of emotional state. Therefore the virtual tutor must be capable of delivering emotive speech content also. However, research in emotive speech regards virtual agent is rare. In behavioral side, most of the designed ITSs follow user defined behavior model. The user defined model contains the collection of virtual behavior model regards students affective state and behavior, user defined behavior modeling works well in a constrained environment and its lacks in reality behavior and it can't able to perform when exceptional cases occurs such as, when new action/behavior found in student which is not encoded by user behavior model . Whereas cognitive behavior modeling works on binding various stimuli, past action which allows virtual tutors to find most appropriate behavior model.

The goal behind in designing of our virtual tutor is to friendly and assistive, in order to maintain friendliness our virtual tutor express only happy, surprise, disgust and sad behavior (i.e., facial expression, gesture and emotive speech) at various levels to express his happy and disappointment corresponding to student action. During adaptive questionnaire section the virtual tutor enables students to practice their learned skills by asking questions regarding content delivered during tutoring and if the student struggles to answer or wrongly answer, the virtual tutor provides some clues regards answer to maintain assistive.

To summarize, this paper examines the various synthesized verbal and nonverbal behavior of the virtual tutor corresponding to various affect states of students from Annamalai University.

Section 2 briefly describes related research work carried out. Section 3 explains the methods involved in designing virtual tutoring agent. Section 4 contains experimental result and discussion of our work. Section 5 is a conclusion section.

2. Related Research Work

Recently, number of studies in multimodal affects recognition such as combined face, body and speech information are available[14-16]. In out method, we used multi modal features for affective state recognition in order to predict the affective state of the student more exactly. Pantic, et al.[14] furnished multimodal approach for the recognition of various emotions that integrates the information from facial expressions, body gestures and speech. They showed a recognition improvement of more than ten percent compared to the most successful unimodal system and the influence of feature-level fusion against decision level fusion.

Human tutor have been expressing various affective state due to their ability to provide students with mentoring getting appropriate feedback, and the interactive manner in which they direct the student towards a solving a problem. This following action performed by the human tutor keeps the student to be engaged during tutoring. Wolcott, et al.[17] encoded the same information in which virtual tutor's nonverbal behavior such as eye contact, facial expressions and body gesture corresponding the affective states of students, which points out the degree of success in the instructional transaction.

Rickel, et al.[13] developed a half-body virtual tutor named Steve, to teach the students with complex appliances such as High Pressure Air Compressor, etc. The learner interact with the Steve through natural language, Steve could understand some small set of preprogramed questionnaires. In responding to the learners affective state Steve can able to show few facial expressions, simple gestures, and moved by floating rather than walking.

Mello and Graesser[22], implemented fully automated affective state, speech-based intelligent tutoring system. The affective tutor can able to detect student's boredom, confusion and frustration by recognizing learner's affective cues from facial features and body gestures. The tutor responds to the learner's query through synthesized speech with corresponding facial expressions. However, noticing the affective dynamics for the students while listening the tutor over time[44,45], postulated that particular affective state are likely to be occur than any other state[44]. Towards justification, exploration and conformation above theory affective system had postulated what are the affective states responsible for positive learning and what are the affective states responsible for poor learning[18,11,12]. Learners can able to experience positive learning outcomes when they are actively engaged in learning activity[24]. Therefore the amount of time user spends with the virtual agent cam be increased by manipulating the agent's behavior[25]. Research works[19-21] uncovers that the communicative interface, appearance and behavior of the virtual agent have more impact on learning activity, if the new interface lacks in original training condition (i.e., similar to human tutor) it leads to have negative affect[23] for the learners.

Improved positive affect have been proved to be have more several benefits on students learning process. Positive affect encourage the greater flexibility and openness to the students in exploring new ideas and possibilities[26]. Positive affect will also lead to improve the interest[27], which in turn leads to the greater long time engagement towards the virtual tutor[28]. Badaracco, et al.[29] had designed an intelligent tutoring system based on competency education in which the analyzing process evaluate, renovate and stores the information in students model, which is achieved by student during learning process. The quality of ITS depends on the amount and accuracy of the information available in the student module which acts as an test tool for evaluating ITS

3. Method

In this paper, we used multimodal features: facial expression, hand gesture and speech, for affect state recognition. After extraction of features from various modalities, features are fused and recognized using fuzzy classifier, depending on the recognized affective states, virtual tutor performs various behaviors. Cognitive behavior modelling is used for finding out the appropriate behavior with respect to recognized affective state.

