Initial Framework for Empathic Synthetic Tutors

Work package	Tutor's empathic behaviour and dialogue generation engine (WP5)
Task	Framework for Empathic Tutors Behaviour Generation (T5.1)
Dissemination Level	Public
Publishing date	Contractual M12 | Actual
Deliverable	5.1
WP / Task responsible	Ana Paiva – INESC-ID
Contact person	Ginevra Castellano, UoB
Contributors	HWU, UOB, JacobsUni, INESC-ID
Short abstract	Starting from a review of the literature on empathy and combining it with a specification of data collected from the studies, the creation of an initial version of the framework for empathic tutors' behaviour generation is presented.
Keywords	Empathy; Verbal and Non Verbal behaviours; Empathic Tutor; Framework

This work was partially supported by the European Commission (EC) and was funded by the EU FP7 ICT-317923 project EMOTE (EMbOdied-perceptive Tutors for Empathy-based learning). 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.

Contents

1. INTRODUCTION **1**

2. EXECUTIVE SUMMARY **2**

3. RELATED WORK **4**
- 3.1. WHAT IS AN EMPATHIC TUTOR? 4
- 3.2. EMPATHY IN VIRTUAL AGENTS AND SOCIAL ROBOTS 4

4. REQUIREMENTS FOR THE CREATION OF SYNTHETIC EMPATHIC TUTORS **7**
- 4.1. VERBAL BEHAVIOUR 10
  - *Teacher dialogue moves* *11*
  - *Student dialogue moves* *12*
  - *Empathic Dialogue Moves* *13*
  - *Game Dialogue Moves* *14*
- 4.2. NON-VERBAL BEHAVIOUR 14
  - *Gestures* *14*
  - *Posture (mood)* *15*
  - *Emotive expressions* *15*
  - *Gaze* *16*
  - *Acoustic emblems* *16*
  - *Colour emblems* *17*
- 4.3. GAME AND APPLICATION DEPENDENT EVENTS 17
- 4.4. DATA ANALYSIS (GAZE STUDY) 19

5. INITIAL FRAMEWORK FOR THE CREATION OF SYNTHETIC EMPATHIC TUTORS **22**
- 5.1. HOLISTIC VIEW 22
- 5.2. LEARNER MODELLING 23
- 5.3. EMPATHIC MODELLING 24
- 5.4. BEHAVIOUR SELECTION AND GENERATION 25

6. CUSTOM TOOLS FOR THE STUDY AND DEVELOPMENT OF TUTORS **26**
- 6.1. THALAMUS 26
- 6.2. NUTTYTRACKS 26
- 6.3. WIZARD-OF-OZ TOOLS 27

7. CONCLUSION AND FUTURE WORK **31**

8. REFERENCES **32**

1. Introduction

A central concept in learning is empathy and emotion. Emotions are a part of human essence, while empathy is often seen as the capability for perceiving, understanding and experiencing another person's emotions (see Deliverable 3.1). Such a notion has been central in the study of human relations and is regarded as one of the major elements in social interactions between humans. In order to embed empathy in learning environments one needs to perceive the learner's affective state, model it, reason about how the learner feels, and respond emotionally to the current situation. Also, empathy is considered to be one of the most important qualities of a good teacher. The empathic qualities that enhance the interaction between teacher and learner are linked with the engagement with the tasks and so, with increased learning performance. Empathic interventions refer to the adaptation of the tutor's behaviour to affect-related states emerging during the learning experience (such as engagement with the task) and with the tutor itself.

This Deliverable serves the purposes of EMOTE project by providing a framework for the construction of empathic tutors, in particular their behaviour. The present document relates to Work Package (WP) 5, in particular to its goals of addressing the problem of creating engaging empathic behaviours for embodied tutors. WP5 comprises Task 5.1 (the initial task here reported) which starts with an initial phase where a generic framework for empathic behaviour generation is developed. The framework embeds empathy in two types of processes. The first process is associated with the notion of empathic appraisal and is the mechanism by which the tutor appraises the learning and emotional state of the learner, assesses it, and at the same time, generates an emotional state. In here, we will present a first computational architecture that will capture this process. Then, the second process is the empathic response generation which will use empathic dialogue strategies, facial expressions and body gestures to convey empathic behaviour. The framework will capture the different characteristics of these two types of processes and will propose a framework for embodied tutors with empathic capabilities.

2. Executive Summary

This document aims to provide a set of initial requirements for the EMOTE project in the development of a general framework for the creation of a synthetic empathic tutor. To design the creation of an synthetic empathic tutor, knowledge from four main areas were addressed: an overview of related work regarding empathic tutors and empathy in social robots; the requirements for the creation of synthetic tutors, comprising both verbal and non-verbal behaviour types; the presentation of a generic framework for the creation of empathic tutors; and the tools for the study and development of tutors. The present document is structured into sections according to the following organization:

Related work

Section 3 presents an overview of literature that will serve to provide a layout of the initial generic framework for empathic behaviour generation. The addressed topics are:

- Mechanisms for empathic intervention that encourage enjoyment of the learning process and facilitate the development and maintenance of social bonding between the learner and the tutor;
- Relating to the extensive literature on Empathy from Deliverable 3.1, the present Deliverable gathers a definition and implications about empathic tutors;
- Conditions for the creation of social agents that exhibit empathic behaviour are mentioned and linked to empathic processes;
- Studies on empathy in social robots are integrated in this overview tailored to EMOTE purposes, with a focus on considerations about creating computerized empathy.

Requirements for the creation of synthetic empathic tutors

Understanding, more than solving a problem, is an important goal in education, De Vries, Lund and Baker (2002). *Section 4* tries to combine requirements for the creation of a synthetic emphatic tutor in a way that will lead learners to ho beyond problem solving. These requirements are organized in different behaviour types, where the source of information regarding both timing and form is included. Section 4 presents the initial results. This way, the main goals of these Sections, are:

- To present a brief summary of the studies done by the EMOTE team to establish a preliminary design of an empathic tutor;
- Encompassing the notion that interactive learning environments have different communication forms, different types of behaviour are explored (Dillenbourg & Baker, 1996);
- Combining different forms of human communication; verbal and non-verbal dialogue schemes are presented as the initial requirements for the creation of a synthetic emphatic tutor;
- To present gaze study results from an EMOTE study as a first set of conclusions that are extracted from the video analysis of different types of gaze behaviour.

Initial framework for the creation of synthetic empathic tutors

In *Section 5*, all the aforementioned requirements are organized into an initial general framework that supports the described empathic artificial tutor. The framework captures the different characteristics of empathic appraisal and empathic response mechanisms for embodied tutors with empathic characteristics. Therefore, in this section we will provide:

- A holistic view of the framework is presented;
- The empathic learner framework is expanded presenting details on concepts such as the learner modelling and the empathic appraisal mechanisms;

Custom tools for the study and development of tutors

In *Section 6*, the Wizard-of-Oz (WoZ) environment is presented, as well as the other tools developed by so far for both tutor creation and future studies. In this Section we will describe:

- Some developed software for interactive agents is described;
- WoZ environment is presented;
- Collected data from the Pilot WoZ study are discussed. The opportunities and the challenges of this study are addressed.

With view to the EMOTE general objectives and the description of work of WP5, future work is considered and conclusions are drawn.

3. Related Work

In EMOTE, the mechanisms for an empathic intervention that encourage enjoyment of the learning process and facilitate the development and the maintenance of social bonds will be explored. In D3.1, the concept of empathy and social-emotional bonding was explored. Also, in D3.1, socio-emotional bonding is examined in the teaching/learning environment, and in the context of artificial agents. In this Section, we will further explore these notions in the form of computational models that lead to such empathy. This related work will help to inform the initial framework for empathic tutors.

3.1. What is an empathic tutor?

An empathic tutor can be understood as a social agent that is able to understand others, their intentions, motivations and feelings (Cooper, Brna & Martins, 2000). Humans can also feel empathy not only for their peers but also for fictional characters, robots or other types of social agents. As such, there are two distinct ways to look at empathy in social agents (Paiva, 2011). The first one is to consider that the empathizer is the user and the agent is what triggers the empathic process in the user. In this case, the social agent does not necessarily need to simulate empathic processes or behave in an empathic manner, but it is designed to evoke empathic responses from the user. These agents have been employed, for example, in scenarios where it is important to persuade the user to select the “right" action (Paiva et al., 2004). In this section, the focus will be on the second form of empathy.

In order to create computational social agents that can demonstrate increasing levels of empathic competence, there is a need to be able to simulate dynamics of the empathic processes computationally. To build empathic agents, they should simulate empathic mechanisms, that is, they should recognize the users’ affective states, context and exhibit empathetic outcomes. The recognition of affective states of others is unanimously acknowledged as the first step of any mechanism of empathy. Additionally, understanding the actions that others are taking is equally relevant, because there are situations where the user deliberately inhibits the expression of certain emotions, which means that empathy can only be achieved by perceiving the overall situation (e.g., user's actions and context). The state-of-the-art systems for user perception still have some limitations, especially in regard to affective states that can be extremely subjective and user dependent. Acceptable performance rates for affect recognition can be achieved by combining several modalities such as vision, audio, physiological signals, context interpretation, and so on. In addition to all that was mentioned, agents should also be able to display their internal states in a way that is understood by others (Zeng, Pantic, Roisman, & Huang, 2009).

3.2. Empathy in virtual agents and social robots

In the past few years, empathic virtual agents have been extensively studied in several domains. In the field of virtual agents, empathy is a research topic that has received a lot of attention (Bickmore & Picard, 2005; Brave, Nass, & Hutchinson, 2005), especially in learning environments (Arroyo et al., 2007; McQuiggan & Lester, 2007; Pour, Hussain, AlZoubi, D’Mello, & Calvo, 2010). In the field of social robots, researchers have only recently started to assess the effects of endowing robots with empathic behaviour. One possible reason for this is that there are less technological constraints when inferring the user's affective state in interaction scenarios with virtual agents than it is with human-robot scenarios. Usually, while interacting with virtual agents, the user is sitting in front of a computer screen and there is often the possibility of selecting predefined dialogues to inform the agent of the user's emotional state. The interaction with robots tends to be more open-ended and thus perceiving the user's affective state is more challenging (Bickmore, & Picard, 2005). In this sub-section, since one of EMOTE's end targets is to build an empathic robotic tutor, a brief overview of the literature on this most challenging form of creating computerized empathy is presented .

The early studies in this field focused on basic principles considered as part of the empathic process such as joint attention or mimicking the user's affective state, a particular aspect of empathy also designated as emotional contagion (Hess, & Fischer, 2013). For example, the work by Kozima, Nakagawa and Yano (2004) takes a developmental perspective and tries to capture some of the more basic empathic processes that are developed during infancy. As such, they have created a model for the development of eye contact and shared attention with two robotic platforms, aspects considered important for the establishment of empathic relations. "Eye contact" activates certain brain regions specialised in processing social information such as the face, gaze, theory of mind and empathy (Kozima, Nakagawa, & Yano, 2004). The model is based on the notion that those mechanisms develop in children at early stages (from 6 months old, to 2 years old) and such developmental processes should be reflected by the robotic platforms. To achieve this ambitious goal, the researchers have explored a scenario with two robotic platforms (Infanoid and Keepon) and studied how children interact with them at different stages of development.

