Modelling Socially Intelligent Virtual Humans

Nadia Magnenat-Thalmann †‡
MIRALab, University of Geneva
Nanyang Technological University, Singapore

Zerrin Kasap ‡
MIRALab, University of Geneva

Abstract

In this paper, we mention about various techniques in order to create "socially intelligent virtual humans" which can interact with people using natural ways of communication. Such characters should have personality and emotions in order to engage the user to a believable interaction and make decisions based on their goals and motivations considering their past interactions and relationships with other characters. We particularly focus on computational models of personality, emotion, social relationships and episodic memory as well as present how those can be linked with decision making mechanisms in dialogue. Finally, we demonstrate our research to create long-term emotional interaction with virtual characters using an emotion model that we developed, which uses past interactions derived from episodic memory and then mention about our future research goals.

CR Categories: I.2.11 [Artificial Intelligence]: Distributed Artificial Intelligence—Intelligent agents; I.3.7 [Computer Graphics]: Three-dimensional graphics and realism—Animation

Keywords: socially intelligent virtual humans, emotion and memory modeling, decision making in dialogue

1 Introduction

Creating socially intelligent virtual humans that can interact with people and other virtual characters in a natural way is a great challenge. These characters should both have their personality and make decisions with respect to their goals, as well as behave emotionally and socially during their interaction with other characters considering their past interactions and relationships. They should also be expressive enough using various modalities such as speech and natural language, facial expressions and body gestures. What is more, during interaction with a real user, they should also have the capability to understand what the user is saying through speech recognition and recognize the user’s state from his/her facial expressions and gestures.

In our research, the goal is to create a virtual character that can interact with users in a long-term manner establishing relationships with them, that can reason/replan based on past exchanges and talk about experiences from its own perspective. We expect our characters to possess two aspects: (1) Having a sense-of-self (2) Having a sense-of-others. In order to realize this, we focus on the computational models of emotion and episodic memory and how these phenomena can be linked with decision making process in dialogue.

In section 2, first we mention about computational models of personality, emotion and social relationships and give examples of emotional virtual characters from the literature. Then, we give a background on episodic memory both from the neuroscience, psychology and artificial intelligence (AI) perspective and present an overview of life-like characters with episodic memory. In section 4, we mention about decision making techniques in conversational characters in order to link the emotion and memory processes to dialogue and natural language. In the last section, we describe our research on creating emotional and intelligent characters that can establish long-term relationships with users and mention about a case study that describes an interaction scenario with a virtual teacher.

2 Modeling Personality, Emotions and Social Relationships

Emotions are one of the important mechanisms to control the behavior of virtual characters both for the reason of giving them a personality as well as automatically producing animations by simulating the internal dynamics of the virtual characters. [Gratch and Marsella 2005] defines two methods for emotion modeling in life-like characters: communicative-driven methods and simulation-based methods. Communicative-driven methods treat emotional displays as a means of communication. These systems do not have an internal calculation of emotion but selects a display of emotion based on the current state of interaction and usually encoded via some scripting language. Simulation-based approaches attempt to model the impact of events on internal emotion dynamics focusing on the cognitive function of emotion. Emotional displays are tied to the virtual character’s emotional state rather than being triggered by their communicative function. The first group of systems usually uses categorical models of emotions such as the six basic emotions introduced by Ekman [Ekman 1982]. Second group of systems model the appraisal of events in the environment and their effects on the internal emotional state. For example, in OCC appraisal model, agent’s concerns in an environment are divided into goals (desired states of the world), standards (ideas about how people should act) and preferences (likes and dislikes) [Ortony et al. 1988].

Moods are average internal affective states of an individual. In [Vinayagamoorthy et al. 2006], mood and emotions are differentiated based on three criteria: temporal, expression and cause. Moods last longer than emotions and they are not associated with a specific event. In other words, emotions modulate actions while moods modulate cognition. The relation between mood and emotions is two-way. Mood affects the appraisal of events and decides which emotion will be triggered and with what intensity. For example, if the person is in an anxious mood, it will get more easily disappointed by bad events and with higher intensity. Emotions can also cause a particular mood to occur. For example, a person in a bored mood can become in a more positive mood after some emotional appraisal from the environment. Moods are usually represented with continuous dimensions rather than being discrete labels. For example, activation evaluation [Whissel 1980] and Mehrabian’s model of PAD [Mehrabian 1996] is used for the modeling of moods.