3.1 Feature Extraction

3.1.1 Facial Feature Extraction

In our method we use Active Appearance Model (AAM) for facial feature extraction, by investigating literature[10,30-32], AAM seems to have more advantage in expression recognition task. AAMs consist of two processes. The first one is modelling part, which consist of parametric shape and appearance model represented as low dimensional feature points followed by fitting process, which fits the AAM with new face image. In our proposed fitting scheme temporal matching constraint is used, the temporal matching enforces an inter-frame local appearance constraint between frames to avoid mismatched points. To make AAM more stable for cluttered backgrounds color based face segmentation as a soft constraint. We calculated raw distance across various 86 facial feature points. For example, distance between mouth corner, upper and lower eye lid, eye brows, etc., were measured. The distance among those facial feature points in each frame are calculated and assigned as features. In our AAM, processing for one face scan takes average 39 ms, i.e., about 25 fps.

3.1.2 Hand Pose and Orientation Estimation

We employ the Camshift technique[33] for tracking hand and bounding rectangle for predicting the pose and orientation of the hand in succeeding frames as shown in Figure Orientation helps in differentiating various poses of hand. The Camshift algorithm returns height, width and orientation of the bounding rectangle of the hand. Using this statistics we can able to analyses the pose of the hand (i.e. The rectangle having width greater than height represents horizontal pose and rectangle having width lower than height represents vertical pose). After determining the pose of the hand it is easy to find the position of the fingers. In our method four types of finger position are estimated: right, left, up and down. Finally by analyzing

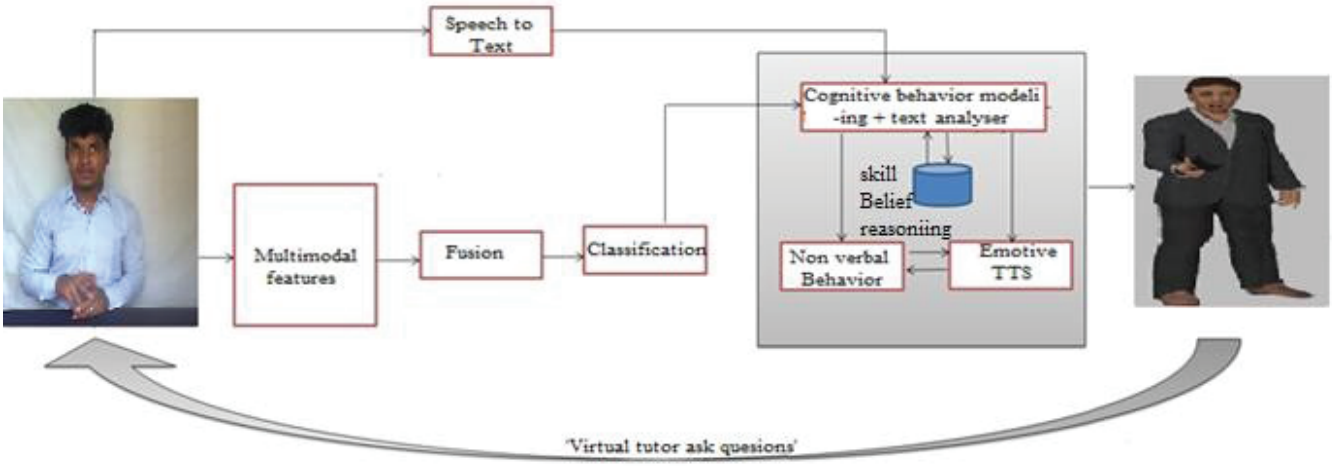

Figure 1. System frame work for Intelligent Tutoring System.

the orientation and pose of the hands and fingers, we can able to classify the hand movements (e.g., arms raised up, arms crossed, hand touching the face, etc.).

3.1.3 Speech Features

From the literature study[34,35], prosody continuous speech features contributes more information in recognizing the affective state of the human. Since our experiment is a real time and evidence is complex and mixed regards speech features, in order to avoid more complexity, we extracted only continuous speech (i.e., fundamental frequency, pitch, formant and energy) and speech qualitative features (i.e., voice level, phoneme).

3.2 Classification

In accord with the literature[36,37], Subtractive Fuzzy Rule Clustering System (SFCM) is used to determine the initial rules for fuzzy inference system. Then, an Adaptive Neuro-Fuzzy Inference System (ANFIS) was applied for adjusting inference parameters obtained from SFCM. ANFIS has proved its ability as an effective classifier in the movement classification area[38].