Tapus and Mataric (2007) proposed another model of empathy to be used in the context of Socially Assistive Robotics. Several capabilities were identified, for example, recognising and interpreting another's emotional state, the capability of processing and expressing emotions and the capability of communicating and perspective taking. These empathic processes were modelled based on the work by Davis (1983), where empathy is seen not only through its processes, but also by the outcomes it leads to.

In the area of emotional contagion, Hegel, Spexard, Wrede, Horstmann and Vogt (2006), conducted a study with an anthropomorphic robot that recognises the user's emotional state through speech intonation and then mirrors the inferred state using a corresponding facial expression. The results suggest that users who interacted with this version of the robot found the robot's responses more adequate both in terms of appropriateness to the social situation and timing, than subjects who interacted with the robot without affective expressions.

In another study (Riek, Paul, & Robinson, 2010), a robot with the form of a chimpanzee head mimics the user's mouth and head movements. When interacting with this robot, most subjects considered the interaction more satisfactory than participants who interacted with a version of the robot without mimicking capabilities. Cramer, Goddijn, Wielinga and Evers (2010), studied how empathy affects people's attitudes towards robots. In a between-subjects design, two groups of participants saw a four minute video with an actor playing a cooperative game with an iCat robot. The experimental manipulation consisted in programming the robot to express empathic behaviour towards the actor in an accurate or inaccurate manner (i.e., incongruent to the situation), depending on the control group. In this study, there was a significant negative effect on user's trust in the inaccurate empathic behaviour condition. Conversely, participants that observed the robot displaying accurate empathic behaviours perceived their relationship with the robot as closer (Cramer, Goddijn, Wielinga, & Evers, 2010)

More recently, a scenario where an iCat robot is able to play chess with children while reacting emotionally to the moves played on the chessboard was developed to study empathy in Human-Robot Interaction (Leite, Castellano, Pereira, Martinho, & Paiva, in press). After every child's move on the chessboard, the robot provides empathic feedback on that move by conveying facial expressions influenced by the child's affective state and the state of the game. To infer children's affective states, an SVM-based Affect Recognition System returns the probabilities of children's valence (positive, negative and neutral). In addition to the empathic facial expressions, if the child's affective state is negative below a certain threshold, the robot also displays social supportive behaviours. The developed empathic model had a positive impact in long-term interaction between children and the robot (Leite, Castellano, Pereira, Martinho, & Paiva, in press). In a repeated interaction study where the same group of children interacted with the robot five times, children's ratings of social presence, engagement, help and self-validation remained similar after several weeks, contrasting with the results obtained in an initial long-term study where the robot was not endowed with the empathic model. Video observation and interviews were also employed to further analyse these results, suggesting that over time, users were even more aware that their actions influenced the iCat's behaviour. The results also showed that children felt supported by the empathic robot to a similar extent to what, in general, children feel supported by their peers. The same authors developed a variation of this scenario in which the iCat robot acts as a social companion to two players in a chess game (Leite, Pereira, Mascarenhas, Martinho, Prada, & Paiva, 2013). However, in this case, the emotions of the human participants were captured using perspective-taking, i.e., putting oneself in the other’s place and imagining how he or she feels. In other words, instead of using the multimodal affect detection system, the robot used the context of the game to “put itself into the shoes of the players". After assessing the game from the perspective of one of the players, the empathic responses of the iCat were modulated by the relationship with the players (companion or opponent). Towards the companion, the iCat reacted in a very empathic way, providing encouraging comments and being enthusiastic towards his/her moves. On the other hand, to the other player, the iCat's relation was mostly neutral, and as such, the comments were mainly factual and less emotive. The results of this study indicate that users towards whom the robot behaved empathically perceived the robot as friendlier, when compared to participants to whom the robot displayed neutral behaviours only.

A relevant finding in empathy research is that empathy is positively linked to prosocial behaviour (Hoffman, 1982; Eisenberg & Fabes, 1990). In another study using the iCat robot, Saerbeck, Schut, Bartneck and Janse (2010), investigated the effects of social supportive behaviours of the robot on student's learning performance and motivation. Among other supportive behaviours, the robot conveyed basic empathic responses through its facial expressions. The results indicate that simple manipulations in terms of the robot's supportiveness, while maintaining the same learning content, increased students' motivation and scores.

4. Requirements for the creation of synthetic empathic tutors

The requirements established by the user studies detailed in Deliverable 2.1 were combined with the presented requirements for the creation of a synthetic emphatic tutor. With that in mind, different modalities for such behaviour (verbal and non-verbal behaviours) were explored. To establish preliminary designs of this type of tutor, a set of studies (mock-ups and WoZ) were designed and conducted. The collected data will be used to define an initial version of these requirements.

The summary of the setups and main goals of each study is shown in **Figure 1**.

Mock-ups			Wizard-of-Oz
Scenario 1 UoB & UGOT	**Scenario 1** UGOT	**Scenario 2** INESC-ID	**Scenario 1** HWU
Study how **teachers adapt their pedagogical strategies** in a learning situation to respond to students' needs.	Test **the influence of the positioning of the tutor** and the student in relation to each other.	Investigate tutor-students and student-student **interaction**; emerging **curriculum topics.**	Test the **WoZ environment** and **interface**.

Figure 1. Summary of studies.

The studies were performed with Scenarios 1 and 2 learning environments. Scenario 1 focus on local map reading as an individual activity, while Scenario 2 is a collaborative one based on the Enercities game. In Deliverable 2.1, a full description of all these studies is included. In the present Deliverable, considerations relating to protocol and studies design are presented as following:

- **Mock-up 1**

The first Mock-up was conducted by UoB and UGOT. The main goal was to study how teachers adapt their pedagogical strategies in a learning situation to respond to students' needs. The developed activity included geography syllabuses topics in a map reading task. Although slightly different tasks were used in England (UoB) and Sweden (UGOT), the activity's main goal was to access map skills and building activities in the context of local maps. The mock-up sample is described in **Figure 2**. Each session lasted 30 minutes and was audio- and film-recorded. A final interview was conducted with the teachers and students to have their input on the activity.

	England	Sweden
Teacher	One teacher	One teacher
Student (11-13 years old)	3 students	2 students

Figure 2. Mock-up 1 sample description.

- **Wizard-of-Oz (WoZ)**

Based on the first mock-up study, a WoZ study was performed in Scotland (HWU) with Scenario 1. The main goal of this study was to test the WoZ environment and interface. This way, the same geography topics were included and the task format was similar. The study environment included a wizard's desk, an interactive touch table, sensors, and the robotic embodiment.

A WoZ study was performed with three students and a wizard. Pre- and post-questionnaires were given to the students in order to assess their knowledge on related topics (e.g., identify map symbols) and evaluate their learning experience. A questionnaire was also given to the wizard to understand his feelings about the student-interface interaction.

The three sessions of the WoZ were audio- and film-recorded and served to get feedback from other teachers about the user-interface, pedagogical approach, and tutor behaviour.

- **Mock-up 2**

A second Mock-up was performed in Sweden (UGOT) under Scenario 1 activity. The mock-up's three main goals were (1) to test the scenario with teachers and students in order to adapt the final application in terms of technical design and appropriate difficulty levels; (2) to gather utterances and behavioural data from the teachers and students in order to adapt the robotic tutor's perceptive capabilities as well as pedagogical approach, and (3) to test the influence of the positioning of the tutor and the student in relation to each other (at two sides of the table opposite each other, or two adjacent sides). The last condition was considerated important after analysing the results from gaze study described in Section 4.4.

The sessions were held under two different testing conditions: In Condition 1, the teacher stood on the student's left side, and in Condition 2, the teacher stood in front of the student. The sessions' mean duration was fifteen minutes. The sample consisted of two teachers and twelve students and all the sessions were video- and audio-recorded.

- **Mock-up 3**

Mock-up 3 was conducted by INESC-ID with Scenario 2. This was the first study performed within the collaborative scenario. The mock-up study's main goals were to analyse the group interaction, possible curriculum links and emerging issues about the game. In order to cover these goals, two conditions were raised: In Condition 1, one the teacher played the Enercities game with two students, and in Condition 2 three students play Enercities game.

The Enercities game was played in the multi-touch table (**Figure 3**) and the study sample included one teacher and thirty-one students. After playing the game, each student answered individually a quantitative questionnaire related to the game and the team. Each student was individually interviewed in a semi-formal format afterwards in order to collect qualitative data. At the end of each session, the teacher also answered a qualitative subjective questionnaire regarding collaborative learning within the sessions with students. The total time per session was one hour, and all the sessions were audio- and video-recorded.

Figure 3. The setup for the Mock-up 3 study in Portugal.

The aforementioned initial studies were carefully designed in order to maximize the amount of useable captured data. The goal is to find out how both teachers and students act during an interactive learning scenario, so recurring and tendential behaviours will be analysed, both in regards to the communication between the participants, and also as to their own lines of thought, both in the student and teacher role.

A link between the revised literature and studies observations will be assessed. Initial conclusions are extracted and explained below. **Figure 4** shows a summary of how the different types of behaviour for the artificial empathic tutor are categorized. For each behaviour type we also present the source of information regarding both timing and form for each of the behaviour types (verbal and non-verbal behaviours). Timing refers to when and how long each type of behaviour should occur, while the form refers to how the behaviour should be modelled.

In an educational context, meaning is constructed through a continuous dialogue, which takes place in the teaching, and learning processes in all its various forms (Vygotsky, 1978). Human tutoring is documented to be a very effective method of instruction due to the combination of different behaviour types (verbal and non-verbal behaviours) (Graesser, Wiermer-Hastings, Wiermer-Hastings, & Kreuz, 1999). Interactive learning environments presuppose communicative acts (Dillenbourg & Baker, 1996), being the most studied form of communication within this context. Nevertheless, the ubiquities of non-verbal communication in an educational environment have been postulated (Deutsch, 1974). Hence, it will be considered that both verbal and non-verbal behavioural expressions address the requirements for the creation of a synthetic emphatic tutor.

Behaviour Type	Timing (source)	Form (source)
Game Dialogue Moves	Videos (T&S)	Videos (T&S)
Teacher Dialogue Moves	Videos (Teacher)	Videos (Teacher)
Student Dialogue Moves	Videos (Students)	Videos (Students)
Empathic Dialogue Moves	Videos (T&S)	Videos (T&S)
Gestures	Videos (T&S)	Literature
Postures	Videos (T&S)	Videos + Literature
Gazing	Videos (T&S)	Videos + Literature
Acoustic Emblems	Relate to Dialogue Moves	Literature
Colour Emblems	Relate to Dialogue Moves	Literature

Figure 4. The different behaviour types for the artificial empathic tutor (T&S represents Teacher & Students).

