Building long-term relationships with virtual and robotic characters: the role of remembering

Kasap, Zerrin; Magnenat Thalmann, Nadia

How to cite

KASAP, Zerrin, MAGNENAT THALMANN, Nadia. Building long-term relationships with virtual and robotic characters: the role of remembering. In: Visual Computer, 2012, vol. 28, n° 1, p. 87–97. doi: 10.1007/s00371-011-0630-7

This publication URL: https://archive-ouverte.unige.ch//unige:112389
Publication DOI: 10.1007/s00371-011-0630-7

Vis Comput (2012) 28:87–97
DOI 10.1007/s00371-011-0630-7

ORIGINAL ARTICLE

Building long-term relationships with virtual and robotic characters: the role of remembering

Zerrin Kasap · Nadia Magnenat-Thalmann

Published online: 23 September 2011

Abstract With the recent advances, today people are able to communicate with embodied (virtual/robotic) entities using natural ways of communication. In order to use them in our daily lives, they need to be intelligent enough to make long-term relationships with us and this is highly challenging. Previous work on long-term interaction frequently reported that after the novelty effect disappeared, users' interest into the interaction decreased with time. Our primary goal in this study was to develop a system that can still keep the attention of the users after the first interaction.

Incorporating the notion of time, we think that the key to long-term interaction is the recall of past memories during current conversation. For this purpose, we developed a long-term interaction framework with remembering and dialogue planning capability. In order to see the effect of remembering on users, we designed a tutoring application and measured the changes in social presence and task engagement levels according to the existence of memory. Different from previous work, users' interest in our system did not decrease with time with the important contributions of remembering to the engagement level of users.

Keywords Long-term relationships · Long-term interaction · Episodic memory · Social presence

Z. Kasap (✉) · N. Magnenat-Thalmann
University of Geneva, Battelle Building A, 7 Route de Drize, 1227 Carouge, Geneva, Switzerland
e-mail: zerrin.kasap@miralab.ch

N. Magnenat-Thalmann
e-mail: thalmann@miralab.ch

N. Magnenat-Thalmann
Institute for Media Innovation, Nanyang Technological University, NTU 50 Nanyang Drive, Research Techno Plaza Foyer, Level 02, Singapore 637553, Singapore

1 Introduction

When we talk about *long-term social interaction*, we intend to describe the multiple interaction of users with a socially interactive character over a long period of time for achieving a task, getting support or for the sole purpose of entertainment. These characters can be virtual humans or robots taking the role of a friend, companion, assistant or trainer, all behaving in social norms and giving the illusion that they have a personality. Kanda et al. [10, p. 65] states that "result of immersing a robot in an environment that demands ongoing participation is likely to be entirely different from that of exhibiting the robot in a public place like a museum, where the people who interact with it are transient".

The idea of having social entities that can have bonds with their users over a long period of time started in 1990s with toy robots such as Hasbro's Furby and Sony's robotic dog AIBO. Such robots are mainly designed for entertainment, taking the role of a pet that can always accompany the users at their homes. Robots such as PaPeRo (Partner-type-Personal-Robot) and Paro (therapeutic seal robot) are designed for child and elderly care and accompany their users at homes and hospitals. Although such robots are useful in certain contexts, interaction with them is not the kind of relationship that occurs in human-human communication which requires more advanced cognitive and communicative skills.

Creating human-like interpersonal relationships with computers is a highly challenging topic. One of the recent attempts to develop a robot interacting over a long period of time was Valerie the Roboceptionist [7]. Bickmore et al. [3] introduced the term "relational agent" and constructed a health advisor agent with social-emotional and long-term relationship-building skills. Leite et al. [15] studied the role of social presence in long-term interaction with the iCat chess player robot.

Previous work on long-term interaction with virtual characters and robots frequently reported that after the novelty effect disappeared, users' interest into the interaction decreased with time. Our overall research goal is to develop a long-term interaction system that can still keep the attention of the users after the first intraction. We consider several aspects in our system such as emotion/personality modeling, dialogue generation based on past exchanges and expressive behaviour. As a case study, we developed a robotic tutor called Eva and designed an experiment to measure the changes in social presence, task engagement and motivation levels. There were four different conditions in this experiment based on memory and personality factors (1) Eva with supportive personality and with memory (2) Eva with supportive personality and without memory (3) Eva with unsupportive personality and with memory (4) Eva with unsupportive personality and without memory. We found that users' interest in our system did not decrease with time and even increased from one interaction to the other. In addition, we looked at the effect of individual components in our system to the measurements. Our results provide first evidence that existence of memory in a long-term interaction system can help to keep the attention of the users as time passes.

The paper is organized as follows: in the next section, we will discuss the background work on long-term interaction systems with virtual characters/robots. Following that, we will describe how we designed the tutoring application considering social-cognitive theories of learning. Section 4 presents the overall architecture of our system and briefly describes the individual components. In Sect. 5, we describe how dialogue is planned using episodic memory and Hierarchical Task Networks. Section 6 presents a case study followed by a summary of the results and discussion. The paper ends with conclusion and possible directions for future work.

2 Related work

Apart from the early examples such as Hasbro's Furby and Sony's robotic dog AIBO, Robovie [10] is one of the earliest example of a robot with the mention of long-term social interaction. In this study, the authors tested the interaction between a humanoid robot and students at an elementary school for a duration of two weeks. They found that the interaction time in the second week decreased indicating that students lost interest in the robot as time passed. Bickmore and Picard [3] investigated human-computer relationships by going deeper in the theories of social psychology using the weight advisor platform with the virtual agent Laura. The dialogue for Laura was scripted using Augmented Transition Networks and was capable of some simple saving/retrieving for the remembering of past events with users. The authors compared a relational and non-relational condition and found out that subjects had higher desire to continue interacting with the relational agent but there were no differences in terms of the change in exercise behaviour.

