

The Eighth International Joint Conference on Autonomous Agents and Multi Agent
Systems

Proceedings of Workshop 20:

Empathic Agents

Organising Committee

Timothy Bickmore, Northeastern University, USA
Stacy Marsella, USC, ICT, USA
Ana Paiva, INESC-ID and Instituto Superior Técnico, Portugal

Program Committee

Elisabeth André, Universität Augsburg, Germany
Ruth Aylett, Heriot-Watt University, UK
Kerstin Dautenhahn, University of Hertfordshire, UK
João Dias, INESC-ID and Instituto Superior Técnico, Portugal
Sibylle Enz, University of Bamberg, Germany
Lynne Hall, University of Suntherland, UK
Ian Horswill, Northwestern University, USA
Eva Hudlicka, Psychometrix Associates, USA
Stacy Marsella, USC Information Sciences Institute, USA
Magalie Ochs, National Institute of Informatics, NII, Japan
Ana Paiva, INESC-ID and Instituto Superior Técnico, Portugal
Catherine Pelachaud, CNRS, France
Paolo Petta, OFAI, Austria
Helmut Prendinger, National Institute of Informatics, NII, Japan

Extra Reviewers

Patricia Vargas, Heriot-Watt University, UK
Daniel Schulman, Northeastern University, USA
Iolanda Leite, INESC-ID and Instituto Superior Técnico, Portugal
Wenji Mao, USC Information Sciences Institute, USA

Preface

Creating characters and robots that give the illusion of life and allow for the user's suspension of disbelief is still a debated and fundamental goal in the area of virtual agents. However, when we watch a film, play a game or read a book, we not only suspend our disbelief and look at the characters as "life", but, most importantly, we establish emotional relations with the characters. We feel sad when they are sad, angry when something unfair is done to our favorite character, and so on. That is, we put ourselves in the shoes of some characters, and feel emotions for what is happening to them. And these emotional empathic relations are part of achieving that suspension of disbelief.

This workshop will be a meeting point to discuss the creation of agents that are both empathic towards their users and foster empathic reactions towards them by their users. This workshop will be the second on this topic; the first one was organized in 2004 in NewYork at the AAMAS conference.

The main goal of this workshop is to bring together researchers from different disciplines to discuss the creation of what we call empathic agents. Empathy has been defined as an observer reacting emotionally because he perceives that another is experiencing or about to experience an emotion. Humans, when interacting with virtual agents or robots can be led to feel empathy, and experience a diverse set of emotional reactions. On the other hand, agents and robots can in a certain, perhaps limited way, also show certain emotions in reaction to human emotions, thus seemingly expressing empathy towards other agents and towards humans. By seeking inspiration in empathic relations established between humans and between humans and animals, in this workshop we expect to explore these two dimensions of empathic agents.

20th, March, 2009

Contents

1	C. Brom & J. Lukavsky, <i>Towards Virtual Characters with a Full Episodic Memory II: The Episodic Memory Strikes Back</i>	1
2	D. Heylen, <i>Sensitive Empathic Agents</i>	9
3	N. Bee, E. André, T. Vogt & P. Gebhard <i>First ideas on the use of affective cues in an empathic computer-based companion</i>	13
4	S.ENZ, C. Zoll, M. Diruf & C. Spielhagen <i>Concepts and Evaluation of Psychological Models of Empathy</i>	19
5	T. Bickmore & R. Fernando, <i>Towards Empathic Touch by Relational Agents</i>	27
6	C. Adam & P. Ye <i>Reasoning about emotions in an engaging interactive toy</i>	31
7	I. Leite, A. Pereira, C. Martinho, A. Paiva & G. Castellano <i>Towards an Empathic Chess Companion</i>	33
8	N. Seiler, D. Benyon & G. Leplâtre <i>An Affective Channel for Companions</i>	37
9	A. Miklósi <i>How to make agents that display believable empathy? An ethological approach to empathic behavior</i>	43

Towards Virtual Characters with a Full Episodic Memory II: The Episodic Memory Strikes Back

Cyril Brom

Charles University in Prague
Dept. of Software and Computer Science Education
Malostranske nam. 25
Prague 1 – 118 00 – CZ
brom@ksvi.mff.cuni.cz

Jiří Lukavský

Academy of Sciences of the Czech Republic
Institute of Psychology
Politických vězňů 7
Prague 1 – 110 00 – CZ
lukavsky@praha.psu.cas.cz

ABSTRACT

Recently, it has been proposed that virtual characters should have a *full* episodic memory storing more or less everything happening around them, as opposed to an *ad hoc*, that is, *special purpose* episodic memory. However, it was not much clear, what exactly this “fullness” should mean. The purpose of this paper is to clarify it and show how it can contribute to the agents’ believability. Later, our work-in-progress applying several aspects of the full episodic memory will be reviewed. At the time of writing, the memory model integrates following parts: a hierarchically organised memory for events, a component reconstructing the time when an event happened, a topographical memory, and an allocentric and egocentric representations of locations of objects. The main functional features include: representation of complex episodes (e.g. cooking a dinner) over long intervals (days) in large environments (house), forgetting based on emotional importance of episodes, and development of search strategies for objects in the environment.

Categories and Subject Descriptors

I.2.11 [Distributed Artificial Intelligence]: Intelligent agents.
I.2.6 [Learning]: Connectionism and neural nets, Knowledge acquisition.

General Terms

Algorithms, Design, Experimentation, Theory.

Keywords

Virtual characters, episodic memory, autobiographic memory, spatial memory, dating of events, allocentric and egocentric representations.

1. INTRODUCTION

A *believable virtual agent* is an autonomous agent who seems lifelike, whose actions make sense to the audience, and who allows them to suspend their disbelief providing convincing portrayal of the personality they expect or come to expect (Loyall, 1997). It contributes highly to the believability of an agent if the

audience is able to establish *empathic relations* with the agent (e.g. Paiva et al., 2004). In other words, the users should be able to spontaneously and naturally tune themselves into the agent’s “thoughts” and “feelings” (Baron-Cohen, 2003, p. 21), to perceive that the agent is experiencing or about to experience emotion (Paiva et al., 2004). Arguably, episodic memory is one of the key components contributing to establishing the empathic relations, because it allows the user to understand better the agent’s history, personality, and internal state: both actual state and past state. It has been already discussed that believable agents (or characters) should have, at least for some applications, episodic memory (Ho & Watson, 2006; Castellano et al., 2008). In our previous work, we have even proposed that they should have a *full* episodic memory (Brom et al., 2007). But what does it mean “a full episodic memory” (FEM)? In the above mentioned paper, we used a vague definition of a memory storing almost everything happening in the proximity of the agent, as opposed to the *ad hoc/special purpose* solutions. Certainly, this full episodic memory *cannot* be a faithful reconstruction of human episodic memory—it can be a model mimicking some of its features, but which ones? And when speaking about empathic characters, are there some features that are more important for them than others?

The main purpose of this paper is to revisit the notion of the FEM, give it a more exact shape and reconcile it in the light of needs of empathic agents. The aim is to arrive a) at a tentative list of features of episodic memory most important for empathic agents, and b) at the definition of the FEM.

We begin our search tapping at the door of people who should be most knowledgeable about real episodic memory: psychologists and neurobiologists. It will turn out, however, that we won’t be much lucky. Then, we will sketch out some cognitive skills requiring some aspects of episodic memory. This step will help us with the objective (a), but only partly with (b). Through another step, we will come very close to the definition of the FEM, but, surprisingly, we will resist the temptation to define it claiming that the definition would be of no use. But we will also arrive at a definition of something else, more important than the FEM.

After this discussion, the paper will give a technical context to some of the ideas sketched out previously reviewing briefly our on-going work on a virtual character that encodes and recalls complex events, including detail information about time and space. An important feature of our model is a gradual forgetting. For the space constraints, the model cannot be detailed here fully, but the reader can find more in (Brom & Lukavský, 2009). The

Cite as: Title, Author(s), *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, Decker, Sichman, Sierra, and Castelfranchi (eds.), May, 10–15, 2009, Budapest, Hungary, pp. XXX–XXX. Copyright © 2009, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

whole paper addresses primarily the audience of developers of empathic virtual characters; it aims at providing them with some hints concerning equipping their agents with episodic memories. However, some points may be of use also for neuro-/psychologists. The discussion will be kept on the conceptual and the methodological levels. This paper extends our original work on characters with the FEM (Brom et al., 2007) and complements our methodological paper on possible utilisation of virtual characters with episodic memory in the field of neuro-/psychological computational modelling (Brom & Lukavský, 2008). The conceptual issues related to virtual characters with episodic memory (not necessarily a full one) have been also discussed by Ho & Watson (2006).

2. TOWARDS FEATURES OF THE FEM

The important concept behind current neuro-/psychological memory research is the idea of multiple memory systems. Episodic memory (Tulving & Donaldson, 1972; Baddley et al., 2001) is an umbrella term for those of these systems that operate with representations of personal history of an entity, which entails encoding these representations, their maintenance, consolidation and recollection. These representations are related to particular places and moments, and connected to subjective feelings and current goals. Fundamentally, the episodic memory is being distinguished from the semantic memory and the procedural memory. The former is conceived, more or less, as systems operating with general facts about the world as viewed from the objective perspective. The latter covers processes related to skill learning and the subjective experience is again not emphasised. The importance of agent's subjective history makes episodic memory an interesting area for empathic agents developers.

However, beyond these general statements the issues become dim. For example, to which extent the systems of episodic, semantic and procedural memory overlap? Many accept the tentative neuro-/psychological taxonomy of memory types developed by Squire & Zola-Morgan (1991) (see also Eichenbaum & Cohen, 2001), but this taxonomy elaborates the notion of procedural rather than episodic memory. Some make a distinction between episodic memories consisting of sensory-perceptual-conceptual-affective information derived from single experiences, and autobiographical knowledge, which is basically personal semantic knowledge, devoid of context in which it was acquired (Conway, 2005; Williams et al., 2008). The terminology does not seem to be settled yet, therefore it is not possible to simply implement the properties of human episodic memory. Think of this example: If a virtual agent remembers that her glasses are at the TV, is this related to episodic memory (a remembrance of the episode of putting this glasses there), semantic memory (the general knowledge about where things tend to be), or procedural memory (an unconscious stimulus-response-like habit)? A neuropsychologist would likely say that all the three alternatives are possible. But which of these properties should FEM possess? It is not only a problem of psychological terminology. Imagine we know how to implement the agent with the ability to recall the position of glasses – what is actually recalled? Think of the first alternative. Should she recall only the relation <at, glasses-23, TV-4>? Or also the features of TV, for instance its colour? Should she also recall that she put the glasses at the TV *because* she wanted to read newspaper, a task she needed different glasses for? What should happen, if, after all, the glasses are not at the TV? To our knowledge psycho-

logical details of these processes sufficient for implementing our virtual agent are not available.

As we are speaking about the needs of believable characters, we can, for obvious methodological reasons, undergo the “user centric” turn and to stop asking questions about the nature of episodic memory and to start asking questions about what users would expect from FEM agents. Assuming they would expect from them the same as from real humans, we are actually asking questions about users’ folk psychology. The problem is, that at least to our knowledge, it is not known much about this issue. Nevertheless, it seems reasonable to expect that most people do not have the concept of episodic memory at all and there are suggestions that humans expect the human memory in general to behave unlike it really behaves (e.g. Loftus, 1979; Friedman, 1993).

Hence, the neuro-/psychology thread helped us to reveal two problems with our hypothetical FEM: that 1) we do not know what features the FEM should possess, and 2) even if we knew it, we would not know how to implement them. It seems that we will have to guess the features and somehow try to implement them, a blind search approach. Luckily, even though neuro-/psychology cannot offer us the technical specification for the FEM, it can constrain our search. It can offer us some interesting general architectures (e.g. Conway, 2005; Zacks et al., 2007), inspiring observations, e.g. the idea of false memories (Loftus, 1979; Brainerd & Reyna, 2005), and some hints such as that one has to distinguish between a short-term and a long-term memory (that is, briefly, between memories from which information fades out quickly vs. not so quickly¹). And of course, this discipline can offer us loads of data, from which are arguably most interesting for our purposes diary studies (e.g. Wagenaar, 1986; Burt et al., 2003), event perception studies (Shipley & Zacks, 2008) and forensic psychology data (e.g. Loftus, 1979). It offers us also some computational models of laboratory tasks such as memorising of words or navigation in the Morris water maze (e.g. Miyake and Shah, 1999; Norman et al., 2008; Krichmar et al., 2005), but we would hardly utilise these for the FEM, unless we aim at engaging our agents in really weird tasks. Finally, we know that we should evaluate our models on users, that is, we should ask whether the models would pass an episodic memory variant of the Turing test.

What next? Perhaps... could we try the luck at the very field of virtual agents? Indeed, several reports have emerged during last years on agents with various episodic memory-like capabilities. Agents have been reported with spatial memory to increase believability of navigation and/or “what-where” judgments (Thomas and Donikian, 2006; Strassner and Langer, 2005; Peters, 2006; Isla and Blumberg 2002; Noser et al., 1995). Other characters have been equipped with a memory for past events for the purposes of debriefing (Johnson, 1994; Rickel and Johnson, 1999; Dias et al., 2007). Also there has been work on robots with a simple episodic memory (Dodd, 2005) and work at the intersection of the field of virtual characters and the artificial life investigating

¹ What exactly means “quickly” depends on the kind of memory one is talking about. One story would be told by a neurobiologist investigating memory mechanisms at a neural level (e.g. Kandel, 2001), another by a psychologist investigating memory for words (Baddley, 1986; Chap. 3). One may also argue that humans do not have one short-term memory and one long-term memory, but many interacting memory systems, each of which keeps information over a specific time interval.

how different types of episodic memories can improve an agent's chances of survival (Ho et al., 2008).

These models depart from computational neuro-/psychological models in one important way. They are aimed at representing complex, rich, human-like episodes, or large spaces such as a city with many landmarks and objects. If a forgetting mechanism is implemented, the models can be used in scenarios lasting long time intervals, e.g. days. However, these models can not be conceived as FEM models; they are technical, special-purpose solutions invented to address a particular issue (and they typically work well for the purposes of that issue). Can they help us to underpin the features of the FEM at the least? Yes, similarly to the neuro-/psychology, we can draw inspiration from them; however, the standpoint is now different. These models force us to think not about the properties of the FEM, but about cognitive skills an agent potentially may have that demand these properties. In other words, we are forced to think about how to utilise the FEM.

2.1 How to utilise episodic memory?

On the one hand, we are still not far from where we begun, on the other hand, we have some vague ideas, hints and constraints, which encourage us to try the good-old-fashion approach: brute-force search. Let us now challenge the notion of FEM during a two-step search. First, we will ask “why”: why we need an FEM agent? We will lay down a tentative list of cognitive skills that demand some kind of episodic memory, not necessarily the FEM, and ask for examples of real world applications that would utilise agents with particular skills (see? this step is motivated by the outcome of that part of our previous debate that concerned itself with virtual characters). Of course, applications featuring FEM agents have to demand *all* the skills, and we will try to identify these applications. Second, we will ask “what”: what requirements on the FEM architecture stem from these skills (this will capitalise on neuro-/psychological inspirations).

Now, let us start with the “why” part—the required agent skills:

A1. Debriefing. Tutoring agents should be able to talk about history of given lessons. As said above, agents with this ability already appeared.

A2. Giving information. This skill extends A1 for the purposes of long-living agents; it is the ability of giving users information about what happened in the virtual world in the past. Arguably, this skill is presently most important for role-playing game (RPG) characters. Predominantly, these agents now tend to inform players about important past happenings by means of pre-scripted dialogs. It would be useful to generate this information dynamically both from the design point of view as well as for believability reasons. Virtual characters living in large yet-to-be-developed social virtual worlds (Goertzel, 2007) would need this ability as well.

A3. Remembering the course of interaction. Agents with conversational abilities, such as virtual companions (Castellano et al., 2008), virtual guides (Kopp et al., 2005; Lim, 2007) or again NPCs need to keep a track of the dialog with a user. Long-term companions may be engaged in dialogs extended over many days. This may demand building information about their users. Think of an agent chatting with an elderly user about her old photographs (Companions, 2006); the agent should remember when the events

portrayed had happened and who they are about.² Note, that this ability is, to a large extent, based also on semantic memory system.

A4. Searching for objects. Think again about the example of searching for glasses. Every agent living in a world that include objects that can change their positions beyond the agents' capabilities must be able to judge reliability of contradictory memory records (unless the agent looks directly to the world map). Where are the glasses: at the TV, or at the bed side table? Suppose the agent needs also a pencil, which may be either at the TV, or somewhere in the study room. Where the agent should go first?³

A5. Topological orientation. Agents embodied in virtual environments (as opposed to speaking heads etc.) should be able to orient themselves, no matter whether they act in a city, a family house, or a country-side. This is an easy issue. However, whenever the topology can change dynamically, the agents have to construct dynamically their internal “topological memories”. Even though there is an abundance of work addressing this issue in robotics (e.g. Kuipers, 2000), and some also in the domain of virtual characters (e.g. Thomas and Donikian 2006), many players of real-time character-based strategies are still witnessing soldiers “hiding” behind a once-existing wall that has been destroyed, for the place was marked as a cover by a designer.

*A6. Mental imagery and predictions.*⁴ Agents employing declarative representations may be already conceived as using imagery. One general example is the usage a graph of way-points during path-planning, but there are also many special purpose imagery-based tasks virtual agents may need to solve in specific applications. For example, some tutoring agents may be required to answer questions such as: “what would happen if I press this button?” While simple answers may be represented in advance, for more complex situations, the agent may need to generate the answer using the imagery (cf. Rickel and Johnson, 1999).

A7. Sharing of knowledge. It is known that sharing of information can improve agent's survival (e.g. Ho et al., 2008; Cace & Bryson, 2007). Generally, this objective is more related to ethological modelling than virtual characters, perhaps with the exception of team-based action games (“the weapons are behind the corner!”). However, someday, long-living agents inhabiting a large virtual societies in RPGs or yet-to-be-developed virtual worlds will need to share information for believability purposes; without sharing,

² We would like to thank to our colleague Jan Hajič for pointing us at this example.

³ For present purposes, we conceive spatial memory systems as a part of episodic memory. Actually, spatial memory is a field of study of its own and its episodic nature is being discussed. For example, one major theory about the role of the hippocampus posits that its main function is spatial, while another theory argues for its role in processing of events. The neurobiological field seems to be interested in convergence of these two main threads of thinking (e.g. Eichenbaum 2004; Morris 2007, p. 581; Hassabis & Maguire, 2007).

⁴ Humans are quite good in employing imagery to solve various problems ranging from path-planning to anticipating consequences of some situations to solving puzzles. Even though the nature of mental imagery is still hotly debated (see e.g. Pylyshyn, 2003; Kosslyn, 2006), it is clear that at least some of its aspects depend on the episodic memory system. Concerning the role of episodic memory in anticipation, see Zacks et al. (2007).

they would start to look as strange, uncommunicative individuals. Think of agents living in a closed area, such as a small city, that share information about a bizarre event that happened in past; an outsider should be recognised immediately for its unfamiliarity with the event. On a long time scale, in large long-running virtual worlds, we may even witness emergence of different “socio-cultural” groups of agents! Note that A7 skill departs from A2 in that A7 is oriented towards other agents while A2 towards users.

A8. Learning. Episodic memories can be exploited for the purposes of learning. For example, they can be used in an off-line manner during tuning of an agent’s behaviour. Another possible use is for problem-solving; when an agent faces a problem, he can try to find whether he had not already solved a similar problem in the past and if he did, he can try to tackle the present problem in the way that worked then. Nuxoll (2007) points out similarity between these usages of episodic memories with case-based reasoning (Kolodner, 1993).⁵

Surely, we have not listed all possible skills capitalising on some facets of episodic memory, but the list is sufficient for the illustration that virtual characters may really need this memory. Arguably, the skills needed directly for interaction with users—A1, A2, A3—are most important for empathic characters. However, all other skills can be vital for some applications with empathic characters as well. Believability of an agent stems not only from user-agent interactions but also from the overall agent behaviour and agent-agent interactions while the agent is observed by the user.

Now, an FEM agent should possess all the skills at the same time. Can we imagine such an agent? What about an agent living in a magnificent yet-to-be-developed MMORPG or in a large social virtual world of the future (Goertzel, 2007)? The defining feature of this agent, besides her longevity, would be conversational abilities. This agent could have a regular “employment” in her virtual world, she could be a museum guide in a virtual museum for instance.

Well, but except of sci-fi examples, do we have really something? Unlikely. Most virtual characters would need some of these skills, but *not all* of them. Nevertheless, let us imagine that we have an FEM agent, that is, an agent with the A1–A8 skills; which requirements on the FEM architecture stem from having these skills?

2.2 Requirements on episodic memory

Let us start with real humans (see? the neuro-/psychology is coming...). Humans tend to segment the external flow into pieces organised around objects, actors, actions and the orders in which these elements combine to achieve specific goals: events (Nelson, 1986). Events take place in scenes: specific combinations of objects and/or situations at specific locations (Tversky et al., 2008). Events have a beginning and an end (Zacks et al., 2007), even though these may be sometimes fuzzy. Events can be either witnessed or communicated via language. It is these concepts—

events, objects, actors, spaces, scenes, time, and language—we propose to design the FEM architecture around (Fig. 1):

B1. The notion of complex events. An FEM should support representation of complex real-world events that involve actors with human-level cognitive abilities. Such events have typically a nested structure – they can be logically decomposed to smaller events, until an atomic level is reached (Zacks & Tversky, 2001). This is in opposition to both merely physical events such as collision of two objects and laboratory events such as a presentation of a world list to memorise. This notion is demanded most by skills A1, A2, A6, A7, and A8.

B2. The notion of the time and the order. An FEM agent should be able to answer questions like “when something happened?”, “what happened sooner?”, or “what happened after something?”. This feature is somehow demanded by all the skills; even for A4, the agent needs to compare the recency of memory records, and for A5, one can argue that some agents may need to remember past topologies (“there was a shop here, but now, there isn’t”).⁶ But the notion of time has also other manifestations. For example, agents should use relative time concepts when speaking, such as “last summer,” or “morning”. This is not just an issue of mapping of absolute time units to a relative scale; relative notions are context depended—Monday morning is typically sooner than Sunday morning. Another sign of the time notion: agents should remember the course of interactions and be able to continue an interaction appropriately if interrupted (even today it may happen that when an RPG player returns to the pub he already visited, the virtual guests show no sign of remembrance him). These signs are most important for A1–A3. Yet another sign of the time notion: a long-living agent can be expected to adapt to a new way of life, e.g. after experiencing a change of a time-zone.

B3. The notion of objects and actors. An FEM agent should understand not only events in which she participates as the actor, but also events that she only observes. She has to understand who is the actor of an observed event (its causal factor) and what are the objects (the entities being manipulated with). Sometimes, there is no apparent causal factor, the case of “raining”; sometimes, there is a kind of “joint actor”, the case of a dancing couple. This notion is most important for A1, A2, A3, A6, A7, and A8 skills. Sometimes, it may be sufficient to understand just affordances of an object (“this object can be used for this and that”), other times, features of an object and their changes may be needed as well (“the glass has been destroyed during that action”). Arguably, the latter is most vital for A8.

B4. The notion of space. An FEM should underpin many facets of spatial cognition. One of them is the topological knowledge about accessible locations, another is the awareness of the actual agent’s surrounding, another is the long-term memory for positions of objects, yet another is the support for usage of linguistic terms describing spatial information, such as “left from” or “in front of me”. Spatial skills go far beyond the A* and steering algorithms. Skills A4 and A5 benefit from the space notion; however, other

⁵ Arguably, humans use past episodic memories to improve some of their problem-solving skills or semantic knowledge. However, the matters are not without controversy; for example, Tulving (2001) hypothesises that humans construct their semantic knowledge first and episodic memories second.

