한국멀티미디어학회지 제6권 제4호 2002년 12월

Generic Model for Emotional Tutoring Agents

Moon Won Choo* · Youngmee Choi*

1. Introduction

Information explosion and widespread availability of computing environments for public require the highly effective and user-friendly interface for accessing, extracting and manipulating valuable information. Software agent seems to promise the advent of new paradigm for intelligent and emotional communication channels for this seemingly uneasy task. Especially emotional software agent is actively focused on its implementational correlate, life-like character, usually called cyber-character or avatar, and is ready to replace traditional desktop metaphor with highly interactive life-like agent-oriented animated metaphor. As multimedia contents are increasingly reconfigured and upgraded for the Web environments, and general publics are accustomed to and becoming active supporters for digital communication, the highly sophisticated models for technology-based intelligent tutoring/learning are developed with unusual enthusiasm [1,3,9,14,15,19,27]. Tutoring agents are entities whose ultimate purpose is to communicate with the student in order to efficiently fulfil their respective tutoring function, as part of the pedagogical mission of the system. The fundamental reason for introducing agents as tutoring knowledge elements is their capabilities of communication and interaction. Hence, the tutoring agents with animated and emotional presentation are actively supported by many researchers[11,25]. These approaches add expressive power to a system's presentation skills and engage students without distracting them from the learning experience. In this paper, the attributes and structures of emotional agents for tutoring environment are introduced and the generic model for emotion generation for tutoring agents are suggested. .

*Division of Multimedia, Sungkyul University

2. The related researches

The number of intelligent tutoring systems featuring emotional agents has dramatically increased in recent years. Clippit, the paperclip agent that appear in various Microsoft Office applications may be one of the primitive tutoring agents. Some example include STEVE, a virtual-reality tutor developed for naval training[23]; COSMO, a program that assists people learning to troubleshoot networking problems[26]; Herman the Bug, which teaches children about botany by helping them to design plants that will survive in different environments[26]; AiA, a navigational aid for Web-based information [1]; and LANCA, a learning program featuring four networked agents that work in concert to provide optimal learning strategies depending on the individual learner's progress[13]. The Tutoring Research Group at the University of Memphis has developed HURA Advisor which is designed to facilitate navigation and information retrieval about ethical concerns in the use of human participants in research[16], and AutoTutor which is an intelligent tutoring systems designed to tutor college-level students in physics and computer literacy by engaging them in a multi-turn conversation in natural language[17]. Among these projects, Adele agent may be a good examplar tutoring agent, which could be utilized in practical settings[12].

Adele(Agent for Distance Learning Environment) developed by ADE (Advanced Distance Education) project, communicates with student using multimodal channels interactively, teaching medical materials on the Web environment.

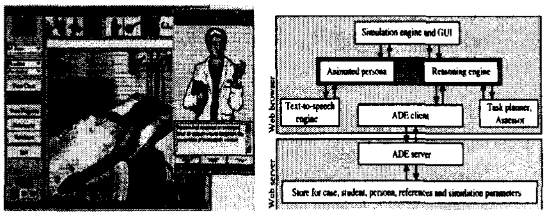

Fig. 2. ADE interface and structure

This system consists of four modules : pedagogical agent, simulation, client- server, and server store. Pedagogical agent includes two sub-modules: animated persona and reasoning engine. Central sever maintains the database for students' learning progress. Reasoning engine performs monitoring and decision-making depending on the initial states, student model, and case task plan. Animated persona generates animations using Java applets. Text-speech module generates the voice by composing the separate phonemes with cooperation with Animated persona. This model may be a good start point for those who want to develop more sophisticated tutoring agents.

Fig.1. Herman the Bug(left), Cosmo(right) interfaces.