Similar to the work of Bickmore and Picard, Autom [13] was a robot developed for the purpose of helping users to lose or keep their weights. The robot used in this research was simplistic in terms of appearance with four degrees of moving head and eyes. Three different phases of relationship between the robot and the user were initial, normal and repair states. The authors compared the robot version with a standalone computer version and found that participants using the robot version of the system felt a closer relationship than the standalone computer. Another long-term social interaction study was done in Carnegie Mellon University with Valerie the Roboceptionist [7] where the authors found that the number of visitors and the duration of interaction were decreased after the novelty effect disappeared. Valerie was a graphical human-like face and was designed to help the visitors in many ways such as giving directions and checking weather forecasts. She was equipped with emotions, moods and attitudes in order to create relationships with users. Over nine weeks, the robot displayed neutral, positive or negative moods and positive version was preferred most and neutral least by the participants. Leite et al. [15] developed a chess player robot based on Philips iCat robot that can play chess with children in a chess club. The authors evaluated the system during five weeks in terms of the change in the perception of the social presence. The result of the study showed that social presence decreased over time, in particular the dimensions related with believability and the users' attention.

Although there have been some efforts to model evolving relationships, memory models for these systems is far less considered. However, almost all the work above, mentioned the importance of memory and remembering in long-term interaction. So far these systems have been developed in very different forms being a cartoon-like or a very low degrees-of-freedom robot or a human-expressive virtual character. For the first time, we use a highly expressive human-like robot for such research with complex facial expressions deriven from research in the field of computer animation. The experiments made so far with these system took 2 weeks to a couple of months. In our case, we kept the experiment duration at 2 weeks. This was mainly due to using an expensive robot and it was not possible to have an individual robot for each participant. Since interaction with the robot was on a basis of individual appointments, it was rather difficult to have a study going on for several weeks.

3 Designing the tutoring application

Social-cognitive theories mention that teaching and learning are highly social activities and cognitive/affective development of the learners are in close relation to their interaction with the learning environment [2]. While traditional intelligent tutoring systems focus only on knowledge acquisition behaving as an expert system, animated pedagogical agents [5, 9] and robotic tutors have the potential to intervene in the learning process by creating deep relationships with the learners and responding to their individual needs and emotions. For example, in [20], Saerbeck et al. describes the development of a socially supportive robotic tutor for language training. Their application is based on the idea that teachers with more social supportive behaviour are expected to achieve higher student learning performances compared to the ones focusing only on knowledge transfer. Similarly, Baylor and Kim [2] tested with three types of agent role: expert, motivator and mentor. While the expert is limited in terms of animations and speaks in a formal way and focuses only on information exchange, the motivator is highly expressive and encouraging and not necessarily knowledgeable. The mentor is a combination of these two roles combining the knowledge with encouragement and emotional feedback.

By nature, learning is a long-term process and a good learning environment requires adaptation to the students' affective and cognitive states by taking into account the cumulative effect of past interactions. In [14], several design constituents are proposed for learning companions based on social-cognitive theories. We focus on two of these constituents; affect and feedback as explained below:

- *Affect*: During the learning process, students can experience a variety of positive and negative affective states such as pleasure, frustration, boredom, anxiety or confidence [27]. Pedagogical agents can help students to control their feelings and increase their self-awareness by providing emotional feedback and emotional non-verbal behaviour. For example, when a student makes several mistakes and start to feel unconfident, a pedagogical agent can support the student with encouraging phrases and facial expressions indicating empathy.
- *Feedback*: A pedagogical agent should explicitly give feedback on the learner's progress in order to increase the learning outcome. This feedback can be in terms of reporting the success and failures of the students accompanied with explanations and reminding/repeating the topics that the learner had the most difficulty before. This can help the learner be aware of his/her progress with respect to the course plan and better understand the challenging parts.

In our system, we designed Eva as a tutor of digital photography. The content of the course was taken from the Cambridge in Color Digital Photography Tutorial.[1] Figure 1

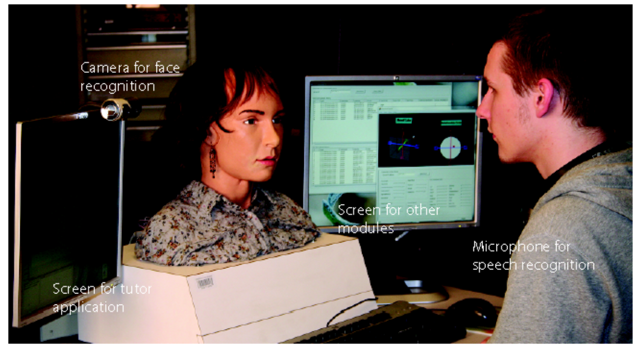

Fig. 1 Interaction setup for tutoring application

shows our system setup. There is a camera on top of the left screen which captures the user's face for recognition. The user also wears a headset for speech recognition. Whenever a new user is recognized through the camera of the robot, a new interaction session starts, otherwise the robot stays in an idle mode.

4 System architecture: theory and implementation

Our architecture is based on the similarities between episodic memory, Belief–Desire–Intention (BDI) architecture and Hierarchical Task Network (HTN) Planning. While the general architecture we use is based on beliefs, desires (goals) and intentions (actions), we use HTN planning to select the actions to be executed. The close relation between BDI and HTN planning was mentioned in some studies recently [8, 21]. Both systems have a representation of the world state and a set of primitive tasks or actions. While BDI model uses plan rules from a plan library to reduce high-level goals, HTN planning uses methods to reduce compound tasks into primitive tasks. BDI model is powerful in terms of dynamic agent representation, however it is limited in terms of not including the advantages of classical planning algorithms [21]. Consequently, we adapt a classical planning algorithm into the BDI architecture in order to support dynamic execution and also link that with an appropriate memory structure that is compatible with the BDI style.