Personality influences the way people perceive their environment and affects their behaviors and actions and distinguishes one from another. [Vinayagamoorthy et al. 2006] define two differences between emotions and personality: duration and focus. Personality is constant and like mood it is not specific to particular events. For example, a person who is stable will have the tendency to behave in a less emotional way in difficult situations. Although there is no universally accepted theory of personality, Five Factor Model or OCEAN model [Mccrae and John 1992] is most widely used in the
Another very important factor that shapes our everyday emotional reactions is the status of our interpersonal relations with other people. This concept becomes important when the conversational partners are coming together multiple times forming a long-term interaction. In [Bickmore and Picard 2005], Bickmore mention about five social relationship models based on social psychology:

- **Dyadic models** define relationship as the interdependency between two people such that a change in the state of one will produce a change in the state of the other.
- **Provision models** are based on what one person provides for the other.
- **Economic models** such as social-exchange theory, model relationships in terms of costs and benefits.
- **Stage models** assume that relationships go through a fixed set of stages.
- **Dimensional models** attempt to abstract a given relationships characteristics to a point in a small dimensional Euclidean space.

From these models, dimensional models are more widely used for the modeling of life-like characters. The most commonly used dimensions are power and social distance [Svennevig 1999] where power refers to the ability of one individual to control the resources of another and social distance refers to how close two people and what is the level of friendship between them. Argyle [Argyle 1998] makes a similar definition and proposes a model based on two dimensions: friendliness and dominance. [Bickmore and Cassell 2001] mentions about two dimensions, self-disclosure and trust, which are important in health and sales applications. In [Bickmore and Cassell 2005], dimensions such as familiarity (degree to which people have exchanged information), solidarity (how similar two people are to eachother in their interests, beliefs) and affect (the degree of liking) are used.

Personality, mood and emotions are used in many applications as a three layer model. [Kshirsagar and Magnenat-Thalmann 2002] uses OCC model for emotions and OCEAN model for personality. Mood is represented as good, bad or neutral and Bayesian Belief Networks are used in order to model the uncertainty in mood change based on different personality. In [Egges et al. 2004], a generic model is described for emotional communication, where different theory of emotion, mood and personality can be applied. ALMA (A Layered Model of Affect) [Gebhard 2005] uses a mood model with three dimensions based on PAD model of [Mehrabian 1996] where OCEAN personality decides the initial mood of the virtual characters. In the embodied conversational agent MAX’s emotion system [Becker et al. 2004], emotion and mood is represented as an orthogonal system where the mutual interaction between them is explained with a mass-spring system. These applications do not take into account the effect of interpersonal relationships on the emotion and mood level. One of the first applications that attempt to model interpersonal relationships is [Bickmore and Cassell 2001] where they introduce the term “relational agents” and mention about a relationship model based on trust. [Predinger and M. 2002] mention about how relationships evolve over time as a consequence of affective interaction histories with other agents based on OCC appraisal model. In BASIC [Romano et al. 2005], a socially intelligent agent is developed considering factors such as emotion, mood, personality, memory of relationships with other agents and social cognitive factors. [Gockley et al. 2006] describe an affective system for a socially-interactive virtual robot based on emotion, mood and attitudes towards users where attitudes are represented with dimensions like familiarity and affect. iCat developed by Philips is a cartoon-like robot cat with facial expressions that can play chess with a human opponent. Each move of the user is interpreted as an emotional input and effects iCat’s emotional system and its response [Leite and Pereira 2007].

### 3 Modeling Episodic Memory

Episodic memory was first described by Endel Tulving [Tulving 1983] who is a neuroscientist. According to his work, memory is composed of long-term and short-term (working) memory and long-term memory is divided into declarative and non-declarative memories. Non-declarative memory is also called procedural memory which is related with learning skills like riding a bicycle or playing a guitar which requires the storage and retrieval of sensory-motor skills. Declarative memory is composed of episodic and semantic memory. Even though both episodic and semantic memory is related with the acquisition and retrieval of knowledge, they differ in many aspects. Episodic memory represents our experiences as points on a timeline associating each event with a point in time. Semantic memory is derived from episodic memory and it is a structured representation of learned facts and concepts. For example, remembering what you have done last summer is an episodic memory and your knowledge about a city is semantic memory.