⁶ Note that there is an inherent plausibility—folk psychology tension in the issue of timing. For example, people are known to be quite poor in dating, but may expect quite the opposite at the same time (Friedman, 1993).

skills may also require at least a rudimentary understanding of space, including A1 and A3.

B5. The notion of scenes. Events do not take place in the abstract space, their stages are scenes; in a sense, scenes extends the notions of space, objects, and events.⁷ While some scenes can be conceived as situation-based, it is the spatial facet that dominates in others. An example of the former is a “queue for something” while of the latter a “kitchen”. Most skills require the notion of scenes, but while an FEM agent should have the ability to learn new scenes based on her interaction in the virtual world, most special-purpose agents act in limited domains, thus can be given the list of all the scenes *a priori*.

B6. The notion of language. In present context, the language is the medium for mediating knowledge about events (in a narrative-like way?). An FEM agent should be able not only to represent the flow of events based on what she directly perceives and feels, but also what someone tells her. This understanding is most important for the social skills A3 and A7. On the other hand, the FEM agent should be able to express her experience via language; the skills A1, A2, A7. Note that language can be actually used for building any declarative knowledge, including semantic knowledge.

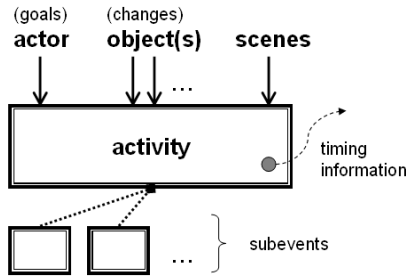


Fig. 1. The hypothetical unit around which episodic memories are organised (cf. Schank & Abelson, 1977; Zacks & Tversky, 2001).

2.3 The “definition” of the FEM

Now, the B-list from previous section is reasonably large and we can return to our original questions: 1) What are the features of the FEM? 2) What are FEM agents good for? We will first answer the former. Then, it will turn out, that we won’t need to answer the latter anymore.

Notice that the list above tells us one important thing: there is no one to one mapping between the skills (As) and the requirements (Bs). For example, the skill A2 somehow underpins all the requirements; though some of them more (e.g. B1) while others less (e.g. B4). This and the fact that many agents need more than one skill (though not all of them) bring us to the hypothesis that many developers of today agents or agents to be built in the near future have to have similar, though not exactly same, requirements on episodic memory systems of their agents. Had every agent with a skill from the A-list demanded just one or two isolated mechanisms and, in addition, were these mechanisms different for every agent, it would make sense to develop these mechanisms during

regular agent development, that is, to produce special-purpose solutions (which is what happens now for the few agents with episodic memory). However, it seems that this is not the case. Instead, there seems to be a large overlap of agent needs, hence there could be many (presently, non-existent) techniques that could be re-used. If this hypothesis is true, it would make sense to start a fundamental research program on generic episodic memory mechanisms, such mechanisms that can be picked by developers and customised for their agents similarly to how A* and steering techniques are now used. This research program would prevent developers to reinvent wheels as well as bring fruits of the integrative approach (when two mechanisms, such as a spatial memory and a memory for events, interact each other with, it is typically advantageous to start to investigate them together at some stage of progression; but this typically does not happen during regular development).

To sum up, it seems that there are strong reasons to start a research program, whose main goal would basically be:

to produce a bunch of ready-to-use mechanisms modelling some functional aspects of episodic memory for believable characters, capitalising on the integrative approach.

The methodology of the program would be as follows: 1) to choose some mechanisms to investigate, 2) to investigate them in isolation *not* in the context of any specific application, 3) to wire them together again *not* in the context of any application, and to investigate how they communicate, influence each other, and hopefully produce emergent phenomena, 4) to customise this amalgamation or its parts for purposes of a specific application, 5) to add a new mechanism, returning somewhere between Stages (2) and (3). Of course, the selection made in Stages (1) and (5) should be well motivated, perhaps with the help of the A- and B-lists.

Now, we may return to Question (1). We have two possibilities how to define the FEM. First, we can say something like “the FEM is a bunch of memory systems that a) underpins the skills A1-A8 and b) is organised around the concepts B1-B6”. Well, but we know that neither of the lists is definite. Imagine we define the FEM as suggested and an agent that needs the skills A1-A8 *plus* a new skill A9 will appear. This would be a silly situation: will we define something like FEM+? What to do if an A10 skill appear? It does not seem that this would be a useful definition.

But we now have also another possibility. Recall that the objective of the abovementioned research program is to produce a body of episodic memory mechanisms. We can define this body as the FEM. However, we think that this would be again a useless definition for this research will unlikely produce an outcome that will be fixed for eternity: the body of mechanisms would likely grow according to the needs of future agents.

What is the conclusion? We propose to resist the temptation to define the FEM for fruitlessness of this concept. Does this mean that the whole discussion was useless? It was not for two reasons. First, it helped us to isolate the A-list and the B-list, which are crucial for empathic characters. Second, it allowed us to formulate arguments for the advantage of integrative approach to the fundamental research on episodic memory for virtual agents. Given this conclusion, we should also resist the temptation to answer Question (2) for we have no definition of an FEM agent. However, this does not mean that the proposed research cannot produce many

⁷ There are arguments based on fMRI experiments that the mental scene reconstruction is the key component process of various episodic and spatial memory abilities (Hassabis & Maguire, 2007).

interesting agents, as side-products in fact. If such an agent is developed and she finds no direct application, would it mean that the agent is useless? It won't for she would help to investigate the mechanisms of episodic memory, which will likely be directly applicable for another agents if the choices made during Stage (1) would be wise.

2.4 Some fruits of the integrative approach

We now illustrate two features of human episodic memory that goes across all or most of the points of the A-list and the B-list. Hence, it does not seem odd to investigate how these features can contribute to various mechanisms, not just to one special-purpose mechanism developed in isolation.

C1. Sparseness of encoding. Human episodic memory does not encode all available information. Some may not pass through the attention, some is likely encoded in an abstract way, without details. This applies for objects, spaces as well as events. For example, one may encode that an event happened at a scene “a place where I usually have breakfast” without encoding the colour of the table cloth. Or one may encode that he was “cooking”, omitting the moment-by-moment course of the event (think of schemas (Bartlett, 1932)). Why episodic memory works in this way? The reasons seem to be “technical”; for instance, it is often argued that the following causes play their parts: the limited resources of our brains and the coding of the information in such a way that the information can be retrieved later easily after being cued.

C2. Forgetting and error susceptibility. Humans are not able to retrieve everything what they have encoded. Something can be retrieved only in the right context, something may be lost. Forgetting includes degradation of the content of episodes, spatial representation as well as temporal information. Its important feature is that it is *gradual* as opposed to *binary*. Different memories are forgotten in different speeds (likely based on their importance and emotional relevance). Similar memories can be eventually blended together. False memories can occur. Again, there are arguments that these “faults” are not faults but functional features of human memory (e.g. Schacter, 2002).

These points bring us to the widely accepted notion that human episodic memory has *reconstructive* nature, according to which the episodic memory is an active process of “constructing the past” that *engrave* memories and *reconstructs* them as opposed to merely *storing* them and *searching for* them in a database-like manner (e.g. Bartlett, 1932; Koriart & Goldsmith, 1996). Notice that the reconstructive nature underpins both C1 and C2.

Even though most present-day episodic memory agents have C1 feature, their memory systems tend to *store* everything what passed through a threshold mechanism of attention, and they typically do not employ forgetting or they use it in a simplified all-none fashion (see Strassner and Langer, 2005 for an exception). Although this approach is sufficient for most present-day applications (Ho & Watson, 2006), it may have two drawbacks from the long-term perspective. First, as suggested, the reconstructive nature of episodic memory is likely functional, it is technically advantageous. We believe that it will be easier to tackle some issues such as blending of episodes or limited computational resources when adopting the reconstructive perspective instead of the storage-based one. The second drawback is that storage-based memories are not psychologically plausible. However, it is not clear

presently to which extent this is really an issue for believable agents need to be *folk* psychologically plausible, but not psychologically plausible. How exactly do humans expect episodic memory to behave? Here, we come to the second objective of the research program proposed above:

to investigate which features of agent episodic memory contribute to agents believability and which do not.

Results of this line of research can also contribute to psychology. However, we have to develop the models first.

3. OUR AGENT

The purpose of this section is to review our on-going work on episodic memory for virtual characters which follows the research program defined in Sec. 2. For brevity, we will only sketch the main features of the model here. The model is detailed in the extended version of the paper (Brom & Lukavský, 2009), which also demonstrates benefits of the integrative research method taking various parts of the model as examples, and which gives some hints to empathic agents developers which parts of the model can be utilised in their applications.

Conceptually, the model integrates following parts: a visual short term memory, a long-term memory for “what-where” information, a life-long episodic memory, a component for timing, and a simple prospective memory. The action selection mechanism of the agent is a derivation of the BDI (Bratman, 1987). The agent features a simple valence-based emotion model. Presently, we have four independent implementations of various parts of the model, three of them employing a 2D grid world, the last one using a 3D world of the action game Unreal Tournament (Epic, 2004).

The key component of the model is the *long-term episodic memory* (LTEM), which has been already published in the paper that had proposed the notion of FEM (Brom et al., 2007). The LTEM represents what happened to the agent in the past, the flow of events. The memory is a hierarchical structure organised around tasks the agent can have in order to achieve some goals. The node of this structure resembles the unit on Fig. 1. The whole structure has some support in psychology (Zacks & Tversky, 2001). The fact that the tasks the LTEM stores have variable grain size allows for *gradual forgetting*: unimportant details of episodes can be forgotten. This memory has two mechanisms for storing timing information. One is based on time tags: when an event happens, an exact time information is added. This mechanism is simple to implement, but not plausible (Friedman, 1993). The second mechanism is a connectionist network that is able to a) acquire time concepts such as “morning” or “after lunch” based on the history of the agent’s interaction, b) to represent timing information approximately, c) to gradually forget the timing information, d) to blend similar episodes that happened at different times.

One of the limitations of the LTEM is that it is not able to answer believably questions on positions of objects that are passive but whose locations can be changed by external forces. For this reason, the LTEM is intertwined with a memory for “what-where” information. This memory stores positional information in three frames of reference: egocentric, allocentric, and associative-based (the last one simply makes weighted associations between objects and places, estimating possible objects’ locations). Our work in progress concerning this component is a mechanism that is able to learn notions of places based on where the agent lives such as “a

kitchen”, “a corner in the kitchen”, “a place in front of the monitor at the table” etc.

Together, the LTEM and the “what-where” memory underpin the skills A2 and A4. For example, if the agent is asked where are her glasses, she is able to answer: “likely at the bedside table, less likely next to the TV, and if they are not there, they might be somewhere in the living room or in the kitchen”. If the agent is asked when she was gardening yesterday, she will answer “after lunch”, not “from 2.13 to 4.12 p.m.”.

There are several important points about this memory model. Most notably, the model is not a monolithic mechanism capitalising on a single representation, instead, it is a bunch of interconnected systems. Another thing is that even though it is not clear whether the agent featuring the whole memory can be directly utilised in a real-world application, the components of the memory can be. For example, virtual companions acting in the context of humans’ flats (and in fact, robotic companions as well) can utilise the “what-where” map. Many long-living characters, such as RPG agents or storytelling agents, can use our LTEM, possibly with the timing mechanism. Finally, some of the mechanisms can be useful in other disciplines, for example in the subfield of psychology studying spatial cognition.

4. CONCLUSION OF EPISODE II

Episodic memory is one of the key components contributing to establishing the empathic relations with virtual agents, because it allows the user to understand better an agent’s history, personality, and internal state. We have started with an idea that the *full* episodic memory might be an important, but yet-to-be defined, component of empathic agents. Now, our view is that it does not make a sense to define this component; instead, it is more fruitful to define a new research paradigm that investigates various episodic memory mechanisms capitalising on the integrative research method. The main goal of this paradigm is twofold: a) to develop a set of special purpose episodic memory techniques for agent developers, b) to investigate the plausibility—believability tension by evaluating the models with respect to real users. The paper also briefly reviewed our on-going work that can be regarded as pursuing this kind of research. Another contribution of this text is that it verbalised several fundamental skills of virtual characters that demand episodic memory and several notions around which episodic memory models should be organised.

To complete the picture it must be said that many issues concerning episodic memory have not been discussed here. For instance, how is the content of episodic memory related to the concept of self? (See Ho & Watson (2006) for more on this issue.) Would it be possible to generate the content of episodic memory automatically, e.g. using HTN-planning? There are many works on spatial cognition abilities in robotics; can we utilise some? Could a hardware chip for episodic memory be developed?

Exciting times seem to be at the horizon. Looking forward to Episode III.

5. ACKNOWLEDGMENTS

This work was partially supported by the Program “Information Society” under project 1ET100300517, and by the Ministry of Education of the Czech Republic (Res. Project MSM0021620838). The authors thank to students developing

parts of the mentioned episodic memory model as their theses: Tomáš Korenko, Tomáš Soukup, Jan Vyhnánek, Jakub Kotrla, Klára Pešková, Rudolf Kadlec, and Ondřej Burkert.

6. REFERENCES

- Baddley, A. 1986/1997. *Human Memory*. Psychology Press in Taylor & Francis Group.
- Baddley, A., Conway, M., Aggelton, J. (eds) 2001. *Episodic Memory: New directions in research*. Oxford Uni. Press.
- Baron-Cohen, S. 2003. *The Essential Difference*. New York: Basic Books.
- Bartlett, F. 1932. *Remembering*. Cambridge Uni. Press.
- Bratman M. E. 1987. *Intention, plans, and practical reason*. Cambridge, Mass: Harvard University Press.
- Brainerd, C. J., Reyna, V. F. 2005. *The Science of False Memory*. Oxford Uni. Press.
- Brom, C., Pešková, K., Lukavský, J. 2007. What does your actor remember. Towards characters with a full episodic memory. In *Proc. of 4th ICVS, LNCS 4871*. 89-101. Berlin, Springer-Verlag.
- Brom, C., Lukavský, J. 2008. Episodic Memory for Human-like Agents and Human-like Agents for Episodic Memory. In *Tech. Reps FS-08-04 AAAI Fall Symposium BICA*, AAAI Press. 42 – 47
- Brom, C., Lukavský, J. 2009. Towards Virtual Characters with a Full Episodic Memory II: The Episodic Memory Strikes Back – unabridged version. Tech. Rep. 3/2009. Dept Software Comp Sci Education. Charles University in Prague.
- Burt, C.D.B., Kemp, S., Conway, M.A. 2003. Themes, events and episodes in autobiographical memory. In *Memory & Cognition* 31(2): 317-325
- Cace, I., Bryson, J.J. 2007. Agent Based Modelling of Communication Costs: Why Information can be Free. In *Emergence and Evolution of Linguistic Communication*. 305-322
- Castellano, G., Aylett, R., Dautenhahn, K., Paiva, A., McOwan, P. W., Ho, S. 2008. Long-term affect sensitive and socially interactive companions. In *4th Int. Workshop on Human-Computer Conversation*.
- Companions: Intelligent, Persistent, Personalised Multimodal Interfaces to the Internet. 2006. The European Community's Seventh Framework Programme. URL: <http://www.companions-project.org/> [12.2.2009]
- Conway, M. A. 2005. Memory and the self. In *Jn. of Mem. Lang.* 53: 594-628
- Dias, J., Ho, W.C., Vogt, T., Beeckman N., Paiva, A., Andre, E. 2007. I Know What I Did Last Summer: Autobiographic Memory in Synthetic Characters. In *Proc. of ACII*. 606–617. Berlin, Springer-Verlag.
- Dodd, W. 2005. The design of procedural, semantic, and episodic memory systems for a cognitive robot. Master thesis. Vanderbilt University, Nashville, Tennessee.
- Eichenbaum, H., Cohen, N. J. 2001. *From conditioning to conscious recollection*. Oxford Uni. Press
- Eichenbaum, H. 2004. Hippocampus: Cognitive Processes and Neural Representations that Underlie Declarative Memory. *Neuron* 44: 109-120.
- Epic Epic Games: UnrealTournament 2004. <http://www.unrealtournament.com> [12. 2. 2009]

- Friedman, W.J. 1993. Memory for the Time of Past Events. In *Psychol Bull* 113(1): 44 – 66
- Goertzel, B. 2007. AI Meets the Metaverse: Teachable AI Agents Living in Virtual Worlds. URL: <http://lifeboat.com/ex/ai.meets.the.metaverse> [12.2.2009]
- Hassabis, D., Maguire, E. A. 2007. Deconstructing episodic memory with construction. In *Trends Cogn Sci* 11(7): 299-306
- Ho, W. C., Watson, S. 2006. Autobiographic knowledge for believable virtual characters. In *Proc. of Intelligent Virtual Agents*, LNCS 4133. 383-394. Berlin, Springer-Verlag.
- Ho, W., Dautenhahn, K., Nehaniv, C. 2008. Computational Memory Architectures for Autobiographic Agents Interacting in a Complex Virtual Environment. *Connection Science*. In press.
- Isla, D., Blumberg, B. 2002. Object Persistence for Synthetic Creatures. In *Proc. AAMAS'02*, 1356-1363
- Johnson, W. L. 1994. Agents that learn to explain themselves. In *Proc. of the 12th Nat. Conf. on Artificial Intelligence*, 1257-1263. AAAI Press.
- Kandel, E. 2001. The Molecular Biology of Memory Storage: A Dialogue Between Genes and Synapses. In *Science* 294. 1030-1038
- Kolodner, J. 1993. *Case-Based Reasoning*. Morgan Kaufmann.
- Kopp, S., Gesellensetter, L., Krämer, N. C., Wachsmuth, I. 2005. A Conversational Agent as Museum Guide – Design and Evaluation of a Real-World Application. In *Proc of IVA*, LNCS 3661. 329-343. Berlin, Springer-Verlag.
- Kosslyn, S. M. 2006. *The Case for Mental Imagery*. Oxford Uni. Press.
- Koriat, A. & Goldsmith, M. 1996. Memory metaphors and the real-life/laboratory controversy: Correspondence versus storehouse conceptions of memory. In *Behavioral and Brain Sciences* 19(2): 167-228
- Krichmar, J. L., Seth, A. K., Douglas, A., N., Fleischer, J. G., Edelman, G. M. 2005. Spatial Navigation and Causal Analysis in a Brain-Based Device Modelling Cortical-Hippocampal Interactions. In *Neuroinformatics* 3(3): 197-221
- Kuipers, B. 2000. The Spatial Semantic Hierarchy. In *Artificial Intelligence* 119: 191-233
- Lim, M. Y. 2007. *Emotions, Behaviour and Belief Regulation in An Intelligent Guide with Attitude*. Ph.D. diss. School of Mathematical and Computer Sciences, Heriot-Watt University.
- Loftus, E. F. 1979. *Eyewitness testimony*. Harvard Uni. Press.
- Loyall, B. A. 1997. *Believable Agents: Building Interactive Personalities*. Ph.D. diss. Carnegie Mellon University.
- McNaughton, B.L. et al. 2003. Off-line reprocessing of recent memory and its role in memory consolidation: a progress report. In *Sleep and Brain Plasticity*. 225-246. Oxford Uni. Press.
- Miyake, A., Shah, P. eds. 1999. *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. Cambridge University Press.
- Morris, R. 2007. Theories of hippocampal function. In *The hippocampus Book*, 581-694. Oxford University Press.
- Nelson, K. (ed.) 1986. *Event knowledge: Structure and function in development*. Lawrence Erlbaum.
- Noser, H., Renault, O., Thalmann, D., Magnenat-Thalmann, N. 1995. Navigation for Digital Actors based on Synthetic Vision, Memory and Learning. *Computer and Graphics*, 19(1): 7-19
- Norman, K. A., Detre, G. J., Polyn, S. M. 2008. Computational models of episodic memory. In *The Cambridge Handbook of Computational Cognitive Modeling*. In press.
- Nuxoll, A. 2007. *Enhancing Intelligent Agents with Episodic Memory*. Ph.D. diss. The University of Michigan.
- Paiva, A., Dias, J., Sobral, D., Woods, S., Hall, L. 2004. Building Empathic Lifelike Characters: the proximity factor. In *Proc of Workshop on Empathic Agents*, AAMAS'04.
- Peters, C. 2006. Designing Synthetic Memory Systems for Supporting Autonomous Embodied Agent Behaviour. In *Proc. RO-MAN06*. 14-19.
- Pylyshyn, Z. 2003. Return of the mental image: are there really pictures in the brain? In *Trends in Cogn Sci* 7(3): 113-118
- Rickel, J., Johnson, W. L. 1999. Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition, and Motor Control. *App. Artificial Intelligence* 13(4-5): 343-382
- Schacter, D. L. 2002. *The Seven Sins of Memory: How the Mind Forgets and Remembers*. Mariner Books.
- Shipley, T. F., Zacks, J. M. (eds) 2008. *Understanding Events*. Oxford Uni. Press.
- Schank, R. C., Abelson, R. P. 1977. *Scripts, plans, goals, and understanding*. Hillsdale, N.J: L. Erlbaum Associates.
- Squire, L. R., Zola-Morgan, S. 1991. The medial temporal lobe memory system. In *Science* 253: 1380-1386
- Strassner, J., Langer, M. 2005. Virtual humans with personalized perception and dynamic levels of knowledge. *Comp. Anim. Virtual Worlds* 16: 331-342
- Thomas, R., Donikian, S. 2006. A spatial cognitive map and a human-like memory model dedicated to pedestrian navigation in virtual urban environments. In *Proc. Spatial Cognition V*, LNCS 4387. 421-436. Berlin, Springer-Verlag.
- Tulving, E. 2001. Episodic memory and common sense: how far apart? In *Episodic Memory: New directions in research*. Oxford Uni. Press. 269-288.
- Tulving, E., Donaldson, W. 1972. *Organization of memory*. New York: Academic Press.
- Tversky, B., Zacks, J. M., Hard, B. M. 2008. The Structure of Experience. In *Understanding Events*. Oxford Uni. Press.
- Wagenaar, W. A. 1986. A study of autobiographical memory over six years. In *Cogn Psychol* 18: 225-252
- Williams, H. L., Conway, M. A., Baddley, A. D. 2008. The Boundaries of Episodic Memories. In *Understanding Events*. Oxford Uni. Press.
- Zacks, J. M., Tversky, B. 2001. Event Structure in Perception and Conception. In *Psychol Bull* 127(1): 3-21
- Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., Reynolds, J. R. 2007. Event Perception: A Mind-Brain Perspective. In *Psychol Bull* 133(2): 273-293

Sensitive Empathic Agents

Dirk Heylen
Human Media Interaction
University of Twente
The Netherlands
heylen@ewi.utwente.nl

ABSTRACT

The Sensitive Artificial Listener is a project in which agents with different personalities engage the user in a dialogue to try to change the user's emotional state. The utterances that the characters can use are divided into four sets depending on the emotional state of the user (aroused/positive, aroused/negative, not aroused/positive, not aroused/negative). In this paper, we look at the utterances from an "interpersonal" perspective. Do characters differ in their interpersonal stance (dominance, affiliation, empathy) and is this reflected in the type of utterance?