A common aspect of most computational models of memory is the organization of memory elements around a specific goal [19, 22, 24]. This property is the basis of our idea for linking episodic memory to the goal-based structure of BDI and HTN. In our system, episodes are represented in terms of context, contents and outcome [24]:

- *Context* involves the initial state and the desired state (goal) to be achieved.
- *Contents* are the events that happen during an episode.
- *Outcome* is the result of the episode and indicates if the desired goal was achieved or violated.

[1] http://www.cambridgeincolour.com/

While the goal in the context dimension of an episode is linked to the goals in BDI model and methods in HTN planning, initial state and outcome are composed of predicates that are similar to the ones in the initial state of HTN and beliefs from BDI. The actions from BDI and primitive tasks (operators) from HTN are stored as events in the content of an episode. Table 1 shows the similarities between HTN planning, BDI model and episodic memory.

Figure 2 shows our system architecture based on the theory described above. A prerequisite for long-term interaction is the recognition of individuals and remembering their names in order to associate events to the users. For this purpose, we have integrated a *Face Recognition* component to the architecture. When a new user appears in front of the robot, the *Face Recognition* module automatically detects this person and updates the belief base about the user's name. Then, the *HTN planner* starts planning the dialogue based on the updated beliefs and selects primitive tasks or sends memory queries to the *Episodic Memory*. The primitive tasks are executed at the *Finite-State-Machine* (*FSM*) *Layer* of the *Dialogue Manager* (*DAM*) where each primitive task from the HTN planner corresponds to a related FSM. *Speech Recognition* captures user speech and active FSMs are checked to figure out if they match the recognized speech. After the FSM execution, beliefs related to the recently executed FSMs are updated and the *HTN planner* produces a new plan. In addition, the *Episodic Memory* can also update the beliefs and trigger re-planning in the *HTN Planner*. At each FSM state change, a memory record is created and sent to *Episodic Memory*. While creating this memory record, several pieces of information are considered such as time and emotional state of the event.

Table 1 Similarities between BDI model, HTN planning and episodic memory

BDI model	HTN planning	Episodic memory
Beliefs	Initial state	Context and Outcome
Goals	Methods	Context
Actions	Primitive tasks	Contents

In addition to the components regarding the dialogue management, the emotional state of the robot is updated by an *Emotion Engine* component. Emotional appraisal of events is handled in FSMs using condition-action rules and emotional state and mood are updated upon receiving an emotional event. In addition, past emotional interactions in *Episodic Memory* are used to calculate a relationship level with the current user and this is remembered in the following sessions and used to update the mood level in the *Emotion Engine*. *Facial Animation & Text-to-Speech* component receives the generated response from the dialogue manager and converts text into speech and generates the facial animation. The animations are first applied and tested on a virtual face and converted to robot animation using a MPEG-4 FAP to robot conversion algorithm. Facial animation is generated based on emotional state and mood coming from the *Emotion Engine*. Emotional state vectors are mapped to a database of facial expression and blended with speech. Mood is used to create eye blinking and idle gaze/-head behaviour.

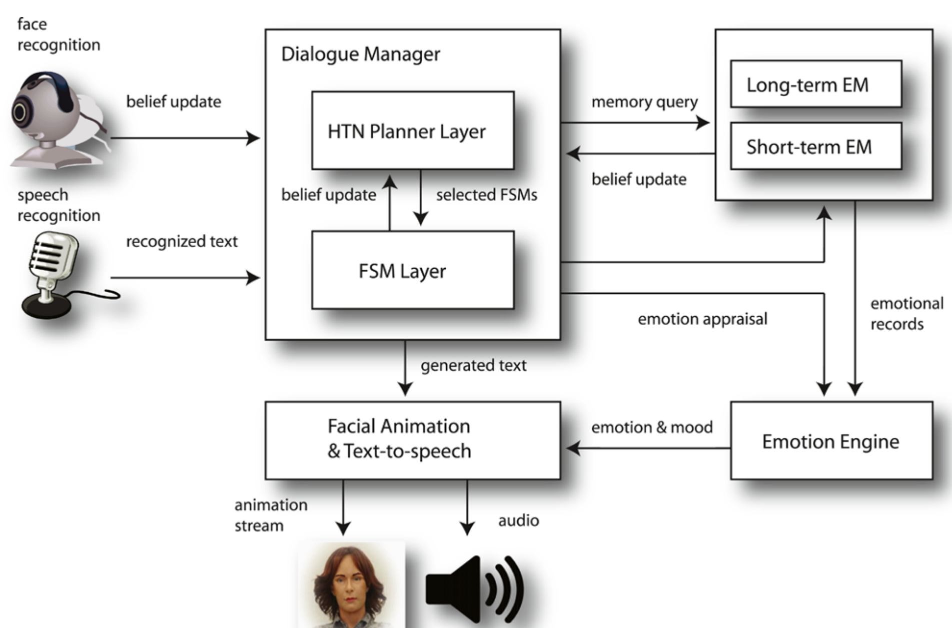

Fig. 2 Overall system architecture

5 Episodic memory and dialogue planning

5.1 Episodic memory model

As we mentioned previously, our episodic memory model is based on the context, contents and outcome sections as originally defined by Tecuci in [24]. Context is composed of initial state and goal. Goal information in the context dimension of each episode originates from the HTN planner. In other words, each primitive task from the HTN planner is a new goal and all actions executed to achieve this goal are stored in the contents section of each episode. Initial state and outcome are belief states composed of predicates and variables. We will describe our episodic memory model considering the three phases of episodic memory: encoding, storage and retrieval.