The following subsections define the established requirements of the coding schemes for the various types of empathic behaviour. It explicitly defines what will be addressed in the studies, and which will be the sources of information for both the timing and form components of each behaviour. Both were distinguished because it was considered that the mock-ups and WoZ studies should give most of the hints regarding when and why some behaviour should happen, while for some types of behaviours, the studies don't give us enough information as to how it should be accomplished. Furthermore, there is already a vast corpus of literature on some of these behaviour types - which were linked.

4.1. Verbal Behaviour

De Vries, Lund and Baker (2002), point out that learning is intimately related to discourse and talking. This way, a salient feature of natural tutoring consists of generating dialogue moves (processes that create dialogue actions) that assist students in active constructions of explanations, elaborations, and mental models of the material. This active process of constructing explanations is seen as critical for learning, having a greater impact comparing to learning by exposing (Graesser, Wiermer-Hastings, Wiermer-Hastings, & Kreuz, 1999).

Teacher dialogue moves

Dialogue moves that tutors generate have been documented (Graesser, Person, & Magliano, 1995; Alexander, 2007) to nurture the collaborative building of explanations in learning context. According to EMOTE project objectives and literature review on the theme (Graesser, Wiermer-Hastings, Wiermer-Hastings, & Kreuz, 1999; Alexander, 2007) the following dialogue moves of the teacher will be assessed:

- Pose Problem

The teacher poses the initial problem (e.g., "What is a contour line?"). This dialog move serves to initiate a dialogue with the student in a learning context discourse.

- Re-question

The teacher re-asks the original main question when the dialog is moving off course or on a tangent (e.g., "Remember that we want to know what a contour line is").

- Feedback

In a teaching context, exchange of feedback not only motivates students, but also gives them the quality of their discourse contribution. Three types of feedback are addressed: Positive (e.g., "Great"), Negative (e.g., "Not quite, no") and Neutral (e.g., "Uh-uh") Feedbacks.

- Pump

The teacher pumps the student for more information (e.g., "Anything else?"). Pumping serves the function of exposing knowledge of the student to construct content by themselves.

- Prompt

The teacher supplies the student with discourse content and prompts him/her to fill in a missing word, phrase or sentence (e.g., "A hill has contour lines that are ...."). Prompting is a scaffolding device for students who are reluctant to supply information. This way, it is expected that students supply more content as they progress in learning the domain knowledge.

- Hint

When the student is having problems answering a question or solving a problem, the teacher gives hints by presenting a fact, asking a leading question, or reframing the problem (e.g., "Try to think back to our previous discussion on contour lines"). Hints are usually indirect speech acts, so they run the risk of being missed by an insensitive student.

- Elaboration

Elaborating is an assertion made by the teacher with the objective to fill in missing information that is regarded as important, but that could be missed by the student (e.g., "Contour lines can show hills and valleys"). In essence, this information is transmitted from the teacher to the student, as opposed to having the student generating the information.

- Splice/Correction

The teacher splices correct information when the student produces a contribution that is obviously error-ridden (e.g., (after misunderstanding) “With valleys the numbers on the contour lines decrease”). The teacher needs to be able to recognize errors and splices in order to do this.

- Summary

The objective of summarizing is to recap an answer to a question or solution to a problem (e.g., “We've looked at how contours and line of sight can be represented on a map”). The function of this dialogue move is to succinctly codify a lengthy, multi-turn, collaborative exchange when a question is answered or problem is solved.

- Link

Between problems, the teacher links the dialog to not disrupt it and de-mechanize the learning discourse (e.g., “Let's skip to the next question").

- Into Reality

Along the dialog the teacher tries to make the learning activity more engaging by adding context and bringing the approach contents to life (e.g., “Do you see that street? Is this one on the map"). This type of dialogue move will facilitate the connection between the curriculum and student reality.

- Silence

Silence time along the interaction will be considerate when in the absence of the other dialogue moves.

Student dialogue moves

Epistemic dialogues involving activities such as explanation and argumentation have been recognized as potential vehicle for conceptual understanding. This type of dialogue is worth attention on educational context once problem solving does not necessarily imply understanding (De Vries, & Lund, & Baker, 2002).

Also, Benjamin S. Bloom (1949) created a taxonomy of educational objectives (Krathwohl, 2002). This taxonomy is a framework for the classification of what is expected of, or intended for, students to learn as the result of instruction. Besides being a measurement tool of learning, it also served as:

- A common language across people about learning goals to facilitate communication in the educational context;
- As a basis for determining a particular course or curriculum;
- As a mean to determine the congruence of educational objectives;
- To determine a panorama of the range of educational possibilities.

Bloom's taxonomy was conceived into two dimensions: The Knowledge Dimension (classifies four types of knowledge that learners may be expected to acquire or construct); and The Cognitive Process Domain (represent a continuum of increasing cognitive complexity) (Krathwohl, 2002).

For the purposes of the EMOTE project, the Cognitive Process Domain will serve as a way to categorize students' dialog moves along the learning activities comprising notions of epistemic dialogue. The categories of this domain are organized in **Figure 5**.

Empathic Dialogue Moves

An important aspect of the EMOTE project is to determine specific ways in which the robot can convey empathy. The set of empathic sound emblems in the next deliverable of WP3 (D3.2) will play a significant role to achieve this goal. At the same time, preliminary phrasings of verbal empathic expressions have been suggested and are currently being tested in some of the ongoing tests on social vs. task engagement at UoB.

Remember	Understand	Apply	Analyse	Evaluate	Create
Recognizing	**Interpreting**	**Executing**	**Differentiating**	**Checking**	**Generating**
Identifying	Clarifying	Carrying out	Discriminating	Coordinating	Hypothesizing
Recalling	Paraphrasing	**Implementing**	Distinguishing	Detecting	**Planning**
Retrieving	Representing	Using	Focusing	Monitoring	Designing
	Translating		Selecting	Testing	**Producing**
	Exemplifying		**Organizing**	**Critiquing**	Constructing
	Illustrating		Finding	Judging	
	Instantiating		Coherence		
	Classifying		Integrating		
	Categorizing		Outlining		
	Subsuming		Parsing		
	Summarizing		Structuring		
	Abstracting		**Attributing**		
	Generalizing		Deconstructing		
	Inferring				
	Concluding				
	Extrapolating				
	Interpolating				
	Predicting				
	Comparing				
	Contrasting				
	Mapping				
	Matching				
	Explaining				
	Constructing models				

Lower order thinking skills → **higher order thinking skills**

Figure 5. The cognitive process dimensions (Bloom, 1949).

Game Dialogue Moves

There are certain kinds of dialogue moves that relate solely to the game. These dialogues are generally questions and assistance in using the interface, or understanding the game rules. They can be things like "this button doesn't seem to work" or "why can't I build this?" or "we can only get that in the next level".

The main difficulty in processing this kind of dialogue is that it requires the understanding and interpretation of everything that the students are saying, mostly without any kind of physical interaction. This type of dialogue is, however, included in the coding scheme as it seems important for us to know how often this type of dialogues occur, and especially, we believe that in a later stage, by associating the collected data on this topic with the rest of the annotations, it may be possible to understand when such kind of difficulty is happening (i.e., if there are 5 consecutive clicks around the same area in a short period of time - that may symbolize a difficulty towards the game interface).

4.2. Non-verbal behaviour

As Johnson, Rickel and Lester (2000) mentioned, one primary role of a tutor is to provide feedback on students' actions. In addition to providing verbal feedback, a tutor also uses nonverbal communication to influence the student, giving an empathic and engaging context to learning. In some learning situations, using nonverbal behaviour constitutes a preferable way to provide feedback because it is less obstructive than verbal comments. Also, in face-to-face dialogues, subjects employ a variety of nonverbal signals to help regulate the conversation, complement their verbal utterances and give a context of empathy to learning (Johnson, Rickel, & Lester, 2000). Therefore, the requirements for non-verbal behaviour are listed and explained below.

Gestures

Gestures have been explored by various authors (e.g., Ekman, & Friesen, 1969; McNeill, 1992; Kendon, 2004). The latter is regarded as one of the most popular sources in literature for categorizing and defining the different types of gestures. He defined that gestures can be grouped into five different categories: iconic, metaphoric, deictic, beat, and emblem gestures. These are better defined in **Figure 6**.

Gesture Type	Description
Iconic	A gesture is iconic if it bears a close formal relationship to the semantic content of speech.
Metaphoric	Metaphoric gestures are similar to iconic gestures in that they present imagery, but as an abstract concept, (e.g., knowledge, language itself).
Deictic	Deictic gestures are pointing movements, among others.
Beat	Beats are defined as movements that do not present a discernible meaning and they can be recognized positively in terms of their prototypal movement characteristics.

Figure 6**.** The gesture coding scheme.

Gestures are generally connected with dialogue, by both extending and complementing what a speaker is saying. It is therefore important that the developed tutor is able to hold a correct relationship between what it is saying and gesturing. Because of this and the fact that the goal is to develop a tutoring player and not just a peer player, gestures will be annotated only in the teacher's videos. This annotation will provide the timing component of the gestures for the tutor. As to the form of the gestures, literature from both on human gesturing (McNeill, 1992), and also on robot gesturing (Cassell, Vilhjálmsson, & Bickmore, 2001; Gibet, Lebourque, & Marteau, 2001; Chafai, Pelachaud, & Pelé, 2007) will be considered.

Posture (mood)

There have been previous studies that relate directly to the future usage of posture in EMOTE, namely the work by Beck and Colleagues (2013), who have been exploring the emotive expression of the NAO robot through postures. In the EMOTE tutor however, posture is planned to be used mainly for the expression of mood. Gebhard (2005), defines mood as "medium-term affect, which is generally not related with a concrete event, action or object. Moods are longer lasting stable affective states, which have a great influence on human's cognitive functions". Mehrabian (1996), has associated some moods (which he calls temperaments) to octants of the Pleasure-Arousal-Dominance (PAD) emotional space and recently the work by Kleinsmith and Bianchi-Berthouze (2007), connects the affective dimensions of Valence (Pleasure), Arousal, Potency (Dominance) and Avoidance with several characteristics of posture. This gives evidence that it is possible to model posture as a continuous function of the mood of the tutor. The aforementioned work by Mehrabian (1996) leads to the suggestion of considering Exuberant, Bored, Dependent, Disdainful, Relaxed, Anxious, Docile and Hostile as a coding scheme for posture data collection.

Posture can also be used for rapport. As referred by Lowman (1995), mimicry in college teaching is an important variable of interpersonal learning behaviour. In the same direction, Bernieri (1988) provides a study on the rapport between high-school students teaching each other a list of imaginary words in high-school classes. Since the target group in EMOTE concerns early high-school students (detailed in Deliverable 2.1), rapport behaviour will be further considered in order to find out how it can co-exist with the expression of mood through posture in the empathic robotic tutor.

Further analysis of the data collected from studies will confirm, refine or withhold the proposed coding scheme for both mood and rapport modelling from posture.

Emotive expressions

It is important to distinguish between emotive expressions and mood from the previous points. Gebhard (2005) defines an emotion as a *"short-term affect, which is usually bound to a specific event, action or object, which is the cause of this emotion"* (p. 31). It is therefore considered that Emotive Expressions in EMOTE are different from the mood expression, presented in the previous section. With this in mind, key emotional facial expressions will be annotated based on the extensive work of Ekman and Friesen (1975).