Keywords

Interpersonal stance, Speech Acts, Personality, Embodied Agents

1. CONTEXT

The goal of the SEMAINE project (<http://www.semaine-project.eu>) is to create four Embodied Conversational Agents with different personalities that engage a human interlocutor into chatting with them. Each agent has a different personality corresponding, *more or less* to the four quadrants of a two-dimensional emotion space (activation/valence). Poppy is cheerful and active (positive valence, full of energy). Obadiah is gloomy (negative, no energy), Spike is full of energy but somewhat aggressive and Prudence is mostly practical (more or less positive but not very active). The goal of the characters is, first of all, to keep the user engaged and talking, and secondly to draw the user into the same emotional quadrant by some general remarks that are typical of chat-bots (see below for examples). The challenges for the project are to build real-time reactive agents that a) can get a minimal sense of what the user is talking about to respond with a reply that doesn't miss the mark completely, b) can get a sense (by analysing facial expressions, head movements, gaze, and speech - prosody and voice quality) of the "emotional" state of the interlocutor to know what to say to change it into their own mood. Therefore, the agents should be able to a) choose the right thing to say (choose from a fairly limited selection of canned phrases) or to communicate nonverbally, b) choose the right time to say it, c) and give adequate backchannel cues to show the appropriate engagement in the dialogue, stimulate further conversation and display the right attitude that fits the personality.

The Semaine project builds on earlier work on the Sensitive Artificial Listener technique:

"The Sensitive Artificial Listener technique (SAL for short) is based on the observation that it is possible for two people to have a conversation in which one pays little or no attention to the meaning of what the other says, and chooses responses on the basis of superficial cues. The point was made long ago by the ELIZA scenario (Weizenbaum 1996). In the SAL technique, system responses are keyed to the emotional colouring of what

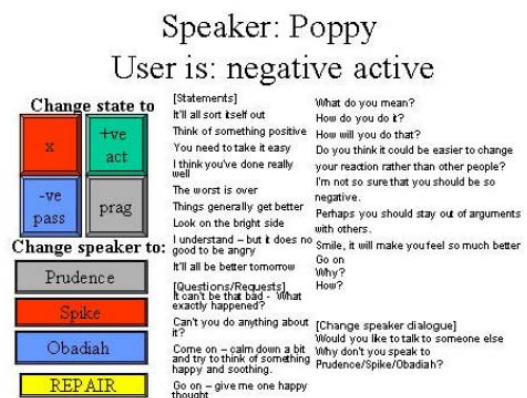
the user says, rather than (as in Eliza) words or phrases. The versions used so far have used Wizard of Oz techniques where a human operator follows a script that specifies possible responses. Because the aim is to evoke emotionally coloured responses, the statements are stock phrases chosen to evoke strong reactions in the listener." [1].

The goal of the SEMAINE project is to replace the Wizard by a machine, a dialogue system that autonomously decides on the basis of verbal and nonverbal cues what facial displays, nonverbal vocalisations and linguistic utterances (accompanied by the appropriate facial displays and head movements) to show.

2. REFRAMING THE PROBLEM

In this paper we would like to take a closer look at the sentences that the various characters can select from and we go back to the way they were created.

"The scripts for the characters were developed, tested and refined in an iterative way. Each character has a number of different types of script depending on the emotional state of the user. So, for example, Poppy has a script for the user in each of four emotional states – a positive active state, a negative active state, a pragmatic state, a negative passive state. There are also script types relevant to the part of the conversation (beginning, main part) or structural state of the conversation (repair script). Each script type has within it a range of statements and questions." [1].



The set of utterances in the original version (and the new versions of SAL) have been created by *imagining* what a specific character representing one of the four quadrants would say to a person in a specific mood to get the person drawn into its emotional quadrant. In this paper we would like to ponder – in retrospect – on what this "imagination" involved. Take, the case of Poppy (positive/active character), illustrated above. The figure shows a

part of the WoZ interface. The user is talking to Poppy and has been “diagnosed” by the Wizard to be in a negative/active state (anger would fit in this quadrant). Now the operator can choose to a) change the assessment of the user state into negative/passive, positive/active or pragmatic (\approx positive/passive) if there is a reason for that, b) change to another character, or c) choose one of the sentences shown in the figure. Suppose, one opts for the third choice. Then, for the Wizard and for the automatic system that is to replace the Wizard, there is the question which of the utterances to choose from. Is each of them always appropriate, or does it depend in a way to what the user was talking about? This is one of the issues we need to solve for the autonomously operating system, and we found out that many utterances of the character can only be selected if the system has minimal knowledge of the content of the user’s utterance. “I think you’ve done really well”, for instance, is appropriate only if the user was talking about something he did. As one can see from the examples, the responses resemble those of the typical ELISA-style chatbot. So to what extent are these four characters different instantiations of the Rogerian style psych-bot?

In looking at the various ways in which to classify the utterances and their “selection restrictions” we found some other taxonomies of speech acts that may be relevant to consider. Let us therefore look now at this issue from another perspective. What the character is trying to do is to evaluate the emotional state of the interlocutor, and – attempting to keep true to character – act in such a way as to modify the state. This is an interpersonal action requiring not just the skill to read the mental/affective state of the interlocutor but also the skill to judge which of the utterances would have the intended effect of causing a change in the interlocutor’s mental state. We can thus rephrase the challenge of the automatic interactive system to have the adequate interpersonal skills (clearly we are entering empathy-and-related-notions grounds here). Reframing the problem in this way, we can also ask the questions whether we have a framework to categorize the utterances on this interpersonal effect and how the original selection of utterances for the different characters and user states fare with respect to this categorization in interpersonal terms. Another way to phrase the question is whether the personality of the character and the fact that it makes its choice of utterance depends on the emotional state of the user, is systematically correlated with taking different interpersonal stances and strategies, reflected in the type of dialogue act? A question we would like to address in the future is how the autonomous system could make use of the interpersonal variables and processes that are going on, rather than make use of a simple look-up table?

Do different characters use different categories of speech acts in different situations? Is a character that tries to get a person into a happy mood more empathic than a character that tries to get the user depressed? How is this reflected in the utterances?

Stiles [2] has described a classification of speech acts based on some fundamental *interpersonal* variables: does one take the point of view of the other or of self, does one talk about the experience of self or other, does one assume to know the experience of the other or of self? The four characters in the SAL scenario differ in terms of their “view on life” and their emotional stance, but is this also reflected in these interpersonal dimensions. The framework introduced by Stiles may provide an interesting classification of the utterances reflecting the imagination skills (and the understanding of human nature) by the authors of the WoZ

system and its worth in providing a possible framework for the future architecture of the automatic system that will take into account these underlying *interpersonal* variables to select an utterance as well besides the classification into the emotional quadrants. In Section 3 we briefly present the schema introduced by Stiles to classify utterances and in Section 4 we will show how the two characters Poppy (positive, energetic) and Obadiah (negative valence, low arousal) fair with their utterances on the dimensions that Stiles proposes.

3. THE FRAMEWORK

The VRM taxonomy is a general-purpose system for coding speech acts. VRM stands for verbal response modes (though it might be applied to non-speech/nonverbal acts as well in some cases and to a certain extent). As Stiles puts it in his book “Describing Talk” [2] “the taxonomic categories are mutually exclusive and they are exhaustive in that every comprehensible utterance can be classified”. The taxonomy distinguishes eight modes based on three binary principles of classification: a) what is the source of experience of the utterance (is one talking about the experiences of self or experiences of other); b) the presumption about experience (does one presume to know the experience or is one wandering about it) and c) the frame of reference (is one taking the point of view of self or the point of view of the other). To give some examples, the utterance “I want to go fishing” refers to the experience (thoughts, feelings, perceptions and intentional actions) of the speaker, whereas the question “Do you want to go fishing?” refers to the experience of the addressee (“other”). With the question the speaker does not presume to know what the other person is thinking, feeling, perceiving or intending, nor was, will be or should be thinking, or intending. But with the statement about his own intention, the speaker does presume knowledge about his experience. In contrast with this, a directive such as “Go fishing” does make the presumption about what the speaker thinks the other should be doing and with it it tries to impose an experience on the other. With the utterance “You want to go fishing.” the speaker does not only talks about the experience of the other, and makes presumptions about it, but also takes the perspective (viewpoint, frame of reference) of the other.

The following table lists the 8 categories introduced by Stiles and their dimensions: do the talk about the experience of other or self; do they presume knowledge about the experience of other or self and do they take the point of view of other or self (in that order)

Reflection	o	o	o
Interpretation	o	o	s
Acknowledgement	o	s	o
Question	o	s	s
Confirmation	s	o	o
Advisement	s	o	s
Edification	s	s	o
Disclosure	s	s	s

We also provide a brief paraphrase.

- Reflection: expressing empathy

- Interpretation: explaining or classifying another's behaviour
- Acknowledgement: providing interpersonal space
- Question: gathering information about other
- Confirmation: expressions of agreement, disagreement, or shared experience
- Advisement: guiding another's behaviour
- Edification: representations about objective reality
- Disclosure: revealing one's personal condition

The four characters for SEMAINE differ in their emotional stance, by definition. One could imagine though that the more aggressive character Spike is a bit unfriendly and may lack certain empathic skills. Obadiah is a bit gloomy, but may empathise about the sad conditions in which the human interlocutor finds himself in. Poppy is cheerful and happy, but does her emotional state leads her to fail to empathise with the emotion of the other? And what about the pragmatic Prudence? In the following section we report on the first statistics regarding the classification of the utterances of different characters in terms of the dimensions introduced by Stiles.

4. COMPARING POPPY AND OBADIAH

We compare the different emotional categories of utterances from one character and the utterances in the same emotional category by different characters. Does the character of Obadiah (the more gloomy, negative/passive character) show in the kinds of Verbal response Modes it produces and how is this different from the cheerful Poppy? Do they only differ in terms of their emotional stance, or does it have an effect on their interpersonal stance?

The following table shows some typical Poppy utterances.

I think you should feel really happy today
 Absolutely
 amazing!
 Are you still there?
 aren't you just great?
 Can't you do anything about it?
 Cheer up
 Come on – calm down a bit and try to think of something happy and soothing.
 Come on – let go a bit. Tell me about the last time you were really happy
 Did things get better?
 Do go on - I love hearing all this happiness
 Do you have any good news to tell me?
 Do you think it could be easier to change your reaction rather than other people?
 Don't feel bad.
 Don't worry, the whole world can't be bad, it's just the way

you feel at the moment...that will change

Every cloud has a silver lining

Everything will work out – you'll see

Fantastic!

Give me just one happy thought and you'll feel better

Go on

Whereas the following is a selection of Obadiah utterances:

Are there any difficulties you can think of?

Are you still there?

As one door closes another one slams in your face, I always say

But don't you get sad sometimes?

But how will you get through this?

But what can you do?

Can it get any worse?

Can you remember feeling more miserable?
 Come down to earth a bit – just think about all those depressing things you have to do

Depends on your point of view

Doesn't sound that hopeful

Don't count your chickens before they hatch

Don't get too carried away

Don't get too excited

Don't know what you've got to feel so cheery about

Don't you get sad sometimes?

Don't you sometimes wish that you could just run away?

Don't you think life wears you down?

Don't take it out on me

Go on

We labeled all of the responses to a user utterance of both Poppy (100 utterances) and Obadiah (109) utterances with the most prominent label of the Stiles VRM scheme and found the following results. The table below shows the percentage of utterances in each class. There is a big difference between the number of Reflections between Obadiah (9) and Poppy (26). Poppy seems to use more expressions of empathy than Obadiah. Poppy also raises more questions, whereas Obadiah gives slightly more advice.

	Obadiah	Poppy
acknowledgement	0	1
advisement	26,5	21
confirmation	8	2

disclosure		
edification	4	0
interpretation	9	2
question	17	10
reflection	26,5	38
	9	26

We can also look at each of the dimensions a) Source of experience (Other = {reflection, interpretation, acknowledgement, question}) Presumption about experience (Other = {reflection, interpretation, confirmation, advisement}) and c) Frame of Reference (Other = {reflection, acknowledgement, confirmation, edification}) in turn.

The following table shows how many utterances of each of the characters are about the experiences of the other. Interestingly, the gloomy character Obadiah talks about as much about its own experience as about the experience of the interlocutor, whereas Poppy, focuses on the experience of the interlocutor in three out of four utterances.

	Obadiah	Poppy	
Source Experience	9	26	R
	0	1	A
	17	10	I
	26,5	38	Q
	52,5	75	(total)
	(other)	(other)	

Looking at the “other” point of view that characters display through their utterance, this appears to be in balance overall: 26 versus 31, though most of these are taken up by reflection utterances in the case of Poppy whereas Obadiah takes the frame of reference of the other in reflection, confirmation and edification utterances. Note that both characters predominantly (1 out of 4 and 1 out of 3) display their own frame of reference in their utterances.

Frame of reference	9	26	R
	0	1	A
	8	2	C
	9	2	E
	26	31	
	(other)	(other)	

Finally, when one looks at whether the characters presume knowledge about the experience of the others in their utterances, one sees, again, the same tendency in both characters.

Presumption Exp.	9	26	R
	17	19	I
	8	2	C
	26,5	21	A
	60,5	68	
	(other)	(other)	

5. CONCLUSION

What do these figures tell us? For now, it is a bit preliminary to draw conclusions. What is obvious though, is that the responses for the various SAL characters do not just differ in terms of their affective dimensions (arousal, valence) but also in terms of interpersonal dimensions. Poppy is talking less about self than Obadiah. Poppy has more empathic utterances at hand.

I have learned, from personal experience, that people have different ways to express empathy or sympathy with your situation. Some people will ask how you feel, whereas others will tell you they fully sympathise with your feeling without inquiring about them first. This personally felt difference in the expression of empathy related to the presumption about experience made me interested in the wider applicability of Stiles’ classification of utterances along some primitive interpersonal dimensions when I came across it recently. The numbers presented in this paper, show that when authors of an interactive character try to come up with the things to say for that character, are not just invoking ideas about the emotional state of the character, but also, make some presumptions about its interpersonal attitudes; or so it seems.

6. ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 211486 (SEMAINE).

7. REFERENCES

- [1] E. Douglas-Cowie, R. Cowie, C. Cox, N. Amier and D.K.J. Heylen The Sensitive Artificial Listener: an induction technique for generating emotionally coloured conversation, in LREC Workshop on Corpora for Research on Emotion and Affect, L. Devillers, J.-C. Martin, R. Cowie, E. Douglas-Cowie and A. Batliner (eds), ELRA, Paris, France, ISBN 2-9517408-4-0, pp. 1-4, 2008.
- [2] W. B. Stiles, Describing Talk Taxonomy of Verbal Response Modes. Sage Publications, 1992

First ideas on the use of affective cues in an empathic computer-based companion

Nikolaus Bee
Institute of Computer Science
University of Augsburg,
Germany
bee@informatik.uni-augsburg.de

Elisabeth André
Institute of Computer Science
University of Augsburg,
Germany
andre@informatik.uni-augsburg.de

Thurid Vogt
Institute of Computer Science
University of Augsburg,
Germany
vogt@informatik.uni-augsburg.de

Patrick Gebhard
DFKI
Embodied Agents Research
Group, Germany
patrick.gebhard@dfki.de

ABSTRACT

This paper describes a system to enhance the interaction between humans and virtual characters with emotional mimicry and role-taking. Such system increases the believability of virtual agents. Mimicking necessitates a model of emotional intelligence to understand and display user's emotions. A more complex processing is however necessary for a reactive behavior, where the virtual agent reacts e.g. in an encouraging way which allows to actively change user's current state.

A virtual character with highly expressive capabilities was created to create a platform to figure out the differences in mimicking and role-taking. As we will concentrate on non-verbal behavior input from users, our agent will not be able to understand what, but how users are speaking.

1. INTRODUCTION

A virtual companion that stays for a long period of time with a user and that learns and knows about the preferences and wishes of its owner continuously, requires emotional intelligence that allows it to observe, to estimate and to manage its and the others emotions. It should be capable to detect users' affective state and respond appropriately to it in real-time [8]. In one situation, it might help if the agent shows that it feels with the user by simply mirroring the user's emotional state. This mimicry (i.e. parallel empathy) is the capability to display the user's emotion in a similar manner to the user's current emotional expression. In contrast, reactive behavior aims to understand the user's affective state and tries to alter or enhance it.

Assume, for example, a situation where your best friend fails an exam and tells this to you. This exam was very important for your friend as it decides about graduating. You might now react in two ways. One would be to completely empathize with your friend about the failed exam. Fortu-

nately you are an attentive listener and finally your friend already feels better, true the motto: a problem shared is a problem halved. The other possibility for you to react, is to understand the current situation and you start to encourage your friend by, for instance, explaining that you also failed some exams but there is always a second chance. In cognitive science, theory of mind enables a person to infer from the users' verbal and non-verbal behavior what they intend to do, desire, think or belief.

While a lot of work has been done in creation of affective output for virtual characters, less work was done in combining the recognition of user's affective state with the affective display of current systems for embodied conversational agents.

Prendergast and colleagues [18] developed an empathic companion that accompanies a user in a virtual job interview. This system measures users' physiological state (skin conductance and electromyography) in real-time and interprets it as emotion. The virtual agent reacts with empathic feedback dependent on the users' current affective state. The reaction is calculated with a Bayesian net, that takes the physiological state and user's job interview answers into account. The Bayesian net is modeled after findings in literature. Gratch et al. [8] describe a system for rapport in human-machine dialogs. They detect speech and head orientation from the user to create continuing dialogs with their system. The speech detection is used for detecting backchannel opportunity points, disfluencies, questions and loudness and the head tracker detects head nods, shakes, gaze shifts and posture shifts. Their system is mainly meant to create contingent nonverbal behavior and neither takes the user's emotional state into account nor intends to react with affective behavior. McQuiggan et al. [11] created a system for empathic virtual agents. They compare parallel with reactive empathy behavior. Using a machine learning approach let's them automatically learn from the users' behavior. The system is trained by users' in-game behavior. Thus, it does not consider the real users' actual emotional state. Ochs et al. [15] focus on creating an empathic model for virtual agents with a combination of analyzing real human-machine dialogs and theoretical descriptions of emotions. This agent

Cite as: , , *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, Decker, Sichman, Sierra and Castelfranchi (eds.), May, 10–15, 2009, Budapest, Hungary, pp. XXX-XXX.

Copyright © 2008, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

helps the users to obtain empathic information about their emails. Their corpus was annotated taking vocal and semantic cues into account. Boukricha [2] proposes a computational model of affective theory of mind for empathic virtual agents. She defines affective theory of mind by sharing their emotions (mimicry) and by understanding the counterparts' emotions cognitively (role-taking). Detection of the users' current emotional state, which will be mapped to a virtual character, is planned by using facial feature detection. There is an ongoing project named SEMAINE [20] that deals with building Sensitive Artificial Listeners which are designed to sustain the communication with a person. Their system will recognize and generate non-verbal behavior in real-time.

Our objective is the creation of an empathic listening agent that analyzes the user's emotive state and responds to it in real-time. To realize such an agent, we have been experimenting with several metaphors, such as that of a virtual pet or that of a virtual butler. While the virtual pet is not able to verbally respond to the user's state, the virtual butler gives both verbal as well as non-verbal feedback. Imagine the user has a rather bad day and is talking to the agent with a depressed voice. In the case of parallel empathy, the agent would simply mimicry the user's emotive state and show depression as well (see Fig. 1 left). In contrast to that, reactive empathy requires the agent to decide which emotion to display as a response to the user's emotive state. Here, the agent's emotions does not necessarily coincide with the user's emotions. That is the agent might, for example, decide to cheer the user up by showing encouragement (see Fig. 1 right). The agent is also able to provide simple verbal feedback, such as "I know how you feel!" or "That is really awful!".

2. THEORY OF MIND

Theory of mind is the cognitive ability to understand what others intend to do or think. It enables us to interpret the counterpart's behavior. Furthermore, it allows us to assume or predict what our counterpart intends, desires and believes. Such characteristic is essential for a virtual agent with emotional intelligence. As we are particularly interested in the interaction between real and virtual world (human-agent), a model that reflects the real and virtual world in an agent's mind is necessary. The attentive capability of our agent model will include both worlds. In contrast to the affective cognitive model, that will be limited to the real world and the user only, as our empathic listener is currently alone in its virtual world.

Children develop a theory of mind with 3-5 years. Typical tests for humans to detect the capability of theory of mind are the appearance-reality or false-belief task. A cognitive model that passes the latter task was, for example, implemented by Bringsjord [3]. As our virtual agents do not have to understand false-beliefs, we will simplify our theory of mind and build the cognitive ability of our virtual agent on an affective theory of mind for mirroring users' emotions and for being aware of users' emotions, which will allow us to react on users' emotions.

Boukricha [2] describes Affective Theory of Mind as a model that shares emotions (being able to mimic a person's emotion) and that understands a person's emotions (being able to alter a person's emotion). The first part of this model's behavioral pattern is innate and the expression of emotional feedback is involuntarily. Such feedback behavior

does not need a high level of cognitive capabilities. Whereas, for the second part of the model for ATOM, i.e. to react with pity or sympathy to the users' emotional state, our system needs a higher level of processing. The virtual agent shall understand in what emotional state the user is to react in an appropriate way. So it necessitates, when the input components detect e.g. sadness from the user that the current state in the emotional model of our virtual character shifts to something appropriate for 'pity for'. This lets our virtual character display the correct emotion for interacting with a user.

3. AFFECTIVE INTERACTION

There are many emotional models around, e.g. EMA [7]. We use ALMA because it describes how emotions evolve over time. We combine the component for affective sensory input with an emotional model for our empathic listener that allows it to act or react to the users' feelings. Although the feedback as listener is limited in a way, timing and understanding the user becomes crucial. Our system detects the users' emotions from voice via the tool EmoVoice [22]. Further, we use a realistic virtual character with highly expressive facial emotions to display appropriate facial expressions.



Figure 1: The virtual character Alfred is designed utilizing FACS to compose facial expressions.

Our architecture provides currently components to sense emotional states from the user using a microphone, a component to process affective states for mimicry or role taking and a component to display affect with virtual characters.

3.1 Affect Sensing

For sensing affect from users' voice, we use EmoVoice. It is a framework that provides support for the acquisition of emotional speech corpora and the training of classifiers. Furthermore, it is suitable both for offline as well as online vocal emotion recognition. The framework is intended to be used by non-experts and therefore comes with an interface to create an own personal or application specific emotion recognizer [22].

3.2 Affect Model - ALMA Bio

For the affect simulation in real-time, we rely on *ALMA Bio* and extended version of the computational model *ALMA*

[6]. ALMA Bio allows processing of bio signals. In this context bio signals are treated as unspecific emotions that contains individual (measured) pleasure, arousal and dominance values. Bio signal emotions are used as input to change or intensify the current mood.

ALMA provides three affect types as they occur in human beings: (1) *emotions* reflect short-term affect that decays after a short period of time; (2) *moods* reflect medium-term affect, which is generally not related to a concrete event, action or object; and (3) *personality* reflects individual differences in mental characteristics and affective dispositions.

ALMA implements the cognitive model of emotions developed by Ortony, Clore and Collins (OCC) [16] combined with the *BigFive* model of personality [10] and a simulation of mood based on the PAD model [12]. The relations between the different affect types is an central part of the affect simulation:

- *A given personality* defines a default mood and influences the intensities of different emotions.
- *The current mood* amplifies or dampens the intensities of emotions.
- *Emotions* as short term affective events influence the longer-term mood.

Elicited emotions influence an individual’s mood. The higher the intensity of an emotion, the higher the particular mood change. Emotions usually do not last forever. Over a specific period the intensity of emotions decays and the influence on the current mood fades.

The current mood also influences the intensity of emotions [13]. This simulates, for example, the intensity increase of *joy* and the intensity decrease of *distress*, when a individual is in an *exuberant* mood. Mood is represented by a triple of the mood traits pleasure (P), arousal (A) and dominance (D). The mood’s trait values define the mood class. If, for example, every trait value is positive (+P,+A,+D), the mood is *exuberant*.