5.1.1 Encoding

The encoding phase is related to when a new episode will be recorded and what kind of information needs to be stored. A new episode is recorded whenever a new primitive task is selected by the HTN planner. Events happening during the achievement of a primitive task are stored in each episode. Episodes are first saved into short-term episodic memory (STEM). When an interaction session ends, episodes are moved to long-term episodic memory (LTEM) selecting only important and emotionally salient events from STEM. These are events that indicate an achievement or violation of a goal and/or that has emotional impact to the robot. After an interaction session ends, episodes in STEM are deleted. Episodic memory stores the important pieces of information related to each event and that could be important for the retrieval time. For each event, we record the following six items: (1) time of the event (2) user name (3) state that results after the occurrence of that event (4) goal status; whether an atomic goal is achieved or violated (4) emotional state (5) emotion intensity (6) recall probability.

5.1.2 Storage

The structure of the episodic memory should allow it to store large numbers of episodes and retrieval performance should not decrease with the increase in the number of episodes [24]. Relational databases are good candidates for implementing episodic memory because of their efficient storage and retrieval capability. In our system, we used SQLite and manipulated the database using SQL queries. Human memory works in a way that newer memories are remembered more preventing the recall of older experiences [25]. In order to support, efficiency and scalability, forgetting functions can be applied to the episodes (e.g. functional decay theory [1] and power law of forgetting [26]). In our model, we applied an exponential memory decay function using the personality factor neuroticism, based on the fact that mood has an effect on memory retrieval and people have the tendency to remember bad things when they are in a bad mood [16].

5.1.3 Retrieval

Retrieval can be done in two ways: spontaneous and deliberate [19] (or automatic and voluntary as mentioned by Tecuci [24]). Spontaneous retrieval occurs in cases where there are emotionally salient events that are related to the current situation. In our case, this occurs by spontaneously retrieving emotionally salient events with a specific user and calculating the relationship level with that user as soon as he/she is recognized. The relationship level is calculated by taking into account the cumulative effect of past emotional exchanges with respect to their recall probability and adding this effect to the mood of the robot as explained in more detail in our previous work [11]. Deliberate retrieval can be done to achieve a task, e.g. in case the robot needs to talk about the user's success level or needs to plan the content of the course. In our system, deliberate retrieval is incorporated into the HTN planning system and triggered by the planner.

5.2 Dialogue planning with hierarchical task networks

The global execution cycle for long-term dialogue planning occurs by the exchange of messages between the HTN and FSM layers of the dialogue manager and episodic memory as shown in Algorithm 1. HTN planner is the main module that updates the goals and produces the next actions to be executed. The planner gets two inputs, goal and beliefs. In our case, several goals and sub goals are produced continuously so we keep the goals in a queue. Then the last element in the goal queue is given as input to the planner at each planning cycle. In addition, there exists a belief base which contains beliefs in terms of predicates. Each time an update is made to the belief base or goal queue, the planner produces a new plan.

The system starts with the initial goal *startsession* and whenever there is an input from the face recognition, the *selectdialoguephase* goal is set. The method and the operator to start a new session is shown below. The variable *?recognizedUser* can either be *unknown* or can be a name of the recognized person from the face recognition module.

```
(:method (startsession)
  ((recognizedUser ?recognizedUser))
  ((!setgoal selectdialoguephase))
)
```

Each session is composed of four dialogue phases: welcome, warm up, teach and farewell. In order to select the

Algorithm 1 Algorithm for long-term dialogue planning

```
1: Belief base is empty, beliefBase.clear()
2: Initial goal is starting a session,
   goalQueue.add("startsession")
3: while TRUE do
4:   Check belief updates and goal updates,
     checkUpdates()
5:   if beliefbase.isUpdated()||goalQueue.isUpdated()
     then
6:     Replan, plan =
       HTNPlanner.replan(beliefbase, goalQueue.end())
7:     Execute the plan, executePlan(plan)
8:   end if
9: end while
```

next dialogue phase, the HTN planner checks which of them are executed before considering the information in the belief base. If all the dialogue phases are executed, the planner ends the session with the task *!setgoal endsession*.

Algorithm 2 shows how the *executePlan* function mentioned in Algorithm 1 works. The produced plan can be one of these three types; either it can be a memory query or a new goal; or it can be a primitive task to be executed by the FSM layer. In the above example, the plan is a new goal, which is *selectdialoguephase*. It is added to the dialogue queue which will then be input to the next planning cycle. Each goal is kept in the goal queue unless an explicit *!endgoal* action is generated.

Algorithm 2 Algorithm for plan execution

```
1: for i = 0 to plan.size() do
2:   if plan[i] is a memory query then
3:     Query the target memory,
       queryMemory(targetMemory)
4:   else if plan[i] is a new goal then
5:     Update the goal queue with the new goal,
       goalQueue.add(plan[i])
6:   else if plan[i] is an executable task then
7:     if plan[i] ==' endsession' then
8:       Save records in STEM to LTEM,
         saveSTEMToLTEM()
9:     else
10:      Execute the task, plan = executeFSM(plan[i])
11:    end if
12:  end if
13: end for
```

The produced plan can also be a memory query. In this case, the query is sent to the episodic memory and the belief base is updated with the returned belief from the memory. For example, in case of the *warmup* dialogue phase, session number is queried with the primitive task *!queryEMfor sessionno* and based on the session number, either a general introduction is done about the course by selecting the *talkaboutcourse* operator or Eva can make an overview of the last session by selecting the *talkaboutlastsession* method.

Algorithm 3 shows the FSM execution cycle for the *executeFSM* function. FSM layer loads and executes the FSM corresponding to the selected task. Each FSM has a start and end state and can have several other states in between. In our framework, FSMs are used mainly for interpreting user input and for producing answers accordingly using the condition-action structures as defined in [6]. We extended this structure in order to save state changes in FSMs into the memory.