In our method after performing the initial rules for each subject, the extracted features from the multimodal feature set were added as input to SFCM inference system for training and testing. Then ANFIS is applied for adjusting obtained SFCM inference parameters because of elevated training rate and toughness of this pooled approach. The obtained inference system is efficient for judging the affective state of the student. The above training procedure is applied continuously with classifier updating. In order to manage excessive classified output regarding continuous segmentation, majority voting is applied as a post processing technique. In out method majority voting includes past and current decisions for a given point to from a new decision.

3.3 Emotive Speech Synthesis

In this paper two software platforms are used in this process FESTIVAL-MBROLA for emotive speech synthesis. We used diphone (two-phone combination) synthesizer with a rule based prosody modification approach for emotive speech synthesize. Our method consists of three main blocks: a Natural Language Processing (NLP) unit followed by Emotional Transfer (ET) unit and Digital Signal Processing (DSP) unit. The NLP unit converts orthographical text into proper phonetic form by analyzing text and predicting prosodic information's, the ET unit will perform various prosody modifications corresponding to emotions and DSP unit takes modified prosody and phonetic transcription as input and transfer them into a speech signal.

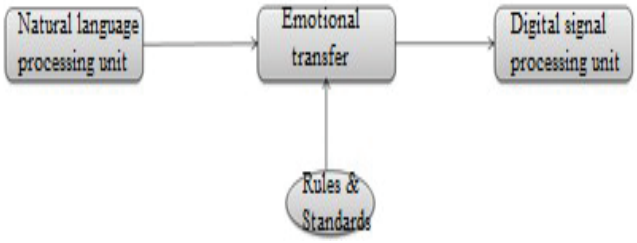

Figure 2. Framework of emotive speech synthesis using diphone synthesizer.

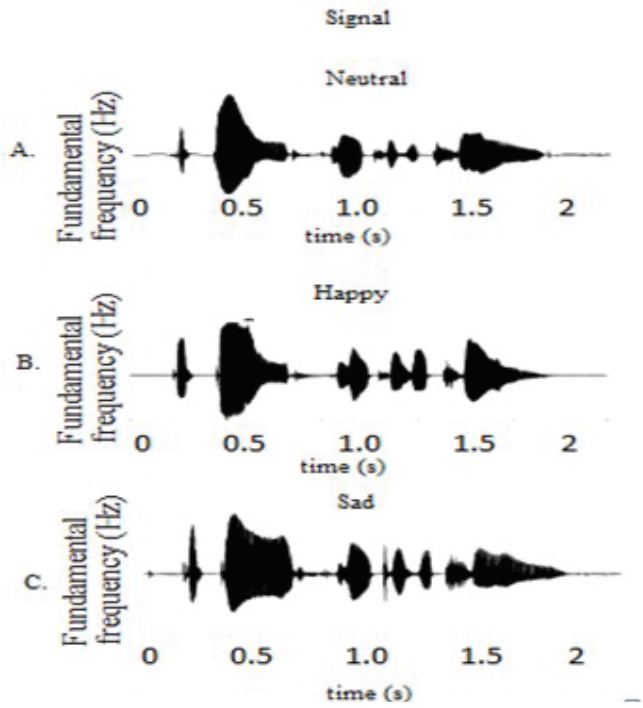

Figure 3. Examples of happy and sad emotive speech regards when the virtual tutor says "This is my environment".

The frame work of our emotive speech synthesizing block is shown in Figure 2. Using Festival[39], open source speech processing software will analyses the input text and predicts neutral prosody for the corresponding text. The outputs of Festival are written in PHO file format, which consists of list of phonemes and their associated prosodic parameters. In PHO format each phoneme and its associated prosodic information are represented in single line.

The ET unit has a set of pre-defined rules and standards in its database, which defines prosody transitions and voice quality corresponding to variable emotions (e.g., for getting sad voice, neutral prosody features are transformed to low values and jitter period are reduced). Output from the ET (i.e., new prosody features with new voice quality information's) are composed in a new PHO file. Finally the PHO file contains emotion-description information's. MBROLA[40], a free-non-commercial diphone synthesizer tool for emotive speech generation. MBROLA reads the phonemes with their prosodic information available in new PHO file and produces speech samples on 16 bits (linear) at the sampling frequency of diphone database. As shown in Figure 3, the prosody of the synthetic speech is quite different from the original speech which is neutral.

