Framework for an Empathic Filipino Embodied Conversational Agent for an Intelligent Tutoring System

Jocelynn Cu
jiji.cu@delasalle.ph

Jannah Del Barrio
jannahdb@gmail.com

Kedren Villena
kedren.villena@gmail.com

Geraldine Ablaza
gemi.ablaza@delasalle.ph

Charo Musni
chai.musni@gmail.com

Merlin Suarez
merlin.suarez@delasalle.ph

ABSTRACT

Studies have confirmed that culture-based embodied conversational agent enhance the effectiveness of intelligent tutoring systems. It also helps sustain the engagement between the student and the virtual tutor. In this paper, we propose a framework where the virtual tutor shares the same cultural background as the learner. The framework is discussed in the context where the virtual agent embodies the characteristic of a human Filipino tutor while teaching English language to elementary school students. We describe how the cognitive, affective and culture aspects of the system are integrated to provide supportive feedback to the student.

Keywords

Embodied conversational agent, intelligent tutoring system, affective computing, culture-based virtual agents

1.0 INTRODUCTION

Embodied conversational agents (ECAs) are dialogue systems that speak with a graphical representation of a human body [13]. They are computer-generated images with life-like facial features, body movements, and auditory qualities that range from human to cartoon [30].

Early ECAs rely on natural language dialogue systems to get a specific task done. One example is *Karin*, a virtual receptionist who manages the information desk of the Virtual Music Center. Its task is to answer questions regarding performances and assist customers in ticket reservations through a dialogue system developed using the Wizard of Oz [23]. Another example is *Autotutor*, a human tutor simulator augmented with natural language capability. It simulates human discourse pattern and employs various tutoring techniques to help students gain deep conceptual knowledge [10]. Yet another example is *Oscar*, a conversational intelligent tutoring system (CITS). It mimics the human tutor by using a conversational agent that adapts to the learning style of the student [18].

Although these ECAs were able to carry out their specific tasks, they still lack the ability to truly "connect" with their human counterparts. ECAs also need to address the non-verbal communication cues and respond appropriately. *Karin* can manipulate its eyebrows, nose, lips and torso to form various facial expressions when talking to its clients. This helped *Karin* complete its task, e.g. ticket reservation, in less amount of time [13]. *Autotutor* can perceive the student's emotion through analysis of facial expression, body movement, and dialogue patterns [10]. Other systems, like the *Sensitive Artificial Listener* (SAL), abandoned the complexities of dialogue systems and instead focused on the emotional and non-verbal response of the person it is talking to [29]. Each SAL character has a personality, whose goal is to draw the person toward its dominant emotion. Based on the person's non-verbal expressions, the SAL character reacts accordingly.

In line with this goal, the study made by [19] discovered that users prefer to interact with virtual agents who share their cultural background. This was demonstrated by CUBE-G, a virtual agent that can infer a user's cultural background by sensing his/her non-verbal behaviour. Having identified the culture, the agent will then dynamically adapt its responses to the user.

Locally, there were studies focusing on the student's affective states while learning (e.g., [3, 28]) and on intelligent tutoring system (ITS) with an embodied agent (e.g., [15, 16]), but none are culture-based ECAs. The ethnicity of an ECA lends credibility to the virtual agent and helps sustain the engagement between the user and agent [4]. This motivates our aim to explore how to instil culture into an ECA for an intelligent virtual tutor.

This paper is organized as follows. Section 2 reviews existing studies on culture-based ECAs, agent-based ITS, and how affect is related to learning. Section 3 introduces the framework of a Filipino ECS for intelligent tutors. Section 4 presents preliminary results on building a student affect model. The last section concludes the paper.

2.0 REVIEW OF RELATED LITERATURE

Computer-generated virtual agents are classified as synthetic characters, animated agents, emotional agents, or embodied conversational agents. *Synthetic characters* are three-dimensional virtual agents designed for video and computer games. *Animated agents* are virtual agents designed to play the role of a tutor or a learning companion [17]. *Emotional agents* are virtual agents with predefined personality types and allow the user to change its mood or emotion [31]. *Embodied conversational agents* are virtual agents designed to interact with users by emulating human conversation [21].