3.3 Alfred – Affect Display

The affective display of our virtual agent consists of an enriched set of facial animations. “Alfred”, Our realistic looking virtual character is able to talk with a text-to-speech (TTS) system or by playing prerecorded audio files.

3.3.1 Facial Expression

Ekman and Friesen developed the Facial Action Coding System (FACS) to classify human facial expressions [4]. FACS divides the face into action units (AU) to describe the different expressions a face can display (e.g. inner brow raiser, nose wrinkler, or cheek puffer). Although FACS was originally designed to analyze natural facial expressions, it turned out to be usable as a standard for production purposes too. That is why FACS based coding systems are used with the generation of facial expressions displayed by virtual characters, like Kong in Peter Jackson’s King Kong [19]. But the usage of FACS is not only limited to virtual characters in movies. The gaming industry with Half-Life 2 by Valve, also utilizes the FACS system to produce the facial expressions of their characters [21].

Alfred (see Fig. 1), a butler-like character, uses these action units to synthesize an unlimited set of different facial

expressions. The action units were designed using morph targets and thus gives the designer the full power in defining the facial expression outlook. The system includes a tool to control the single action units. The tool allows to store the result in a XML file for later usage in our agent system [1].

We chose the FACS-based approach for our facial animation system, because of the Facial Expression Repertoire (FER) [5], which maps over 150 emotional expressions to the action units of FACS. Not only does it explain in detail, which action unit must be activated for certain facial expressions, it further provides a rich dataset of videos which show how the action units ought to be designed.

Alfred’s mesh has a resolution of about 21.000 triangles. For displaying more detailed wrinkles in the face, normal maps baked from a high-resolution mesh are used [14]. The morph targets for the action units are modeled using the actor’s templates from the FER. For rendering the character and its animations the Horde3D GameEngine [9] is used.

3.3.2 Speech

The system interfaces the Microsoft Speech API to synchronize the audio output with the lip movements. This allows us to use any text-to-speech that supports SAPI 5. As the quality of common TTS systems may not be satisfactory, we integrated a module to synchronize prerecorded audio speech files with the lip movements of the virtual character. This allows us to use highly emotional sentences or affect bursts to be spoken through a virtual character. As FACS defines several action units involving mouth muscles (e.g. lip funneleer, lip tightener, mouth stretch), we utilize the FACS system for lip movements. The approach is similar to displaying facial expressions. The output from the editor to modify the single action units is stored in a XML file. Reusing the FACS approach for visemes enables Alfred to display facial expressions and lip movements parallel.

4. EMPATHIC FEEDBACK

We designed an empathic model which is responsible to interact with mimicry and to react with reactive empathy. While the model for parallel empathy goes along with the emotional model in ALMA, the model for reactive empathy needs some further discussion. Both systems generate the facial expression by picking up the according facial expression dependent on the current emotional state from an emotional model.

4.1 Parallel Empathy

For mimicry (empathy or emotional contagion) our system (see Fig. 2) does not need a complex model to understand and interpret users’ emotions. It is sufficient to map users’ recognized emotions directly to the affective display of a virtual character. We created an emotional model in ALMA to represent the user’s emotion in our system. We simply take the current state in the model and display it with the virtual character to mirror the user’s emotional state. EmoVoice maps the classification result directly into ALMA. The same mapping between PAD values and the facial animation system is used to display emotional expression with Alfred, our virtual character. While in parallel empathy, we use ALMA to describe how emotions evolve over time and simply use it to “store” the current user’s emotional state.

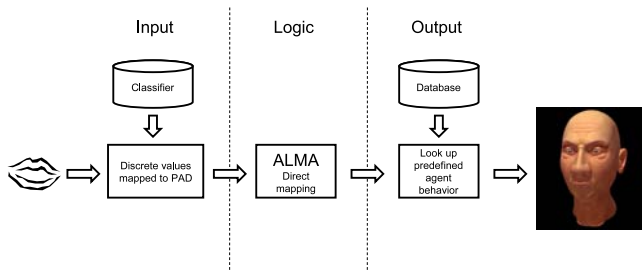


Figure 2: Model for mirroring

4.2 Reactive Empathy

Here, the agent does not display emotions it feels, but tries to express deliberately emotions that might help to put the user into a better emotional state. Reactive empathy requires the agent to put himself into the shoes of the user in order to decide on an appropriate emotional display. For example, if the user is afraid of failing in an exam, the agent might tell the user that he will manage due to his good preparation. That is the agent praises the user which is perceived by the user as a positive event. In addition, the agent utters a statement which decreases for the user the likelihood of the negative event (namely failing in the exam). Further, in reactive mode, it is crucial for the virtual agents to be able to utter vocally. Especially if its emotional display does not coincide with the user's emotion. For example, if the user is sad and the agent is smiling, the user might not perceive this as encouragement, but as gloating (see Fig. 1 right).

5. CONCLUSION

While developing the current architecture, we came up with some issues that need to be further discussed.

Reactive empathy cannot be modeled so far by ALMA Bio. While ALMA is at least capable to analyze the user's state. For example, is the user sad, is it an unpleasant event. Is the user the agent's friend, will the agent feel "pity for". But is the user the agent's enemy, the agent will feel gloating. In this case the reaction would not deal with empathy.

Another not trivial issue is the decision when a virtual agent with the ability of empathic reasoning should respond parallel or reactive [11]. Our system provides the possibility to test different behavior variants to figure out when a parallel empathic reaction or a reactive empathic reaction for a virtual character in a human-machine interaction is appropriate.

Conversational feedback is missing in our approach. Nevertheless, it is essential in human-human interaction to keep a conversation running and to show understanding to what somebody says. A virtual agent that is not able to show that it is following the conversation with giving appropriate feedback to the users will hardly be accepted as an empathic listener. Engender rapport with a virtual agent can be as effective as a human listener [8]. Feedback without understanding the content of what is said is called envelope feedback. Although our agent will not understand what users say, a method for responding in an appropriate manner is necessary to keep a conversation running. Such signals must transmit engagement, interest, understanding, agreement and of course emotional feedback [17].

6. ACKNOWLEDGMENTS

This work has been funded in part by the European Commission via the CALLAS Integrated Project. (ref. 034800, <http://www.callas-newmedia.eu/>).

7. REFERENCES

- [1] N. Bee, B. Falk, and E. André. Simplified facial animation control utilizing novel input devices: A comparative study. In *International Conference on Intelligent User Interfaces (IUI '09)*, pages 197–206, 2009.
- [2] H. Boukricha. A first approach for simulating affective theory of mind through mimicry and role-taking. In *The Third International Conference on Cognitive Science, Symposium: Emotional Computer Systems and Interfaces*, 2008.
- [3] S. Bringsjord, A. Shilliday, M. Clark, D. Werner, J. Taylor, A. Bringsjord, and E. Charpentier. Toward logic-based cognitively robust synthetic characters in digital environments. In *Proceedings of the First Conference on Artificial General Intelligence (AGI-08)*, 2008.
- [4] P. Ekman and W. Friesen. *Unmasking the Face*. Prentice Hall, 1975.
- [5] Facial Expression Repertoire. Filmakademie Baden-Württemberg. <http://research.animationsinstitut.de/>.
- [6] P. Gebhard. Alma - a layered model of affect. In *Proc. of the 4th Int. Joint Conference on Autonomous Agents and Multiagent Systems*, pages 29–36. ACM, June 2005.
- [7] J. Gratch and S. Marsella. A domain-independent framework for modeling emotion. *Cognitive Systems Research*, 5(4):269–306, December 2004.
- [8] J. Gratch, N. Wang, J. Gerten, E. Fast, and R. Duffy. Creating rapport with virtual agents. In *Intelligent Virtual Agents (IVA 2007)*, pages 125–138, 2007.
- [9] Horde3D GameEngine. University of Augsburg. <http://mm-werkstatt.informatik.uni-augsburg.de/projects/GameEngine/>.
- [10] R. McCrae and O. John. An introduction to the five-factor model and its applications. *Journal of Personality*, 60:175–215, 1992.
- [11] S. W. McQuiggan, J. L. Robison, R. Phillips, and J. C. Lester. Modelling parallel and reactive empathy in virtual agents: An inductive approach. In *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, pages 167–174, 2008.
- [12] A. Mehrabian. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology: Developmental, Learning, Personality, Social*, 14:261–292, 1996.
- [13] R. Neumann, B. Seibt, and F. Strack. The influence of mood on the intensity of emotional responses: Disentangling feeling and knowing. *Journal of Cognition & Emotion*, 15(6):725–747, 2001.
- [14] C. Oat. Animated wrinkle maps. In *SIGGRAPH '07: ACM SIGGRAPH 2007 courses*, pages 33–37, New York, NY, USA, 2007. ACM.
- [15] M. Ochs, C. Pelachaud, and D. Sadek. An empathic virtual dialog agent to improve human-machine

- interaction. In *7th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, pages 89–96, 2008.
- [16] A. Ortony, G. L. Clore, and A. Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, Cambridge, MA, 1988.
 - [17] C. Peters, C. Pelachaud, E. Bevacqua, M. Ochs, N. E. Chafai, and M. Mancini. Social capabilities for autonomous virtual characters. In *International Digital Games Conference*, pages 37–48, 2006.
 - [18] H. Prendinger, H. Dohi, H. Wang, S. Mayer, and M. Ishizuka. Empathic embodied interfaces: Addressing users’ affective state. In *In Proceedings Tutorial and Research Workshop on Affective Dialogue Systems, LNAI 3068*, pages 53–64, 2004.
 - [19] M. Sagar. Facial performance capture and expressive translation for king kong. In *SIGGRAPH ’06: ACM SIGGRAPH 2006 Sketches*, page 26, New York, NY, USA, 2006. ACM.
 - [20] M. Schröder, R. Cowie, D. Heylen, M. Pantic, C. Pelachaud, and B. Schuller. Towards responsive sensitive artificial listeners. In *Fourth International Workshop on Human-Computer Conversation*, Bellagio, Italy, 2008.
 - [21] Valve. Facial Expressions Primer from Half-Life 2 by Valve. http://developer.valvesoftware.com/wiki/Facial_Expressions_Primer.
 - [22] T. Vogt, E. André, and N. Bee. EmoVoice - a framework for online recognition of emotions from voice. In *Proceedings of Workshop on Perception and Interactive Technologies*, 2008.

Concepts and Evaluation of Psychological Models of Empathy

Sibylle Enz

Otto-Friedrich-Universität Bamberg
Kapuzinerstraße 16
D-96045 Bamberg
+49 (0)951 863 1958

sibylle.enz@uni-bamberg.de

Carsten Zoll

Otto-Friedrich-Universität Bamberg
Kapuzinerstraße 16
D-96045 Bamberg
+49 (0)951 863 1965

carsten.zoll@uni-bamberg.de

Martin Diruf

Otto-Friedrich-Universität Bamberg
Kapuzinerstraße 16
D-96045 Bamberg
+49 (0)951 863 1964

martin.diruf@uni-bamberg.de

Caroline Spielhagen

Otto-Friedrich-Universität Bamberg
Kapuzinerstraße 16
D-96045 Bamberg
+49 (0)951 863 1956

caroline.spielhagen@uni-bamberg.de

ABSTRACT

This paper provides an overview over contemporary empathy research, including concepts and definitions as well as descriptions of empathic processes and outcomes. Based on these theoretical foundations, three different approaches to model empathy are described: a low-level computational approach, an OCC-based approach, and an empathy model inspired by PSI, a general psychological theory of psychic functioning. Ideas on how these models could be implemented in agents are discussed and preliminary efforts to evaluate the plausibility and believability of the empathic processes and outcomes are drafted.

Categories and Subject Descriptors

D.3.3 [Programming Languages]: none

General Terms

Theory

Keywords

Empathy, Psychological Modeling, Evaluation

1. INTRODUCTION

In the field of social and emotional learning and intercultural education, virtual learning environments provide users with the opportunity for learning in a safe environment that is inhabited by emotionally expressive, autonomous agents (e.g. FearNot! [1]). Social and emotional learning with such agents is allowed for through empathic reactions in the user towards the virtual agents on the screen, a reaction that is enforced by the emotional expressivity of the agents. However, the true power of social relations towards artificial entities (such as agents in virtual worlds or as robots in the real world) can only be discovered if we manage to provide the user or learner with companions that

show interest in the user and react sensitively towards their needs and intentions, hence, that react empathically towards the user.

1.1 Empathy concepts and definitions

Empathy is defined by contemporary researchers as a construct that comprises two components: affective and cognitive aspects. While some researchers embrace both aspects in their empathy definitions [2,3], others emphasize either the one or the other, e.g. according to Hogan [4] "...empathy means the intellectual or imaginative apprehension of another's condition or state of mind without actually experiencing that person's feelings..." (cognitive empathy), whereas Hoffman [5] posits that "...empathy [is] a vicarious affective response to others..." (affective empathy). For the present study, we want to define empathy as an observer's understanding of the internal state of a target (cognitive empathy) as well as the observer's emotional reaction to what he/she perceives as being the internal state of a target (affective empathy).

Cognitive empathy means, the observer has to focus his/her attention on the target, reading expressive signals as well as situational context cues, and to try to understand – based on what he/she knows about emotional expressions in general, meanings of situations in general, and previous reactions of the target – the current reactions of the target. In general, for the empathic reaction to even start, the observer needs to be motivated and able to perceive and interpret correctly the expressive and situational cues indicating the reaction / internal state of the target. To be able to do this, the observer needs knowledge about emotional states and other reactions, how they are expressed, and what elicits them, and he/she needs to either know the target person in order to understand his/her internal state or perceive the target person as similar to themselves.

Affective empathy relates to the general way of how emotions emerge in a person. In the case of affective empathy, the emotions in the observer emerge due to the (conscious or unconscious) perception of internal states in a target (either emotions or thoughts and attitudes). Affective empathy thus can be the result of cognitive empathy, but can also grow out of the perception of expressive behavior that immediately transfers emotional states from one individual to another (emotional contagion). In this case, qualitatively highly similar affective states are evoked in the observer, resulting from a direct link or transfer of emotional states between individuals through verbal, para-verbal and non-verbal cues. This mechanism serves the biological function of fostering social identity and adaptation to the group, e.g. when it is vital for a herd of animals to react quickly to a predator that is only detected by one or few members of the group. In case of reactive affective empathy emerging due to cognitive (empathic) processes, a more complex conglomeration of affective states (like gloating) may result as opposed to the highly similar emotional states that result from emotional contagion.

1.2 Empathic processes and outcomes

Another important conceptual distinction is made between internal processes involved in empathy and the outcomes of these empathic processes. According to Davis [3], empathic outcomes have to be distinguished from processes that are “empathy-related, because they frequently occur during episodes in which an observer is exposed to a target, and because they often result in some empathy-related outcome” (p. 15). However, these processes are not specific for empathy; they occur in other contexts as well and can then also produce other but empathic outcomes. Empathic outcomes can be further divided in intra- and interpersonal. Referring to Hoffman’s developmental theory of empathy [11], Davis distinguishes between non-cognitive, simple cognitive, and advanced cognitive processes that can be involved in an empathic episode.

Non-cognitive processes These processes rely on the direct link between emotional states perceived in a target and the evocation of according or similar emotional states in the observer as described above. This direct, pre-reflexive and pre-verbal link can be observed very early in the human development, e.g. as “*primary circular reaction*” of newborns that cry if they perceive the crying of other infants. Also, imitation of simple expressive gestures (or *motor mimicry*) can create an according emotional state (see James / Lange theory on emotion [12]) which can be interpreted as a rudimentary form of empathy in very small children. Although empathic abilities improve with the development of cognitive abilities in the child, motor mimicry can also be part of the empathic experience in later life.

Simple cognitive processes Due to progressing cognitive development, more and more complex cognitive processes can add to the empathic experience. First, *classical conditioning* in a given situation or event allows for reinforcing affective reactions when the observer is simultaneously experiencing an emotion evoking situation (UCS) and an intense emotional expression of a target. The perceived emotional expression of the target can serve as a conditioned stimulus later (CS), thus leading to the activation of the emotion in the observer, even in other

situations; e.g. a toddler on her father’s arm in an emotion-arousing situation. Related to this process is *direct association*, a process of associating perceived expressive or situational cues of a target with memory representations of similar expressions or situations experienced earlier by the observer, eventually resulting in similar affective states in the target and the observer. During the very similar process of *labelling* simple representations about the meaning of situations or events are used to infer the internal state of a target experiencing this situation or event (e.g. a funeral implies for people to feel sad).

Advanced cognitive processes On top of the rather simple associative processes described above, associations can also be triggered by *language expressions*, e.g. witnessing a target saying “I’ve been laid off” alone suffices to trigger an understanding and maybe even the associated feeling of somebody who has been laid off (even in the absence of nonverbal gestures; this mode is working when empathizing with fictional characters, e.g. when reading a book). Also *elaborated cognitive networks* are at work when it comes to interpreting other situational cues, apart from language. Both processes rely on feelings and experiences the observer has acquired before being faced with the language or situational cues that trigger empathy. The most advanced cognitive process involved in empathy is *role-taking*, “the attempt by one individual to understand another by imagining the other’s perspective” ([3] p. 17). It involves not only associations to own feelings or experiences collected in the past, but also the effortful suppression of the egocentric perspective and the willingness to experience the situation or event explicitly from the target’s perspective. Hence, it is the only process involved in empathy that lives up to the criterion of consciously distinguishing the Self from the Other and can be regarded as the most mature and developed empathic process.

While the empathic processes can be interpreted as stages in the development of empathy, with role-taking developing latest, all processes can be part of an empathic experience in later life, e.g. processes of emotional contagion, association with memory representations and role-taking may all result in a complex and rich empathic experience within the observer. Also, the single processes may have an impact on each other. Even though there is a lack of empirical investigations into the interactions of different processes that contribute to an empathic episode, it is highly plausible to assume that more than one of them can operate simultaneously. Regarding the outcomes of the empathic processes described above, Davis distinguishes between intrapersonal and interpersonal outcomes [3]. Interpersonal outcomes can be influenced by intrapersonal empathic outcomes.

Intrapersonal outcomes Intrapersonal outcomes are changes in the internal state of the observer that can be either affective or non-affective. Affective outcomes are emotions that emerge in the observer, and can be either parallel or reactive in nature. *Parallel affective outcomes* produce the same or similar emotion as the emotion of the target, e.g. through motor mimicry, whereas *reactive affective outcomes* rely on associative and role-taking processes and merge with own reactions to the perceived situation and reaction of the target (the resulting affective states can be a blend of different emotions rather than an actual copy of the target emotion, e.g. personal distress, sympathy, or gloating). Non-affective outcomes are e.g. the *accuracy* with which the

observer perceives the situational and expressive cues, and the resulting quality of the assessment of thoughts, feelings, and attitudes of the target [13]. According to [3] empirical evidence suggests that similarity and familiarity between observer and target play an important role in interpersonal accuracy. Additionally, the reasons for the target's behavior that the observer attributes to the target are influenced by empathic processes: what has been termed actor-observer-difference describes the empirical finding that one usually refers to situational forces to explain one's own behavior (particularly if the behavior is not successful) while observers tend to explain the behavior of others with the help of personality characteristics or traits [14]. Empathy influences these tendencies by resulting in more *actor-like attributions* (referring to situational forces) in the observer; again, similarity, familiarity, and also sympathy or affection for the target person are additional factors that influence attribution biases apart from empathy.

2. THREE DIFFERENT APPROACHES TO MODELING EMPATHY

distinguished regarding whether the consequences of an event impact the agent itself (desirability for the self) or other agents (desirability for others). For example, when someone wins in the lottery, it is desirable for them, but won't necessarily affect others. Ortony et al. [15] posit that different appraisals lead to qualitatively different types of emotions; figure 1 outlines the appraisals and the resulting emotions for the appraisal of events. Some of these emotions can be interpreted as affective outcomes of empathic processes (happy-for, resentment, gloating and pity). The cognitive empathic processes are the appraisals of events regarding the consequences for the others.

Figure 1. Example appraisals from the OCC model of emotions [15].

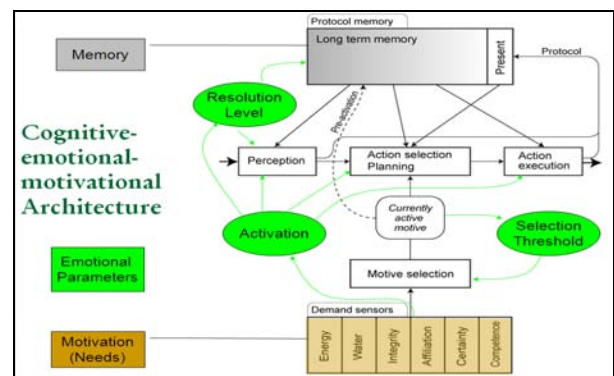


Figure 2: PSI model of the human psyche [7], [8].

Thus, emotions serve as quick adaptations of the organism to a specific situation. E.g. fear is experienced under conditions that produce high need for certainty and competence and – as a result – is characterized by a high arousal level (preparedness for quick reaction), low resolution level (inaccurate perception and planning), and low selection threshold (organism is easily distracted by other cues within its environment in an attempt to detect dangers in it). It is the model of emotion that is embedded in a broad and comprehensive model of action regulation that makes the PSI approach particularly interesting when it comes to modeling empathy. Due to the “perception” of parameter settings in the other agent, a similar emotional state in the empathic agent can emerge, taking the pre-empathic state of the empathic agent into consideration (affective empathy). Knowledge about the internal state of another agent (cognitive empathy) is acquired through the model’s learning mechanism.

The third approach is a simple “if-then” computational approach, sparing computing capacity by being based on structures or processes that are already implemented or that need implementation in any case (see fig. 3). Given that the emotional states of the agents can be described by some type of “emotional parameters”, the empathic agent adopts the emotional parameters of the other agent using an attenuation factor (affective empathy). The empathic agent’s emotion resulting from the empathic process is a mixture of the two agents’ emotional states. Knowledge about the feelings of another agent in a given situation (cognitive empathy) is implemented through “if... then...”-relations. The difference to the OCC-model is that this approach specifies how information about another person’s internal states are stored in and retrieved from memory whereas the OCC model describes the actual process of “reasoning” about the internal state of another person.

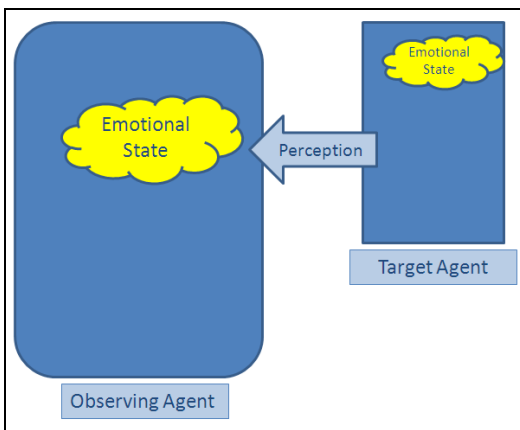


Figure 3: Low-level approach to modeling empathy.