The annotation will provide us with data that will be used to model when the emotive expression animations should be played. These animations should override the current posture, and then, once the animation has finished, the tutor should fall back into its underlying posture. This kind of blending between emotive expression animations and mood postures has already been done by Pereira, Prada and Paiva (2012). Ribeiro and Paiva (2012), has also already provided some clues about how such animations should be designed in order to be used in robots.

Gaze

Previous experiments in human-robot interaction have emphasized on the importance of the gazing behaviour in empathic robots (Kozima, Nakagawa, & Yano, 2004). Also, Pereira, Prada and Paiva (2012), studied and developed a gaze system for a multi-party interaction with a robot via a multi-touch table, which shares many similarities with the scenarios that will be developed in EMOTE project. Their work highlights how gazing should shift between two environments: the social environment (players, external events) and the game environment (the multi-touch table). The shifting between these depended on the two main agent's affective states (focusing on its own move versus observing the move of other players), and also on interaction events with both the multi-touch table, and the surrounding environment (through auditory and visual stimuli).

In the EMOTE project, a study focused on collection of data for modelling the gazing behaviour had already been conducted. The results of the study are detailed further in this document, in Section 3.4. The coding scheme for annotating the gazing behaviours is defined in **Figure 7.**

Student	Teacher
Looks at Teacher	
Looks at other Student	Looks at Student
Looks at Task	Looks at Task
Mutual Gaze	
Gaze Elsewhere	

Figure 7. Gaze coding scheme.

It is, of course, difficult to isolate gazing as a sole expression. Gazing is in fact used for many different purposes, and it is, both functional and expressive behaviour (Takayama, Dooley, & Ju, 2011). These authors have emphasized on this dualism of functional versus expressive motion in robots, and the fact that in a lot of cases they actually work together. Gazing can not only be used for expressing thoughts, engagement, or some emotional expressions, but is also used for the functional process of observing the environment whenever the observing hardware (camera) is rigidly attached or contained in the head or eyes of the subject.

With this in mind, process of Gaze behaviour annotation regarding collected data will be combined with other coding schemes (like from dialogue and emotion expression).

Acoustic emblems

Acoustic emblems are included in the set of empathic emblems previously mentioned. These emblems will be further defined and explained in the next Deliverable of WP3 (D3.2).

Colour emblems

Colour is considered to be the most salient, resonant, and meaningful visual feature of those seen in early vision (Hilbert, 1987). Kaya and Epps (2004) pointed that colour is an inseparable part of everyday life and its presence is evident in everything that is perceived. Among essential roles that colours implies, the role in communication is (Suk, 2006). Regarding human communication, colours have been associated across cultures with emotion expression and emotion recognition. Nevertheless, due to culture specific variables (e.g., language), cross-cultural differences in the meaning of different colours can occur. (Hupka, Zaleski, Otto, Reidl, & Tarabrina, 1997).

Regarding social robot interaction, especially with robots that do not have an expressive face for affective expression and must rely on other methods of expression, multimodal methods of expression, including colour, have been implemented (Bethel, 2009). In the context of interaction with robots, some authors have performed studies using NAO's eye colour to strengthen the emotions shown by the robot (Greczek, Swift-Spong, & Mataric, 2011; Haring, Bee, & André, 2011; Kaya & Epps, 2004).

For the EMOTE project, colours will be considerate, nevertheless, due to the aforementioned literature, colour emblems will need to be further discussed.

4.3. Game and Application dependent events

The final type of annotation we are defining is events regarding to the application (the game or the simple map reading tool) actions and flow. We need to be able to create a context within the application activity in which to fit all the previously mentioned annotations.

These events, shown in **Figure 8,** can easily be given by the application log, which registers everything any learner does. Some of the events are common for both scenarios, and others are specific for each game of the scenarios.

Event Types	MapActions	*User interacts with the map interface.*
	TaskActions	*User performs actions that advance the progress in the task execution.*
	MapTools	*User interacts with extra map tools.*
	DialogueActions	*Users click on buttons that are used for trigger a dialogued interaction with the tutor.*
	EnercitiesExamine	*User is examining construction, upgrades and policy features in the Enercities game.*
Common Events	TaskActions.StartActivity	*User starts the task.*
	TaskActions.EndActivity	*User terminates the task.*
	MapActions.Click	*User clicks anywhere on the map.*
	MapActions.Zoom	*User changes the zoom level of the map.*
Scenario 1	MapActions.Walking	*User drags map to simulate the stick man walking.*
	MapActions.WayPointCreated	*User dropped a pin on the map.*
	MapActions.DragTo	*User drags object/marker/character to location.*
	MapActions.Select	*User selects an object on the map.*
	MapTools.CompassShow	*User activated the Compass tool.*
	MapTools.CompassHide	*User deactivated the Compass tool.*
	MapTools.DistanceShow	*User activated the Distance tool.*
	MapTools.DistanceHide	*User deactivated the Distance tool.*
	MapTools.AddOverlay	*User activates information layer on the map.*
	MapTools.RemoveOverlay	*User deactivates information layer on the map.*
	MapTools.MapKeyShow	*User activated the Map Key which explains what each of the map symbols mean.*
	MapTools.MapKeyHide	*User deactivated the Map Key.*
	DialogueActions.RequestNextInstruction	*User clicks on the NextInstruction button.*
	DialogueActions.RequestRepeatInstruction	*User clicks on the RepeatInstruction button.*
	DialogueActions.RequestHelp	*User clicks on the Help button.*
	DialogueActions.RequestTutorial	*User clicks on the Tutorial button.*
	DialogueActions.RespondWithInfo	*User sends text info to the robot.*
	TaskActions.ChooseInformant	*User selects an informant.*
Scenario 2	MapActions.Pan	*User moves the viewpoint of the map around.*
	EnercitiesExamine.UrbanMenu	*User opens the Urban building menu.*
	EnercitiesExamine.ElectricMenu	*User opens the Electric building menu.*
	EnercitiesExamine.GardensMenu	*User opens the Gardens building menu.*
	EnercitiesExamine.EconomyMenu	*User opens the Economic building menu.*
	EnercitiesExamine.MayorMenu	*User opens the Mayor building menu.*
	EnercitiesExamine.ConstructionType	*User selects a building type.*
	EnercitiesExamine.ConstructionOnCell	*User selects an empty cell of the map for construction.*
	EnercitiesExamine.Cell	*User selects an occupied cell of the map for examination.*
	EnercitiesExamine.Policy	*User selects a City Hall policy for examination.*
	EnercitiesExamine.Upgrade	*User selects a feature upgrade in a building for examination.*
	TaskAction.ConfirmConstruction	*User confirms construction of a building in a specific cell.*
	TaskAction.ImplementPolicy	*User confirms implementation of a City Hall policy.*
	TaskAction.PerformUpgrade	*User confirms upgrade of a building feature.*
	TaskAction.SkipTurn	*User skips its turn.*
	TaskAction.EndOfLevelStart	*Current level ends, and the End-of-Level menu is displayed.*
	TaskAction.EndOfLevelEnd	*The End-of-Level menu disappears, and the next level Starts.*
	TaskAction.EndOfGameStart	*The End-of-Game menu is displayed.*
	TaskAction.EndOfGameEnd	*The End-of-Game menu disappears.*

Figure 8. Definition of the different Game Events.

4.4. Data Analysis (Gaze Study)

As mentioned previously, a first mock-up with scenario 1 with an initial version of map reading activity was held in England (UoB) and Sweden (UGOT). The aim of the mock-up was to study how teachers adapt their pedagogical strategies in a learning situation to respond to student's different skills and capabilities. A side aim was to determine the teachers' and students' gaze during the activity.

All sessions in both Sweden (UGOT) and England (UoB) were filmed, transcribed, and annotated. Although the sample size is quite small, eye-gaze analyses was assessed to understand patterns of gaze during the interaction in the activity. The description of the participants is shown in **Figure 9** and the results are presented below.

	Sweden	**England**
Teacher	1 (female)	1 (male)
Student	2 (females)	3 (1 male, 2 females)

Figure 9. Mock-up 1 participants' description.

In order to establish gaze patterns, video analysis was held to extract conclusions. For this, each video interaction was analysed accordingly to the following gaze coding scheme (see **Figure 10**).

Student	Teacher
Looks at task	Looks at task
Looks at teacher	Looks at student
Mutual Gaze	

***Figure 10*.** Analysed gaze coding scheme.

According to the analysis and results, it can be concluded that in both countries the average percentage of gaze interaction target the map-reading task both with teachers and students. **Figure 11** translates these results, showing that both teacher and student from both countries spent most of the time looking at the task when comparing to other types of gaze coding (e.g., teacher look at student).

Figure 12 shows the average frequency of annotations in each coding scheme category in both teachers and students from each country. It can be seen that teachers look more frequently to the task and the student during map reading activity. It can also be extracted that students establish low frequency of gaze with the teacher and looking at the task. Regarding mutual gaze, its frequency is shown to be the lowest in both countries.

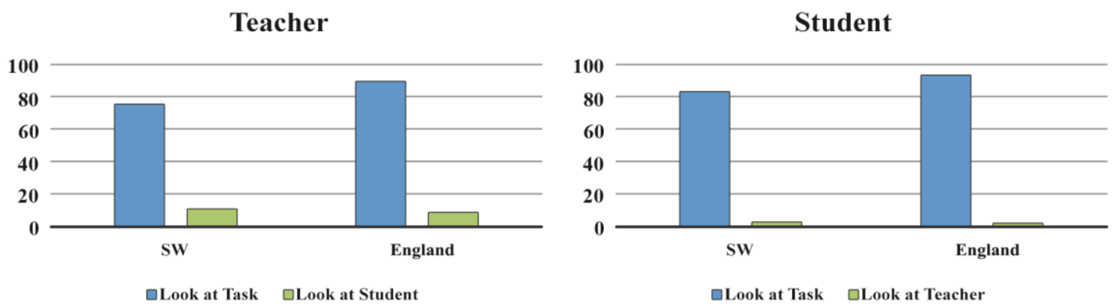

Figure 11. Teacher & student gaze interaction.

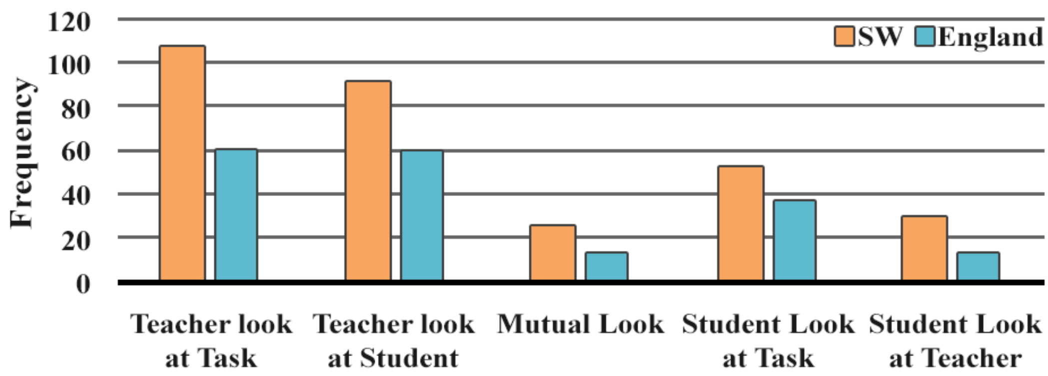

Figure 12. Average percentage of annotations in each category.