Algorithm 3 Algorithm for FSM execution

```
1: for i = 0 to selectedFSM.noOfStates() do
2:   Run the state, run(state[i])
3:   Create memory record for state,
     memrecord = createMemRecord(state[i])
4:   Save to short-term episodic memory,
     saveToSTEM(memrecord)
5: end for
6: Create belief from the executed FSM,
   belief = createBelief(state[i])
7: Update belief base with executed FSM,
   beliefbase.update(belief)
```

6 Case study and results

We developed a case study in order to measure the effect of developed components to the long-term interaction experience. In particular, we wanted to see how memory and emotional aspects of the robot effect the users' experience. We expect that while the existence of memory will create a higher level of social presence and task engagement, emotional aspect of the robot will rather contribute to the perception of robot personality and user motivation. Since our application is in the teaching domain, we consider supportive and unsupportive personality for the robot. Additionally, we consider the effect of time on the experience. We expect that social presence will not decrease over time. We enumerate below the four general hypothesis regarding our system:

1. Existence of memory in the system will cause a higher level of social presence.
2. Social presence level will not decrease by time.
3. Existence of memory in the system will cause a higher task engagement.
4. Participants will discriminate between the different robot personalities and supportive robots will be considered as more motivating.

Fig. 3 Dialogue sample for supportive Eva with memory

Eva: Hello Joachim. Great to see you! You were doing quite good in the last lesson. I remember you got a bit confused at the end but this is normal. Making mistakes is part of learning.
User: *I feel more confident today.*
Eva: That's great. It helps you learn better. Now let's make an overview of your incorrect answers in the last lesson. Regarding shutter speed, I asked you three questions and you only couldn't answer my question related to the images with slow and fast shutter speed. I remind you that the image on the left is with slow shutter speed because slow shutter speed is required for low-light photos. In order to freeze the motion of extremely fast subjects such as the one on the right, fast shutter speed is required. Now, let's start with today's topic. In this lesson we will talk about aperture setting. Can you tell me what aperture is?
User: *Controls the area over which light can enter your camera.*
Eva: Yes, well done! It is specified in terms an f-stop value, which can at times be counterintuitive, because the area of the opening increases as the f-stop decreases. In photographer slang, when someone says they are "stopping down" or "opening up" their lens, they are referring to increasing and decreasing the f-stop value, respectively. Can you tell me the relation between these f-stop values and amount of incoming light?
User: *Every time the f-stop value halves, the light-collecting area quadruples.*
Eva: Bravo! For example, while for f/22, relative light is 1X, for f/11, relative light becomes 4X. You can see examples in the table. Now the next question. Which of the images on the screen is with wide and narrow aperture?
User: *The one on the left is narrow and the one on the right is wide.*
Eva: Wrong answer this time, sorry. But don't worry. I will explain it to you. Actually, the one on the left is wide and the one on the right is narrow. Because aperture determine's a photograph's depth of field and the image on the left has a shallower depth of field. Lower f-stop values, in other words wide aperture correlate with a shallower depth of field.

Fig. 4 Dialogue sample for supportive Eva without memory

Eva: Hello. In this lesson we will talk about aperture setting. Based on the instructions on the screen, can you tell me what aperture is?
User: *Controls the area over which light can enter your camera.*
Eva: Yes, well done! It is specified in terms an f-stop value, which can at times be counterintuitive, because the area of the opening increases as the f-stop decreases. In photographer slang, when someone says they are "stopping down" or "opening up" their lens, they are referring to increasing and decreasing the f-stop value, respectively. Can you tell me the relation between these f-stop values and amount of incoming light?
User: *Every time the f-stop value halves, the light-collecting area quadruples.*
Eva: Bravo! For example, while for f/22, relative light is 1X, for f/11, relative light becomes 4X. You can see examples in the table. Now the next question. Which of the images on the screen is with wide and narrow aperture?
User: *The one on the left is narrow and the one on the right is wide.*
Eva: Wrong answer, sorry. But don't worry. I will explain it to you. Actually, the one on the left is wide and the one on the right is narrow. Because aperture determine's a photograph's depth of field and the image on the left has a shallower depth of field. Lower f-stop values, in other words wide aperture correlate with a

6.1 Study design

We developed four different versions of the robot based on the factors memory and personality: (1) supportive Eva with memory, (2) supportive Eva without memory, (3) unsupportive Eva with memory and (4) unsupportive Eva without memory, as shown in Figs. 3, 4, 5 and 6. While a supportive Eva is helpful to the user and encourages in case of mistakes, unsupportive Eva behaves in a more dominant way and is not emphatic. While Eva with a memory can remember past emotional exchanges and previous incorrect answers, Eva without a memory cannot make links to previous sessions.

6.1.1 Measures

In this study, several measurements were used related to each of the expected effects described for the purpose of the study. Regarding social presence, we have considered three measurements: *engagement*, *user responsiveness* and *affective understanding*. While engagement is related to the extent the participant is involved into the experience, user responsiveness considers how much the user responded to the robot using his/her facial expressions or gestures. Questionnaire items related to these two sections are selected from social presence questionnaire of Lombard and Ditton [17]. Affective understanding is related to the mutual emotional exchange between the robot and the user and it is based on the social presence questionnaire of Biocca [4]. Regarding task engagement, we considered measurements such as *adaptability* of the robot to the user's learning progress, *motivation* and *success level*. While adaptability and motivation were measured with questionnaire items, success of the user was measured by calculating the correct/incorrect answers of the user through the interaction. Similar to [20], we considered extrinsic and intrinsic motivation measurements. While *extrinsic motivation* comes from outside of the individual, *intrinsic motivation* comes from inside. In our case the motivating or unmotivating manner of the robot