3. Computational model for emotion

Life-like characters, visually anthropomorphic, emotional agents, play an important roles in instructional multimedia contents by offering user-friendly, user-oriented, and highly learning-motivating interactions[11,12]. Emotional agents are recently exploited based on the concept of believability[4,7,26]. The believability conceptually includes the entities like personality, emotion, self-motivation, social relationships, and illusion of life[26,27]. The structure of emotional agents comprise of the domain knowledge representation, instructional strategies, maintenance of consistent instructional dialog, and user-interface module which recognizes input stimuli and affects environments [18].

The computational model for emotion should provide the machinery necessary to account for affect and emotion. Most of researches related to this area are affected by cognitive psychology[2.9]. The representative work in this category is OCC cognitive theory suggested by Otorny and his colleagues[2]. In this theory, emotions arise from valenced reactions to situations consisting of events, actors and objects. The valence of one's emotional reaction depends upon the desirability of the situation for oneself, which in turn is defined by one's goals and preferences. For instance, an event that matches one's goal can generate positive emotions such as joy for the event, admiration toward the actor that caused the event, or pride if the actor is oneself. Corresponding negative emotions for an event that interferes with one's goals are distress, reproach and shame. The model proposed in this paper is mainly based on this theory.

AVS Emotional Model is a simple and reasonably powerful method to invest agent characters with a sufficient range of emotional depth to be convincing and enhance user experience. This model is based on dimensional theories of emotion and is primarily inspired by the International Affective Picture System and the International Affective Digitized Sounds system (IADS)[6]. The IAPS is a standard set of images of emotional scenes that have been used in numerous studies of human emotional response to media. The resultant scatter plot of IAPS and IADS responses along the axes of arousal and valence reveal roughly the form of a parabola, an interaction between arousal and valence. There are relatively few if any stimuli that elicit responses categorized as high arousal and neutral valence, likewise for high valence (positive or negative) and low arousal responses. In this model of emotion, three dimensions(arousal, valence and control) define the emotional space. Any given point in this three-dimensional space describes a unique emotional state. Few studies concern themselves with the control dimension as it accounts for little variance, however it is very useful in terms of distinguishing between two negative valence, high arousal emotional states such as grief and rage.

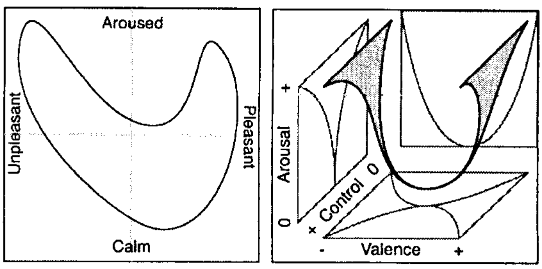

Fig.3. An arousal-valence emotional space (left) and a view of the arousal-valence -control emotion space (right).

The control dimension is necessary to achieve the range of expression needed for a fully expressive character. In the AVC model, there are three components- temperament, mood and emotion, each of which occupy a point in the emotional space. The temperament is a fixed coordinate which defines the at rest state of the system, defining a rudimentary personality for the agent. The emotion, mood and temperament coordinates define the agent's persona and current emotional state at any given time. Although temperament is static and constrained to the surface, mood and emotion coordinates can move freely throughout the space. The position of the emotion coordinate in AVC space gives rise to the expressions of the agent; it is the emotional state of the agent. Emotion can exist anywhere in the AVC space, but gravitates toward the position of the mood. Mood acts much like emotion, only it plods through the AVC space slowly, pulled by emotional events, but continually gravitating toward the temperament value. The dynamics of this system are such that the mood is influenced both by the current emotional state as well as the agent's temperament. The position of the emotion coordinate can be set by environmental factors or direct scripting and is influenced by the mood. Temperament is static and set when the agent is first run.

Many emotion theorists have argued that cognitive appraisal is central to emotion. One of the most salient aspects of the experience of emotions is that they vary a great deal in intensity both within and between people. This means that a theory of emotion must address the question of what determines intensity. The intensity of emotions is influenced by a number of variables, all of which are present in the construal of the situation that gives rise to the emotion in the first place. Thus, in order to address the question of intensity, we first need to consider the mechanism whereby emotion-inducing stimuli are appraised.