Tulving [Tulving 1983] defines the following properties of episodic memories:

- **Autonoetic**: remembering entails the conscious re-experience of past memories, but the retrieved memories are distinguished from the perception of the persons current state.
- **Autobiographical**: a person remembers episodes from his/her own perspective.
- **Variable Duration**: the time period that is spanned by an episode is not fixed.
- **Temporally annotated**: the person has a sense of the time when an episode has occurred.
- **Imperfect**: the memory is incomplete and can have errors.
- **Primed**: recall occurs more quickly when it is primed by repetition or by recall of related information or similar state.
- **Forgetting**: memory performance declines with time or intervening events; this behavior is well fit by a power function (the power law of forgetting).
- **Level of Activation**: exposure frequency and recency affect the speed and probability of recall.

One of the first computational models of memory was introduced by Schank [Schank 1982]. He proposed various memory structures such as scripts (sequences of actions), scenes (group of actions that happen at the same time) and memory organization packets (MOPs) (ordered set of scenes directed toward the achievement of a goal not inferable from individual scenes). His model of dynamic memory was the basis for the Case-Based Reasoning (CBR) from the AI literature, which is a technique to solve problems based on the solutions of similar past problems [Kolodner 1993]. Case-based reasoning is very close to the idea of episodic memory, however they also differ in significant ways [Laird 2005]. For example, the reasoning mechanism in CBR is included in the method itself and realized with case reuse and adaptation. However, episodic memory can be integrated with other applications and reasoning mechanisms. It is more general and application independent which means that once it is constructed, it could be used for multiple purposes such as action selection, learning, problem solving and storytelling.
SOAR and ACT-R are the two general purpose cognitive architectures that have declarative and procedural memory but episodic memory was relatively ignored. SOAR architecture has recently been extended with a robust, general purpose episodic memory functionality [Nuxoll 2007]. Episodic memory is added to this architecture by the way of storing everything in working memory except the working memory elements with very low activation. [Tecuci and Porter 2007] present a generic episodic memory module where an episode is represented in terms of context, content and outcome. While SOAR-EM uses a feature match algorithm for computing the similarity between the cue and a prior episode, they use a semantic matcher.

Since the attention of the AI community to the episodic memory has arrived very recently, AI approaches that aims to integrate computational models of episodic memory into the virtual characters and social robots are very limited. Thus, creating life-long characters that has the capability of long-term interaction in virtual and real environments is a challenging and state-of-the-art topic.

An early example is the pedagogical agent Steve [Rickel and Johnson 1999] that uses an episodic memory structure in order to rationalize his own actions during an after-action review with the trainees. [Ho and Dautenhahn 2008] mention about autobiographic memory and integrate it with IVAs (Intelligent Virtual Agents) with story-telling capabilities where agents can remember significant past experiences and reconstruct their life stories from these experiences, in an emotion-driven planning architecture. [Brom et al. 2007] propose a virtual role-playing game (RPG) actor with a full episodic memory. The authors mention that in RPGs, it is important that non-player characters (NPCs) can tell believable stories about themselves during interaction with the player characters (PCs). Therefore, there is a need for a full episodic memory and a linguistic module transferring the outcome of the memory to natural language. In [Dodd and Gutierrez 2005], an episodic memory system is represented for the humanoid robot ISAC (Intelligent Soft Arm Control). ISAC’s cognitive architecture is designed to be able to learn appropriate behaviour through interaction with its environment. Jockel et al. [Jockel et al. 2007] investigate the use of episodic memory (EPIROME) in the multi-modal service robot TASTER which serves drinks to guests in an office environment. In EPIROME, episodic memories are considered as sequences of events and each event carries a time information and can be assigned to one of the three major event classes: perceptual, command and executive. [N. Mirza and te Boekhorst 2008] use the term interaction histories instead of episodic memory, which is defined as "the temporally extended, dynamically constructed, individual sensorimotor history of an agent situated and acting in its environment, including the social environment, that shapes current and future action". In their study, humanoid robot Kaspar2 learns to engage with a human partner in a simple "peekaboo" interaction game based on past interactions. In [Endo 2008], anticipatory behaviour of a navigation robot is computed based on the episodic-memory. In this study, robot learns to localize itself using past sensory and behavioral information in a partially observable environment.

4 Decision Making and Dialogue

Natural language dialogue is a significant component for socially intelligent virtual humans. In addition to the modeling of conversational functions such as handling of interruptions, repairing of dialogue, feedback and turn-taking, a dialogue system should realize dynamic response generation based on the emotional state, memory and motivations of the intelligent character. In this section, we will mention about decision making strategies in dialogue.