3.4 Speech to Text

In order to recognize the word spoken by the student, Microsoft speech sdk 5.1[46] is used to convert speech to text. Microsoft speech sdk 5.1 is open source win32 API available for speech to text conversion. The converted text is passed to the text analyzer, the text analyzer will checks the available text is a right answer or not with respect to the question asked by the virtual tutor, which is briefly discussed in Section 4.

3.5 Cognitive Expressive Behavior Modeling

In order to make the behavior of the virtual tutor to be more realism and believability in the environment, we use cognitive concepts to model virtual tutor based on reasoning. The virtual tutor represents its beliefs, intention, knowledge and desire in modular data model and performs explicit handling on those models to carryout behaviour planning (e.g., walking, folding hands, etc.). Based on its own beliefs and intention, from the students affective state, the virtual tutor can able to understand the needs and intention of the student. Our cognitive virtual tutor design is based on three principles:

- Beliefs are assigned to be in mutually uniform.
- Virtual tutor plans a behavioural action only while it beliefs the action is possible.
- Virtual tutor need not plan something that
- Won't happen anyway.

To accomplish this, the virtual tutor architecture should be combined with current stimuli, existing behavioral model stored in its memory representing a variety of information, mutual information, past experience, needs and knowledge of the world.

After finding the behaviour of the virtual tutor with respect to student affective state, a decision has to be made for how much extent should the virtual tutor perform its action? Hartmann, et al.[41] proposed a technique to control the behavior of the virtual agent based six parameters: Overall Activation, Spatial Extent, Temporal Extent, Fluidity, Power and repetivity. In our system as proposed by[42], we consider only four parameters: Spatial Extent, Temporal Extent, Fluidity and Power which is enough to control the overall behavior of the virtual tutor.

3.5.1 Spatial Extent

This controls the fullness of the gestures to be performed by the virtual tutor. This parameter values ranges in-between [-1, 1], where -1 represents small and narrow movements, while 1 represents large and wide movements. Spatial extent corresponding to zero resumes the virtual agent back to its initial state (i.e., without any gesture activity). The value of this parameter depends on strength of the affective state observed from the student behavior.

3.5.2 Temporal Extent/Speech and Visual Synchronization

Starting from synchronicity constraint on the end of the gesture stroke to coincide with the stressed affiliate in speech for speech and visual synchronization. This parameter determines the overall persistence of the gestures and responsible for speed of the entire gesture to be performed with respect to the speech time. Its value ranges from [-1, 1], where-1 represents gestures should be performed in slow manner; while 1 represents gestures should be performed in faster manner. Temporal extent corresponding to zero, the behavior the virtual tutor is generated without any gestural control.

3.5.3 Fluidity

This control maintains the smoothness and continuity between various gestures. This can be done through by varying the continuity parameter of Kochanek-Bartels splines[43]. This parameter value ranges from [-1, 1], where the higher value of this parameter reduce the pause and

guarantees the continuity, while lower value produces discontinuity between various gestures. Therefore in order to avoid discontinuity between various gestures, the value of the fluidity is maintained high. Fluidity is equal to zero when there is no gesture to perform.

3.5.4 Power

This determines the movement of the gesture to be appeared as stronger or weaker. Stronger movements are expected to have higher acceleration and deceleration magnitudes. This can be done through by varying the tension and bias parameter of Kochanek-Bartels splines. Tension parameter is used for reduce the stroke phase and bias parameter is used to generate undershoot at the end of stroke. Figure 4 shows various behavior of the virtual tutoring agent.

4. Results and Discussion Experiment I

In this experiment, we used simple facial expression (Boredom, Fear, Confused, Engaged, Happy and Neutral). For this study, 36 new participants were recruited. In terms of gender distribution 60% were male and 40% were female. The tutorial delivered by our virtual tutor contains basic image processing techniques. The length of the tutorial carries 15 min. Each student is allowed to take test alone in a specially designed room, in which the camera is hidden in bookcase. It is well know that students experience more freely when they feel that they haven't observed. During the tutoring, the participant's emotions are recorded.

Our experiment contains two sections: listening section and questionnaire section. In listening section virtual tutor presents the tutorial and students observe only the tutorial presented by the virtual agent and in this case there will be no speech signal available from students therefore null value is set for speech features (i.e., affective state of the students is recognized from facial and hand gesture modality). During questionnaire section the virtual tutor tries to asks question from the tutored subject and in this case student speech signal is measured (i.e., affective state of the students is recognized from facial, hand gesture and speech signal) in addition answer replied by the student regarding the question are also analyzed and if the answer was wrong the virtual provide some clues for the corresponding question, in order to build the student interaction more effective.