4. Proposed Model

In this section the proposed model is described. Woolf observes that typical Intelligent Tutoring Systems consists of [5]:

* **Domain Knowledge Module**, which contains the information that the tutor is teaching,
* **Expert Module**, is a model of how an expert human teacher would present the Domain knowledge,
* **Student Module,** which maintains information that is specific to each user and how far they are progressed,
* **Tutor/Pedagogical Module**, which is responsible for deciding how and when the domain knoledge is presented,
* **Diagnostic/Misconception Module**, which contains the rules used to identify misperceptions, gaps and misunderstandings on the part of the user,
* **Communication Module**, which is the user interface.

Based on this observation, we propose the model for evaluating the student's affect during the interaction with the educational materials.

Because of lifelike tutoring agents' abilities to combine sophisticated communicative functionalities with engaging and emotional personae, they can take advantage of humans' inherent propensities to anthropomorphize software and play a central role in students' problem-solving activities[21]. To handle the modeling task of clarifying the relations between student emotional states, their causes and effects, the model should consider a variety of static random variables, decision variables representing an agent's deliberate actions, the agent's preferences over the possible outcomes of the actions, and the evolution of variables over time. The proposed model consists of four major processes: the user and agent model, the emotional process, goal appraisal and the animation process. Several main modules are detailed in this section.

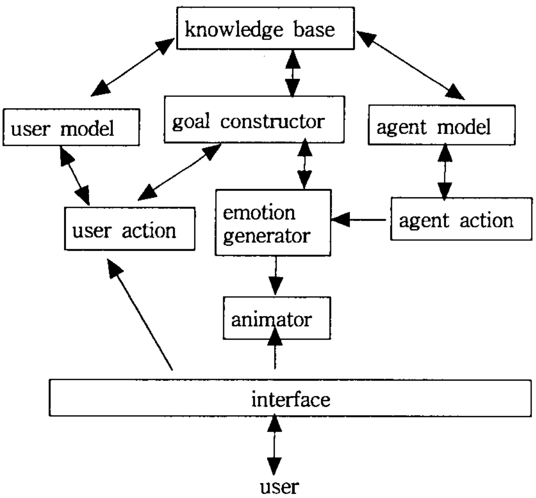

Fig. 4 The proposed emotion generation model (the directions of arrows represents the events/data flows).

Emotion generator

The perceived event taken from the goal constructor and agent action is first evaluated. The evaluation process consists of two sequential steps. Firstly, the user model determines which goals are affected by the event and the degree of impact that the event holds on these goals. Secondly, it proceeds to get a desirability level of the event according to the measure calculated by the first step and the importance of the goals involved. The event evaluation process depends on two major criteria; the importance of the goals affected by the event and the degree by which the event affects these goals. The desirability measure, once calculated, will be passed to next phase to further determine the emotional state of the agent. An emotion or a mixture of emotions will be triggered using the event desirability measure. The mixture will be filtered to produce an emotional state. In essence, the emotional state will be a list of emotions that apply at a specific time given a certain situation. The emotional state is passed to the behavior selection phase. Though this phase, a behavior is chosen according to the situation assessment and the emotional state. The emotional state will then be decayed and fed back to the system to the next iteration.

Goal Constructor

Most of the things that people do are motivated. People rarely engage in random actions devoid of goals and purposes. In some sense, therefore, people must have a structure of goals, interests, and beliefs that underlie their behavior. In tutoring environment, goals are described by students in terms of abstract operations they would like to see play out in the simulated world. After the goal has been completely specified, high-level goals are decomposed to create the goal-space to be usually shaped as tree structure whose leaves are action specification for the agent to perform. This module actively interacts with knowledge base and user model which supports and reflects user's learning process.