Decision making occurs in dynamic environments consisting of intelligent agents and animats. Russell and Norvig [S. and P. 2002] defines three types of agents in dynamic environments: (1) Reflex agents (2) Goal-based agents (3) Utility-based agents. Reflex agents make decisions considering only the current percept in a reactive manner. Some reflex agents also have an internal state information and combines this with the current percept in order to generate the new state. The best-known reactive architecture is the sub-sumption architecture of Brooks [Brooks 2001] where behaviors are represented as simple if-then like structures and take place in a hierarchical structure where each agent has precedence over another. Finite State Machines (FSMs) are also used as reactive architectures where the effect of a condition on the current state results in a new state. Decision-making based on FSMs was recently used in many computer games such as Halo 2 and Quake III. Current state information can not be enough to decide what to do next in most cases. Goal-based agents have a description of what is desirable and combine this with their perception in order to choose actions that achieve the goals. While in FSMs, one has to hand-author the actions to be taken at each step, goal-based architectures provides an intelligent way of decision making using AI planning techniques. For example, intelligent story characters in [R. Aylett and Paiva 2006] and emotional agent in [T.D. Bui and Nijholt 2003] combines emotional appraisal models with STRIPS-like planning. While the former uses a partial-order continuous planner, the latter uses the GraphPlan algorithm. Conversational life style agent Nabatazg [C. Smith and Hakulinen 2008] integrates dialogue and planning using HTN (Hierarchical Task Network) planning algorithm. In [Strong and Mateas 2008], discourse is generated dynamically combining FSMs with HTN planning in order generate dialogue for non-player characters. Belief-Desire-Intention (BDI) architecture is also a widely used goal-based architecture. Belief stands for the agent’s knowledge about the world, desires are the objectives to be accomplished and intentions are what the agents have chosen to do ([Viviana Mascardi 2005]). BDI is based on practical reasoning [Bratman 1987] which is the process of deciding what to do in each moment of time in order to achieve the goals. An overview of BDI-style agents in the literature such as PRS, dMars and JACK can be found in [Viviana Mascardi 2005]. Utility-based agents measures the achievement of goals by mapping a state into a real number which describes the degree of happiness and satisfaction of reaching that state. Utility functions are useful in cases where there are multiple ways to reach a goal or when there are conflicting goals. Decision-theoretic agents combine probability and utility theory and updates current state based on beliefs and by the way of choosing the action with highest expected utility e.g. [Mei Si and Pynadath 2006] and [Yu and Terzopoulos 2007].

5 A Socially Intelligent Virtual Human for Long-term Emotional Interaction

In this section, we mention about our research on creating emotion and memory-enabled virtual characters. The aim is to build long-term relationships between the virtual characters and multiple users. In order to realize our goals, we focus on three aspects: personality/emotion simulation, episodic memory and decision making in dialogue.

For the personality and emotion simulation, we have developed an emotion model based on OCEAN personality [McCrae and John 1992] and OCC emotional appraisal model [Ortony et al. 1988]. For the implementation of mood, we use the PAD temperament model proposed by [Mehrabian 1996]. The three traits in the PAD model are pleasure (P), arousal (A), and dominance (D) which are independent from each other forming a three dimensional space. Mehrabian analyzed the relationship between OCEAN personality factors and the PAD scale and presented their conversion as linear.
Figure 1: Virtual character for long-term emotional interaction

regressions. We use this conversion to set up initial moods for the characters based on different personality. Their emotional status and mood is updated with emotional impulses from the environment. After some time, the characters have the tendency to turn back to their initial mood which represents their true personality.

In our model, we also mention about the concept of long-term memory since we want our virtual characters to look more intelligent to people by storing information about past interactions over time and retrieving them back when necessary. For this reason, we developed an episodic memory structure in order to keep track of people that have been interacted and calculate relationship levels with them. As a relationship model, we use Argyle's model [Argyle 1998] based on two dimensions: friendliness and dominance. The relationship with a user affects the emotional reactions of the virtual character. For example if the user is saying something bad happened to him/her and the character has positive impressions of the person, the resulting emotion will be sorry-for but if the person is not liked, the character will experience the gloating emotion. Relationship update is based on the cumulative evaluation of the individual’s overall interaction with the user. For each interaction, positive and negative impulses from user help to construct a relationship between the interactants. In order to associate events with different users, a face detection and recognition module is integrated into our system. The details of the above mentioned research and applied scenario with a virtual teacher can be found in [Kasap et al. 2009].