3. MODEL EVALUATION

3.1 Text-based approach

It was decided to take a text-based approach to evaluate the empathic outcome of the models, i.e. we produced text-based outputs for the models in an iterative approach: First, the expert

answers by Dr. Dr. Rainer Erlinger, an expert to moral dilemmas who regularly gives advice to the readers of a weekly German magazine [9], were reviewed. Second, we carefully selected four questions from readers seeking advice that allowed for the emergence of emotions and used them as scenarios for the evaluation (cinema, hair stylist, car parking, and antenna). In the following, two examples are provided¹:

“I lately went to the cinema, where only few viewers wanted to watch the movie. Shortly after the beginning of the film, a man sat down on the seat right beside me. I felt upset but didn’t have the heart to change the seat because I didn’t want to be rude. In the end, I felt angry during the whole movie. Was my behavior polite or rather foolish?” (cinema)

“One year ago, my relationship to my boyfriend ended in a terrible way, after I found out that he has been cheating on me for years. He now gave me three gift coupons for my incredibly expensive hair stylist as a birthday present. Even though I don’t have as much money as he does, I didn’t want to benefit from the voucher –because I felt too proud to do so. My hair stylist deemed me totally crazy, especially because my ex-boyfriend already had paid for the vouchers, and convinced me to use them. Now my haircut is amazing, but every time that I look in the mirror, I can’t feel happy about it. I always ask myself whether I am bribable or not. Should I maybe forfeit the remaining vouchers?” (hair stylist)

The two remaining scenarios discuss the potential dangers of a radio antenna on a family home and whether the reader should mention them to her friend who has recently moved in with her children (antenna) and the waiting inside the own car on a public car parking in the highly frequented area in front of the main station, blocking the parking space for others (car parking).

Third, two trained psychologists adapted the answers of Dr. Dr. Erlinger separately to the model conceptions outlined above. This was done by applying the empathy processes and outcomes, as specified by the respective models, to the answers the expert provided in the newspaper, thus changing them slightly and ending up with four “expert” answers for each of the four scenarios. The two psychologists then discussed their respective solutions and adapted them iteratively until they negotiated a common solution for each scenario-model-combination. These scenarios served as the pool of relevant model “behavior” for the evaluation study.

To reduce inter-model variance that was not produced by the differences between the models but by different use of language when adapting the answers, only those words and phrases of the original answer by Dr. Dr. Erlinger were changed that were directly related to empathic processes or outcomes (e.g. speculating on the thoughts and feelings of others, discussing emotions, intentions, needs, wishes, etc.); all other phrases and passages were kept identical between the adaptations of one scenario. This resulted in highly similar adaptations with only marginal variance produced by the models which is illustrated by the following answer displaying the OCC-inspired version of Dr.

¹ Translated into English by the authors.

Dr. Erlingers answer to the cinema scenario (original expressions in brackets):

“To ostentatiously change seats in a cinema is a bit unfriendly. However, there is no need to *feel shame* as long as the other person was unfriendly towards you. [Dr. Dr. Erlinger: „That would be justified if the other person was unfriendly towards you”]. Was this the case? In fact, the other person just chose a seat in which to sit, a right earned by purchasing the ticket. But the mere right is not what is at stake here. The problem here is one of personal space or interpersonal distance: each person has an area around themselves, marked by invisible borders which should not be violated by strangers. The classic reference in behavioral research is “The Hidden Dimension” by Edward T. Hall. Hall introduced the term “proxemics” for the study of perceived behavior in space. He distinguished between several zones, e.g. intimate zone, personal zone, and public zone. The situation in a cinema falls into the personal distance (0.45-1.20 m) which can be intruded only if both interaction partners want to be close or if there is a limited amount of space. Both conditions are not true in the present situation. Among the functions of personal space that have been discussed recently are: keeping control over one’s freedom to act and communicate, maintaining a feeling of safety (from threats or distress), and the possibility of retreat and recreation. Thus, personal space ensures a feeling of contentment which was impaired by your neighbor. Despite the apparent tranquility, he acted not just somehow strange, but *distressed you* like a true provocateur transgressing your personal boundaries [Dr. Dr. Erlinger: „...as a true provocateur attacked you with socially unacceptable behavior”]. You do not have to endure this. Faced with such insolence, it seems remarkable that you have managed to control yourself. To silently change seats is not impolite, but rather noble.”

Due to the relatively high inter-scenario similarity, three decisions were made regarding our evaluation design:

- We decided to include the original answer provided by Dr. Dr. Erlinger as a baseline.
- We decided against providing each participant of our evaluation study with all four scenarios (total workload of 16 scenarios: three adaptations plus the original answer for each of the four scenarios).
- Instead, we decided to randomly assign the 3+1 model adaptations to the scenarios and include two experimental groups with different assignments (see table 1).

Table 1: Scenarios rated by the two experimental groups.

Scenario	group A (N=12)	group B (N=14)
Cinema	Dr. Dr. Erlinger	OCC-inspired approach
Hair Stylist	PSI-inspired approach	Dr. Dr. Erlinger
Antenna	OCC-inspired approach	Low-level approach
Car Parking	Low-level approach	PSI-inspired approach

3.2 Sample and Procedure

26 subjects (21 female, aged 20 – 44 yrs., M = 26 yrs.) were asked to imagine themselves in the role of a newspaper editor who wants to hire an expert for a moral-dilemma-column such as the one in [9]. They were then exposed to four “as-if”-answers to given moral dilemmas of applicants to the job that they had to rate on a list of adjectives in order to assess their qualification for the job. For the ratings, we used a German adaptation of Davis, Luce and Kraus’ adjective list [10] to assess empathy; the resulting empathy scale’s internal consistency in the present study was $\alpha = .93$. As described above, from the 16 possible answers two sets of scenario-model-combination were chosen, of which 12 of the subjects rated one, and 14 of the subjects rated the other one (see table 1). The cover story, dilemmas, answers, and adjective lists were all presented in electronic format.

3.3 Evaluation Results & Discussion

The results of the evaluation suggest that the differences between the scenarios cause differences in the empathy rating, not the underlying model characteristics. A two-way ANOVA yielded a significant main effect for scenario ($F=43.24$; $df=3$; $p=.000$), but neither a significant main effect for model nor an interaction effect. Particularly the answers to the cinema-scenario were rated significantly less empathic than the other answers, independent of the underlying model (see figure 4). Regarding the differences between the models, there are significant differences between the low-level model on the one hand which is rated with the highest empathy scores and the OCC-inspired model and the original answer by Dr. Dr. Erlinger on the other hand which obtain the lowest empathy scores ($t=2.77$; $p=.008$; $t=-3.47$; $p=.001$).

Obviously, while the design tried to carefully minimize variance produced within one scenario through the process of adapting the answers to the model approaches (e.g. by using different words, phrases, varying length of the answer), the variance produced by the differences among the scenarios overruled the subtle differences that were caused by the model adaptations. The scenarios seem to be differentially prone to evoke empathic reactions. While the moral dilemma of the cinema scenario discusses the intrusion of personal space in the situational context of leisure time, the antenna scenario analyses the dangers posed by a radio antenna for mobile phones on the roof of a family home in which a friend plans to raise her small children. Obviously, the latter scenario provides much more potential for emotional involvement than the first. Thus, the results of the present study provide valuable insight into which of the scenarios is more appropriate for empathy research than the others.

However, the main research question could not be answered with the present design due to the fact that (1) the scenarios produce much stronger differences in empathy ratings than the models at stake, and (2) there is no complete set of empathy ratings for all of the four models within one scenario. Definite conclusions about the qualitative difference between the empathic outcome of the models thus have to be addressed by more elaborate data collection, including more subjects and all three (four) competing models within each scenario. In order to reduce the work load for the participants (the main reason for the present design), only one scenario should be used and adapted to the three different models, e.g. “antenna” which seems most appropriate to trigger

empathy according to the present results. Thus, variance between the different scenarios would be controlled for. Furthermore, the approach would benefit from a supplementary validation of the text-based model implementation, e.g. through setting up a small group of trained experts to test whether the empathic “behavior” of the models can be reproduced.

Discussing the (insignificant) differences between the models, the low-level model yielded the highest empathy scores. This might be due to the relatively simple “if-then”-rules which could be easily implemented in the expert answers, providing these answers with an easy-to-read structure that was “rewarded” by the participants of our study. In contrast, the OCC-inspired approach is based on rather complex appraisals resulting in a set of numerous qualitatively different emotional states; only few of these emotional states were applicable to the present scenarios: the complexity of the model was obviously not recognizable for the participants of our study, given the limited character of the “behavior” they had to rate. Dr. Dr. Erlingers on the other hand, who writes the original expert answers in a weekly magazine, does not only provide empathy for the reader who sends in the dilemma, but also wants to entertain his readers, a motivation that might interfere at times with his display of empathy, as opposed to a face-to-face situation between e.g. a client and a consultant or a therapist.

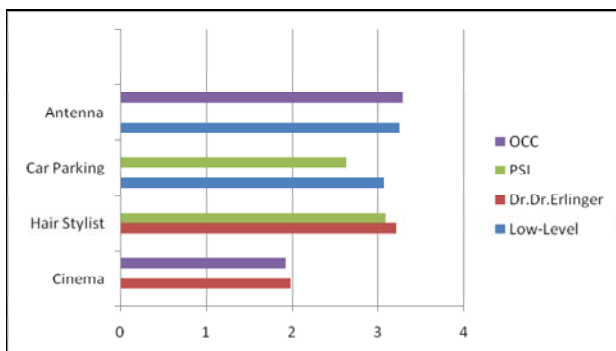


Figure 4. Empathy scores for models and scenarios.

In sum, the text-based evaluation yields differences in empathic outcomes, providing insights in the working mechanisms of different theoretical conceptions of empathy. The scenarios clearly differ in their empathic potential, contrasting topics from leisure time with the safety of a family home. However, taking into account the minimal model adaptations made to the answers by keeping all non-emotional content of the originals untouched, this methodology might be of further use in the forefront of programming agents if the model characteristics and differences would be implemented more clearly than in the present study.

4. ACKNOWLEDGMENTS

This work was partially supported by European Community (EC) and is currently funded by the eCIRCUS project IST-4-027656-STP as well as by the LIREC project, Grant agreement no. 215554. The authors are solely responsible for the content of this

publication. It does not represent the opinion of the EC, and the EC is not responsible for any use that might be made of data appearing therein.

5. REFERENCES

- [1] Paiva, A., Dias, J., Sobral, D., Silva, C., Aylett, R.S., Woods, S. & Zoll, C. 2004. Caring for Agents and Agents that Care: Building Empathic Relations with Synthetic Agents. In Proceedings of the Third International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS04).
- [2] Holz-Ebeling, F. & Steinmetz, M. „Wie brauchbar sind die vorliegenden Fragebogen zur Messung von Empathie? Kritische Analyse unter Berücksichtigung der Iteminhalte“, Zeitschrift für Differentielle und Diagnostische Psychologie 1995, 16, 11-32.
- [3] Davis, M. H. (1994). Empathy – a social psychological approach. Brown & Benchmark Publishers, Madison, Wis.
- [4] Hogan, R. “Development of an empathy scale”, Journal of Consulting and Clinical Psychology 1969, 35, 307–316.
- [5] Hoffman, M. L. 1977. Empathy, its development and prosocial implications. In Nebraska Symposium on Motivation, vol. 25, C.B. Kaesy, Ed. University of Nebraska Press, Lincoln.
- [6] Dias, J. and Paiva, A. 2005. Feeling and Reasoning: a Computational Model. In EPIA 2005. LNCS (LNAI), vol. 3808, C. Bento, A. Cardoso, G. Dias, Eds. Springer, Heidelberg.
- [7] Dörner, D 2001. Bauplan für eine Seele. Rowohlt, Reinbek.
- [8] Dörner, D 2003. The mathematics of emotion. The Mathematics of Emotions. In The Logic of Cognitive Systems – Proceedings of the Fifth International Conference on Cognitive Modeling (ICCM 2003), F. Detje, D. Dörner, D. & H. Schaub, Eds. Universitätsverlag, Bamberg.
- [9] Süddeutsche Magazin, DOI= <http://sz-magazin.sueddeutsche.de/hefte>.
- [10] Davis, M. H., Luce, C., Kraus, S. J., “The heritability of characteristics associated with dispositional empathy”, Journal of Personality 1994, 62, 369-391.
- [11] Hoffman, M. L. 1984. Interaction of affect and cognition in empathy. In Emotions, cognition, and behavior, C. Izard, J. Kagan, & R. Zajonc Eds. Cambridge University Press, New York.
- [12] James, W. 1890. The principles of psychology. Holt, New York.
- [13] Ickes, W., Stinson, L., Bissonnette, V., & Garcia, S. „Naturalistic social cognition: Empathic accuracy in mixed-sex dyads”, Journal of Personality & Social Psychology 1990, 59 (4), 730-742.
- [14] Jones, E. E. & Nisbett, R. E. 1972. The actor and the observer: Divergent perceptions of the causes of the behavior. In Attribution: Perceiving the causes of behaviour, E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins & B. Weiner Eds. General Learning Press, Morristown, NJ.

- [15] Ortony, A., Clore, G., and Collins, A. 1988. The cognitive structure of emotions. University Press, Cambridge.
- [16] Newell, A. 1990. Unified Theories of Cognition. Harvard, Cambridge, MA.
- [17] Anderson, J. R. "Spanning seven orders of magnitude: A challenge for cognitive modelling", *Cognitive Science* 2002, 26, 85-112.

Cite as: Concepts and Evaluation of Psychological Models of Empathy, Enz. S., Zoll, C., Diruf, M., *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, Decker, Sichman, Sierra, and Castelfranchi (eds.), May, 10–15., 2009, Budapest, Hungary, pp. XXX-XXX.

Copyright © 2009, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

Towards Empathic Touch by Relational Agents

Timothy Bickmore, Rukmal Fernando
Northeastern University College of Computer and Information Science
360 Huntington Ave, WVH 202
Boston, MA USA
+1-617-373-5477
{bickmore,rukmal}@ccs.neu.edu

1. INTRODUCTION

Empathy—the process of attending to, understanding, and responding to another person's expressions of emotion—is a prerequisite for providing emotional support which, in turn, is a key element for establishing most kinds of meaningful social relationships between people. Within healthcare, for example, provider empathy for the patient has been widely acknowledged as being an important prerequisite for the establishment of a therapeutic alliance relationship, which is associated with improved health outcomes [13]. Empathy alone can also be important: in physician-patient interactions, physician empathy for a patient plays a significant role in prescription compliance, and a physician's *lack* of empathy for a patient is the single most frequent source of complaints [10].

An essential element of empathic interaction is that the empathizer must clearly communicate their understanding of their partner's emotional state [17]. An important channel for communicating empathic understanding of distress is through physical touch as an acknowledgment of the distress and a message of comfort and caring.

We are developing a conversational agent that has the ability to touch the user at appropriate points in dialogue for the same reasons that people use this modality—to comfort, emphasize, or display or establish social bonds. One embodiment of such a “touchbot” would be a device that hospital patients can hold in their hospital beds, capable of sensing touch (squeezing, stroking, etc.) by the patient and able to use these same communicative signals in conjunction with a speech-based dialogue system for comforting, counseling, and educating the patient.

The importance of physical touch between a health provider and client in face-to-face interaction has been widely documented. For example, hospital patients who are touched by providers have been found to be more satisfied with their experience overall compared to non-touched patients [9]. Touch has also been found to be effective for providing comfort for terminally ill older adults [4] and effective in improving pain and mood in patients with advanced cancer [14]. Health providers—nurses in particular—have been found to frequently use comforting touch with patients. One study of 30 critical care nurse-patient dyads in a hospital setting found that caring touch was used by the nurses twice per

hour on average (with a range of 0-17) [18].

Additional therapeutic forms of touch, such as massage, have also been widely used within healthcare to effectively reduce pain, anxiety, depression and fatigue across many conditions ranging from labor pain during childbirth to pre-debridement anxiety for burn patients [7]. Although many such kinds of touch within the healthcare context have been identified (e.g., [2]), we have been primarily concerned with “affective” and “simple” touch that is used by a provider to intentionally deliver a message of comforting to a patient in pain or distress.

2. RELATED WORK

A few researchers have developed systems that use touch as a mediated form of communication between users, relaying hugs [15], strokes [6], or touch dynamics [3] between users. A few have also explored autonomous systems that touch users for affective or therapeutic purposes, such as therapeutic massage [19]. However, we are aware of no prior work that attempts to simulate conversational touch, that is, touch employed as part of an interaction with an embodied conversational agent or conversational robot.

3. THE “TOUCHBOT” AGENT

Based on observational studies of where nurses touch patients, as well as studies of where people are comfortable being touched by strangers [16], we decided to construct an agent that would touch users on their hands. We also wanted to ensure that the touch felt comfortable and organic, so our initial design for the haptic output



Figure 1. Pneumatic Haptic Glove

Cite as: Title, Author(s), *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, Decker, Sichman, Sierra, and Castelfranchi (eds.), May, 10–15, 2009, Budapest, Hungary, pp. XXX-XXX. Copyright © 2009, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

device uses a glove with an air bladder sewn into the palm (Figure 1). The bladder is inflated or deflated via two valves, one connected to a 25 psi compressed air tank and the other venting to the atmosphere. The valves are controlled by a GadgetMaster II controller board, and our embodied conversational agent dialogue engine [1] was extended with primitives that allow the valves to be controlled within dialogue scripts and synchronized to word boundaries during an agent utterance.

Based on pilot testing and results from a study of affective touch-based communication between people [11], we settled on a simulated stroking pattern of 2 slow inflations (200ms duration), 750ms apart, to represent an empathic touch used during an agent utterance. Pilot observation studies of naturally occurring touch in human-human conversation indicated that touch typically occurs at the beginning of the utterance it is semantically related to, so in all dialogue content we have developed for evaluation, the empathic touch is aligned with the beginning of the corresponding agent utterance.

Preliminary testing of the glove used in combination with an animated head on a desktop monitor indicated that users felt that the glove was not being controlled by the agent. To enhance the feeling of connectedness, we subsequently introduced a mannequin to visually connect the glove to the talking head (Figure 2). Users sit facing the mannequin with their hand in the glove, resting on the mannequin’s hand during a conversation (the glove is on the user, not the mannequin). To remove any complications arising from users using their hands for input control during an interaction, a wizard-of-oz control [5] was developed for pilot evaluation so that users could talk to the agent using speech.

4. PRELIMINARY EVALUATION

We are currently conducting an evaluation study to assess the ability of the TouchBot agent to establish a therapeutic alliance

with users during a single brief counseling dialogue about cancer, comparing this functionality to the same apparatus but with the haptic modality disabled. We hypothesize that the touch modality will lead to significantly greater working alliance, and ratings of liking, trust and naturalness of the agent compared to the control condition.

4.1 Apparatus

A dialogue script was developed consisting of a greeting, introduction, several turns of social chat, a discussion about how the user feels about cancer, and a closing. A single, brief glove inflation was commanded during the greeting to simulate a handshake for all participants. Empathic feedback, including touch, is provided during the cancer discussion at appropriate points (e.g., Agent: “How do you feel about cancer?” User: “I hope I don’t get it.” Agent: *with empathic touch, concerned facial display* “I know, it can be very scary.”). This dialogue lasts approximately two minutes. The only manipulation between the two conditions of the study (TOUCH and NOTOUCH) was that in NOTOUCH the haptic glove was not sent the commands to inflate during empathic dialogue—the treatments were identical in all other respects.

4.2 Measures

In addition to demographics, therapeutic alliance was assessed using the bond subscale of the Working Alliance Inventory, a validated 12-item self-report scale [12]. An additional six items assessed other aspects of the user’s attitude towards the agent, including enjoyment, naturalness, desire to continue, etc. User introversion/extroversion was assessed using a 16-Likert-item self report scale [20]. Touch receptivity (how a user feels about being touched) was assessed using a new 10-Likert-item composite self report scale. User heart rate and galvanic skin conductivity were recorded continuously at 256 Hz, using finger-clip sensors from Thought Technology, Ltd.



Figure 2. Experimental Setup with Mannequin

4.3 Protocol

Prior to the arrival of study participants, the compressed air tank was charged to 25psi using an air compressor, and the compressor was then turned off during the study. There is sufficient capacity in the tank to inflate the glove 8-10 times, and the loud noise of the compressor would have been disruptive.

Participants were consented, then filled out the demographic, personality and touch receptivity questionnaires. Next, they were randomized into a TOUCH or NOTOUCH condition of the study, seated in front of the TouchBot, and their right hand placed in the haptic glove. They were instructed to rest their right hand on the mannequin's hand throughout the interaction, and told that while they were talking to the agent "the agent can occasionally inflate [the glove] to give you the sensation of a slight squeeze." (they were not told the intended meaning of the touch). Finger-mounted galvanic skin response and heart rate sensors were attached to their left hand, which they were then instructed to rest in their lap. Participants were then told they could talk to the agent via a microphone mounted on the desk next to them, but that it could only recognize one of the options displayed along the right side of the screen (dynamically updated during each conversational turn). At this point the experimenter left the observation room and the agent began the dialogue with the participant. Following the conversation, the experimenter re-entered the room, removed the sensors and glove from the participant, and administered the working alliance and attitudinal questionnaires. A semi-structured interview was then conducted to obtain impressions of the experiment and agent. Participants were then debriefed, paid and dismissed. The entire study session was videotaped.

4.4 Subjects

Twenty-one subjects have participated in the study to date, 40% male, age 34.3 (SD 14.8), 80% single, 52% students.

4.5 Preliminary Quantitative Results

There are few significant effects of study condition on outcome measures at this time. However, general trends are emerging on the attitudinal measures indicating a interaction between participant gender and study condition, such that women have generally more positive attitudes towards the agent in the TOUCH condition, while men have generally more negative attitudes towards the agent in the TOUCH condition. The only item in which this interaction is currently significant is for ratings of the agent's friendliness, $F(1,17)=4.75$, $p<.05$, (Figure 3).

Data from the physiological sensors is still being analyzed.

4.6 Preliminary Qualitative Results

When asked for their overall impressions, the most frequent responses were "weird" (3 of 9 respondents) and "awkward" (2 of 9 respondents).

Most participants felt that the agent was communicating empathy, sympathy or comforting with its touch (11 of 15 respondents):

- "I saw it as an expression of sympathy or empathy"
- "Probably sympathy, compassion..."
- "I guess if it was like a real situation, I would interpret it as caring, and you know, really being in to the conversation, and not like talking to me, but talking with me."

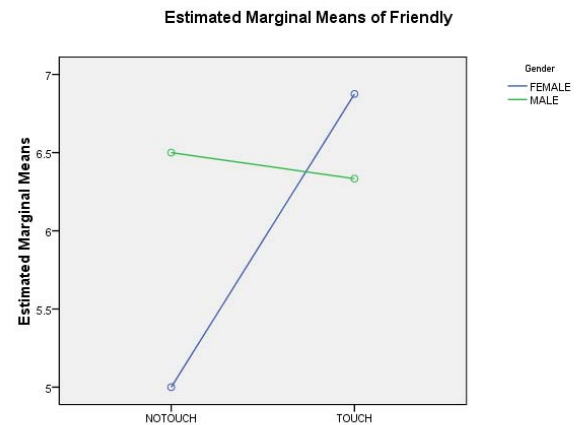


Figure 3. Interaction of Condition and Gender on Perceived Friendliness of Agent

- "Definitely felt.. like a hand squeeze... like sympathy. No, I guess not sympathy, not empathy, sort of - reassuring. Reassuring is the word."

When asked if they felt the touch was natural, respondents gave mixed reactions (roughly half said yes):

- "I thought it felt very natural, as if somebody was holding my hand while he or she was talking to me. I didn't think it was forced"
- "Felt natural towards the end I think. I think I just got more used to it".
- "The way she squeezes the hand is a little different from what normally humans do."

Most still felt that the glove was separate from the agent, even with the mannequin:

- "I thought it was weird to have the body"
- "It seems more separate, but I was trying to connect it."