4.1 Listening Section

Table 1 shows the mean, and standard deviation regarding Boredom, confused, fear, engaged, happy and neutral of the participants while listening the 15 min virtual tutor. From the Table 1 it is clear that most of the students experience happy emotion while listening virtual tutor. The student expressed various emotions while listening the tutor but expression happy occurs to be more compared to other expressions.

Figure 4. Examples of various behavior expressed by the virtual tutor.

4.2 Behaviour and Speech Verification

Figure 4 represents some of the behavior of our virtual tutoring agent. In order to investigate the impact of various facial expressions, gesture and emotive speech synthesis performed by the virtual tutor, statistical analysis was performed using Friedman test. The Friedman test is a non- parametric test used to compare dependent samples that are repeated on same subjects.

Friedman test is conducted separately for facial expression, gesture and emotive speech. Each Friedman test is aimed to investigate whether there is a significant difference in the facial expression, gesture and emotive speech performed by the virtual tutor. In order to test facial expression, gesture, and emotive speech: virtual tutor with neutral facial expression (i.e., with no facial expression) as baseline against modulated facial expression, virtual tutor with neutral gesture (i.e., with no gesture) as baseline against modulated gesture and for emotive speech default prosodic values from the file PHO (i.e., prosodic values obtained from Festival) are compared with modified prosodic values PHO new file (i.e., modulated prosodic values from emotional transfer unit).

4.3 Overall Affect Transition

4.3.1 Case 1

When the student seems to be boredom and answering wrong, the virtual tutor expressed disgust facial, gesture and Speech expressions. The first test analyzed the difference in ratings of facial emotion between the state F0 and F2, the ratings of disgust is significantly greater than the ratings of neutral facial expression ($x^2 = 245.20$; $p = 0.0000001$). The second test analyzed the difference in ratings of gesture between the state G0 and G2, the rating of disgust is significantly greater than the rating of neutral gesture ($x^2 = 300.15$; $p = 0.0000002$). The third test analyzed the difference in ratings of speech between the state S0 and S2, the ratings of disgust is significantly greater that the ratings of neutral speech ($x^2 = 289.20$; $p = 0.00001$).

Inference: When the student was boredom this means he/she not listened the tutor fully, the tutor express stronger disgust in order to express his strong disappointment.

4.3.2 Case 2

When the student seems to be confused and answering Wrong, the virtual tutor expressed sad facial expression, gesture and speech. The first test analysed the difference in ratings of facial emotion between the state F0 and F2, the ratings of sad is significantly greater than the ratings of neutral facial expression ($x^2 = 228.44$; $p = 0.0000001$). The second test analyzed the difference in ratings of gesture between the state G0 and G2, the rating of sad is significantly greater than the rating of neutral gesture ($x^2 = 215.31$; $p = 0.0000002$). The third test analyzed the difference in ratings of speech between the state S0 and S2, the ratings of sad is significantly greater that the ratings of neutral speech ($x^2 = 200.16$; $p = 0.0000001$).

Inference: When the student was confused this means he/she listened the tutor partially, the tutor express medium sad in order to express his little much disappointment.

4.3.3 Case 3

When the student seems to be fearful and answering wrong, the virtual tutor expressed sad facial expression, gesture and speech. The first test analyzed the difference in ratings of facial emotion between the state F0 and F2, the ratings of sad is significantly greater than the ratings of neutral facial expression ($x^2 = 186.31$; $p = 0.0000001$). The second test analyzed the difference in ratings of gesture between the state G0 and G2, the rating of sad is significantly greater than the rating of neutral gesture ($x^2 = 192.21$; $p = 0.0000002$). The third test analyzed the difference in ratings of speech between the state S0 and S2, the ratings of sad is significantly greater that the ratings of neutral speech($x^2 = 118.24$; $p = 0.0000002$).

Inference: When the student was fear this means he/she listened the tutor partially, the tutor express weaker sad in order to express his disappointment.