According to the scenario two students interacts with our virtual character Eva (Fig 1), one playing a good student and the other playing the role of a difficult student. The face recognition module automatically detects a new person whenever somebody arrives and greets the student. Recognition of a new user starts a new interaction session which is a new episode in the episodic memory. With the good student Eva stays calm and polite while with the difficult student she is more aloof in her responses. After some time, the two students have another interaction with Eva. Face recognition module recognizes the students automatically and Eva remembers their faces. She also remembers past relationship levels with the students and updates her emotional state according to the relationships and answers accordingly. Fig 2 shows snippets from the four different dialogues.

Figure 2: Example dialogues of Eva showing two interaction sessions with good and difficult students

<table>
<thead>
<tr>
<th>First Interactions</th>
<th>Good Student</th>
<th>Difficult student</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eva: Nice to meet you Zerlin. I will start with a few simple concepts first. Do you know about different landforms on the Earth?</td>
<td>Student: I don't know anything about them. Boring topic!</td>
<td>Eva: Ok, let me explain some of them to you. It is not that boring, trust me.</td>
</tr>
<tr>
<td>Eva: Today we will talk about three types of landforms: cape, strait, and gulf. Then we will have a small test to check if you understood the concepts. Is it ok for you?</td>
<td>Student: Yes, fine for me.</td>
<td>Eva: Ok, let me explain them to you. Do you understand the concepts? Is it ok for you?</td>
</tr>
<tr>
<td>Eva: Good. Let's start with the cape. A cape is a pointed piece of land that sticks into the sea, ocean, lake, or river. A good example is the Cape of Good Hope to the south of Cape Town. Do you know what a strait is?</td>
<td>Student: No.</td>
<td>Student: Yes teachers always want to take tests.</td>
</tr>
<tr>
<td>Eva: It is important for you to learn these concepts. Let’s start with the cape. A cape is a pointed piece of land that sticks into the sea, ocean, lake, or river. A good example is the Cape of Good Hope to the south of Cape Town. Do you know what a strait is?</td>
<td>Student: No.</td>
<td>Eva: Ok.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Second Interactions</th>
<th>Good Student</th>
<th>Difficult student</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eva: Hi Mahir, great to see you! Student: Hi Eva! Eva: You were very good in our last class. Student: Thank you. You are a nice teacher too. Eva: You really think so? Thank you. Are you ready to learn more things today?</td>
<td>Student: I don't feel well today. Maybe it is better if we continue later on. Eva: Oh, I am really sorry to hear that. Ok, see you another time. Bye.</td>
<td>Eva: Hi Zerlin! You are back again! I thought you were not very interested in geography. Student: Hi Eva! I am sorry for last time, I will try my best from now on. Eva: Well, I hope so. Ok, let's start the lesson then.</td>
</tr>
</tbody>
</table>

6 Conclusion and Future Work

In this paper, we have presented various components which are important for creating socially intelligent virtual humans such as emotion, memory and decision making. Then we have presented our own research in order to create long-term emotional interactions with virtual characters, using social relationships derived from past interactions in episodic memory.

Our future goal is to improve our episodic memory model and decision making process so that the virtual character can not only remember overall relationships after each interaction session but also remember specific events in its memory, reason on them and talk about past experiences from his/her own perspective.

Construction of episodic memory is very much related with personality and emotion of the character since emotionally salient events are remembered more and have more effect on the character’s life. In the episodic memory process it is important to decide on the following questions:

- When an episode will be recorded?
- What information will be stored?
- Which cues can be chosen to store the episode for use in the retrieval process?
- How long an episode will stay in memory, how does it change over time?
- What triggers the retrieval of the episode?
- How does the current situation linked to the episode for retrieval?

After the construction of the episodic memory and personality/emotion simulator, the next step is to link them with the dialogue
manager so that the virtual character can give appropriate responses based on the information received from memory and current emotional state. In our current architecture, we are using a FSM based dialogue manager based on the work of [Egges 2006]. Our future goal is to add a planning layer on top it using a HTN planning algorithm so that we can realize plan updates based on the past knowledge in episodic memory. Besides remembering the past events related with the current dialogue context and giving appropriate responses, we want our character to tell his/her own experiences by combining multiple entries in the episodic memory so that he/she can answer questions like "What did you today, did you have a nice day?" and start telling how her they has passed, which people she has interacted with, which of them she liked/disliked and their levels in terms of the pedagogical goals of Eva.

Acknowledgements

This study is partly funded by the EU Project INDIGO (IST-FP6 045388).

References


LAIRD, J. 2005. Episodic memory vs. case-based reasoning. In Invited talk at the Sixth International Conference on Case-Based Reasoning (ICCBR).


WHISSEL, C. M. 1980. The dictionary of affect in language.