[5] designed ECAs to be able to interact naturally and emotionally. With the recent advances in 3D graphics, the simulation of virtual agents to behave like humans is now possible. Furthermore, rather than verbal communication, non-

verbal communication like facial expression and body movement, is given more emphasis since it provides more information. Greta is an embodied conversational agent who has a personality and a social role. It has the capability to express emotions consistently within the context of the conversation. It also has the capability to adapt her behaviour to some user features relevant to the selected application domain [26]. Greta is composed of the Agent's MIND which includes the Emotional Mind and the Dialog Manager, a generator of the agent's body including a signal generator, signal wrapper and the chosen body that renders the specified behaviour.

Challenges in this field involve building real-time reactive agents that can get a minimal sense of what the user is talking about and to get a sense of the emotional state of the speaker so that the agent can respond accordingly and steer the conversation back to the agent's own mood. Such is the goal of the SAL character, that is, to keep the user engaged in talking and to draw the speaker into the same emotional quadrant as the character [13].

2.1 Culture-based Embodied Conversational Agents

The studies of [20] aims to glean appropriate theories and pragmatic findings on human culture and explore ways in which these can be embedded on the decision making process of advanced ECAs. [14] defined culture as a way of life accepted by a group of people, generally, without really thinking about them. These are passed along by communication and imitation from one generation to the next. Culture has a high impact on human through patterns and decision-making process. Modelling culture-based embodied conversational agents can improve the simulation of human behaviour.

An example is the Beergarden application of [9]. The application is a multi-agent system that simulates culture-specific dialogues and interaction. It shows culture-specific agents inhabiting the system, where the character's appearance and non-verbal behaviour were adapted to their cultural and ethnic background. This application was built to allow users to learn about culture-specific non-verbal behaviours.

2.2 Agent-based Intelligent Tutoring Systems

Intelligent tutoring systems are now employing ECAs as an added value to the learning capability of users. Aside from the conversational skills of ECAs, the non-verbal behaviour and the appearance of virtual tutors are becoming more and more realistic. Thus it serves as an ideal tool for engaging users in a social interaction, for example, tutoring. [7] defined engagement as the holistic sensation that people feel when they act with total involvement.

[12] explored the possibility of enhancing the quality of experience for the student using an online tutoring system by integrating an ECA as a virtual tutor. The main feature of their system is the affective behaviour of the tutor. For example, the tutor shows empathy with the student when the student is struggling, or happiness when the student is doing well. Victor, the affective tutor, reacts to a student's actions and acts affectively, encouraging the student during the learning process. Oscar, a virtual tutor developed by [18], estimates and adapts to a student's learning style. It mimics a human tutor and offer intelligent solution analysis and conversational problem solving support in the domain of database Structured Query Language (SQL).

2.3 Culture-based Pedagogical Agents

Pedagogical agent design had placed greater emphasis on the importance of agent as an actor rather than a tool [25], thus focusing on the agent's social relationship with the student. [22] confirmed that students exposed to an environment with a pedagogical agent demonstrated deeper learning and higher motivation than students in an environment without an agent. In their study, students in a voice-plus-agent environment performed better compared to those in a text-only or voice-only environment in the process of learning. According to [2], a great deal of psychological modelling occurs when learners see their everyday associates as similar to themselves. In addition, it was said that attribute similarities between a social model and a learners, such as gender, ethnicity and competency, often have predictive significance for the learner's efficacy, beliefs and achievements.

Survey Sam, developed by [4], whose pedagogical agent varies in ethnicity, designed to investigate the effects of the type of instruction, deictic gesture and facial expression on student perception of pedagogical agent persona and attitude towards content and learning.

The Wayang Outpost, developed by [1], is a web-based ITS in designed to prepare students for the mathematical section of the SAT. It uses a dynamic multimedia technique that combines sound and animation to inspire better learning. It incorporates the student's knowledge into the tutor and used it as profile information to help guide the tutor. It also maintains a data-centric student model by integrating interactions with past users. The Wayang Outpost has two affective animated agents namely Jake and Jane, who offer advice and encouragement depending on the student's affect.