Two male participants indicated that they did not feel comfortable being touched:

- "I'm more uncomfortable on the whole touching while having a conversation thing."
- "I think it's a little different for guys and girls. Being a guy, I definitely find it a bit weird. You know, if a doctor reached out and squeezed my hand as he gave me bad news, I'd you know...I would find that more strange than anything else"

Finally, several participants actually seemed to enjoy the conversational touch:

- "I found that it was amazing that a computer can actually respond to another human being's hand by squeezing it."
- "Enjoyable, very different, very comfortable"

4.7 Discussion

The interaction between gender and touch on attitudes towards the agent is not too surprising, since in American culture women are touched more than men, both as infants and adults [8], leading to greater comfort with touch. This also carries over into healthcare contexts. One study showed that in a hospital setting female

patients who were touched reported less anxiety about surgery compared to women who were not touched, but men who were touched reported more anxiety [9]. There is also a trend in our data for females to have higher touch receptivity scores compared to males.

5. FUTURE WORK

We are continuing to run study participants and manipulate elements of the protocol and apparatus to understand the best way for a conversational agent to administer empathic touch.

We have found from debriefing interviews that study participants still feel that the hand is not being controlled by the agent. For this reason, and to gain finer control over the touch dynamics (e.g., to replicate the results in [11]), we are in the process of fabricating a mechanical hand that will be covered in foam (Figure 4). We feel that by having the agent's physical hand administer user touch, users will feel more inclined to attribute the touch behavior to the agent.

We also have a study underway to understand the role of conversational touch in emphasizing important information during tutorial and counseling dialogues.

Conversational touch represents an important and unexplored modality for conversational agents, especially those deployed in the healthcare environment.

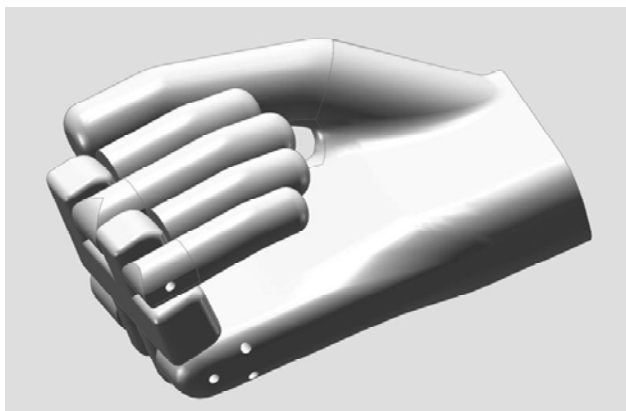


Figure 4. Mechanical Hand Design

6. ACKNOWLEDGMENTS

Thanks to Thomas Brown for his assistance with the evaluation study and Christine Lee for developing much of the hardware used in the system. This work was supported by NSF CAREER IIS-0545932.

7. REFERENCES

- [1] Bickmore, T., Gruber, A. AND Picard, R. 2005. Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Educ Couns* 59, 21-30.
- [2] Bottorff, J. 1993. The use and meaning of touch in caring for patients with cancer. *Oncol Nurs Forum* 20, 1531-1538.
- [3] Brave, S., Dahley, A., Frei, P., Su, V. AND Ishii, H. 1998. inTouch. In *SIGGRAPH'98*.
- [4] Bush, E. 2001. The Use of Human Touch to Improve the Well-Being of Older Adults *Journal of Holistic Nursing* 19, 256-270
- [5] Dahlback, N., Jonsson, A. AND Ahrenberg, L. 1993. Wizard of Oz Studies: Why and How. In *IUI 93*, 193-199.
- [6] Eichhorn, E., Wettach, R. AND Hornecker, E. 2008. A Stroking Device for Spatially Separated Couples. In *MobileHCI*.
- [7] Field, T. 2000. *Touch Therapy*. Churchill Livingstone, Edinburgh.
- [8] Field, T. 2003. *Touch*. MIT Press, Cambridge, MA.
- [9] Fisher, J. AND Gallant, S. 1990. Effect of touch on hospitalized patients. In *Advances in Touch*, N. GUNZENHAUSER, T. BRAZELTON AND T. FIELD Eds. Johnson & Johnson, Skillman, NJ, 141-147.
- [10] Frankel, R. 1995. Emotion and the Physician-Patient Relationship. *Motivation and Emotion* 19, 163-173.
- [11] Hersteinstein, M. AND Keltner, D. 2006. Touch Communicates Distinct Emotions. *Emotion* 6, 528-533.
- [12] Horvath, A. AND Greenberg, L. 1989. Development and Validation of the Working Alliance Inventory. *Journal of Counseling Psychology* 36, 223-233.
- [13] Horvath, A. AND Symonds, D. 1991. Relation Between Working Alliance and Outcome in Psychotherapy A Meta-Analysis. *Journal of Counseling Psychology* 38, 139-149.
- [14] Kutner, J., Smith, M., Corbin, L., Hemphill, L., Benton, K., Mellis, K., Beaty, B., Felton, S., Yamashita, T., Bryant, L. AND Fairclough, D. 2008. Massage Therapy versus Simple Touch to Improve Pain and Mood in Patients with Advanced Cancer. *Annals of Internal Medicine* 149, 369-379.
- [15] Mueller, F., Vetere, F., Gibbs, M., Kjeldskov, J., Pedell, S. AND Howard, S. 2005. Hug over a distance. In *CHI'05*.
- [16] Nguyen, T., Heslin, R. AND Nguyen, M. 1975. The meanings of touch: Sex differences. *Journal of Communication* 25, 92-103.
- [17] Reynolds, W. 2000. *The measurement and development of empathy in nursing*. Ashgate Publishing, Aldershot.
- [18] Schoenhofer, S. 1989. Affectional touch in critical care nursing: a descriptive study. *Heart Lung* 18, 146-154.
- [19] Vaucelle, C. AND Abbas, Y. 2007. Touch-Sensitive Apparel. In *CHI'07*.
- [20] Wiggins, J. 1979. A psychological taxonomy of trait-descriptive terms. *Journal of Personality and Social Psychology* 37, 395-412.

Reasoning about emotions in an engaging interactive toy

(Extended Abstract)

Carole Adam
RMIT University, School of CS& IT
Melbourne, VIC 3000, Australia
carole.adam.rmit@gmail.com

Patrick Ye
RMIT University, School of CS& IT
Melbourne, VIC 3000, Australia
ye.patrick@gmail.com

1. INTRODUCTION

In this work¹ we report on an emotional dialogue module for an intelligent interactive toy being developed in collaboration with an industry partner, as part of an ongoing project aiming at designing toys able to engage children in long-term relationships.

A number of conversational agents now have emotional abilities that have been shown to improve interaction with them [2, 9, 13]. Such emotionally expressive agents encourage the user to express its own emotions [3] so they should then be responsive to these emotions. Existing agents have a limited range of reactions: empathy [5], predefined responses to each emotion [7], or domain-specific adaptation of behaviour [10]. However we need the strategies of our toy to be varied enough to keep the child engaged, and generic enough to adapt to a wide range of interaction situations. We have thus formalised a large set of generic strategies comprising not only empathy but also coping strategies.

We use an existing BDI framework (PLEIAD, [1]) allowing to design agents able to reason about emotions from Ortony, Clore and Collins' (OCC) theory [11]. We integrate this framework with the dialogue manager of the Toy so that its answer to the child depends on emotional data.

2. ARCHITECTURE

The child's utterances are translated into speech acts [12] using hand-coded grammar rules, and interpreted thanks to a semantic based on [6] to deduce the child's mental attitudes. The emotional module comes from the PLEIAD framework and provides the Toy with logical definitions of emotions in terms of mental attitudes. The Goal Selection Module is responsible for computing appropriate communicative goals (*i.e.* answers) depending on the child's input and emotion, and adopt the best one as the Toy's intention. The Action Selection Module then selects a template of natural language sentence from a library and fills its slots.

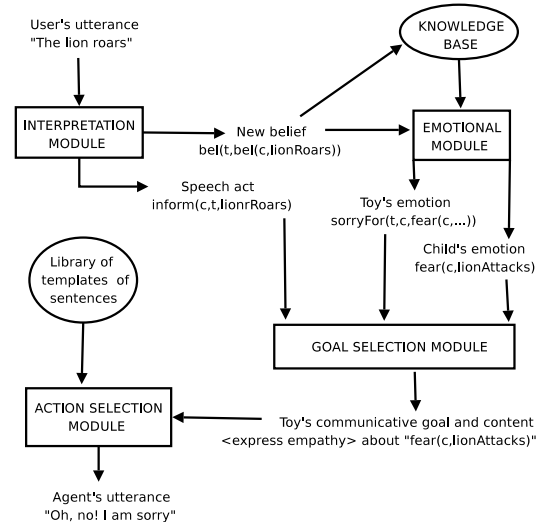
3. EMOTIONAL STRATEGIES

The following "classical" types of response strategies are integrated in the dialogue module of our intelligent toy:

¹A full version of this paper has been submitted to IJCAI'09

Cite as: Reasoning about emotions in an engaging interactive toy (Short Paper), C. Adam and P. Ye, *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, Decker, Sichman, Sierra and Castelfranchi (eds.), May, 10–15, 2009, Budapest, Hungary, pp. XXX-XXX.

Copyright © 2009, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.



- **Empathy:** when the toy is aware of the child's emotion it can adopt the goal to express empathy thanks to the corresponding good-will *fortunes-of-others* emotion of the OCC theory, *i.e.* either *Happy-for* if it is a positive emotion, or *Sorry-for* if it is a negative one;
- **Expression:** if the toy is itself feeling an emotion relevant to the subject at hand, it can adopt the goal to express this emotion;
- **Curiosity:** when the toy cannot infer the child's emotion about the current subject, it has to update the child's profile by adopting the goal to get information from the child about his attitude toward the subject (desires about an event, ideals about an action, or likings related to an object). Such a behaviour also shows interest and care for the child;
- **Confirmation:** even if the toy can infer the child's emotion from his profile information, it can still adopt the goal to confirm its deduction by asking the child if he is really feeling the deduced emotion, which may show understanding of the child's emotion, or allow update of the profile information.

Moreover we have added a new type of response: coping strategies. Coping is a psychological process by which individuals try to lower their negative emotions [8]. Here we focus on the particular manifestation in dialogue of some coping strategies from Carver *et al.*'s classification [4], and

we study how the Toy can try to cope with the child's emotions. Our agent can use the following strategies:

- **Emotional support** by expressing empathy and understanding of the child's emotion, as explained above;
- **Positive reinterpretation** consists in using the toy's commonsense knowledge base to find positive consequences of the stressor (the stimulus having caused the child's negative emotion), and informing the child about them to lessen or remove his negative emotion;
- **Active coping** consists in taking active steps to remove the stressor. The Toy can either encourage and help the child to act, or perform a useful action itself;
- **Mental disengagement** consists in changing the child's focus to divert him from the stressor, for instance by changing the subject of the dialogue, playing a game or telling a story or a joke.

4. CASE STUDY

In this section we show an example of dialogue that can be produced by our agent. The dialogue begins with the toy asking the child about what he did today, and then the toy actively keeps the conversation going by asking questions. This comes from a default goal of the agent to know about the child's activities.

Toy: "What did you do today?"

Child: "I have been to the zoo"

Toy:

- Profile: the child likes going to the zoo
- Emotion module: the child is happy about having visited the zoo
- Empathy strategy: "Oh! That's great!"

Toy: "Which animals did you see?"

Child: "I saw a tiger"

Toy:

- No profile information
- Curiosity strategy: "Do you like tigers?"

Child: "Yes, I love them!"

Toy:

- Update profile: the child likes tigers
- Emotion module: the child felt attraction for the tiger
- Empathy strategy: "I am very happy for you that you saw a tiger"

Child: "I saw kangaroos, koalas..."

Toy:

- Toy personality: the toy likes koalas
- Emotion module: the toy feels affection for koalas
- Expression strategy: "I love koalas!"

Child: "And I also saw a lion, it was roaring at me"

Toy:

- Profile information: the child believes that a roaring lion will attack, and does not want it
- Emotion module: the child was afraid of a possible attack by the lion

- Confirmation strategy: "Were you afraid that it would attack you?"

Child: "Yes..."

Toy:

- Deduction from commonsense knowledge: a lion in a zoo is in a pen and cannot attack visitors
- Active coping strategy (correct false beliefs): "It could not attack you, it was in a pen"

5. CONCLUSION

We have designed an emotional dialogue module implementing various emotional responsiveness strategies, and applied it to the design of an intelligent toy able to respond to the child's emotion in order to keep him engaged in the interaction. The initial dialogue module is implemented in Prolog and interfaced to the Java based dialogue engine of the toy. Initial evaluation suggests this is sufficiently fast for real-time interaction, although more comprehensive testing of the efficiency is planned. This work is part of an ongoing project aiming at developing such kinds of intelligent toys, in collaboration with an industry partner. In future work this dialogue module will be further integrated into the global architecture of the toy, including in particular the speech recognition and speech generation engines, making the interaction more natural. It will then be possible to conduct thorough evaluations of the real impact of these emotional strategies, by having children interact with the prototype and assessing their reactions in collaboration with psychologists.

6. REFERENCES

- [1] C. Adam. *Emotions: from psychological theories to logical formalization and implementation in a BDI agent*. PhD thesis, INP Toulouse, France, July 2007.
- [2] J. Bates. The role of emotion in believable agents. *Communications of the ACM*, 37(7), 1994.
- [3] C. Breazeal. Robot in society: Friend or appliance? In *Agents 99 Workshop on Emotion-based Agent Architectures*, pages 18–26, Seattle, 1999.
- [4] C. S. Carver, M. F. Scheier, and J. K. Weintraub. Assessing coping strategies: a theoretically based approach. *Journal of Personality Psychology*, 56(2):267–283, 1989.
- [5] A. Cavalluzzi, B. De Carolis, V. Carofiglio, and G. Grassano. Emotional dialogs with an embodied agent. In *User Modeling*, pages 86–95, Johnstown, USA, 2003.
- [6] FIPA (Foundation for Intelligent Physical Agents). FIPA Communicative Act Library Specification. <http://www.fipa.org/repository/aclspecs.html>, 2002.
- [7] D. Heylen, A. Nijbolt, and R. op den Akker. Affect in tutoring dialogues. *Applied AI*, 19(3):287–311, 2005.
- [8] R. S. Lazarus and S. Folkman. *Stress, Appraisal, and Coping*. Springer Publishing Company, 1984.
- [9] J. C. Lester, S. A. Converse, S. E. Kahler, S. T. Barlow, B. A. Stone, and R. S. Bhogal. The persona effect: affective impact of animated pedagogical agents. In *SIGCHI conference on Human factors in computing systems*, 1997.
- [10] M. Y. Lim and R. Aylett. Feel the difference: A guide with attitude! In *IVA*, volume 4722/2007 of *LNCS*, 2007.
- [11] A. Ortony, G. Clore, and A. Collins. *The cognitive structure of emotions*. Cambridge Univ. Press, 1988.
- [12] J. R. Searle and D. Vanderveken. *Foundation of illocutionary logic*. Cambridge University Press, 1985.
- [13] D. Traum, S. Marsella, and J. Gratch. Emotion and dialogue in the MRE virtual humans. In *Affective Dialogue Systems*, vol. 3068 of *LNCS*, pp. 117–127. Springer, 2004.

Towards an Empathic Chess Companion

Iolanda Leite, André Pereira, Carlos Martinho,
Ana Paiva
INESC-ID and Instituto Superior Técnico
Av. Prof. Dr. Cavaco Silva - Taguspark
2744-016 Porto Salvo, Portugal
{iolanda.leite, andre.pereira, carlos.martinho}@
tagus.ist.utl.pt, ana.paiva@inesc-id.pt

Ginevra Castellano
Department of Computer Science, Queen Mary,
University of London
Mile End Road London E1 4NS
United Kindom
ginevra@dcs.qmul.ac.uk

ABSTRACT

In this paper, we propose an empathic model for a social robot that acts as a chess companion for children. The model will attempt to recognize some of the user's affective states (interest, boredom and frustration), by combining information retrieved from facial and body expression recognition systems with contextual features of the game (e.g., who is winning, for how long...). We further present a set of possible empathic behaviours that the agent can perform when the user is experiencing such affective states.

Keywords

Empathy, affect, long-term interaction, companions.

1. INTRODUCTION

The interaction paradigm in synthetic characters is changing. Seminal work in this field has considered agents that interacted with users for short periods of time, but we are now moving towards a new paradigm in which characters are able to relate to us, assist us and engage us in a long-term basis [1]. The LIREC Project (Living with Robots and Interactive Companions) aims to create a new generation of interactive and emotionally intelligent companions that are capable of establishing long-term relationships with different users. Research focuses on both virtual agents and physically embodied agents such as robots.

To build agents that are successful in establishing and maintaining long term meaningful interactions with users, some social and cognitive abilities are needed. One of such abilities is empathy, which involves role taking, the understanding of nonverbal cues, sensitivity to the other's affective state and communication of a feeling of caring, or at least sincere attempts to understand in a non judgemental manner [7]. Research shows that empathic agents are perceived as more caring, likeable, and trustworthy than agents without empathic capabilities, and that people feel more supported in the presence of such agents [2]. Therefore, we believe that if a character is endowed with empathic behaviours, the interaction with the user will be more natural, believable and engaging, which can be of extreme relevance for our long-term goal.

Our application scenario includes a social robot, the iCat [3], which plays chess with children using an electronic chessboard. The iCat acts as a peer tutor, helping children to improve their chess skills [14]. While playing with the iCat, children receive feedback from their moves on the chessboard through the iCat's facial expressions, which are generated by an affective system influenced by the state of the game. The affective system is self-oriented or competitive, i.e., when the user plays a good move the iCat displays a sad facial expression and when the user plays a

bad move the iCat displays positive reactions (for more details in the affective system please see [15]). We have adopted this approach instead of a more cooperative behaviour because, from our observations of children playing against each other in a chess club, such reactions are more consistent with what they might expect about their opponents. Nevertheless, after performing experiments with the iCat in a chess club for several weeks [13], we realized that sometimes children felt uncomfortable with the iCat displaying intense happy expressions when they were losing, especially in front of their other colleagues. If the iCat could understand their affective state and react in a more empathic manner, situations like this one could be avoided. Our main challenge is thus to create an empathic chess playing companion that is able of helping children to improve their chess skills, while at the same time behaves in a way that the users will want to continue interacting with it without feeling embarrassed or stressed.

Another interesting finding from the experiment conducted at the chess club was that sometimes users reacted in an empathetic way towards the iCat. We have witnessed some moments in which users were imitating the iCat's sad expressions, as if they were sharing that same emotion. Likewise, some users also demonstrated empathetic behaviour through sentences such as "Oh, the iCat is sad...", with a sad intonation in their voice.

Although there is not a common agreement on the definition of empathy, in most of the proposed definitions the ability to understand another's affective state, either due to a pure cognitive or affective process, appears to be the foundation for the human's empathic behaviours. As such, empathy can be seen as "an observer reacting emotionally because he perceives that another is experiencing or about to experience an emotion" [22].

In this paper, we propose a model for recognizing the user's affective states in a turn-based game. The document is organized as follows. After a brief overview of existing work on empathic agents and recognizing the user's affective state, we present our model, which is composed of two main parts: visual and contextual features. We then present some of the empathic behaviours that the agent might perform in response to those user's affective states. Finally, we draw some conclusions and future work.

2. RELATED WORK

There are two main branches of research when studying empathic agents: agents that simulate empathic behaviour towards the users and agents that foster empathic feelings on the users. The work presented in this section is focused on the first topic.

One of the functions of human emotions is to elicit adaptive social responses from others. It was shown that when we detect personal distress in another person we tend to empathise and display the prosocial behaviour of sympathy [5]. This behaviour can often lead to a decrease or relief of the other person's distress. Reeves and Nass [19], in a series of empirical studies, reported that humans behave naturally and socially towards machines as they do with other humans. In this line of thought, we can hypothesise that a computer with an empathic behaviour can also simulate the prosocial behaviour of empathy, and therefore relieve users of personal distress.

This hypothesis began to be addressed by Klein et al. [12]. Their studies were designed to relieve user frustration caused by an intentional faulty computer application, through the use of a text based agent. This agent used active listening, empathy and sympathy with the intention of helping to relieve the user's negative state. However there were no significant results to prove the hypothesis that a computer program could really help users feel less frustrated only by the use of an empathic agent.

Meanwhile, a study presented years later by Hone [9] continued Klein et al.'s work and tested the same hypothesis. In this new study, the author suggested that the above referred possibility could be right. It was shown through a series of three experiments that a text based agent with empathetic behaviour could indeed help users to successfully relieve their frustration. This study also showed that a virtually embodied character is even more successful at achieving the same purpose. The author reflects on this result explaining that "there is a good match between the characteristics of the feedback strategy (human-human) and the characteristics of the entity delivering that feedback". It remains unknown if a social robot could outperform a virtual agent in this task, even though in our previous work [18] there was evidence that a robotic agent does provide greater feedback than a virtual agent in human-machine interaction.

Ochs et al. [17] showed that a virtual agent is perceived more positively when it expresses empathic emotions than when it expresses no emotions. This work also raised a preeminent challenge in the creation of empathic agents, as it showed that if the same agent expresses the empathic emotions in an inconsistent way, the opposite effects occurs (i.e., the agent is perceived more negatively than another agent without empathic behaviour). These results suggest that recognizing the right affective state of the user (to be able to display the appropriate empathic behaviour) is of extreme relevance.

Therefore, research on empathic companions needs to take into account the design of an affect recognition framework. It is important to stress that a companion's affect recognition abilities must go beyond the detection of prototypical emotions and be sensitive to application-dependent affective states, such as, for example, interest, boredom, frustration, willingness to interact, etc. [4].

Some efforts in this direction have been reported in the literature. Kapoor and Picard [11], for example, proposed an approach for the detection of interest in a learning environment by combining non-verbal cues and information about the learner's task (level of difficulty and state of the game) Kapoor et al. [10] designed a system that can automatically predict frustration of students interacting with a learning companion by using multimodal non-verbal cues including facial expressions, head movement, posture, skin conductance and mouse pressure data. El Kaliouby and

Robinson [6] proposed a computational model for the detection of complex mental states such as *agreeing*, *concentrating*, *disagreeing*, *interested*, *thinking* and *unsure* from head movement and facial expressions.

3. RECOGNIZING THE USER'S AFFECTIVE STATE

As discussed in the related work section, understanding the user's affective state is the ground for empathic behaviour. Initially, we intend to endow our agent with the ability to recognize a limited set of the user's affective states. Taking into account the domain in which the agent is immersed as well as its role, we have chosen to start focusing on interest and boredom. In the future, we will attempt to model the recognition of frustration.

To identify the affective states mentioned above, we propose a model divided in two main parts: (1) recognition of user's facial and body expressions and (2) contextual features of the game. The affective states recognized by the model will work as the basis for the iCat's empathic behaviours. The remaining of this section describes in more detail the approach that we intend to follow.

3.1 Visual Features

During the whole interaction, the user sits in front of the iCat (see Figure 1), separated only by the chessboard. Since both the iCat and the user are in a fixed position, it is possible to use a camera to capture some expressions displayed by the user.



Figure 1. User playing with the iCat at the chess club.

We intend to employ new and existing vision libraries to analyze a set of non-verbal cues, including:

- Head gestures (e.g., head nods, shakes)
- Facial expressions (e.g. smiles)
- Eye gaze (e.g., fixed at the iCat, fixed at the chessboard or looking away)
- Lateral Posture (e.g., approach versus avoidance)

To validate which non-verbal cues are relevant to the affective states that we aim to recognize, as well as to our specific scenario of interaction, we are going to use both results from studies regarding body and facial expression of emotion (such as [21]) and video observation and annotation of interaction sessions conducted at the chess club. We plan to have two different groups of annotators: a first group to annotate the user's affective states (interest, boredom or neither), and a second group to annotate the user's expressions. With this approach, we intend to come up with a set of visual cues that are statistically significant in the discrimination of the defined set of affective states for our specific scenario. Our final aim is to build an affective recognition system that can work in real-time, in a real game scenario.