Regarding gaze video analysis, the average duration of each annotation in each category for both countries was also analysed. **Figure 13** shows these results in seconds. According to this Figure it can be said that the categories with less average duration of gaze, correspond to mutual gaze: teachers looking at the student and student looking at the teacher. On the other hand, the categories with more average duration correspond to the time in which both the teachers and students looked at the task.

It can be concluded that although students look at the task with less frequency compared to teachers, the average duration of this look is higher. This means that students look at the task for long periods of time. It can also be said that teachers look more frequently at the task for shorter periods of time, making more gaze shifts.

	Sweden	Portugal
Teacher looks at Task	12.7	23.1
Teacher looks at Student	1.8	2.0
Mutual Gaze	1.1	1.0
Students look at Task	29.6	50.1
Students look at Teacher	1.1	0.6

Figure 13. Average duration of annotation categories (in seconds).

Summarizing the video analysis of gaze, it appears that both teachers and students from England (UoB) and Sweden (UGOT) may look more frequently and longer at the task compared to mutual gaze and each other gaze interaction. Due to this, mock-up 2 in Sweden (UGOT) was held changing both student and teacher physical position in the game activity to see if any change occurs regarding gaze patterns. Future video analysis will enable to extract conclusions regarding gaze behaviour study.

5. Initial framework for the creation of synthetic empathic tutors

This Section addresses an initial framework for empathic behaviour generation. This framework embeds the definition of two types of processes:

- **Pedagogical empathic strategy selection**: This is the mechanism by which the tutor appraises the learning (learner modelling) and emotional state (empathic appraisal) of the learner, assesses it and at the same time generates an emotional state;
- **Empathic response generation**: Empathic behaviour that conveys the selected pedagogical empathic strategies through dialogue acts with adaptive nonverbal behaviour will be generated by the tutor.

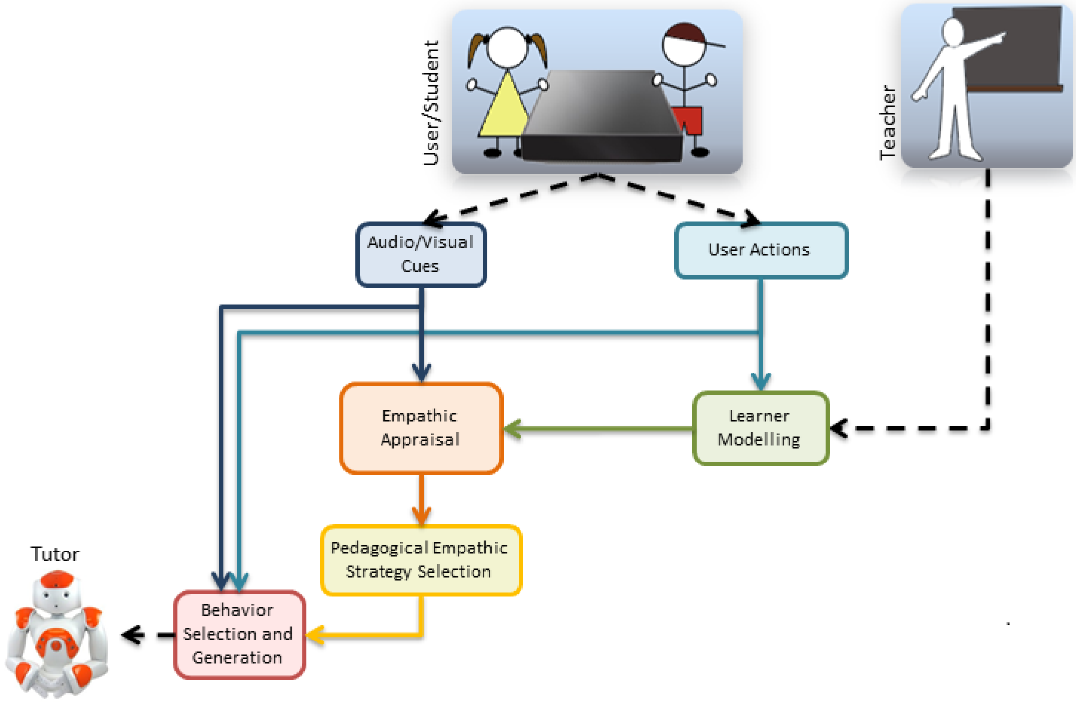

Figure 14. The Empathic Learner Framework - A Holistic View.

5.1. Holistic View

A holistic view of the framework is presented in **Figure 14**. The users that interact with the tutor system are the students and the teacher. The next paragraphs briefly describe the components of the framework:

Auditory/Visual Cues are considered to be all the stimuli that arise from the sensing system of the robot. These can include both visual and auditory stimulus.

User Actions are all the events that are triggered through the interaction with the learning application. It can include not only game events but also information that is inputted by the user through the application (including answers to yes/no questions).

Learner Modelling is the process of modelling the student and what the student has learned. It contains the description of the learning scenario (tasks), the corresponding skills and knowledge that are expected to be acquired, and also the progress and assumed knowledge of each student. This component can be held in memory so that its information can be used on successive interactions.

Empathic Appraisal this process uses the perceived audio-visual cues and the current student model to appraise an empathic emotion that is used to steer the empathic pedagogical strategies (Rodrigues, Mascarenhas, Dias, and Paiva, 2009).

Behaviour Selection and Generation takes input from both audio and visual cues and the empathic pedagogical strategies. The strategies define the kind of behaviours and expressions that are meant to be exhibited at each moment, while making use of the visual cues to provide physical synchronization with the student.

On a typical usage of the system, the teacher would first define the Learning Scenario to be used by the application, which includes both the learning tasks and the corresponding acquired skills. Next, the Students interact with the system, by playing the Learning Scenario. They provide input to the system via Audio and Visual Cues, captured by microphones and cameras, and also by User Actions. These User Actions correspond to interactions with the application, and could both be application actions, and input from text lists or yes/no questions. The actions that the students perform are used for progressing in the learning scenario, by solving learning tasks. During this progression, the learner model is used by an Empathic Appraisal module, which first predicts what the student's emotional state should be based on the current performance in the learning scenario. This predicted emotion is integrated an emotion that the tutor visually recognizes in the user through its perceptual system. The resulting empathic emotion is then used to select and configure the Pedagogical Empathic Strategy, which will, in turn, be used to generate the appropriate Empathic Behaviour.

5.2. Learner Modelling

This learner modelling mechanism (**Figure 15**) is first based on *Teacher* input, providing the learning scenario, and the corresponding Tasks and Domain Knowledge (TDK). The TDK represents both the Tasks that compose the learning scenario and the knowledge and skills that the student is expected to acquire for each completed task.

On each User Action (events generated by interaction with the learning application), the Student Task Performance (STP) and Student Domain Knowledge (SDK) are updated, taking into account the defined TDK. At any point, the Teacher can also provide direct input to the Student Model in order to make some adjustments or corrections to what the tutor is modelling.

Both the STP and SDK can be stored in memory so that the tutor can use the acquired information in successive interactions. The STP is especially useful to recall to the specific tasks that were completed or not, and information about how they were completed (time, hints needed, etc.). The SDK is kept separate because, as it contains the skills and knowledge that the tutor assumes the student has acquired, it can actually be used with different learning scenarios (and thus, with different STPs).

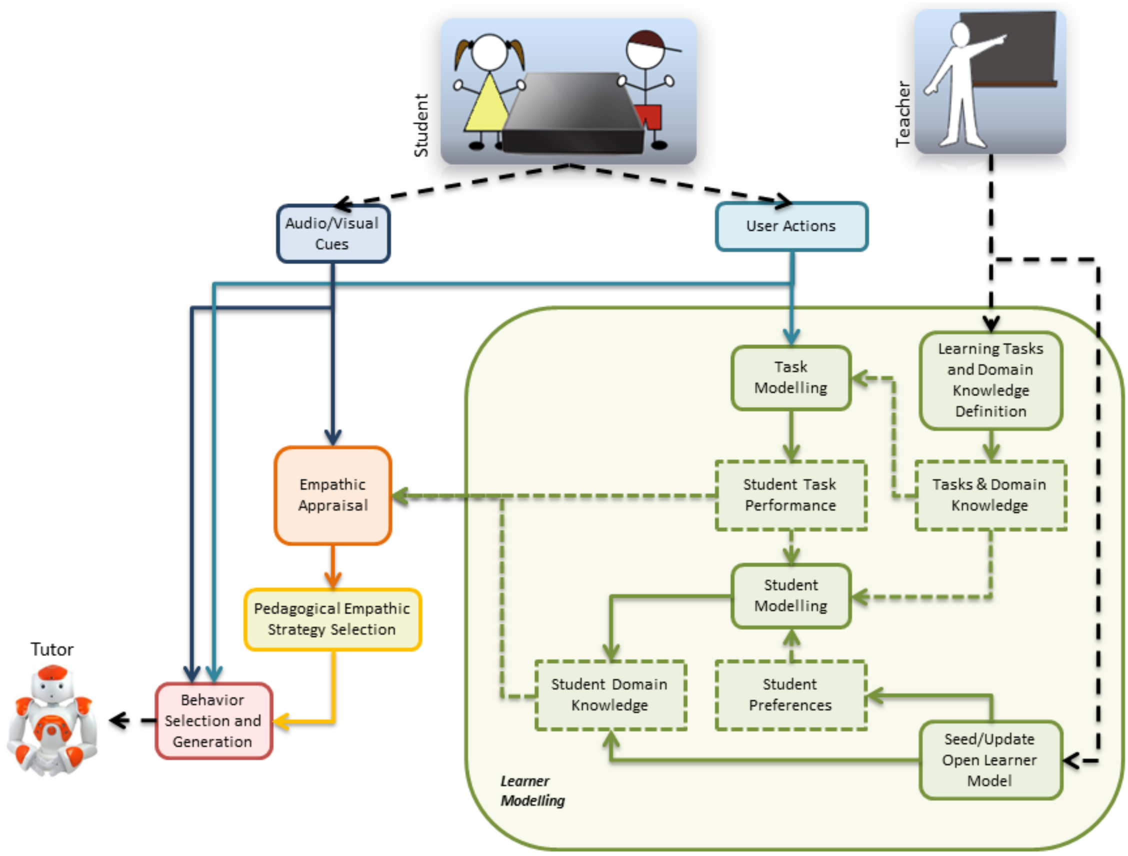

Figure 15. The Empathic Learner Framework - Detail on the Learner Modelling mechanism.