Fig. 5 Dialogue sample for unsupportive Eva with memory

Eva: Hi Joachim. I remember you had problems last time answering my questions correctly. You should pay more attention. I will challenge you more today.
User: *I will try my best.*
Eva: Ok, let's continue. First, let's make an overview of your wrong answers in the last lesson. Regarding shutter speed, I have asked you three questions and you answer my question related to the images with slow and fast shutter speed incorrectly. Actually the image on the left was with slow shutter speed because slow shutter speed is required for low-light photos. To freeze the motion of extremely fast subjects such as the one on the right, fast shutter speed is required. Now, let's start the today's course. In this lesson we will talk about aperture setting. Can you tell me what aperture is?
User: *Controls the area over which light can enter your camera.*
Eva: Correct answer. You know, I think you can not find such a good teacher like me elsewhere. Anyway, you need some more information on this topic. Let's continue. Aperture is specified in terms an f-stop value, which can at times be counter-intuitive, because the area of the opening increases as the f-stop decreases. In photographer slang, when someone says they are "stopping down" or "opening up" their lens, they are referring to increasing and decreasing the f-stop value, respectively. Can you tell me the relation between these f values and amount of incoming light?
User: *Every time the f-stop value halves, the light-collecting area quadruples.*
Eva: Yes, correct. However, it is not a big achievement I would say. Most students can answer that correctly. For example, while for f/22, relative light is 1X, for f/11, relative light becomes 4X which means it quadruples. You can see examples in the table. Can you tell me which of the images on the screen is with wide and narrow aperture?
User: *The one on the left is narrow and the one on the right is wide.*
Eva: Wrong answer this time Joachim. You dissappointed me, you are not as good as I think. Actually, the one on the left is wide and the one on the right is narrow. Because aperture determine's a photograph's depth of field and the image on the left has a shallower depth of field. Lower f-stop values, in other words wide aperture correlate with a shallower depth of field.

Fig. 6 Dialogue sample for unsupportive Eva without memory

Eva: Hi. In this lesson we will talk about aperture setting. Based on the instruction on the screen, can you tell me what aperture is?
User: *Controls the area over which light can enter your camera.*
Eva: Correct answer. Can you tell me the relation between these f values and amount of incoming light?
User: *Every time the f-stop value halves, the light-collecting area quadruples.*
Eva: Yes, correct. However, it is not a big achievement I would say. Most students can answer that correctly. For example, while for f/22, relative light is 1X, for f/11, relative light becomes 4X which means it quadruples. You can see examples in the table. Can you tell me which of the images on the screen is with wide and narrow aperture?
User: *The one on the left is narrow and the one on the right is wide.*
Eva: Wrong answer. You dissappointed me, you are not as good as I think. Actually, the one on the left is wide and the one on the right is narrow. Because aperture determine's a photograph's depth of field and the image on the left has a shallower depth of field. Lower f-stop values, in other words wide aperture correlate with a shallower depth of field.

was the extrinsic motivation and was measured by questionnaire items compiled by the researchers of this study. Intrinsic motivation questionnaire items were based on the Interest/Enjoyment scale of Intrinsic Motivation Inventory [18]. We also considered bipolar adjectives such as (polite–rude, friendly–hostile) in order to assess the personality of the robot considering *friendliness* and *dominance* dimensions. For all questionnaire items, we used a 7-point Likert scale. Table 2 summarizes the measurements used in this study.

Table 2 Measurements

Measurement groups	Measures
Social presence	Engagement
	User responsiveness
	Affective understanding
Task engagement	Adaptability to learning
	Success level
Motivation	Extrinsic motivation
	Intrinsic motivation
Perception of personality	Friendliness
	Dominance

6.1.2 Participants and experiment protocol

The experiment was organized at the computer science (CS) and media design (MD) departments of University of Geneva. 52 participants were recruited being 12 females and 40 males. Participants were mainly students and staff from these two departments at bachelor (30 participants), master (19 participants) and Ph.D. (3 participants) levels of education and with a mean age of 25 ranging from 18 to 46. The participants were of 18 different nationalities. While

Fig. 7 Example participants from the user study

the number of participants from computer science department was 32, the number of participants from media design department was 20. Figure 7 show example images captured during the user study.

Before the interactions, individual appointments were made with each participant by collecting consent forms. It was explained to the participants that the user study is composed of four interactions distributed over two weeks and each interaction will take around 15 minutes at maximum and additional 15 minutes maximum for filling the questionnaire after the second and fourth interactions. Participation was voluntary and a representative amount of gift is given to the students after the experiments.

The robot was set up in an isolated room over a 4 weeks period at the computer science department and 7 days at the media design department being 5 weeks at total. The participants were randomly assigned to each of the four experimental conditions and interacted with the robot four times distributed over two weeks period. During the interaction days, the experimenter welcomed the participants and ran the system to make them start the interaction. After the second and fourth interactions, users were asked to fill in the questionnaire. We applied pre- and post-tests in order to measure the effect of time. We have chosen the second interaction since remembering effect can be understood better in the second session and chose the fourth session as it is the end of all interactions. Although the experiment started with 56 participants, since two of them could not participate after the pre test, we dropped four participants from the experiment to have an equal number of participants for each condition. We had eight participants from CS department and five participants from MD department being 13 at total for each group. In the first interaction, participants were asked to fill in their background information, basically their age, gender, nationality, level of education and their knowledge of social robots and affective computing. The majority of the participants had low or little experience with social robots and affective computing.