A crucial aspect of the representation of goal structure is that the representation is constantly changing as old goals are realized or abandoned and as new ones are introduced. Such changes are not limited just the addition and deletion of nodes or branches from the structure, rather, the entire configuration can change. In the efforts to build computer models of processes of this kind, the subjective transition probabilities with all inter-goal links could be associated(more about goals, refer to [2]). The goal model suggested above might be utilized to set up computationally tractable algorithms to help us to evaluate, modify, and improve the account of computational model. That is, we need a system of rules and representations about the elicitation of emotions.

Once a desirability degree is set for an event, some rules can be fired to determine the emotional state given an expectation and a desirability measure of the perceived event. The expectation value is taken from the user model. A variation of Ortony's model[2,22] is used to define the rules used. These rules are documented in Table 1. To apply the rules shown in the table, we need a few more components. These components are summerized as follows: (1) the desirability measure of the event, which is taken from the event evaluation process, (2) standards and event judgement, which are taken from the learning process, (3) expectation of events to occur, which are also taken from the learning process. The formula calculating the intensity of emotions are

Table 1. The rules to determine the emotion types.

emotion	rule
joy	occurrence of a desirable event
sad	occurrence of undesirable event
disappointment	occurrence of a disconformed desirable event
relief	occurrence of a disconformed undesirable event
hope	occurrence of a unconformed desirable event
fear	occurrence of a unconformed undesirable event
pride	action done by the agent and is approved by standards
shame	action done by the agent and is disapproved by standards
reproach	action done by the other and is not approved by the agents standards
admiration	action done by the other and is approved by the agents standards
anger	complex emotion → sad + reproach
gratitude	complex emotion → joy + admiration
gratification	complex emotion → joy + pride
remorse	complex emotion → sad + shame

illustrated in Table 2[10]. This table shows the method by which intensities are calculated for various emotions given an expectation value and an event desirability measure.

Table 2. The formula for calculating intensities of emotions

emotion	intensity
joy	(1.7*expectation$^{0.5}$)+(-0.7*desiability)
sad	(2*expectation^{2})-desiability
disappointment	hope*desirability
relief	fear*desirability
hope	(1.7*expectation$^{0.5}$)+(-0.7*desiability)
fear	(2*expectation^{2})-desiability
pride	value(event(x), standards)
shame	value(event(x), standards)
reproach	value(event(x), standards)
admiration	value(event(x), standards)

User model

The agent will need to know what event to expect, how much to expect it and how bad or good it is. As explained earlier, the identification of emotions and emotional intensity heavily rely on expectations. Furthermore, events are normally measured by their impact on a set of goals. It is often the case that a given event does not have any impact on any specific goal directly, but some sequence of events may eventually have an impact on some goals. Thus identifying the link between an event and the corresponding goals was noted to be a very complex task to accomplish. The agent can potentially learn this by using a reinforcement learning algorithms, namely the Q-learning algorithm[24], for example. It is often the case that an agent does not know the consequences of a given action until a complete sequence of actions is finished. The agent, therefore, faces the problem of temporal credit assignment, which is defined as determining which of the actions in its sequence are responsible for producing the eventual rewards. To illustrate the solution that reinforcement learning offers to this problem, we will look at reinforcement learning in more detail. The agent represents the hypothesis or the problem space using a table of Q-values in which each entry corresponds to a state-action pair. The table can be initially filled with default values. The agent will begin from a state s. He will take an action, a, which takes him to a new state s'. The agent may obtain a reward r for his action. If it receives a reward it updates the table above using the following formula:

$$Q(s,a) \leftarrow r + \gamma \max_{a'} Q(s',a'),$$

where r is the immediate reward, γ is a discount factor, s' is the new state, and a' is an action from the new state s'. Thus the Q value of the previous state-action pair depends on the Q value of the new-state action pair. Since the agent is interacting with the user, the agent will have to learn about the user's patterns of actions. The agent's mind may not handle more than a length of 7 $\pm$2 consecutive action comparisons[22].

Animator

There are many approaches available to

animate the tutoring agents. One of them is to use video clips pre-recorded, and another method is to model the facial expressions by synthesizing the visemes, morphing, or wire-framing in real-time.