4.3.4 Case 4

When the student seems to be engaged and answering correct, the virtual tutor expressed surprise facial expression, gesture and speech. The first test analysed the difference in ratings of facial emotion between the state F0 and F1, the ratings of surprise is significantly greater than the ratings of neutral facial expression ($x^2 = 218.14$; $p = 0.0000001$). The second test analyzed the difference in ratings of gesture between the state G0 and G1, the rating of surprise is significantly greater than the rating of neutral gesture ($x^2 = 238.37$; $p = 0.0000001$). The third test analyzed the difference in ratings of speech between the state S0 and S1, the ratings of surprise is significantly greater that the ratings of neutral speech ($x^2 = 205.34$; $p = 0.0000001$).

Inference: When the student was engaged, this means he/she listened the tutor fully, the tutor express stronger surprise in order to express his happiness.

4.3.5 Case 5

When the student seems to be happy answering correct, the virtual tutor expressed happy facial expression, gesture and Speech. The first test analyzed the difference in ratings of facial emotion between the state F0 and F1, the ratings of happy is significantly greater than the ratings of neutral facial expression ($x^2 = 201.31$; $p = 0.0000002$). The second test analyzed the difference in ratings of gesture between the state G0 and G1, the rating of happy is significantly greater than the rating of neutral gesture ($x^2 = 195.19$; $p = 0.0000001$). The third test analyzed the difference in ratings of speech between the state S0 and S1, the ratings of happy is significantly greater that the ratings of neutral speech ($x^2 = 212.51$ $p = 0.0000001$).

Inference: When the student was happy, this means he/she willing in interacting with tutor, therefore virtual tutor express medium happy in order to express his happiness.

4.3.6 Case 6

When student seems to be neutral and answering correct, the virtual tutor expressed happy facial expression, gesture and speech. The first test analysed the difference in ratings of facial emotion between the state F0 and F1, the ratings of happy is significantly greater than the ratings of neutral facial expression (x^2 =190.42; p = 0.0000001). The second test analysed the difference in ratings of gesture between the state G0 and G1, the rating of happy is significantly greater than the rating of neutral gesture (x^2 = 189.84; p = 0.0000001). The third test analyzed the difference in ratings of speech between the state S0 and S1, the ratings of happy is significantly greater that the ratings of neutral speech (x^2 = 166.10; p = 0.0000002).

Inference: When the student was neutral this means he/she simply interact with tutor, the tutor express medium happy in order to express his happiness.

Table 1. Confusion Matrix for affect observation

	BOR	CON	FE	ENG	HA	N
BOR	1336	-	-	-	-	-
CON	-	1498	-	-	-	-
FE	-	-	1430	-	-	-
ENG	-	-	-	1882	-	-
HA	-	-	-	-	1900	-
N	-	-	-	-	-	1789

BOR = Boredom CON = Confusion FE = Fear; ENG = Engaged; HA = Happy; N = Neutral

5. Experiment II

Later the participants were asked to report the experience and usage of our interactive cognitive virtual tutoring system. Of the maximum score of 10, from the Table 2 participants reported mean score of 7.58 and S.D 1.61 for performance satisfaction, mean score of 7.61 and S.D 1.87 for friendliness of virtual tutor, mean score of 7.72 and S.D 1.56 for content delivery during tutorial time is audible and understandable, mean score of 7.66 and S.D 1.76 for physical appearance, mean score of 8.14 and S.D 1.43 for assistance provided by our virtual tutoring system, mean score of 7.98 and S.D 1.49 for how much our virtual tutor learns student intention.

6. Conclusion

This paper presents a cognitive intelligent virtual tutoring system based on the student's affective states. In addition, based on the obtained results and the feedback questionnaire's, it is well clear that our proposed virtual tutor improves the functionality through cognitive behavior modelling and interactive environment. Experiment 1 examined the transition states of various facial expression, gesture and speech behavior of our virtual tutor. Experiment 2 examined the experience obtained by the students. This feedback can be utilized to upgrade the intelligent virtual tutoring system.

The students reported our virtual tutor experiences less boredom and more engaged with the virtual tutor. Hence, the students reason in liking virtual tutor is due to human like interaction provided by our virtual tutor. Another possible for liking virtual tutor will be educationally helpful. During some exceptional cases such as student smiles and struggles for answering the question asked by virtual tutor, due to cognitive modeling from the past experience virtual tutor performs weak happy behavior with clues regarding answer for the question. Since our virtual tutor performs one to one communication it should be extended to large group of interactions in various virtual environments such as classroom, conferences hall, exhibition or health fairs. Following studies addressing these limitations will be beneficial for forthcoming pursuits.

For future work, we also aim to improve the experiment with various emotional states other than seven basic emotions to meet the self-assessment and to extend the work in complex virtual environments, and higher degree of behavioural and visual realism.