6.2 Results and discussion

For the analysis of data, we applied a repeated measures Multivariate Analysis of Variance (MANOVA) in order to see the interaction between several independent and dependent variables and repeated measures based on time. In the between-subjects tests, we found that both memory and support factors had significant effects on the model. Results from Pillai's trace are $F(10,39) = 2.709$, $p < 0.013$ for memory and $F(10,39) = 2.309$, $p < 0.03$ for support. Within-subjects results show that time had a significant effect on the model (Pillai's trace is $F(10,39) = 8.191$, $p < 0.01$). However, combined effects of time and memory and time and support are not significant. Pillai's trace is $F(10,39) = 0.519$, $p < 0.866$ for time and memory and $F(10,39) = 0.65$, $p < 0.762$ for time and support. The results show that memory has significant effects on engagement, user responsiveness and adaptability variables and although not significant, noticeable effects on affective understanding and success level. The support factor has effects on the friendliness and dominance variables but no significant effects on extrinsic and intrinsic motivation. More details on the experimental results can be found in [12]. Figure 8 shows the relationship between memory and engagement, user responsiveness, affective understanding and adaptability to learning measurements, considering pre and post tests.

The results show that social presence level did not decrease with time and even increased from one interaction to the other. This aspect is one of the important findings of our study since previous work reported a loss of interest in long-term. Second, we found out that memory had a significant effect on social presence and it is an important component to keep the users' attention as time passes. Although our results do not significantly support that change in social presence through time was based on the existence of memory, they provide enough evidence about the importance of the memory component for long-term interaction.

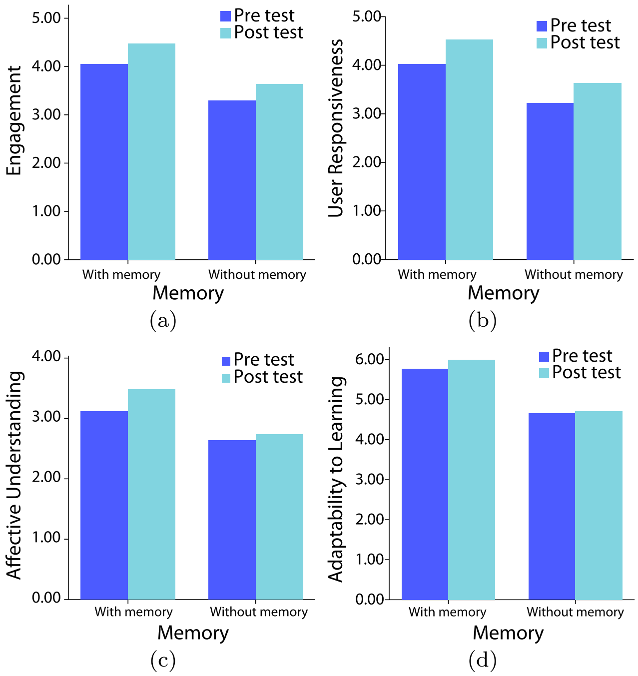

Fig. 8 Effects of memory

Overall impressions from users also supported the statistical results. While users interacting with a memory enabled robot noted that they found the robot's character interesting and it was becoming more interesting with time, participants in the without-memory conditions lose their interest quickly.

Regarding the support factor, the results show that users perceived the supportive personality more friendly and less dominant compared to the unsupportive personality. This result was not surprising but was important in order to validate whether the created personalities were correctly perceived by the participants. One unexpected result from the experiment was the relation between support and motivation of the users. This can be explained by the interesting personality of the unsupportive robot, as some participants liked the dominant manner of the robot while teaching and perceived a stronger personality effect and scored the likeness of the course higher on the intrinsic motivation scale. In other words, matching between the user and robot personality can be important and can effect the human task performance as also mentioned in [23]. For example, while extrinsic users may prefer extrinsic robot personality, intrinsic users may prefer intrinsic robot personality. In our experiment, we observed that some users liked the unsupportive robot more because it was using a more challenging manner of dialogue (extrinsic) rather than nurturing phrases (intrinsic). This was especially the case where the user was successful and the robot was unsupportive. For the extrinsic motivation, some users noted they understood that the robot was trying to motivate but they are not really convinced about that because they do not believe a robot can really motivate them and can care about their learning progress. With respect to our initial hypothesis, we can conclude that:

– Existence of memory in the system caused a higher level of social presence.
– Social presence level did not decrease as time passed and even increased in the post-test. However, although related, increase in the social presence through time was not significantly due to memory.
– Existence of memory in the system caused a higher level of task engagement.
– Participants discriminated between the different robot personalities but they were not motivated more by the supportive robot personality.

7 Conclusion

In this paper, we presented a system for creating an expressive robotic tutor for long-term interaction. We presented the overall architecture of our system based on the theory of Edisodic Memory, HTN Planning and BDI architecture and explained in detail that memory-based dialogue can be generated based on this architecture. The evaluation of the system with several participants showed that the effect of memory created a higher level of social presence on participants and memory is an important component for building long-term relationships with artificial characters.

As a future work, we can consider two aspects: First, in order to create a more flexible system, plans generated by the HTN planner can be integrated with a more sophisticated natural language understanding and generation system instead of using FSMs. Second, regarding the user study, since we could not find a significant difference between the memory conditions through time, this aspect can be explored more to find the underlying reasons as well as the effect of different robot personality on motivation.

Acknowledgements We would like to thank to Nedjma Cadi-Yazli for preparing the illustrations and video file accompanying this paper. This research is partly funded by the EU Project 3DLife (ISTFP7 247688).