Table 1 summarizes the virtual agents reviewed in this section.

Table 1. Summary Table for ECAs.

Agent	Culture-based	Affective Feedback	Modality of Tutor Response
Oscar [18]	×	✓	Textual
Victor [12]	×	✓	Facial expression, gesture, audio
SAL [29]	×	✓	Facial expression, audio
AutoTutor [10]	×	✓	Facial expression, gesture
Greta [26]	×	✓	Facial expression, gesture, audio
Survey Sam [4]	✓	×	Facial expression, gesture
Wayang Outpost [1]	×	✓	Facial expression, gesture

2.4 Learner's Academic Affective States and Intelligent Tutoring Systems

According to [24], a learner's optimism fuels his/her motivation towards accomplishing a goal. As the chances of success increases, so is the willingness of a learner to engage in the same behaviour again. The more emotional the student feels toward a learning material, the more likely they will remember it.

AutoTutor [6, 10] is an ITS that helps students learn by holding a conversation using a mixed initiative dialog. It has an animated conversational agent that uses synthesized speech, gestures and facial expressions to display emotions. It tracks the learner's emotions while an interaction with the ITS is on-going. The emotions, also known as affective academic states (i.e.,

boredom, confusion, flow, delight, frustration, and neutral), are then correlated with the learning outcome of the learner.

[27] studied how these academic affective states are associated with the use of intelligent tutoring systems such as the Aplusix – an algebra learning assistant and The Incredible Machine (TIM) – a simulation problem solving game. Results in this study revealed that positive affect is associated more with the Aplusix rather than the TIM.

[16], on the other hand, focused on student feedback and its effect in learning. The intervention mechanism was designed to be affective, that is, providing praise when the student does well and reassurance when the student encounters difficulty in solving a problem.

3.0 A FILIPINO ECA FOR AN ITS

It was established that culture-based ECAs can be very useful for ITS [2, 4]. However, a Filipino ECA is yet to be built for an ITS. Figure 1 shows the proposed framework of a Filipino embodied conversational agent for an intelligent tutoring system. It consists of the Domain Model, the Student Model, the Tutor Model, and the Filipino ECA.

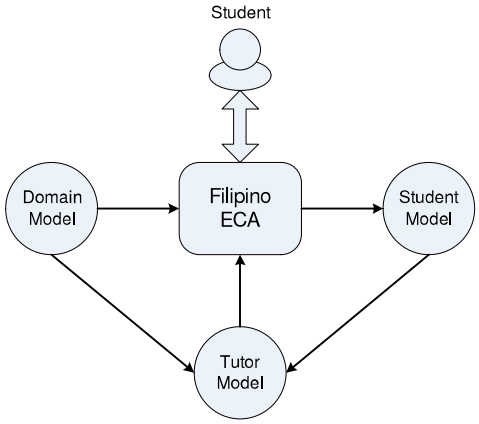

Fig 1. Framework of a Filipino ECA for ITS.

To demonstrate how the entities interact with each other, this framework will be discussed in the context of an ITS designed to help elementary students learn English as a second language (L2), focusing on sound production and not to make the learners sound like a native English speaker. The system has four main modules: the domain model, the Filipino ECA, the student model, and the tutor model.

The Domain Model, also known as the expert system, stores the content knowledge of the system. In L2 language learning for elementary school children, for example, it will contain the vowel sound inventory of Filipino school children. It will contain sound production per category, vocabulary learning, spelling, and comprehension activities arranged in a lesson plan.

The Filipino ECA is basically the user interface that interacts directly with the human user. It is the Filipino virtual agent who serves as the language tutor. It delivers the content based on the lesson plan provided by the Domain Model. It has several submodules, namely speech analyzer, speech synthesizer, and facial expression synthesizer. The speech analyzer accepts the student's vocal utterances as input and identifies the word/s or phrase/s spoken by the student. Speech features extracted from the utterances is compared to stored speech template. The result of the comparison and the characteristic of the utterances are then forwarded to the Student Model.