Table 2. Overall mean and standard deviation for the performance of the system

	Mean	Standard deviation
Q(1)	7.58	1.61
Q(2)	7.61	1.87
Q(3)	7.72	1.56
Q(4)	7.66	1.76
Q(5)	8.14	1.43
Q(6)	7.98	1.49

7. References

1. Brusilovsky P, Peylo C. Adaptive and intelligent web-based educational systems. International Journal of Artificial Intelligence in Education. 2003; 13:159–72.
2. Nass C, Moon Y. Machines and Mindlessness: Social Responses to Computers. J Social Issues. 2000; 56(1):81–103.
3 Reeves B, Nass. The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places. Cambridge Univ Press; 1996.
4. Aleven V, et al. Automated, unobtrusive, action-by-action assessment of self-regulation during learning with an intelligent tutoring system. Educational Psychologist. 2010; 224–33.5.
5. Graesser AC, Chipman P, Haynes BC, Olney A. AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. IEEE Transactions on Education. 2005; 48(4):612–8.6.
6. Litman DJ, Silliman S. ITSPOKE: An intelligent tutoring spoken dialogue system. Demonstration Papers at HLT-NAACL: Association for Computational Linguistics; 2004. p. 5-8.
7. Hwang G-J. A conceptual map model for developing intelligent tutoring systems. International Journal of Computers and Education. 2003 Apr; 40(3):217–35.8.
8. Graesser AC, Conley MW, Olney A. Andrew O. Intelligent tutoring systems. APA Handbook of Educational Psychology. Washington, DC: American Psychological Association; 2012.
9. Henry SG, et al. Association between nonverbal communication during clinical interactions and outcomes: A systematic review and meta-analysis. Patient Education and Counseling. 2012; 86(3):297–315.
10. Liu X. Video-based face model fitting using adaptive active appearance model. Image and Vision Computing. 2010 Jul; 28(7):1162–72.
11. Rodrigo MMT, Baker RSJD, Agapito J, Nabos J Repalam MC, Reyes SS, et al. The effects of an interactive software agent on student affective dynamics while using; an intelligent tutoring system. IEEE Transactions on Affective Computing. 2012; 3(2):224–36.
12. D'Mello SK, Taylor R, Graesser AC. Monitoring affective trajectories during complex learning. Proceedings of 29th Annual Cognitive Science Society; 2007. p. 203–8.
13. Rickel J, Johnoson W. Animated agents for procedural training in virtual reality: Perception, cognition and motor control. Appl Artif Intell. 1999; 13:343–82.
14. Pantic M, Rothkrantz JM. Towards an affect-sensitive multimodal human-computer interaction. L.J.M. Proceedings of the IEEE; 2003 Sep. p. 1370–90.
15. Camurri A, Volpe G, De Poli G, Leman M. Communicating expressiveness and affect in multimodal interactive systems. IEEE MultiMedia. 2005; 12(1):43–53. doi:10.1109/MMUL.2005.2.
16. Kollias S, Karpouzis K. Multimedia and Expo. ICME. IEEE International Conference on Digital Object Identifier; 2005. p. 779-83. Doi:10.1109/ICME.2005.1521539.
17. Wolcott RS, Bishop BE. Cooperative decision- making and multi-target attack using multiple sentry guns. IEEE 41st Southeastern Symposium on System Theory; 2009. p. 261–5
18. McQuiggan S, Robison J, Lester J. Affective transitions in narrative-centered learning environments. Proceedings of 9th International Conference Intelligent Tutoring Systems; 2008.
19. Nam CS, Johnson S, Li Y, Seong Y. Evaluation of human-agent user interfaces in multi-agent systems. International J Industrial Ergonomics. 2009; 39:192–201.
20. Cao A, Chintamani KK, Pandya AK, Ellis RD. NASA TLX: Software for assessing subjective mental workload. Behavior Research Methods. 2009; 41(1):113–7.
21. Blakemore SJ, Frith C. Self-awareness and action. Current Opinion in Neurobiology. 2003; 13(2):219–24.
22. D'Mello SK, Graesser AC. Auto-Tutor and Affective AutoTutor: Learning by Talking with Cognitively and Emotionally Intelligent Computers that Talk Back. ACM Transactions on Interactive Intelligent Systems. 2012 Dec; 2(4):23.
23. Koedinger KR, Vincent A. Exploring the assistance dilemma in experiments with cognitive tutors. Educational Psychology Review. 2007; 19(3):239–64.
24. Fredricks JA, Blumenfeld PC, Paris AH. School engagement: Potential of the concept, state of the evidence. Rev Educational Research. 2004; 74(1):59–109.
25. Cap M, Heuvelink A, Bosch K, Doesburg W. Using agent technology to build a real-world training application. In Agents for Games and Simulations II. Berlin Heidelberg: Springer; 2011. p. 132–47.
26. Conati C, Manske M. Evaluating adaptive feedback in an educational computer game. Proceedings of 9th International Conference on Intelligent Virtual Agents; 2009. p. 146–58.
27. Hidi S. Anderson V. Situational interest and its impact on reading and expository writing. In: Renninger KA, Hidi S, Krapp A, editors. The Role of Interest in Learning and Development. 1992; p. 215–38.
28. Krapp A. Structural and dynamic aspects of interest development: Theoretical considerations from an ontogenetic perspective. Learning and Instruction. 2002; 12(4):383–409.
29. Badaracco M, Martinez L. Intelligent tutoring system architecture for competency-based learning. In: Konig A, Dengel A Hinkelmann K, Kise K, Howlett R, Jain L, editors Knowledge-based and Intelligent Information And Engineering Systems. Berlin/ Heidelberg: Springer; 2011; 6882:124–33