5.3. Empathic Modelling

The *Empathic Appraisal* mechanism follows the model proposed by Rodrigues, Mascarenhas, Dias, and Paiva (2009), and is presented in more detail in **Figure 16**. The Visual Cues captured by the tutor's perceptual system are used to infer what the student is feeling, which is represented as a set of candidate emotions. These emotions are then integrated with the emotion elicited from the Self-Projection Appraisal mechanism. This mechanism elicits an emotion by simulating what is expected for the user to feel, by considering the current situation and the current event. The resulting Empathic Emotion is then used to select an adequate Pedagogical Strategy that takes into account not only the result of the learner modelling process, but also the empathic appraisal.

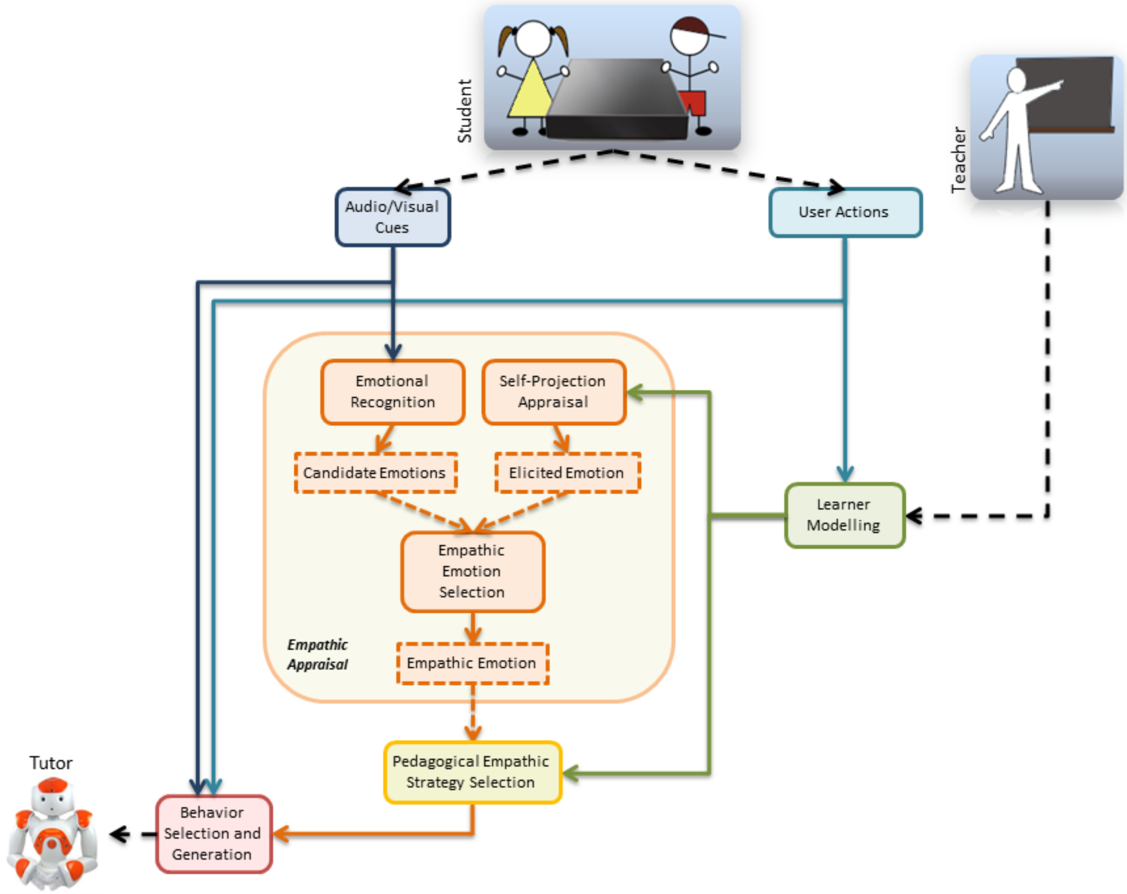

Figure 16. The Empathic Learner Framework - Detail on the Empathic Appraisal mechanism.

5.4. Behaviour Selection and Generation

The final result of the framework that is presented to the students is actually the output of a Behaviour Generation system. This system will have to manage both the verbal and non-verbal communication of the tutor, keeping it interactive with the students, and making it feel like the tutor is actually present. This is currently summed up as Behaviour Selection and Generation, as part of the behaviour may be pre-designed (animations, pre-scripted speech, postures, etc.) while other may be synthesized during the interaction (gazing behaviour, adaptive dialogue, etc.). The Dialogue Management mechanism is further detailed in Deliverable 6.1. It will output behaviour actions into Thalamus, which will be used as an integrative component to make it possible to render simultaneous multi-modal behaviours on the robot using Nutty Tracks for the final animation synthesis and mixing. Both Thalamus and Nutty Tracks are better described in Section 6 of the current document.

6. Custom tools for the study and development of tutors

6.1. Thalamus

Thalamus is the backbone of the tutoring agent that is being developed for EMOTE. It is a framework based on the Censys model for creating modular interactive characters, which means that it makes it easier to build an interactive character which can be composed of different modules (Ribeiro, Vala, & Paiva, 2013a). Following on the gathered requirements, Thalamus makes it easier to develop modules that can individually deal with the different types of functions and behaviour (like those presented in *Section 4*), while still acting in synchronization with each other. This tool is being used as the underlying mechanism of the EMOTE architecture, and is therefore better described in Deliverable 6.1.

6.2. NuttyTracks

NuttyTracks is a symbolic real-time animation system that can be used to design and run animation on an interactive character (our tutor), independently of its embodiment, and using animation tools commonly used by professional animators (**Figure 17**). It is inspired by the traditional art of animation, and its goal is to bring this secular art closer to the technologies currently used on interactive characters by having animation artists work closely with programmers in the quest for creating the illusion of life in robotic characters. It is meant to address the *interactivity* versus *animation quality* dichotomy that we can find in different types of artificial characters. On one hand, one can find high quality animations in characters from animated movies (Thomas, & Johnston, 1981). However, the animation in these characters is designed specifically for the timing, environment, and performance that were hard-coded by the movie's scriptwriters. On the other hand there are robots that can interact with humans in the real physical environment, which exhibit a level of interactivity much higher that an on-screen virtual character. However, the animation of such robotic characters is, in most cases, either very poor in terms of quality, or poor in terms of interactivity.

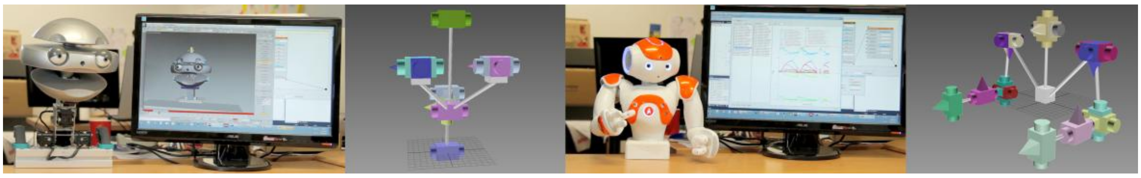

Figure 17. NuttyTracks being used with the EMYS and the NAO robots, and Autodesk 3dsmax.

The advantage of using NuttyTracks in EMOTE is that the robotic tutor will be able to use a mix between pre-designed and procedural, automated animation, in which the blending and parameterization of all the animation sources and variations can be controlled by high-level processes in the tutoring agent. Furthermore, as this is achieved through a unique symbolic animation pipeline inspired by the well-succeeded computer-graphics pipelines, it also supports the use of the same animation programs with different embodiments, along with seamlessly transitions between robotic and virtual bodies.

All this technical backline of NuttyTracks was thought to bring interactive robot animation up to the level of professional animation, by integrating with the animation tools commonly used by animation artists like Autodesk 3dsmax or Autodesk Maya. This makes it possible to have animation artists not only immediately animate the robot using their own tools, but also have them take part in the way that the animation is shaped by the tutor’s empathic artificial intelligence (Ribeiro, Dooley, & Paiva, 2013b).

6.3. Wizard-of-Oz Tools

Wizard-of-OZ (WoZ) is an effective technique in Human Computer Interaction (HCI) studies where an interactive agent, which is not yet fully autonomous, is fully or partially controlled by a remote human wizard. However the participants who are interacting with the agent are not told that the agent is being remotely controlled. The wizard may be tasked to control one or many parts of the agent such as speech recognition and understanding, affect recognition, dialogue management, utterance and gesture generation and so on. Studies have shown that users “go easy'” on computers during interaction and therefore interaction with “wizarded” systems are at the level of complexity that can be learned and emulated (Pearson, Hu, Branigan, Pickering, & Nass, 2006). The basic WoZ setup and a pilot study were presented at SIGDIAL 2013 conference and ECGI 2013 workshop (see related publications below). A short video of the pilot study can be found at http://www.macs.hw.ac.uk/~amol/emote/icsr-Final.mp4.

Wizard-of-Oz environment

The WoZ environment consists of the wizard's desk, the interactive touch table, sensors, and the robotic embodiment as shown in **Figure 18**. The wizard is seated in a different room away from the learner.

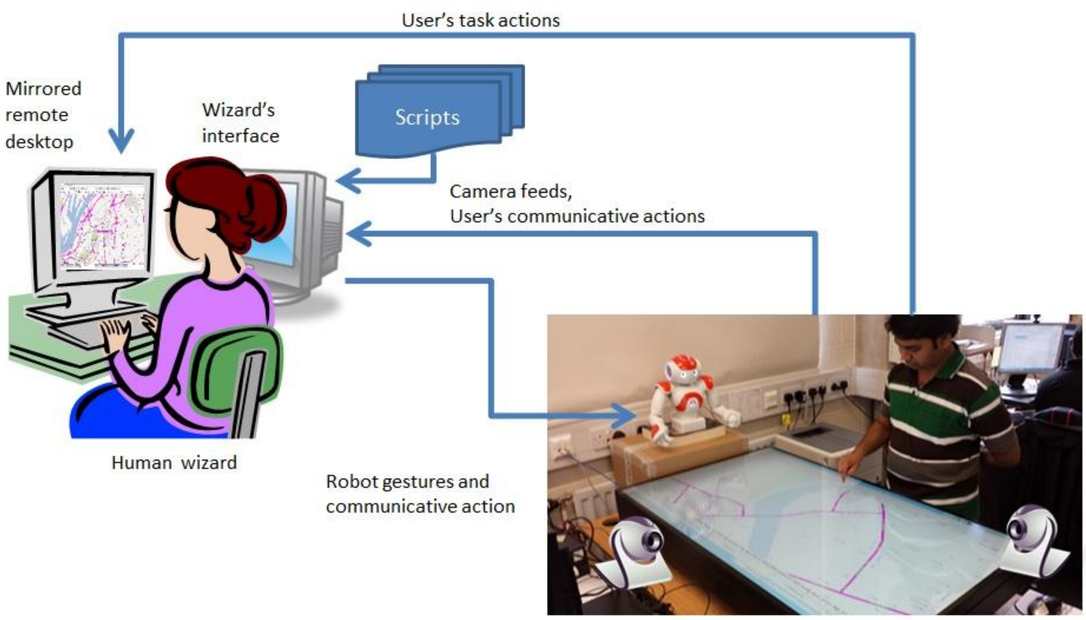

Figure 18. Wizard-of-Oz environment.