The Student Model is made up of the cognitive and affective models of the student. The cognitive model of student describes the student's knowledge on the subject matter, for example the correct pronunciation of the vowel sound /e/ in the word /bait/. The level of comprehension and mispronunciation, for example, is tracked and monitored statistically. The cognitive model also contains a list of common errors encountered by L2 language learners, for example, sounds that Filipino school children have difficulty learning and producing. The affective model determines the affective state of the student through his/her facial expression and voice while learning. The model will determine whether the student is *bored*, *confused*, *flow*, *delighted* or *frustrated*. The *neutral* state will serve as the baseline for which the other academic affective states will be compared with.

The Tutor Model has the task of comparing the student's knowledge with that of the expert systems' and recommend appropriate correction action. It chooses the correct level of response, the set of dialogue, and the corresponding facial expression for supportive feedback. It is a function of the affective state of the student and the type of error found in the student's utterance. If the spoken word is correct, for example in terms of pronunciation, the system will reply with encouraging words like "*Ang galing mo*!" (You are good!) or "*Tama*!" (Correct!). If the spoken word is incorrect, the system will support the student through several levels of feedback. For example, for the first level, the system will repeat the word with correct pronunciation, and encourage the student to say the word/s again. For the second level, the system will say the word slowly and will encourage the student to repeat after it on a per syllable basis. For the third level, the system will say the word again and explain where the emphasis should be. This allows the student to have several opportunities to get the pronunciation right.

The culture aspect of the virtual agent is integrated in its physical appearance, mannerisms, and feedback to establish affinity with the student. The ECA will have the physical characteristics of a Filipino, for example, brown skin tone, black hair and eyebrows. The speech synthesizer works synchronously with the facial expression synthesizer. The animation produces the correct sequence of facial expression, with all parts of the head and mouth moving in time with the words spoken by the virtual agent. An example of a Filipino ECA is shown in Figure 2.

Fig 2. Filipino ECA (a) shy, (b) happy, (c) neutral.

4.0 PRELIMINARY RESULTS: BUILDING THE STUDENT AFFECT MODEL

Initial work was carried out to build the affect model of the student. Data was collected from three students showing various spontaneous academic affective states while using POOLE III, shown in Figure 3. A total of 28,019 useful instances were collected and annotated using FEELTRACE. Table 2 shows the breakdown of instances per emotion. Each instance of emotion is represented by 68 facial points (represented by x-, y-coordinates on different regions of the face) extracted from the frontal face image using the Active Appearance Model and 18 prosodic-spectral points (i.e., pitch, formants, MFCCs) extracted from the audio using signal processing techniques.

Fig 3. Affect expressed by the learner, *left to right, top to bottom*, (a) boredom, (b) confusion, (c) delight, (d) engagement, (e) frustration, and (f) neutral.

Table 2. Number of instances per emotion.

Emotion	Instances
Boredom	3,026
Confusion	4,067
Frustration	1,621
Delight	2,643
Engagement	2,495
Neutral	14,167

All instances are used to create the affect model. Two machine learning algorithms are tested: the Support Vector Machine (SVM) and the k-Nearest Neighbour (kNN, where k=3). Shown in Table 3 is the result of modelling, validated using 10-fold cross validation. Between SVM and kNN, the latter can better classify academic affect: it has higher accuracy, higher kappa and lower mean absolute error.

Table 3. Results of affect modelling using SVM and kNN.

Algorithm	Accuracy	MAE	Kappa
SVM	68.76%	0.312	0.536
kNN (k=3)	90.59%	0.111	0.862

Legend: SVM – Support Vector Machine, kNN – k-Nearest Neighbour, MAE – Mean Absolute Error

5.0 CONCLUSION

The main objective of this paper is to propose a framework for a culture-based embodied conversational agent for an intelligent tutoring system. Several studies had confirmed that virtual agents built with the same cultural background as the student enforces a strong foundation of beliefs and intentions to the benefit of the student. Knowing the academic affective state of the student reinforces the agent's capability to provide empathic responses.