30. van der Maaten L, Hendriks E. Action unit classification using active appearance models and conditional random fields. Cognitive Processing. 2012; 13(2):507–18.
31. Rodomagoulakis I, et al. Experiments on global and local active appearance models for analysis of sign language facial expressions. 9th International Gesture Workshop on Gestures in Embodied Communication and Human-Computer Interaction; 2011. p. 96–9.
32. Gross R, Matthews I, Cohn J, Kanade T, Baker S. Multi-pie. Image and Vision Computing. 2010; 5:807–13.
33. Bradski GR. Computer vision face tracking for use in a perceptual user interface. Intel Technology Journal Second Quarter. 1998; p. 214–9.
34. Ginevra C, et al. Affect recognition for interactive companions: Challenges and design in real world scenarios. Journal on Multimodal User Interfaces. 2010; 3:89–98.
35. Donald G, et al. Toward a minimal representation of affective gestures. IEEE Transactions on Affective Computing. 2011; 2(2):106–18.
36 Rezazadeh IM, Wang X, Wang R, Firoozabadi SMP. Toward affective handsfree human-machine interface approach in virtual environment-based equipment operation training. Proceedings of 9th International Conference Construction Applications of Virtual Reality; 2009.
37 Priyona A, Ridwan M, Alias A, Atiq R, Rahmat R, Hassan A, et al. Generation of fuzzy rules with subtractive clustering. Universiti Teknologi Malaysia, Journal Technology. 2003 Dec; 43:143–53.
38. Khezri M, Jahed M. Real-time intelligent pattern recognition algorithm for surface EMG signals. Biomedical Eng. 2003; 3(6):45.
39. The Festival Project [Internet]. Available from: http://www.cstr.ed.ac.uk/projects/festival/
40. The MBROLA Project [Internet]. Available from: http://mambo.ucsc.edu/psl/mbrola/
41. Hartmann B, Mancini M, Buisine S, Pelachaud C. Design and evaluation of expressive gesture synthesis for embodied conversational agents. Proceedings of 3rd International Joint Conference AAMAS; Utretch. 2005. p. 1095–6.
42. Castellano G, Mancini M Peters C, McOwan PW. Expressive copying behavior for social agents: A perceptual analysis. IEEE Transactions on Systems and Humans. 2012; 42(3):776–783.
43. Kochanek DHU, Bartels RH. Interpolating splines with local tension, continuity and bias control. In: Christiansen H, editor. Computer Graphics (SIGGRAPH'84 Proceedings); 1984. p. 33–41.
44. Kort B, Reilly R, Picard RW. An affective model of interplay between emotions and learning: reengineering educational pedagogy-building a learning companion. Proceedings of International Conference Advanced Learning Technologies; 2001.
45. Rodrigo MMT, Rebolledo-Mendez G, Baker RSJd, Boulay B, Sugay JO, Lim SAL, et al. The effects of motivational modeling on affect in an intelligent tutoring system. Proceedings of International Conference Computers in Education; 2008. p. 49–56.
46. Microsoft Speech to text software [Internet]. Available from: http:/www.microsoft.com/enin/download/details.aspx?id=10121