References

1. Altmann, E., Gray, W.: Managing attention by preparing to forget. In: Human Factors and Ergonomics Society Annual Meeting Proceedings, Cognitive Ergonomics, vol. 1, pp. 152–155 (2000)
2. Baylor, A.L., Kim, Y.: Simulating instructional roles through pedagogical agents. Int. J. Artif. Intell. Ed. **15**, 95–115 (2005)
3. Bickmore, T.W., Picard, R.W.: Establishing and maintaining long-term human-computer relationships. ACM Trans. Comput.-Hum. Interact. **12**, 293–327 (2005)
4. Biocca, F., Harms, C., Burgoon, J.: Towards a more robust theory and measure of social presence: review and suggested criteria. Presence: Teleoperators and Virtual Environments **12**, 456–480 (2003)

5. D'Mello, S., Picard, R.W., Graesser, A.: Towards an affect-sensitive autotutor. IEEE Intell. Syst. **22**, 53–61 (2007)
6. Egges, A.: Real-time animation of interactive virtual characters. Ph.D. thesis, University of Geneva, Geneva, Switzerland (2006)
7. Gockley, R., Forlizzi, J., Simmons, R.: Interactions with a moody robot. In: Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction, HRI'06, pp. 186–193. ACM, New York (2006)
8. Harbers, M., van den Bosch, K., Meyer, J.J.: A methodology for developing self-explaining agents for virtual training. In: Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems, AAMAS'09, vol. 2, pp. 1129–1130. International Foundation for Autonomous Agents and Multiagent Systems, Richland (2009)
9. Johnson, W.L., Rickel, J., Stiles, R., Munro, A.: Integrating pedagogical agents into virtual environments. Presence: Teleoper. Virtual Environ. **7**, 523–546 (1998)
10. Kanda, T., Hirano, T., Eaton, D., Ishiguro, H.: Interactive robots as social partners and peer tutors for children: a field trial. Hum.-Comput. Interact. **19**, 61–84 (2004)
11. Kasap, Z., Magnenat-Thalmann, N.: Towards episodic memory based long-term affective interaction with a human-like robot. In: IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), pp. 479–484. IEEE Press, New York (2010)
12. Kasap, Z., Magnenat-Thalmann, N.: Long-term social interaction with an expressive robot. In: Computer Graphics International (2011)
13. Kidd, C., Breazeal, C.: Robots at home: understanding long-term human-robot interaction. Nice, France (2008)
14. Kim, Y., Baylor, A.L.: A social-cognitive framework for pedagogical agents as learning companions. Educ. Technol. Res. Dev. **54**, 569–596 (2006)
15. Leite, I., Martinho, C., Pereira, A., Paiva, A.: As time goes: by long-term evaluation of social presence in robotic companions. In: Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication, pp. 669–674 (2009)
16. Lewis, P., Critchley, H.: Mood-dependent memory. Trends Cogn. Sci. **17**, 431–433 (2003)
17. Lombard, M., Ditton, T.B., Crane, D., Davis, B., Gil-Egui, G., Horvath, K.: Measuring presence: a literature-based approach to the development of a standardized paper-and-pencil instrument. In: IJsselsteijn, W., Freeman, J., Ridder, H. de (eds.) Proceedings of the Third International Workshop on Presence (2000)
18. McAuley, E.: Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: a confirmatory factor analysis. Res. Q. Exerc. Sport **60**, 48–58 (1989)
19. Nuxoll, A.M.: Enhancing intelligent agents with episodic memory. Ph.D. thesis, University of Michigan, Ann Arbor, MI, USA (2007). AAI3287596
20. Saerbeck, M., Schut, T., Bartneck, C., Janse, M.D.: Expressive robots in education: varying the degree of social supportive behavior of a robotic tutor. In: Proceedings of the 28th international conference on Human factors in computing systems, CHI'10, pp. 1613–1622. ACM, New York (2010)
21. Sardina, S., de Silva, L., Padgham, L.: Hierarchical planning in bdi agent programming languages: a formal approach. In: Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS'06, pp. 1001–1008. ACM, New York (2006)
22. Schank, R.C., Abelson, R.P.: Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures. Erlbaum, Hillsdale (1977)
23. Tapus, A., Tapus, C., Mataric, M.: User-robot personality matching and robot behavior adaptation for post-stroke rehabilitation therapy. Intell. Serv. Robot. J. **2**, 169–183 (2008). Special Issue on Multidisciplinary Collaboration for Socially Assistive Robotics
24. Tecuci, D., Porter, B.: A generic memory module for events. In: Proceedings to the 20th Florida Artificial Intelligence Research Society Conference (FLAIRS20), Key West, FL (2007)
25. Tulving, E.: Episodic and semantic memory. In: Tulving, E., Donaldson, W. (eds.) Organization of Memory, pp. 381–403. Academic Press, New York (1972)
26. Wixted, J.: Analyzing the empirical course of forgetting. J. Exp. Psychol. Learn. Mem. Cogn. **16**, 927–935 (1990)
27. Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., Picard, R.: Affect-aware tutors: recognising and responding to student affect. Int. J. Learn. Technol. **4**, 129–164 (2009)

Zerrin Kasap is a computer scientist specialized in social interaction with virtual characters and humanoid robots. She obtained her Ph.D. in Computer Science at MIRALab, University of Geneva in Switzerland in March 2011. Her research is about all aspects of social interaction with intelligent autonomous agents. Some of the research topics she is interested in are emotion/personality modeling, artificial memory, discourse and expressive behaviour generation. Since she joined MIRALab in 2006, she has been actively involved in many European projects such as INDIGO, HUMAINE, PlayMancer and 3DLife. She was also Editorial Assistant of the Computer Animation and Virtual Worlds Journal for five years.

Nadia Magnenat-Thalmann is currently Director of the research lab MIRALab at the University of Geneva and Director of the Institute for Media Innovation (IMI) at Nanyang Technological University in Singapore. She has pioneered research into virtual humans over the last 30 years. She obtained several Bachelor's and Master's degrees in various disciplines (Psychology, Biology and Biochemistry) and a Ph.D. in Quantum Physics from the University of Geneva in 1977. She was the coordinator of several European Research Projects, among them the European Center of Excellence InterMedia and the European Center of Excellence 3D Anatomical Human. She is also Editor-in-Chief of the Visual Computer Journal published by Springer-Verlag, Co-Editor-in-Chief of the journal Computer Animation and Virtual Worlds published by Wiley and Associate Editor of IEEE Transactions on Multimedia. Together with her Ph.D. students, she has published more than 550 papers on virtual humans and virtual worlds.