There are several experiments and impact studies that can be carried out using this framework. For example, one can investigate the impact and effectivity of using ITS in language learning, the capabilities of ECAs in learning, the best methodology or approach in L2 learning, possibilities of embedding stories as teaching and learning tools among others.

However, to build such system, one would need several types of useful databases, for example (1) a database of young students learning English, to accurately model their academic affective states; (2) a database of human tutors teaching young students, to accurately model their facial and vocal expression when providing supportive feedback; (3) a speech database of correctly pronounced English words or phrases, to accurately model the vocal characteristic of each words or phrases, among others. There is also a need to plan the lessons, taking special notice on the course content, instructional strategy, and assessment tools. In addition, the dialogue and conversational exchange between the learner and the virtual tutor has to be planned. Although the area of speech processing has matured enough to allow the synthesis of non-robotic conversational speech, it is still a challenge to develop speech synthesizers that incorporates culture-specific speech patterns, e.g., Filipino English.

This is an interdisciplinary project that calls for contribution from experts in the field of affective computing, intelligent tutoring systems, speech analysis and synthesis, dialogue management, English language learning, Filipino culture and society. Different aspects of this project can be further refined and developed as issues are encountered and solved.

6.0 ACKNOWLEDGEMENT

We would like to acknowledge the De La Salle University – University Research Coordination Office (DLSU-URCO) for the support funds; the College of Computer Studies (CCS), the Center for Empathic Human Computer Interactions (CEHCI), the Center for Language and Translation (CeLT), the Department of English and Applied Linguistics (DEAL), the Filipino Department, and the Psychology Department for providing us the theoretical background and raising research issues concerning this project. This research is supported in part by the Department of Science and Technology – Philippine Council for Industry, Energy and Emerging Technology Research and Development (DOST-PCIEERD).

7.0 REFERENCES

[1] Arroyo, I., Beal, C., Murray, T., Walles, R., and Woolf, B. (2003) Wayang outpost: intelligent tutoring for high stakes achievement tests. University of Massachusetts Amherst.

[2] Bandura, A. (2001) Social cognitive theory: an agentic perspective. Annual review of psychology, 52, 1-26.

[3] Bate, A. E. C. and Rodrigo, M. M. T. (2010) Automatic detection of student off-task behaviour while using an intelligent tutor for algebra. Philippine Computing Journal.

[4] Baylor, A. and Kim, Y. (2004) Pedagogical agent design: the impact of agent realism, gender, ethnicity, and