Wizard's desk

The wizard's desk consists of two display screens. The touch table display at the user end is mirrored on to one of the displays at the wizard's desk using which the wizard can observe the learner's activities related to the educational application. Another display contains the wizard Interface, a software application that allows the wizard to interact with the learner (see **Figure 19**). The Wizard Interface consists of four panels: task control, information, learner response and operations. In the task control panel, the wizard is able to choose a task plan for the learner and access the tool and curriculum scripts (XML file). The tool script contains information on how to use the tools that are at the disposal of the learner. For instance, to create a marker on the map, one has to click on the appropriate tool and click on the map and so on.

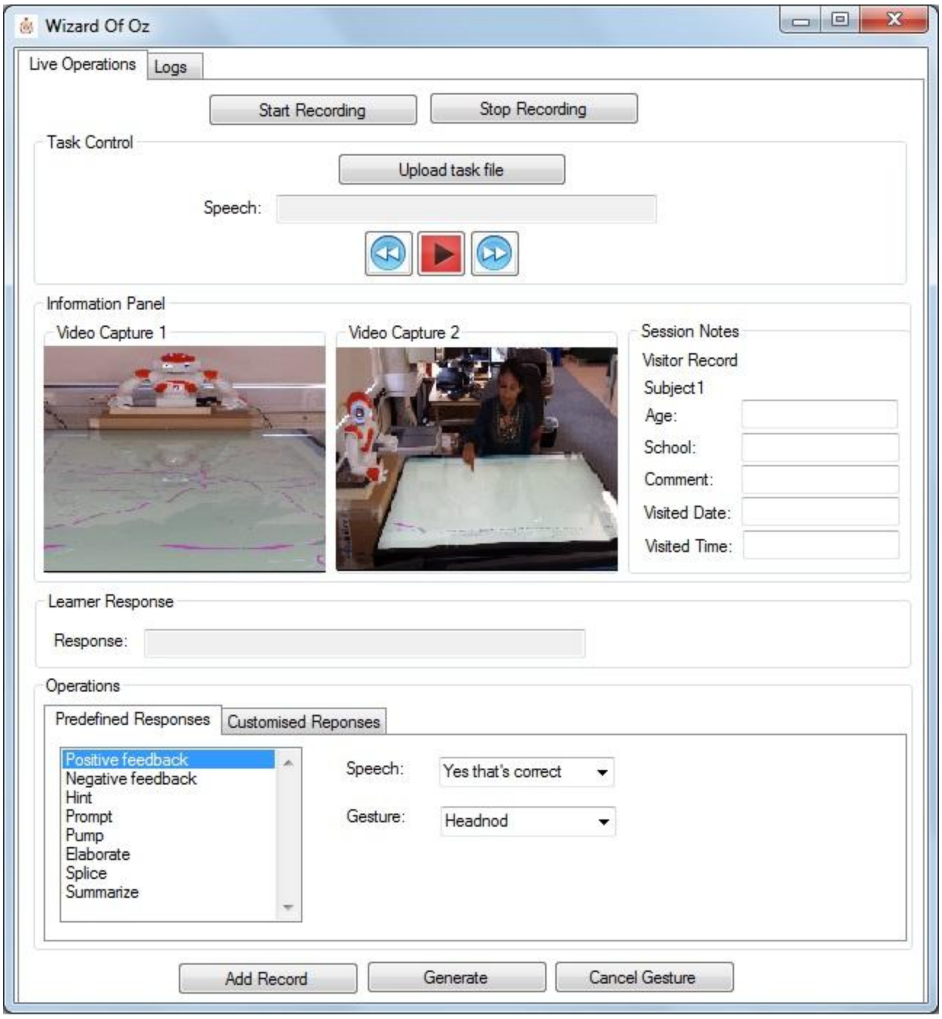

Figure 19. Wizard's Interface.

The curriculum script contains information on the skills that the learner needs to exercise or develop during his interaction with the system. For instance, in order to identify the right compass direction, the system will present the mnemonic phrase "**N**aughty **E**lephants **S**quirt

Water" in various forms such as a hint, question, pumping move, etc. to provide support to the learner. The *information panel* contains the video feed from two cameras. This will allow the wizard to determine the affective state of the learner. The learner's response to the agent's utterances (such as answering questions in the curriculum scripts) is displayed in the *learner response panel*. Finally, the *operations panel* provides options for the wizard to respond to the learner based on the tools and curriculum scripts. These responses are either customised or predefined. The customised responses help the wizard to execute robot movements at a finer granularity (individual head, arm movements) and predefined responses contain a list for combined predefined speech, sounds and behaviours (Bhargava, Janarthanam, Hastie, Deshmukh, & Aylett, 2013).

Touch table

The interactive touch table is a 55-inch Multitaction[1] table capable of sensing multiple touch events simultaneously. The educational application is displayed on the table surface. A map based application has been developed to teach learners basic and advanced map reading skills (see **Figure 20**). The touch interface allows the learner to use touch to click, drag and zoom the map. The application has two panels of GUI objects such as buttons and text boxes namely, the tools panel and the interaction panel. The tools panel consists of tools that the learner can use to manipulate the map, while using the interaction panel the learner can interact with the tutor. Some of the tools that are currently available are to get grid references for a position on the map, dropping markers on the map, change map types, etc. For instance, if the tutor asks a yes/no question, the learner can respond by pressing the yes or the no button. The learner can answer the tutor's questions by typing into the text box in the interaction panel.

Robotic embodiment

The robotic embodiment used for the WoZ study was a Nao robot (torso version) that sits on the side of the touch table. It is capable of head, arm and body gestures in addition to synthesised speech. The robot receives the text and gestures selected by the wizard through the wizard Interface. Tutor's utterances were synthesized into speech using the in-built text to speech (TTS) engine while the gestures are realised using appropriate head, arm and body motions. To increase naturalness, the robot also had idle movement in-between wizard selections.

Sensors

The environment had an array of sensors such as two video cameras and a Kinect sensor. A Kinect sensor and a video camera are placed in front the learner. Another camera was placed in front of the robot (as shown in **Figure 18**).

Data collection

In this section, the collected data during the Pilot WoZ study is discussed using the set-up described above. The subjects' age ranged from 8-10 year old and the wizard's role was played

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

by a schoolteacher. The task for the learner was to carry out an expedition using the map application that he or she was provided with. In order to solve the steps of the expedition, the learner had to exercise his/her map reading skills. This way, map reading skills such as compass directions, and ordnance survey symbols were tested. The wizard's role was to observe the learner responses (both verbal and physical) and respond to them appropriately using the interaction panel in the wizard Interface application. Simultaneous video feeds from two cameras recorded the tutor-learner interaction.

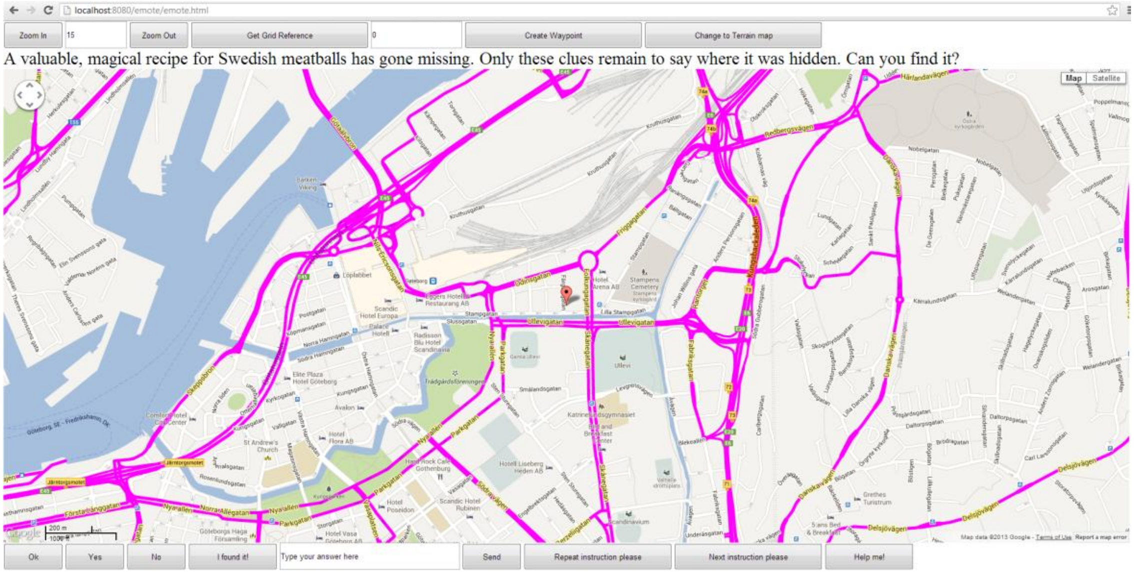

Figure 20. Map application.

Opportunities and Challenges

There are clear challenges involved in such a WoZ data gathering experiment. In this section, we describe these challenges and discuss lessons learned from our pilot studies and the requirements that are going to be considered. Some situations occurred that need to be address in further experiments. These situations refers to the map application faltering; the challenge for the robot to point out where some of the map features are (e.g., train station) or provide directions when users were confused; and the robot (i.e., wizard) response time that were perceived as too long evidenced by the children's “blank” expressions after giving an answer and waiting for a response

All of these challenges presents an opportunity to overlap some future situations such as the utilization of multi-modal outputs through the application running on the touch-table (e.g., pointing out map features by overlaying shapes such as circles, bounding boxes and arrows on top of the map). Also, presented the challenge of effective turn management wherein the tutoring system needs to decide how to stall during diagnosis (e.g., using backchannels or encouragement), which dialogue move to select and when to intervene by continuously monitoring the state of the map application (Deshmukh, Janarthanam, Hastie, Bhargava, & Aylett, 2013).

7. Conclusion and Future Work

The main purpose of the present Deliverable was to create a framework for empathic tutors' behaviour generation. Starting from a review of the literature on empathy and combining it with a specification of data collected from the studies, an initial version of the framework is presented.

This combination provides a coding scheme that is designed to fit into the framework so that the collected data can further be used to train simulations and interaction strategies. The data is thus well designed to address the requirements to generate a synthetic empathic tutor, involving different types of behaviour in a verbal and nonverbal communication.

8. References

Alexander, S. T. V. (2007). *An effect-sensitive intelligent tutoring system with an animated pedagogical agent that adapts to student emotion like a human tutor*. Master thesis from Massey University, Albany, New Zealand.

Arroyo, I., Ferguson, K., Johns, J. Dragons, T., Meheranian, H., Fisher, D., Barto, A., Mahadevan, S., & Woolf, B. P. (2007). Repairing Disengagement with Non-Invasive Interventions. In *Proceedings of the 2007 conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts that Work*, 195-202, 2007.