First approach enjoys many benefits from traditional animation products and technologies. But it lacks flexibility in terms of dynamic presentations in real-time. Fig. 5 shows the illustrative behavioral space which can maintain all possible scenarios of emotional presentations, behavioral key frames, speech, and background music as well[14,15].

There are many researches for animating facial movement synchronized with audio, video, and text-based inputs[19]. MPEG4 specifies a set of face animation parameters (FAPs), each corresponding to a particular facial action deforming a face model in its neutral state. The FAP value for a particular FAP indicates the magnitude of the corresponding action, e.g., a big versus a small smile. A particular facial action sequence is

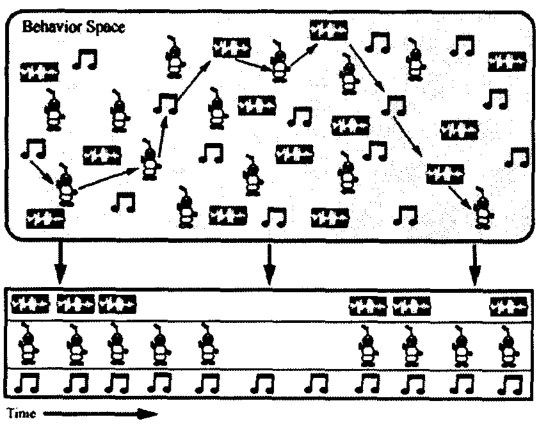

Fig. 5. Behavior space

generated by deforming the face model in its neutral state according to the specified FAP values for the corresponding time instant. Then the model is rendered onto the screen. The head in its neutral state is defined as follows (Fig. 6): Gaze is in direction of Z axis; all face muscles are relaxed; eyelids are tangent to the iris; the pupil is one third of iris; lips are in contact; the line of the lips is horizontal and at the same height of lip corners; the mouth is closed and the upper teeth touch the lower ones; the tongue is flat, horizontal with the tip of tongue touching the boundary between upper and lower teeth. For the renderer to interpret the FAP values using its face model, the renderer has to have predefined model specific animation rules to produce the facial action corresponding to each FAP. Since the FAPs are required to animate faces of different sizes and proportions, the FAP values are defined in face animation parameter units (FAPU). FAPU are defined as fractions of distances between key facial features. These features like eye separation, eye-nose separation, mouth nose separation, and mouth width, are defined for the face in its neutral state. They allow interpretation of the FAPs on any facial model in a consistent way, producing reasonable results in terms of expression and speech pronunciation. MPEG-4 defines the animation rule for each FAP by specifying feature points and their direction of movement. The renderer can either use its own animation rules for his proprietary model or download the face model and the FaceDefTables that define the animation rules for the model.

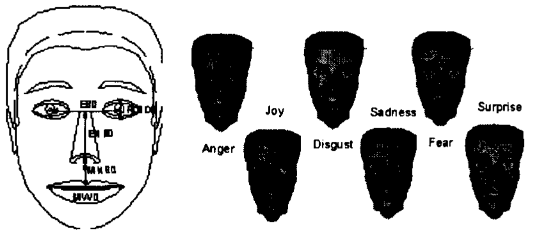

Fig. 6. FAPUs and primary facial expressions

Another example is using AVC model mentioned above. The expression of the agent character across its feature set is by and large a linear mapping of one or two emotional dimensions to a body feature. For instance, the amount the character's eyebrows raise is directly proportion to its arousal and changes its shape based on the valence of its current emotional state. Table 3 below presents the mappings of the agent's expressive features from the underlying AVC model. A wide range of meaningful emotional expressions can be derived from the combination of the AVC model and the feature mapping.

Table 3. Mapping emotional dimensions to agent features.