instructional role. LNCS Intelligent Tutoring Systems, 3220, 268-270.
[5] Bevacqua, E., Mancini, M., Pelachaud, C. (2004) Speaking with emotions. In Society for the Study of Artificial Intelligence and the Simulation of Behaviour, Leeds.
[6] Craig, S. D., Graesser, A. C., Sullins, J. and Gholson, B. (2004) Affect and learning: an exploratory look into the role of affect in learning in AutoTutor. Journal of Education Media, 29(3), 241-250.
[7] Csikszentmihalyi, M. (2000) Beyond boredom and anxiety: experiencing flow in work and play.
[8] D'Mello, S., Graesser, A., and Picard, R. (2007) Toward an affect-sensitive autotutor. IEEE Intelligent Systems, 22(4), 53-61.
[9] Endrass, B., Rehm, M., and Andre, E. (2009) Culture-specific communication management for virtual agents. In Proc 8th Int'l Conf Autonomous agents and multiagents systems, 1, 281-287.
[10] Graesser, A., Chipman, P., Haynes, B., and Olney, A. (2005) Autotutor: an intelligent tutoring system with mixed initiative dialogue. IEEE Trans Education, 48(4), 612-618.
[11] Graesser, A., McDaniel, B., Chipman, P., Witherspoon, A., D'Mello, S., and Gholson, B. (2006) Detection of emotions during learning with autotutor. In Proc 28th annual meetings of the Cognitive Science Society, 285-290.
[12] Grujic, Z., Kovacic, B., and Pandzic, I. (2009) Building Victor – a virtual affective tutor. In Int'l Conf on Telecommunications, 185-189.
[13] Heylen, D., Theune, M., Akker, R., and Nijolt, A. (2009) Social agents: the first generations. In Proc ACII 2009, 2, 114-120.
[14] Hofstede, G. and Hofstede, G. (2005) Culture and organizations: software in mind. McGraw-Hill, USA.
[15] Inventado, P. S., Legaspi, R., Bui, T. D. and Suarez, M. (2010) Predicting student appraisal of feedback in ITS using previous affective states and continuous affect labels from EEG. In Proc Int'l Conf on Computer and Education, 71-75.
[16] Inventado, P. S., Legaspi, R., Suarez, M. And Numao, M. (2011) Predicting student emotions resulting from appraisal of its feedback. In Research and Practice in Technology Enhanced Learning.
[17] Johnson, W. L., Rickel, J. W. and Lester, J. C. (2000) Animated pedagogical agents: face-to-face interaction in interactive learning environments. Int'l Journal of AI in Education, 11, 47-78.
[18] Latham, A., Crockett, K., McLean, D., Edmonds, B., and O'Shea, K. (2010) Oscar: an intelligent conversational agent tutor to estimate learning styles. In IEEE Int'l Conf Fuzzy Systems, 1-8.
[19] Lee, E. and Nass, C. (1998) Does the ethnicity of the computer agent matter? An experimental comparison of human-computer interaction and computer-mediated communication. In 1st Workshop of Embodied Conversational Characters.
[20] Mazadi, Z., Ghasem-Aghaee, N., Oren, T. I. (2008) Prelude to cultural software agents: cultural backgrounds in agent simulation. In Proc of 2008 Spring Simulation Multiconference, 135-142.
[21] Mcbreen, H., Anderson, J. and Jack, M. (2001) Evaluating 3D embodied conversational agents in contrasting VRML retail applications. In Workshop on Multimodal Communications and Context in Embodied Agents.
[22] Moreno, R., Mayer, R. E., Spires, H. And Lester, J. (2001) The case for social agency in computer-based teaching: do students learn more deeply when they interact with animated pedagogical agents? Cognition and Instruction, 19, 177-213.
[23] Nijolt, A. and Heylen, D. (2002) Multimodal communication in inhabited virtual environment. Int'l Journal of Speech Technology, 5(4), 343-354.
[24] Omrod, J. E. (2006) Educational psychology: developing learners. Upper saddle river, N. J., Pearson.
[25] Persson, P., Laaksolahti, J., and (2002) Socially intelligent agents: creating relationships with computers and robots. Kluwer Academic Publishers, Norwell, MA.
[26] Poggi, I., Pelachaud, C., Rosis, F. de, Carofiglio, V. and Carolis, B. de (2005) GRETA: a believable embodied conversational agent. In P. Paggio and B. Jongejan (Eds) 27-45.
[27] Rodrigo, M. M. T., Baker, R. S. J., D'Mello, S. (2006) Comparing learner's affect while using an intelligent tutoring system and a simulation problem solving game. Intelligent Tutoring Systems, 5091(5091), 40-49.
[28] Rodrigo, M. M. T. (2011) The dynamics of student affective transitions while using a pre-algebra game. Simulation and Gaming, 42(1), 85-99.
[29] Schroder, M. et. al. (2009) A demonstration of audiovisual sensitive artificial listeners. In Proc ACII 2009, 263-264.
[30] Ventura, J. Ventura, M., and Olabe, J. (2005) Embodied conversational agents: developing usable agents. In Proc SouthEastCon 2005, 663-669.
[31] Yoon, S. Y., Blumberg, B., Schneider, G. (2000) Motivation driven learning for interactive synthetic characters. ACM Press.

8.0 AUTHOR'S AFFILIATIONS

Center for Empathic Human-Computer Interactions, De La Salle University, 2401 Taft Avenue, Manila, Philippines, 1004.