Beck, A., Cañamero, L., Hiolle, A., Damiano, L., Cosi, P., Tesser, F., & Sommavilla, G. (2013). Interpretation of Emotional Body Language Displayed by a Humanoid Robot: A Case Study with Children. *International Journal of Social Robotics*, *5*(3), 325–334. doi:10.1007/s12369-013-0193-z

Bernieri, F. J. (1988). Coordinated movement and rapport in teacher-student interactions. Journal of Nonverbal Behavior, 12(2), 120–138

Bethel, C. L. (2009). *Robots without faces: Non-verbal social human-robot interaction*. PhD thesis from University of South Florida, USA

Bhargava, S., Janarthanam, S., Hastie, H., Deshmukh, A., & Aylett, R. (2013). Demonstration of the Emote Wizard of Oz Interface for Empathic Robotic Tutors. In *Association for Computational Linguistics: Proceedings of SIGDIAL 2013 Conference*, 363-365, August 2013.

Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *12*(2), 293-327. doi:10.1145/1067860.1067867

Brave, S., Nass, C., & Hutchinson, K. (2005). Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International Journal of Human-Computer Studies*, *62*(2),161-178. doi:10.1016/j.ijhcs.2004.11.002

Cassell, J., Vilhjálmsson, H. H., & Bickmore, T. (2001). BEAT: the Behavior Expression Animation Toolkit. In *Computer Graphics: Proceedings of the 28th annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, 477-486, 2001. doi: 10.1145/383259.383315

Chafai, N. E., Pelachaud, C., & Pelé, D. (2007). Towards the specification of an ECA with variants of gestures. *Intelligent Virtual Agents*, 4722, 366-367

Cooper, B., Brna, P., & Martins, A. (2000). Effective Affective in Intelligent Systems–Building on Evidence of Empathy in Teaching and Learning. In *Affective Interactions*, 21-34. Springer Berlin Heidelberg.

Gibet, S., Lebourque, T., & Marteau, P-F. (2001). High-level Specification and Animation of Communicative Gestures. *Journal of Visual Languages and Computing*, *12*(6), 657–687. doi:10.1006/jvlc.2001.0202

Cramer, H., Goddijn, J., Wielinga, B., & Evers, V. (2010). Effects of (in)accurate empathy and situational valence on attitudes towards robots. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction (HRI)*, 141-142, 2010.

Davis, M. H. (1983). The effects of dispositional empathy on emotional reactions and helping: A multidimensional approach. *Journal of Personality*, *51*(2), 167-184. doi:10.1111/j.1467-6494.1983.tb00860.x

De Vries, E., & Lund, K., & Baker, M. (2002). Computer-mediated epistemic dialogue: Explanations and argumentation as vehicles for understanding scientific notions*. The Journal of the Learning Sciences*, *11*(1), 63-103.

Deshmukh, A., Janarthanam, S., Hastie, H., Bhargava, S., & Aylett, R. (2013). WoZ Pilot Experiment for Empathic Robotic Tutors - Opportunities and Challenges. In *International Conference on Social Robotics: Proceedings of workshop on Embodied Communication of Goals and Intentions*, October 2013.

Deutsch, B. G. (1974). The structure of task oriented dialogs. In *Proceedings of the IEEE Speech Symposium*, 1974.

Dillenbourg, P., & Baker, M.J. (1996). Negotiation Spaces in Human-Computer Collaboration. In *Actes du colloque COOP'96, Second International Conference on Design of Cooperative Systems*, 187-206, June, 1996.

Eisenberg, N., & Fabes, R. A. (1990). Empathy: Conceptualization, measurement, and relation to prosocial behavior. *Motivation and Emotion*, *14*(2), 131-149. doi:10.1007/BF00991640

Ekman, P., & Friesen. (1969). *The Repertoire Of Nonverbal Behaviour - Categories, Origins, Usage and Coding.*

Ekman, P., & Friesen, W. V. (1975). *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*. Prentice-Hall.

Gebhard, P. (2005). ALMA: A layered model of affect. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, 29-36, 2005. doi:10.1145/1082473.1082478

Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialog patterns in naturalistic one-on-one tutoring. *Applied Cognitive Psychology*, *9*, 359-387

Graesser, A. C., Wiermer-Hastings, K., Wiermer-Hastings, P., & Kreuz, R. (1999). Auto-Tutor: A simulation of a human tutor. *Journal of Cognitive Systems Research*, *1*(1), 35-51.

Greczek, J., Swift-Spong, K., & Mataric, M. (2011). Using eye shape to improve affect recognition on a human robot with limitated expression. *Technical report*.

Haring, M., Bee, N., & André, E. (2011). Creation and evaluation of emotion expression with body movement, sound and eye color for humanoid robots. In *RO-MAN, 2011 IEEE*, 204-209, 2011.

Hegel, F., Spexard, T., Wrede, B., Horstmann, G., & Vogt, T. (2006). Playing a different imitation game: Interaction with an empathic android robot. In *IEEE-RAS International Conference on Humanoid Robots (Humanoids'),* 56-61, 2006. doi:10.1109/ICHR.2006.321363

Hess, U., & Fischer, A. (2013). Emotional mimicry as social regulation. *Personality and Social Psychology Review*, *17*(2), 142-157. doi:10.1177/1088868312472607

Hilbert, D. R. (1987). Color and Color Perception: Study in Anthropocentric Realism. *Csli Lecture Notes*. University of Chicago Press: Chicago.

Hoffman, M.L. (1982). Development of prosocial motivation: Empathy and guilt. In N. Eisenberg-Berg, (Ed.), *Development of Prosocial Behavior*. New York: Academic Press, 281-313.

Hupka, R. B., Zaleski, Z., Otto, J., Reidl, L., & Tarabrina, N. (1997). The colors of anger, envy, fear, and jealousy: A cross-cultural study. *Journal of Cross-Cultural Psychology, 28*(2), 156-171. doi:10.1177/0022022197282002

Johnson, W. L., & Rickel, J. W., & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education, 11*, 47-78

Kaya, N., & Epps, H. H. (2004). Relationship between Color and Emotion: A Study of College Students. *College Student Journal, 38*(3), 396-405

Kleinsmith, A., & Bianchi-Berthouze, N. (2007). Recognizing affective dimensions from body posture. *Affective Computing and Intelligent Interaction, 4738*, 48-58.

Krathwohl, D. R. (2002). A revision of Bloom's Taxonomy: An overview. *Theory Into Practise, 41*(4), 212-218. doi:10.1207/s1543042tip4104_2

Kozima, H., Nakagawa, C., & Yano, H. (2004). Can a robot empathize with people? *Articial Life and Robotics, 8*(1), 83-88. doi:10.1007/s10015-004-0293-9

Leite, I., Castellano, G., Pereira, A., Martinho, C., & Paiva, A. (in press). Empathic robots for long-term interaction: Evaluating social presence, engagement and perceived support in children. *International Journal of Social Robotics*.

Leite, I., Martinho, C., Pereira, A., & Paiva, A. (2008). iCat: an affective game buddy based on anticipatory mechanisms. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 3*, 1229-1232. International Foundation for Autonomous Agents and Multiagent Systems.

Leite, I., Pereira, A., Mascarenhas, S., Martinho, C., Prada, R., & Paiva, A. (2013). The influence of empathy in human robot relations. *International Journal of Human-Computer Studies, 71*(3), 250-260. doi:10.1016/j.ijhcs.2012.09.005

Lowman, J. (1995). *Mastering the techniques of teaching (2nd Ed.)*. San Francisco: Jossey-Bass.

Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in Temperament. *Current Psychology, 14*(4), 261–292. doi:10.1007/BF02686918

McNeill, D. (1992). *Hand and Mind: What gestures reveal about thought*. University of Chicago Press: Chicago.

McQuiggan, S. W., & Lester, J. C. (2007). Modeling and evaluating empathy in embodied companion agents. *International Journal of Human-Computer Studies, 65*(4), 348–360. doi:10.1016/j.ijhcs.2006.11.015

Paiva, A. (2011). Empathy in social agents. *The International Journal of Virtual Reality, 10*(1), 65-68

Paiva, A., Dias, J., Sobral, D., Aylett, R., Sobreperez, P., Woods, S., Zoll, C., & Hall, L. (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 Multiagent Systems (AAMAS), 1*, 194-201, 2004. doi:10.1109/AAMAS.2004.82

Pearson, P., Hu, J., Branigan, H. P., Pickering, M. J., & Nass., C. (2006). Adaptive language behavior in HCI: How expectations and beliefs about a system affect users' word choice. In *Beliefs and Affect: Proceedings of the SIGCHI Conference on Human Factors in computing systems*, 1177-1180, April 2006.

Pereira, A., Prada, R., & Paiva, A. (2012). Socially present board game opponents. In Advances in *Computer Entertainment: Proceedings of the 9th international conference on Advances in Computer Entertainment (ACE)*, 101-116, 2012.

Pour, P. A., Hussain, M. S., AlZoubi, O., D'Mello, S., & Calvo, R. A. (2010). The impact of system feedback on learners' affective and physiological states. In *Proceedings of the 10th international conference on Intelligent Tutoring Systems (ITS) - Volume Part I*, 264-273. doi: 10.1007/978-3-642-13388-6_31

Ribeiro, T., Dooley, D., & Paiva, A. (2013b). Nutty tracks: symbolic animation pipeline for expressive robotics. In *ACM SIGGRAPH 2013 Posters (SIGGRAPH)*, 2013b. doi:10.1145/2503385.2503394

Ribeiro, T., & Paiva, A. (2012). The illusion of robotic life: Principles and practices of animation for robots. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction (HRI)*, 383-390, March 2012.

Ribeiro, T., Vala, M., & Paiva, A. (2013a). Censys: A Model for Distributed Embodied Cognition. In *Proceedings of the 13th International Conference on Intelligent Virtual Agents (IVA), 8108*, 58-67, 2013a.

Riek, L. D., Paul, P. C., & Robinson, P. (2010). When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry. *Journal on Multimodal User Interfaces, 3*(1-2), 99-108. doi:10.1007/s12193-009-0028-2

Rodrigues, S. H., Mascarenhas, S. F., Dias, J., & Paiva, A. (2009). "I can feel it too!": Emergent empathic reactions between synthetic characters. In *Proceedings of the 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, 1-7, 2009. doi:10.1109/ACII.2009.5349570

Saerbeck, M., Schut, T., Bartneck, C., & Janse, M. D. (2010). Expressive robots in education: Varying the degree of social supportive behavior of a robotic tutor. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1613-1622, 2010. doi:10.1145/1753326.1753567

Suk, H. J. (2006). *Color and Emotion: A Study on the Affective Judgment of Color*. PhD Thesis from University of Mannheim, Mannheim, Germany.

Takayama, L., Dooley, D., & Ju, W. (2011). Expressing Thought: Improving Robot Readability with Animation Principles. In *Proceedings of the 6th international conference on Human-robot interaction (HRI)*, 69–76, 2011. doi:10.1145/1957656.1957674

Tapus, A., & Mataric, M. J. (2007). Emulating empathy in socially assistive robotics. In *American Association for Artificial Intelligence (AAAI) Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics*, 93-96, 2007.

Thomas, F., & Johnston, O. (1995). *The illusion of life: Disney animation*. Hyperion.

Vygotsky, L. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press, Cambridge, MA.

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31*(1), 39-58. doi: 10.1109/TPAMI.2008.52