Feature	*Aspect*	*Corresponding Dimension(s)*
Eyebrow	Vertical Position Eyebrow Shape	Arousal and Valence Valence
Eyelid	Vertical Position	Arousal
Mouth	Position Mouth Shape	Control Valence and Control
Pods	Vertical Position Horizontal Spacing Wiggliness Color	Control Arousal Arousal Valance and Control

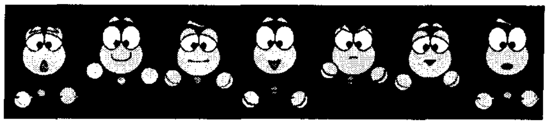

Fig.7. A sample of AVC derived expressions from an emotional agent.

5. Issues

As emotional tutoring agents become more prominent, it becomes increasingly important to understand the factors and efficacy that affect the quality of interactions between agents and students. Emotional agent can potentially facilitate the interaction by serving as a navigational guide, helping the user decide what to do next. They can enhance the user's enjoyment of the tutoring programs and some type of learning. Nevertheless, agents can also annoy or even discourage the user, thereby obstructing human-computer interaction. Futhermore, agents may not facilitate all types of learning. It is therefore important to understand the factors that affect user's reactions to emotional agents. Some issues related to these factors are summarized.

The agents present great variety in terms of characteristics that may affect their interactions with users. Some agents resemble humans, whereas others are based on non-human organisms or even inanimate objects. Agents may vary in their gender, age, race, and other demographic variables. They can be programmed to simulate different personality traits, such as humor, supportiveness, or irritability. One question of interest is whether users are sensitive to differences among agents. Do users perceive agents as differing on basic personality dimensions? Do such differences affect learning of materials? System developers should answer these questions fully before plunging into designing and coding[20].

Modeling student affect during the interacation with an educational materials is a task frequently permeated with uncertainty. One cause for this uncertainty is that often the same situation can induce a variety of different emotional states in different students, depending upon student perperties that are not easily observable. A second cause of uncertainty is that the bodily expressions that are symptoms of emotional arousal can be difficult to assess precisely and seldom support a one-to-one mapping with emotional states[8].

6. Conclusion

We have presented very abstract model for agent's emotion generation for tutoring environment during the interaction with an education materials. This model was built by integrating the OCC cognitive theory of emotion. It assesses affect by predicting how student goals, profile such as personality and domain knowledge. The model contains several modules to appraise the emotional states of students. However, each module can be implemented by adopting any kind of approaches available depending on the nature of applications.

Several are the directions for future work. First, the model should be refined to be more practical to be referenced. Second, the issues mentioned above should be answered properly. Third, the proper size of application should be built based upon this model.

References

[1] Andre, E. et. al., "Employing AI Methods to control the behavior of animated interface agents," Applied Artificial Intelligence, 1999.

[2] Andrew Ortony, G.L. Clore, and Allan Collins, "The Cognitive Structure of Emotions," Cambridge University Press, 1988.

[3] Badler, Phillips, & Webber, "Simulating Humans," Oxford University Press, 1993.

[4] Barbara Hayes-Roth and Robert van Gent, "Improvisational Puppets, Actors, and Avatars,"

[5] Barry Kort and Rob Reilly, "An Affective Module for an Intelligent Tutoring System," ITS 2002, LNCS 2363, 2002.

[6] Bradley, M.M. and Lang, P.J., "International affective digitized sounds (IADS)," Technical manual and affective ratings, University of Florida, Center for Research in Psychophysiology, 1991.

[7] C. Elliot & J. B., "Autonomous Agents as Synthetic Characters," AI Magazine, 1998, summer.

[8] Cristina Conati and Xiaoming Zhou, "Modeling Students' Emotions from Cognitive Appraisal in Educational Games," Intelligent Tutoring Systems 2002, LNCS 2363, pp. 944-954, 2002.

[9] Daniel Rousseau & Barbara Hayes-Roth, "A Social-Psychological Model for Synthetic Actors," KSA lab. report no. KSL 97-07, Dept. of Computer Science, Stanford University.