3.2 Contextual Features

Even though facial and body expressions are very important means of non-verbal communication, sometimes they can be misleading. People may want to dissimulate their facial expressions [20], particularly in a situation of a turn-based game in which two players play against each other. Moreover, affect recognition through visual cues may return the same patterns for different affective states, or people may express the same affective states in slightly different manners. These are some of the reasons for which we believe that situational context is very important when recognizing the user's affective state. As such, we will use contextual features either to disambiguate or to strengthen the confidence of the affective states identified by the vision system.

We start assuming that, when the user is playing with the iCat, many of the experienced affective states may be related to the events happening in the game, or with the behaviour and expressions displayed by the robot. The following list contains the contextual features that may influence the user's affective state:

- *Who has advantage/disadvantage in the game:* this information is obtained by the chess evaluation function, which also works as the main input for the iCat's affective model. Information such as which pieces were captured both in the user's side and in the iCat's side can also be retrieved.
- *Robot's facial expressions:* there may be a correlation between the user's affective state and the iCat's expressions, especially the ones displayed after each user's move.
- *Time the user takes to play a move:* this feature may vary among different users, and therefore it will only be helpful after some interactions. For instance, if the user usually takes about two minutes to play each move, and at some moment of the game he/she starts looking away more often and taking much more time to play, that might be a signal of boredom. But boredom might not be always associated to taking too much time to play. Another different user might feel bored if the exercises proposed by the iCat are very easy for him/her, and in that case the user does not need much time to play.

In addition to these features, we can also use the mechanism that the iCat uses to generate its affective reactions, but with the information from the user's perspective, i.e., taking into account the user's position in the game. The *emotivector* (Figure 2) is an anticipatory system that generates an affective signal resulting from the mismatch between the expected and the sensed values of the sensor to which it is coupled to [16].

In the iCat's affective system, the emotivector is coupled to values received from the chess evaluation function (for more details see [15]). When the user plays a new move, the chess evaluation function returns a new value, updated according to the new state of the game. The emotivector system captures this value and, by using the history of evaluation values, an expected value is computed (applying the moving averages prediction algorithm [8]). Based on the mismatch between the expected and the actual sensed value (i.e., the new value received from the evaluation function), the emotivector generates one of the nine different affective signals for that percept (see Figure 2). Each one of these

nine sensations will result in a different affective reaction in the iCat's facial expression.

	more R	as expected	more P
expected R	stronger R (S+) 	expected R 	weaker R (S+)
negligible	unexpected R 	negligible 	unexpected P
expected P	weaker P (S-) 	expected P 	stronger P (S-)

Figure 2. Emotivector mechanism. “R” means reward and “P” stands for punishment.

For instance, after three moves in the chess game, if the iCat has already captured an opponent's piece, it might be expecting to remain in advantage in the game (i.e., expecting a “reward”) after the next user's move. So if the user plays an even worse move than the one that iCat was expecting (e.g., by putting her queen in a very dangerous position), the elicited sensation will be a “stronger reward”, which means “this state of the game is better than I was expecting”. In the presence of a “stronger reward”, the iCat displays a facial expression of happiness.

Now we will present the same example, but from the user's perspective. After three moves in the game, the user has lost one piece, so he/she might be expecting the iCat to keep the advantage (i.e., expecting another “punishment”). If the user plays a terrible move, and acknowledges that by looking at the iCat's expression of happiness, he/she might be experiencing something closer to a “stronger punishment” sensation. At this time, and taking into account the game history, the iCat may assume that the user is experiencing frustration.

This example attempts to show the kind of reasoning that the iCat can perform about the user, to infer his/her affective experiences. Of course such results need to be verified, either by other contextual features or by information from the vision system.

4. EMPATHIC BEHAVIOUR

After recognizing the user's affective states, the agent should use that information to behave in a more empathic manner. Some of the empathic behaviours that might be employed are the following:

- *Boredom:* when the agent detects that the user is bored, it can ask him/her to start over the game, propose a new exercise or, at extreme conditions, suggest the ending of the interaction. If the game is balanced, the iCat can propose a stalemate, which may increase the user's interest to continue the interaction. Small talk about the game, or about previous games the iCat and the user played together, is another technique that could be used to prevent or remediate boredom. Finally, if the user is bored for being constantly in an advantageous position in the game, one can increase the chess engine's difficulty, and the iCat will become a stronger opponent. The opposite may also occur (the user getting bored because the game is too difficult), and it can be amended as well.
- *Interest:* if the user is currently on this state, the agent can assume that it is on the right track and so it should

continue with the same behaviours and playing with the same difficulty level.

- *Frustration*: when the user is frustrated for being in disadvantage in the game, he/she might become even more frustrated with the iCat expressing very happy emotions. Therefore, one of the empathic behaviours that we suggest to deal with user's frustration is for the iCat to inhibit some of its happy facial expressions, or display them with a lower intensity. Another alternative to reduce frustration might be to reduce the difficulty of the chess game engine.

Most of these empathic behaviors are context dependent. Even so, the same strategies, if proved to be successful, could be applied in other contexts of interaction. This can be particularly true for the behaviors related to the expression/inhibition of emotions.

5. CONCLUSIONS AND FUTURE WORK

In long-term interactions, social robots need to be capable of more than just displaying emotions and social cues towards the user. They need to be socially aware, interactive and empathic, by taking into account the user's intentions and affective states. In fact, previous research on virtual agents has shown that one of the main aspects that breaks the user's suspension of disbelief in such interactions is the restricted way in which agents are receptive to the social cues displayed by the user [1]. This also happened in our scenario, as the iCat only perceived the game events, and was unable to "understand" the affective cues displayed by its opponent.

In this paper, we presented a model for detecting the user's affective state with the purpose of endowing a social robot with more empathic capabilities. In the near future, we intend to validate the proposed model by performing another field trial to collect new data, so we can validate the results obtained with the existing data. After completing this step, we plan to implement the empathic behaviour mentioned in Section 4, and evaluate if such behaviour has impact on the user's long-term interaction with the agent. As a ground for comparison, we will use the results obtained from our previous long-term experience with the iCat [13].

6. ACKNOWLEDGMENTS

The research leading to these results has received funding from European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement n° 215554.

7. REFERENCES

- [1] Bickmore, T. and Picard, R. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction*, **12**(2), (2005), 293-327.
- [2] Brave, S., Nass, C., and Hutchinson, K. Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. **62**(2), (2005), 161-178.
- [3] Breemen, A., Yan, X., and Meerbeek, B. iCat: an animated user-interface robot with personality. in *Autonomous Agents and Multiagent Systems, AAMAS'05*. Pechoucek, Steiner and Thompson (eds.): ACM Press, New York, NY, USA, (2005), 143-144.
- [4] Castellano, G., Aylett, R., Paiva, A., and McOwan, P. W. Affect recognition for interactive companions, in *Workshop on Affective Interaction in Natural Environments (AFFINE)*, *ACM International Conference on Multimodal Interfaces (ICMI'08)*. 2008: Chania, Crete, Greece.
- [5] Eisenberg, N., Fabes, R., Miller, P., Fultz, J., Shell, R., Mathy, R., and Reno, R. Relation of sympathy and personal distress to prosocial behavior: a multimethod study. *Journal of personality and social psychology* **57**(1), (1989), 55-56.
- [6] El Kaliouby, R. and Robinson, P. Generalization of a Vision-Based Computational Model of Mind-Reading., in *1st International Conference on Affective Computing and Intelligent Interaction*. 2005: Beijing, China.
- [7] Goldstein, A. and Michaels, G. *Empathy : Development, Training and Consequences*. 1985: Hillsdale, N.J. : L. Erlbaum Associates.
- [8] Hannan, E., Krishnaiah, P., and Rao, M. *Handbook of Statistics 5: Time Series in the Time Domain* 1985: Elsevier.
- [9] Hone, K. Empathic agents to reduce user frustration: The effects of varying agent characteristics. 2006, Elsevier Science Inc. 227-245.
- [10] Kapoor, A., Burleson, W., and Picard, R. W. Automatic prediction of frustration. 2007, Academic Press, Inc. 724-736.
- [11] Kapoor, A. and Picard, R. W. Multimodal affect recognition in learning environments, in *Proceedings of the 13th annual ACM international conference on Multimedia*. 2005, ACM: Hilton, Singapore.
- [12] Klein, J., Moon, Y., and Picard, R. W. This computer responds to user frustration, in *CHI '99 extended abstracts on Human factors in computing systems*. 1999, ACM: Pittsburgh, Pennsylvania.
- [13] Leite, I., Martinho, C., Paiva, A., and Pereira, A. Social Presence in Long-Term Human-Computer Relationships in *Fourth International Workshop on Human-Computer Conversation*. 2008: Bellagio, Italy.
- [14] Leite, I., Martinho, C., Pereira, A., and Paiva, A. iCat: an affective game buddy based on anticipatory mechanisms. in *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*. Padgham, Parkes, Müller and Parsons (eds.). Estoril, Portugal: International Foundation for Autonomous Agents and Multiagent Systems, (2008), 1229-1232.
- [15] Leite, I., Pereira, A., Martinho, C., and Paiva, A. Are Emotional Robots More Fun to Play With?, in *IEEE RO-MAN 2008*. 2008: Munich, Germany.
- [16] Martinho, C. and Paiva, A. Using Anticipation to Create Believable Behaviour. in *Proceedings of the 21st National Conference on Artificial Intelligence and the 18th Innovative Applications of Artificial Intelligence Conference*. Boston MA, USA: AAAI Press: Stanford, California, USA, (2006), 175-180.
- [17] Ochs, M., Pelachaud, C., and Sadek, D. An empathic virtual dialog agent to improve human-machine interaction, in *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems - Volume 1*. 2008, International Foundation for Autonomous Agents and Multiagent Systems: Estoril, Portugal.
- [18] Pereira, A., Martinho, C., Leite, I., and Paiva, A. iCat, the chess player: the influence of embodiment in the enjoyment of a game. in *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*. Padgham, Parkes, Müller and Parsons (eds.). Estoril, Portugal: International Foundation for Autonomous Agents and Multiagent Systems, (2008), 1229-1232.
- [19] Reeves, B., Nass, C. *The Media Equation. How People Treat Computers, Television, and New Media Like Real People and Places*. 1998: Cambridge University Press.
- [20] Strayer, J. Affective and cognitive perspectives on empathy, in *Empathy and its development*, Eisenberg and Strayer (eds.) Cambridge University Press. (1987).
- [21] Wallbott, H. G. Bodily expression of emotion. *European Journal of Social Psychology*, **28**(6), (1998), 879-896.
- [22] Wispé, L. History of the concept of empathy, in *Empathy and its Development*, Eisenberg and Strayer (eds.) Cambridge University Press. (1987).

An Affective Channel for Companions

Néna Roa Seiler

Centre for Interaction Design

Napier University

Edinburgh, EH10 5DT

n.roa-seiler@napier.ac.uk

David Benyon

Centre for Interaction Design

Napier University

Edinburgh, EH10 5DT

d.benyon@napier.ac.uk

Grégory Leplâtre

Centre for Interaction Design

Napier University

Edinburgh, EH10 5DT

g.leplatre@napier.ac.uk

ABSTRACT

Companions represent a new form of human-computer interaction. Companions know their owners; they provide personalized forms of interaction in an intelligent way. The interaction between people and Companions is multimodal, including speech and natural language. Companions are often represented as an onscreen avatar and because of the anthropomorphic communication this creates, Companions are also expected to be affective interfaces. Empathy is an essential component of the interaction between people and Companions. However in qualitative research with people, it was clear that there was a difficulty with interaction with Embodied Conversational Agents (ECA). In order to improve their performance, and to encourage the development of relationships between people and their companions, interactions need to take into account the people's emotional state.

Companions must be skilled in understanding this state and to respond in an empathetic way. The 'affective channel' is the emotive capability of Companions, which contains voice, prosody, facial expression, body posture, and semantic information according to people's emotional state. The affective channel of Companions has to adapt in order to react to every person's attitude. In this paper we present the basis for the implementation of this channel in Companions. We argue that the implementation of this channel is required to enable affective engagement between people and Companions.

Categories and Subject Descriptors

H.5.2, [User Interfaces]: Natural language, Prototyping, Evaluation, Methodology.

I.2.11 [Distributed Artificial Intelligence]: Intelligent agents

General Terms

Measurement, Human Factors.

Keywords

Agents, companions, new interfaces, affective interaction, emotional design, affective applications.

1. INTRODUCTION

The importance of long-term emotions has not been understood, many believe, because of Descartes' legacy. 'I think, therefore I am' foregrounded cognition as reason. However, recent advances in neurobiology show that emotions and reason are interdependent [9]. Emotions seem to be a major scientific paradigm of this new century and in the last few years there has been a growing interest in feelings and emotion within the field of HCI [12,14].

In the present study, Companions represent the next generation of (ECA) with a robust dialogue capability. ECAs alter the interaction between peoples and computers to a more natural setting: face-to-face communication.

The Companions considered in this paper are personalized conversational interfaces to the Internet that know their 'owners'. They are implemented on indoor and nomadic platforms based on integrated high-quality research into multimodal human-computer interfaces, intelligent agents, and human language technology [17]. It is envisaged that Companions will act as managers for a myriad of services offered by the Internet. Considered as emotional interfaces, the impact and feelings elicited by Companions on people is unknown [15]. However, people's attachment to their Companion seems crucial if it is to achieve its purpose. In this paper we focus on eliciting the emotional reaction of people to Companions and the empathetic response of Companions toward people: the 'Affective Channel'.

This paper is organized as follows: first we describe the context of the Companion project and the characteristics of Companions. Secondly we present the early exploratory work conducted thus far in order to understand perceptions through Companions. This work leads to the proposed theoretical framework. Finally our approach to the implementation of 'the affective channel' in the human-companion interaction is presented.

1.1 The Companion Project

Companions is a 4-year, EU funded Framework Programme 6

Cite as: An Affective Channel for Companions, Roa-Seiler, Benyon, and Leplâtre. *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, Decker, Sichman, Sierra, and Castelfranchi (eds.), May, 10–15, 2009, Budapest, Hungary, pp. XXX-XXX. Copyright © 2009, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

Project, involving a consortium of 16 partners across 8 countries. Its aim is to develop a personalized conversational interface that can act as an alternative access point to resources on the Internet.

Companions stay with their owners for long periods of time, developing a relationship and 'knowing' their owners' preferences and wishes. Companions use technologies such as touch screens, sensors or RFID. They glean the most important information about people from conversation with them. This is used to assist carrying out specific Internet tasks [17].

2. CHARACTERISTICS OF COMPANIONS

Companions are an evolution of ECAs. As defined by Cassel [8], these new interfaces are not only lifelike, with human or animal embodiment, but also specifically conversational. They need to use their bodies in a conversation using rules that humans utilise into a face to face-conversation. The complex rules, which lead our face-to-face interaction, express several human conditions such as social attitudes, relationship status and affective status. People use these protocols to navigate in a social world. Patterns of bodily movement, posture, patterns of visual interaction with the listener, facial expressions, physical proximity, language and speech are part of people's behaviour [1,6]. To be able to engage the user in a conversation and maintain it, ECAs must be skilled in recognizing and producing verbal and non-verbal behaviors, showing emotional states and maintaining a social relationship with people. Utility, form, personality, emotion, social aspects and trust are the characteristics of Companions if they are designed for relationships [4]. Personality and trust are key issues if Companions are to gain the confidence of people. Other authors such as Bates or Creed [2,9] think believability is another important aspect to consider when working with synthetic characters. In particular expressivity or the expression of emotions and empathy are essential to achieve some degree of believability and to improve tasks such as learning or health coach [13,8].

3. PERCEPTIONS OF COMPANIONS

Previous studies have investigated users' response to interfaces like Companions. Some have remained at a conceptual or simulation level, while a few others have evaluated fully functional prototypes. (See for example work done on the Rea system [7]). However, more work is required in order to devise a channel of communication between humans and companions that involves affectivity. Little is known regarding the response of people to interfaces like Companions, as they remain either fictional or theoretical. Gathering more information on this question is a crucial first step towards the design of an affective channel of communication between humans and their Companion. Tests were implemented in order to understand people's global perception of Companions as interfaces. The data gathered was analyzed in order to identify the dimensions used by people to conceptualize Companions. Moreover we investigated terms used by respondents to assess the function of a companion and the relationship between the embodiment of Companion and its function. We have used the repertory grid or Kelly's grid [10]. According to Kelly's Theory, people are observers of the world around them. Like scientists, they draw up hypotheses, which they check with life experiences to elaborate their own theories, to construct their own vision of the world. In other words the user have concepts or references (called constructs), which allow them

to make sense of the world. These constructs also help the users' environment become more predictable to them. This process has an important impact on the users decisions. The repertory grid is a technique that is helpful to uncover people's concepts, the values they call on to understand something, which dimensions people are attached to and what their influences are. In short, how their mental constructs work.

Because of the rich variety of Companions available as part of the project, this study only selected nine of them as shown in Figure 1.

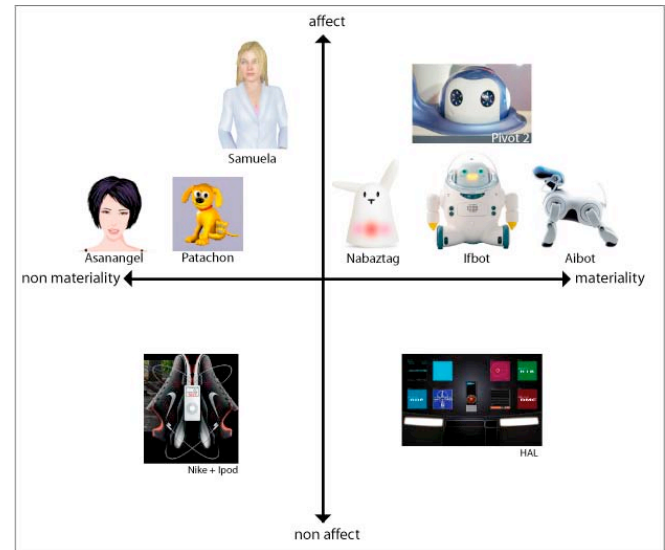


Figure 1: The chosen Companions for experiment

The classification of Companions was made from two dimensions: materiality (the embodiment), and affectivity (feelings) based on users' response.

In the experiment materiality was considered as the tangible package of Companion and not materiality as the opposite space.

One of our goals was to evaluate users' affective responses to different embodiments, while using features such as voice, lighting, facial expressions, and gestures. This would enable us to verify whether common assumptions such as 'Asanangel Companion', which is in the immaterial part of square because she is behind a screen and she doesn't propose any interactivity with the user, are founded.

This exploratory work was undertaken in three languages: French, English and Spanish, because of the need to know whether language and culture have an impact on perception and if so, what is its involvement in the people's mental construct of a Companion.

Figure 2a shows a panel of nine selected Companion images, each linked to a short video presenting the Companion in a real-life context. Participants were presented with this panel and were free to watch the videos as many times as they liked.

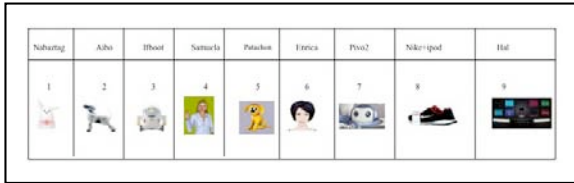


Figure 2a: top of proposed questionnaire

Then participants were asked to choose sets of three companions (triads) as showed in Figure 2b.

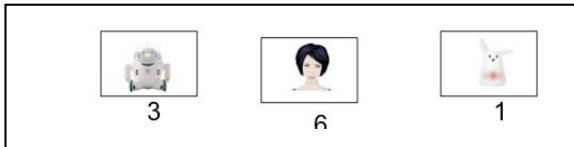


Figure 2b : example of triad association of Companion

For each chosen triad (as illustrated by figure 2b), participants were asked to provide an adjective to express how two Companions of the triad were similar and how the third one was different as showed in Figure 2c. 24 respondents took part in the experiment; as a result 95 grids were collected. Using the same video material showed in Kelly's grid experiment, an open-ended face-to-face interview, after or before the grid test session, depending on the peoples' familiarity with the technologies, was conducted with each participant.

Now, associate two of them for instance : marck this on your paper grille






 3	 1	 6
HOW THEY ARE SIMILAR		HOW IT IS DIFFERENT WITH REGARD TO THE LEFT COMPANIONS
They are : 1 luminous 2 have expressions 3 are cute 4 are endearing 5 interact with you 6 can sing 7 make signs with his body 8 speak		She is 1. Virtuality 2. 2D 3. human female appearance 4. nice eyes 5. nice voice 6. ...


Figure 2c : example of adjectives given to combination of Companions

Nabaztag	Aibo	Iibo	Samuel	Patachon	Enrica	Pivo2	Nike+ipod	Hal
1	2	3	4	5	6	7	8	9

Please choose 3 Companions from this grid, as follow :


3


6


1

Now, associate two of them for instance : marck this on your paper grille




 3	 1	 6
HOW THEY ARE SIMILAR		HOW IT IS DIFFERENT WITH REGARD TO THE LEFT COMPANIONS
They are : 1 luminous 2 have expressions 3 are cute 4 are endearing 5 interact with you 6 can sing 7 make signs with his body 8 speak		She is 1. Virtuality 2. 2D 3. human female appearance 4. nice eyes 5. nice voice 6. ...

Figure 3: Model of grid for experiment.

Figure 3 shows the model of grid proposed to participants.

3.1 FINDINGS

An important result revealed by the survey was that people have some difficulty allocating adjectives to Companions that they had just watched in a video presentation. People seem to need time to speak freely about this approach to technology. Analyzing the examples participants gave, provided an insight into the societal impact and the new relationships people want to develop with Companions as a new interface involving emergent technology. People described their social interaction with Companions and drew a singular approach to how the "Companion's hierarchy" could work (my Companion, your Companion, the Survey Companion that belongs to a company and so on), illustrating their expectations of the 'technology promises' in which the future becomes an object of desire. They also described the level of technology and multimodal exchange they wanted with Companions.

The arrival of Agents endowed with human attributes (voice, recognition abilities) and different embodiments (robots, screen personas, communicating things) are changing the hierarchy people have given to objects in the past. This is evident in the interviews even if peoples cannot explain it directly. For this reason they used a lot of metaphors. This is probably because these changes are very diffuse and perhaps because we do not yet have the words to define the feelings elicited by these new artifacts. As a result of these interviews we are able to present this shift. Companions are expected to have a human behaviour and an agent behavior, like an artifact. It means that people may try to invent the relationship with a Companion, which is somewhere between the human relationship and object relationship.

The participants' reactions around Samuela (Companion developed by Telefonica used in this project) are a good example of this evolution, mostly of the personification of these emergent technologies. During her interviews, a 25-year-old participant stated that she would like Samuela (as her Companion) to live with her. She imagined Samuela (inside her screen, no matter which screen: computer or mobile phone or both) at parties she would organise at home. For example, they would be able to choose dresses together, as well as the music, and she may ask Samuela to perform several tasks at the same time, something which humans cannot do. Samuela would also be expected to disappear (by herself) when her owner does not need her anymore. This participant said: Samuela must 'feel' when the right moment is to appear and to disappear. In their interactive relationship, the user will extend and copy the structured behaviour as regards people and objects. So Samuela is expected to have a human behaviour and an agent behaviour like an artefact.

This seems to be a mental model people have of an ECA able to be a Companion. The semantics of these artifacts is emerging; only a long-term relationship with a companion would be able to explain this. The 'human side' of Companions expected by people seems to be similar to human strategies to capture audience involvement such as: interaction, humor and contextualization. Other elements of human strategies of communication and involvement such as body language, stance, facial expressions, use of space, and gesticulations, appear to be likely. It confirms recent theories concerning body communication, which consider the system of gestures as a complement to speech production [3].