[10] Donald D. Price, et. al, "A Quantitative Experiential Analysis of Human Emotions," Motivation and Emotion, vol. 9, no. 1, 1985.

[11] Elliot, C., "Affective Reasoner Personality Models for Automated Tutoring Systems.

AI-ED97: English world Conference on AI in Education- Workshop V: Pedagogical Agents, 8., 1997

[12] Erin Shaw, et al., "Pedagogical Agents on the Web," www.isi.edu /isd/ ADE/papers /agents99/ agents99.htm

[13] Frasson, C et. al, "LANCA: A distance learning architecture based on networked cognitive agents," Intelligent Tutoring Systems, pp. 593-603, 1999.

[14] Gael Sanner, et al, "VHD: a system for directing real-time virtual actors," miralab www.unige.ch/

[15] Gene Ball et al, "Lifelike Computer Characters: the Persona project at MS Research," www. research.microsoft.com/ research/ui/ persona/ related.html

[16] Graesser, A.C., et. al, "Teaching with the help of talking heads," Proceedings of the IEEE Internatinal Conference on ALT, Los Alamitos, CA, IEEE Computer Society, pp. 460-461, 2001.

[17] Graesser, A.C, et. al, "Teaching tactics and dialog in AutoTutor," International Journal of AI in Education

[18] Henry Lieberman, "Autonomous Interface Agents," lieber.www.media. mit.edu/people/ lieber/Lieberary /Letizia/ AIA/AIA.html

[19] Hiroshi Dohi and Mitsuru Ishizuka, "Visual Software Agent: An Internet- Based Interface Agent with Rocking Realistic Face and Speech Dialog Function," Working Notes of AAAI-96 Workshop on Internet-Based Information Systems, pp.35-40, Aug. 1996.

[20] Kristen N.M., et. al, "Perceived Characteristics and Pedagogical Efficacy of Animated Conversational Agents," Intelligent Tutoring Systems 2002, LNCS 2363, 2002.

[21] Lester, J, et. al, "The Persona effect: Affective impact of animated pedagogical agents," Proceedings of CHI, 1997.

[22] Magy Sief El-Nasr, "Modeling Emotion Dynamics in Intelligent Agents," thesis for master of science, Texas A&M Univ., May 1998.

[23] Rickel, J., Johnson, W.L, "Animated agents for procedural training in virtual reality: Perception, cognition, and motor control," Applied AI 13, pp. 343-382, 1999.

[24] Tom M. Mitchell, "Machine Learning," New York, MacGraw-Hill Co., 1996.

[25] Tomoko and Pattie Maes, "Agents with Faces: Effect of Personification," MIT Media Lab.

[26] W. Lewis Johnson & Jeff W. Rickelll, "Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments," www.csc. ncsu.edu/eos/users/www/imedia/apa-ijaied-2000.htm

[27] W. Scott Neal Reilly, "Believable Social Emotional Agents," Ph.D thesis, Dept. of Computer Science, Carnegie Mellon University.

주 문 원

- 1986년 San Jose State University, 수학과 전산전공 수료
- 1987년 New York Institute of Technology(전산학석사)
- 1988년~1991년 삼성전자 시스템연구소 연구원
- 1996년 Stevens Institute of Technology(전산학박사)
- 1997년~현재 성결대학교 멀티미디어학부 조교수
- 관심분야 : 컴퓨터비전, 멀티미디어에이전트
- E-mail : mchoo@sungkyul.edu

최 영 미

- 1979년 이화여자대학교 수학과(이학사)
- 1981년 이화여대대학원 수학과 전산학전공(이학석사)
- 1989년 University of Sydney 전자계산학과 (Visiting Scholar)
- 1993년 아주대학교 컴퓨터공학과(공학박사)
- 2001년 University of Pittsburgh 정보과학과(객원교수)
- 1994년~현재 성결대학교 멀티미디어학부 부교수
- 관심분야 : 지능형교수시스템, HCI, 멀티미디어에이전트
- E-mail : choiym@sungkyul.edu