This study also reveals that people attach particular importance to subtle signs like eye gaze or intonation, facial expressions and body gestures working together. For example if the Companion reacts when the user touches the screen, it is indicative of the Companion's human-like interest in the user. There is an expectation for this form of human behavior from the people.

The Companion proactive involvement in the interaction is perceived as a sign of empathy in this context. But since they remain objects usefulness is unsurprisingly also crucial to the acceptance of a Companion.

4. USER EMOTIONAL STATE AND AFFECTIVE CHANNEL.

Emotions are complex, and sometimes difficult to interpret. Humans are programmed to share their emotions; their body and face are the mediums they use intentionally or not to transmit information concerning their feelings. These are powerful signs of the emotional state of each communicating party in the protocols of face-to-face conversation. Some are more perceptible than others. When interacting with their owners, Companions should comply to these rules.

In Figure 4 the overall process of interaction is shown: first owner's emotional state is detected, then the Affective Channel of The Companion is put into action in order to respond appropriately to the user's emotional state. Physical expressions, physiological responses, gaze, and tone of voice express the user's emotional state. So do facial expressions: different facial muscles, eyes, mouth, eyebrows and forehead positions give useful

information about the people's emotional state. Changes in facial expressions during a conversation, are synchronized with what happens in the conversational exchange, giving extra information about the feelings of people regarding the experiences they are sharing. Facial expressions do not only enhance communication through speech, they can also compensate for breakdowns of the auditory channel, and for example when one of the parties has difficulties hearing the speaker. Gaze is an important feature of this process too; it has a regulatory function in the flow of conversation [7]. By following the user's gaze while interacting with a Companion the spatial organization of the interface can be optimized over time. Furthermore as a Companion resides on the screen where are also shown other elements requested by the user, gaze delineates the user involvement in various activities represented by elements presented on their screen. Physiological changes in skin conductivity permit measurements of arousal – one of the main components of emotions- the skin momentarily becomes better conductor of electricity when external or internal stimuli take place. It provides information about very subtle emotional changes. Measuring other physiological elements like blood pressure, pulse and heart rate also contributes to the identification of emotional states. Every factor must be analyzed in order to accurately specify the user's emotional state.

Recognizing the emotional state of people is useful, but it is not enough. Companions must be able to recognize it as a part of their face recognition protocol but it must also respond in a way which is easy, friendly, and above all relevant to the situation the people is experiencing. Companions must be able to interpret the whole emotional state. In the final system, it is envisaged that Companions will be able to capture their owner's involvement in a task with a sense of empathy.

In order to capture these ideas and facilitate their implementation, we introduce the concept of 'Affective Channel', defined as follows: The Affective Channel is the ability of a Companion to recognise and interpret people's emotions and respond to them empathetically through the use of appropriate voice qualities, facial expressions (including gaze), body language and semantics.

Because of the dynamism of the user's emotional state, the ability to quickly adapt and change its behavior is required, it contributes to the naturalness of the interaction and consequently to the success of the system. It means that the Companion must be able to learn the strategies humans use in their communication: verbal and non-verbal. As claimed by Mehrabian, among all human communication: 7% happens in spoken words, 38% happens through voice tone and 55% happens via general body language [17].

A big challenge for affective interfaces is to find ways to verify that emotions that emotions expressed by humans are correctly interpreted by Companions .

We believe that the implementation of AF is necessary to update an avatar to be a Companion. The next steps in this research are to investigate how people perceive the simulated emotions created by this channel, how they contribute to people engagement and what the impact of the emotional connection with a Companion on the performance of various tasks is.

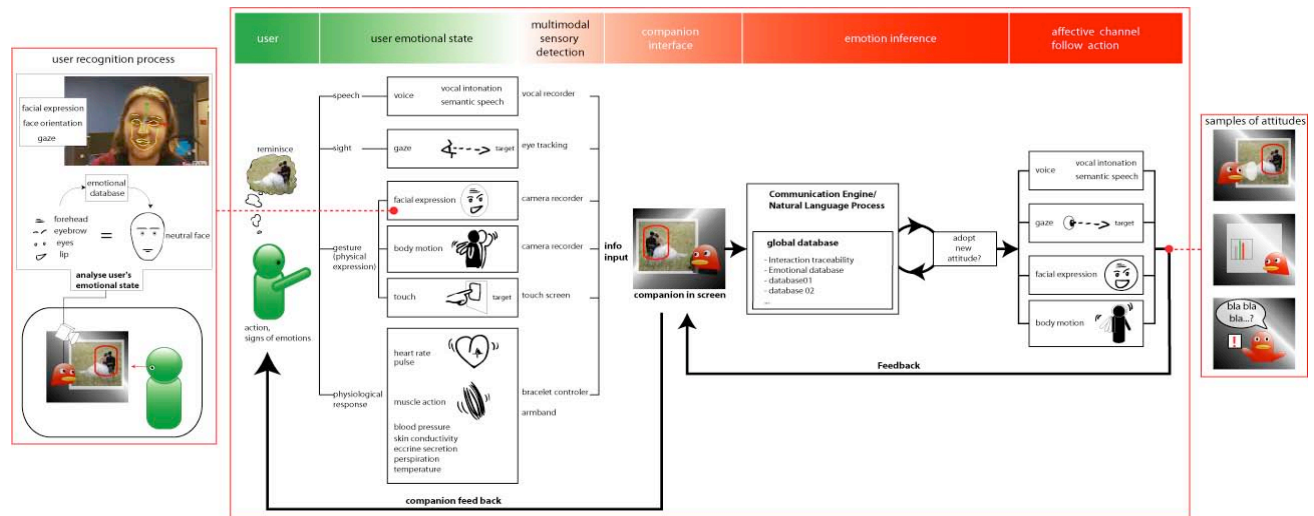


Figure 4 : Overall Emotional Process: People Emotional State detection by multimodal system, emotion inference to permit a variation on Companion attitude according to people behavior.

ACKNOWLEDGMENTS

This work is funded by the European Commission under contract IST 034434.

REFERENCES

- [1] Argyle, M., Dendom, A. (1967): The experimental analysis of social performance, Advances in Experimental Social Psychology, 3, 55-98.
- [2] Bates J. (1994) The role of emotion in believable agents Communication of the ACM, 37(7), 122-125.
- [3] Beattie, G. (2005). Visible Thought: The New Psychology Of Body Language. Routledge: London.
- [4] Benyon, D. and Mival, O. (2007). Introducing the COMPANIONS project: Intelligent, persistent, personalised multimodal interfaces to the internet. In Proceedings of Artificial Intelligence and Simulation of Behaviour Convention on Artificial and Ambient Intelligence, Newcastle University, United Kingdom
- [5] Bickmore, T. (2004) "Unspoken Rules of Spoken Interaction" Communications of the ACM 47(4): 38-44.
- [6] Bickmore, T., Cassell, J. (2005) "Social Dialogue with Embodied Conversational Agents" In J. van Kuppevelt, L. Dybkjaer, & N. Bernsen (eds.), Advances in Natural, Multimodal Dialogue Systems. New York: Kluwer Academic.
- [7] Cassell, J. (2000). "More than Just Another Pretty Face: Embodied Conversational Interface Agents." Communications of the ACM 43(4): 70-78
- [8] Creeds C. (2008) Affective Agents for Long –Term Interaction unpublished thesis Birmingham.
- [9] Damasio, A. (1994). Descartes'Error, Emotion, Reason, and the Human Brain , Putnam Publishing. .
- [10] Kelly George, 1991. The Psychology of Personal Constructs, Volume One : Theory of Personality. London : Routedge.
- [11] Mehrabian,A.(1971) Silent messages. Wadsworth, Belmont, California.
- [12] Norman, D. (2004). Emotional Design: Why we love (or hate) everyday things,. New York. Basic Books.
- [13] Paiva, A., Dias, J., Sobral, D., Aylett, R., Woods, S., Hall, L.E. and Zoll, C. (2005) Learning By Feeling: Evoking Empathy With Synthetic Characters. Applied Artificial Intelligence 19:235-266..
- [14] Picard, R. W.(1997). Affective Computing. Cambridge : MIT Press
- [15] Roa Seiler N., Benyon, D., & Mival O(2007) Emotional Companions. 3rd Workshop on Emotion in HCI of the Annual Human Computer Interaction Conference, Lancaster, United Kingdom.
- [16] Truchet Ph. (2004) La synergologie,Paris, Paperback.
- [17] Wilks Y., (2006, October) Artificial Companions as a new kind of interface to the future internet. Oxford Internet Institute, Research Report 13.

How to make agents that display believable empathy? An ethological approach to empathic behavior (Extended Abstract)

Ádám Miklósi
Eötvös Lóránd University
Pázmány P s 1c
Budapest, 1117 Hungary
+36 1 382 27 79
amiklosi62@gmail.com

ABSTRACT

The plan to engineer “empathic agents” is very ambitious, specifically because many researchers resist attributing such ability to any animal other than humans. Thus it seems to be paradoxical to have empathic agents but no empathic animals. This review suggests that affective computing may be boosting force for developing a unified approach to the evolution in empathic behaviour in living systems, and the knowledge gained could be utilised for designing machines that produce empathic behaviour which is believable for the human partners.

General Terms

Design, Human Factors, Theory

Keywords

Empathy, animals, evolution, inter-specific

1. INTRODUCTION

The scientific interest in empathic behaviour has a long story in the psychological sciences. Although it was often used as an explanatory term for many aspects of human behaviour, specific research was lacking. Among other factors the so called „cognitive revolution” in psychology facilitated research in this topic, especially by studying the developmental aspects of empathic behaviour in human children.

As animals (e.g. rats) have been often utilised as models of human behaviour, already in the 60ies researchers demonstrated „empathy-like” behaviour in rats. If a rat had observed a stressful con-specific that was suspended in the air by a harness, it moved to press the bar in order to lower the rat back to ground [9]. Although such laboratory investigations of „animal models” documented many situations when the behaviour of the observer animal could be interpreted as being driven by „empathy”, researchers were reluctant to argue for basic human-animal similarities in the underlying mechanisms. Even today many researchers avoid referring to empathy altogether when explaining some social behaviour, or they put the word in quotations.

Based on the arguments put forward by Darwin [2] on the continuity of „mental abilities and emotional expression” in evolution, interest has emerged to look for phylogenetic roots of human empathy in animals (for comparative review see 8).

2. DEFINITION OF EMPATHY

The definition of empathy suffers from problems that are common with terms that are used in everyday situations, and which are associated with specific human abilities. Even if researchers try to be objective, they have difficulties to avoid a human-centred view (anthropocentrism) that is often combined with „unconscious” introspective tendencies. Thus for many researchers empathic ability equals the „capacity for putting oneself in somebody’s place”. This approach is in many ways analogous to what is attributed to “mind reading”. It is not surprising that psychologists prefer to talk about understanding another’s emotional state, and refer to unobservable cognitive states when explaining the mechanisms controlling empathic behaviour in humans. This attitude is problematic because it is difficult to utilise such a research agenda in a comparative perspective if one is interested in the evolutionary origin of empathic behaviour.

For example, how can be utilise Hoffman’s [3] widely cited definition of empathy („any process where the attended perception of the object’s state generates a state in the subject that is more applicable to the object’s state or situation than to the subject’s own prior state or situation”) in the case of animals or especially artificial agents? It would be very difficult to argue for empathy in animals, and researchers would be accused of anthropomorphism, because there is no objective method for the comparison of inter-specific or inter-agent inner states.

In line with this criticism Preston and de Waal [8] use a somewhat extended definition for their „Perception-Action Model” of empathic behaviour. They argue that the „attended perception of the object’s state automatically activates the subject’s representations of the state, situation, and object, and that activation of these representations automatically primes or generates the associated autonomic and somatic responses, unless inhibited”. This definition is more useful because it refers not just to states but also to the behaviour (at least on the part of the subject). It is still problematic that in the discussion of empathy researchers move to quickly to the underlying (and unobservable) mental states of the mind and pay much less attention to investigate the mechanisms at the behavioural level. This situation creates often a terminological confusion in the use of categories; especially because researchers have a tendency of re-use value-loaded verbal expressions of human behaviour features (e.g. „sympathy”).

In the following we will follow Tinbergen's [10] receipt and look for possible functions of behaviours that might be interpreted as being „empathic“. Ideas based on evolutionary considerations will help us in this case.

3. EVOLUTION OF EMPATHIC BEHAVIOUR

Already Darwin attempted to explain the evolutionary origin of empathic behaviour. He and later other argued that such interactions might be very important in the mother-infant relationship [2], especially in mammals in which we find a very intensive and often long-lasting parental care. Empathic behaviour could mutually strengthen this bond and contribute to the survival of the offspring.

Interest in altruistic (“unselfish”) behaviour among animals [11] led to the assumption that inclusive fitness and reciprocal altruism could explain the evolution of empathy. In this model empathy is the mechanism, which facilitates the mutual relationship between the interacting partners. Thus this is an extension of the empathic aspects of mother-infant bond to relatives or even unrelated group members.

More recently, Preston and de Waal [8] argued for an even more general evolutionary function for empathic behaviour. They suggest that the phylogenetic explanation of empathic behaviour can be found in social animals in which the synchronic activities of the group are of vital importance. According to this scenario social animals would be at an advantage to display similar behaviours, that is, if one animal responds with a matching action after having perceived the behaviour of the other. They imagine a “perception-action mechanism” that is one of the basic features of neural organisation, and which provides the necessary “hardware” for the evolution of empathy. Thus behavioural matching is seen as a key to all phenomena that rely on state-matching or social facilitation, including empathy. It also follows that mammals, and more specifically group-living mammals should be able to show the basic features of empathic behaviour.

4. EMPATHIC BEHAVIOR IN ANIMALS

Although, animals have been often credited with some capacity for empathy in some scientific circles these ideas have not found their place in main stream research, and very often any claim for empathic behaviour was dismissed as being anthropomorphic.

Recent work on mice indicates, however, that animal models of empathy might have some general validity. After having observed object mice that received electric shock paired with a tone stimulus, subject mice displayed various forms of distress to the same tone and also to the tone-shock presentation [1]. This suggests that the behavioural (including vocalisation and odours) cues displayed by the objects were powerful stimuli in evoking similar inner state in the subjects. The same study also provided some evidence that the observed tendency to show empathic behaviour was associated with the general social attitudes of the mice. Mice from a strain with more social affiliative tendencies displayed also more empathic behaviour. In another experiment it was demonstrated that observing object mice in pain intensifies the response of subjects to pain [4].

Similar studies were also run with rhesus monkeys. Subject monkeys learnt to stop shocking object monkeys by pressing a bar, and this behaviour could be also evoked by showing pictures

of shocked monkeys [6]. Subject monkeys also withhold pulling chain for food if this also resulted in object monkeys being shocked [12].

Not surprisingly chimpanzees are in the focus of many studies on empathy. They also react empathically to pictures or videos showing con-specifics who display emotional behaviour (e.g. 7). Importantly, they also react to objects (e.g. needles used to injection) and to positive emotions when presented on pictures. However, in the case of the former the role of direct experience with needles cannot be excluded. In contrast to other animals studies so far only chimpanzees were found to respond also empathically to “positive” stimuli (e.g. play), that is, they displayed matching emotions.

Reading emotional expression of a group mate could also provide more direct information about the environment. In a social learning situation infant monkeys will avoid novel objects if they observe that the mother is looking fearfully at these objects [5]. In similar lines younger monkey can also learn the novel objects are not dangerous. In a reverse case infant monkeys encountering a novel object might look at the face of their mother. The phenomenon described as “social reference” provides some evidence that the emotion displayed by the adult influences the future behaviour of the infant toward the object. Both types of interactions play a major role in learning about the environment in human infants.

5. THE EMPATHIC CIRCLE

Research on empathy differentiates the “object” and the “subject”. Empathy is attributed to the subject if it matches its inner state to that of the object. However, this view is too simplistic for many reasons.

Both Hoffman's [3] and Preston and de Waal's [8] definition of empathy is problematic because they refer to the “*the attended perception of the object's state*”. Importantly, the subject has no means to perceive the object's “state”. It can only observe the behavioural cues which are associated with the actual inner state of the object, and can only infer the underlying inner state. This distinction is important because the aforementioned authors envision a deterministic relationship between the inner state and the behavioural cues. In reality however the relationship is more complex. First, there is no evidence that inner states are matched directly on a set of behavioural cues. Some inner states may be never revealed at the behavioural level. Second, behavioural cues are probably constrained in revealing exactly any inner state, and thirdly, “information” is also lost in the perceptual process. Thus the subject can only infer, judge, or approximate the inner state of the object through attending behavioural cues (visual, acoustic, chemical etc.).

Importantly, the “object-subject” view is based on a third person perspective, and empathy is visualised as a uni-directional process. However, based on the above definition it is very difficult to discriminate “empathy” from “communication”. Communication is also defined as having a sender, which by the means of specific behavioural cues, influences the behaviour of the receiver. This is especially problematic if we find that showing pictures of playing object animals releases playful behaviour from the subjects. What are the distinctive features of this interaction that differentiate communication from empathic behaviour?

A further problem is that it is not clear how the previous experience of the subject influences empathic behaviour. For example, seeing a needle could also release fear because own experience with a pain. In many experiments it is also not clear that the subject is exposed only to the actual emotional behaviour cues or they also witness how the object actually arrived at a given emotional state.

Finally, models of empathy reflect only rarely on the problem whether the object recognises the empathic behaviour of the subject. If empathy has an important role in inter-subjective relationships then there is a need of mutual recognition of empathic behaviour. This also follows from describing empathy as a form of altruism. One would expect that the behaviour of the subject gains a further advantage (also from an evolutionary point of view) if the object can recognise the empathic component. Only in this case can one assume that empathy provides a foundation for inter-subjective relationships.

6. CATEGORIES AND FUNCTIONALITY OF EMPATHY

Preston and de Waal [8] distinguished 6 levels of empathic behaviour (emotional contagion, sympathy, empathy, cognitive empathy, prosocial behaviour. They used three aspects to differentiate among these levels. They asked whether the empathic behaviour reflects a matching of the inner state, whether the subject actively acts on the object (e.g. “helping”), and whether there is some evidence for self-other distinction. As indicated above this and similar types of categorisations put an emphasis on the inner state matching and thus fail to distinguish some simpler forms of empathy from communicative interactions. Consider the case for the empathy of pain in mice cited above. One could assume that behaviour associated with pain functions in the same way as alarm signals. Alarm signals are produced by animals that witness some danger in their environment. They not only affect the behaviour of the other members in the group but also change their inner state. Visual, auditory, olfactory cues associated with pain could also have a similar effect on the subject. Interestingly, there are such alarm systems in fish. The attacked and physically harmed individual releases pheromones which initiate flight reactions from the group members.

In our view empathic behaviour can be separated from communication if we include that the subjects should pay some cost for being empathic. Thus “mirroring” or “matching” behaviour or “inner states” does not seem to fulfil criteria for empathic behaviour. By “cost” we mean that the actual matching of behaviour (or change in the behaviour) and/or inner state may not be in the own interest of the subject or, in reverse, it can be shown that by being empathic the subject investments in a personal relationship. Such cases usually involve interaction are often referred to as “consolation”, “helping”.

7. AFFECTIVE COMPUTING AND EMPATHIC AGENTS

Research on information technology has explored for long time how emotional interaction may facilitate human-computer or human-robot (=human-machine) interaction. This led to the emergence of a field called “affective computing” which draws its theories from the psychology of human of emotion and communication. Given the fact that the scientific understanding of

human emotions is quite limited, affective computing has a very ambitious research goal when it tries to explore the possibilities of mutual communication between humans and machines based on stimuli and behavioural cues that have emotional valence. “Empathic agents” are often defined as artificial systems that are able to engage in mutual empathic communication of humans. Today it seems that there are both theoretical/conceptual and practical problems in achieving this goal. Space does not permit to reflect on all issues, however, a few important aspects derived from the above discussion on evolutionary are listed.

For many researchers empathy equals mirroring of inner states. Importantly, this is not the case for human-machine interaction because the inner state of the artificial agents and humans do not match. The common origin of species (at least for the case of mammals) provided an important argument for common processes that underlie animal and human emotions and empathic behaviour. Thus the artificial system must rely on the ability to mimic both emotional states and empathy by displaying behavioural cues for the communicative interaction. (Since this discussion is based on evolutionary comparison no attempt is made to include the utilisation of linguistic interaction in empathic interactions.). Importantly however both the design of these communicative behaviours and the recognition of the human equivalents are problematic technically.

Affective computing relies on models of human emotions. The trend is also to utilise human-like behavioural cues for the interaction which might actually decrease believability, especially in the case of robots.

The evolutionary model of empathy is based on a similarity relationship between the object and the subject at the ecological level and on familiarity at the individual level. According to Preston and de Waal [8] the bodily similarity between the interacting partners and the perceived familiarity to the object individual increases the tendency for empathic behaviour in subjects. Both arguments seem to make difficult the design of human-machine empathic interaction.

Previous discussion also indicated that empathy is more than just noticing emotions of the other and reflecting on them. For example, the subject has to have some means to infer that the object is in the position to have similar past experience. In this case the situation is very different in the case of virtual agents are robots. Humans may attribute (“fantasy”) very similar capacities to a virtual agent, which looks and behaves very similar to them. In this case they do not only perceive the bodily similarity but also potential similarity in personal experience by mental attribution. However, even this may not be enough because virtual agents are probably never being confused with “real” agents. In the case of robots the believability is very questionable because the discrepancy between bodily similarities between objects and subjects, the mutual recognition (and display) of emotions, and the understanding that present robots are not in the position to have similar past experiences as their human partner.

8. ACKNOWLEDGMENTS

This research is supported by the EU FP7-ICT-2007 project (LIREC: 105554).

9. REFERENCES

- [1] Chen, Q., Panksepp, J. B., Lahvis, G.P. (2009). Empathy is moderated by genetic background in mice. *PLoS*, 4, 1-14.
- [2] Darwin C. (1999/1872). The expression of the emotions in man and animals. FontanaPress. London
- [3] Hoffman, M. L. (2000). Empathy and moral development: Implications for caring and justice. Cambridge University Press.
- [4] Langford DJ, Crager SE, Shehzad Z, Smith SB, Sotocinal SG, et al. (2006). Social modulation of pain as evidence for empathy in mice. *Science* 312, 1967–1970.
- [5] Mineka, S., Cook, M. (1993). Mechanisms involved in the observational conditioning of fear. *Journal of Experimental Psychology: General* 122, 23-38.
- [6] Mirsky, I. A., Miller, R. E., Murphy, J. V. (1958) The communication of affect in rhesus monkeys. *Journal of the American Psychoanalytic Association* 6, 433-441.
- [7] Parr, L.A. (2001). Cognitive and physiological markers of emotional awareness in chimpanzees. *Animal Cognition*, 4, 223-229.
- [8] Preston, S.D., de Waal F. B.M. (2002). Empathy: its ultimate and proximate bases. *Behavioural Brain Sciences*, 25, 1–72.
- [9] Rice, G. E. J., Gainer, P. (1962). "Altruism" in the albino rat. *Journal of Comparative & Physiological Psychology* 55, 123-125.
- [10] Tinbergen N. (1963). On aims and methods of ethology. *Z. Tierpsychol.* 20, 410–433.
- [11] Trivers RL. (1971). The evolution of reciprocal altruism. *Q. Rev. Biol.* 46, 35–57
- [12] Wechkin, S., Masserman, J. H., Terris, W., Jr. (1964). Shock to a conspecific as an aversive stimulus. *Psychonomic Science* 1, 47-48.