Department of Computer Science and Software Engineering
University of Canterbury

June 2007

Affect Recognition and Support in Intelligent Tutoring Systems

A thesis submitted in partial fulfilment of the requirements for the Degree of Master of Science in Computer Science in the University of Canterbury by

Konstantin Zakharov

Supervisor: Associate Professor Dr. Antonija Mitrović[1]
Associate supervisor: Associate Professor Dr. Lucy Johnston[2]
Examiner: Associate Professor Cristina Conati[3]

[1]Department of Computer Science and Software Engineering, University of Canterbury
[2]Department of Psychology, University of Canterbury
[3]Department of Computer Science, University of British Columbia

Abstract

Empirical research provides evidence of strong interaction between cognitive and affective processes in the human mind. Education research proposes a model of constructive learning that relates cognitive and affective processes in an evolving cycle of affective states. Intelligent Tutoring Systems (ITSs) are capable of providing comprehensive cognitive support. Affective support in ITSs, however, is lagging behind; the in-depth exploration of cognitive and affective processes in ITSs is yet to be seen.

Our research focuses on the integration of affective support in an ITS enhanced with an affective pedagogical agent. In our work we adopt the dimensional (versus categorical) view of emotions for modelling affective states of the agent and the ITSs users. In two stages we develop and evaluate an affective pedagogical agent. The affective response of the first agent version is based on the appraisal of the interaction state; this agent's affective response is displayed as affective facial expressions. The pilot study at the end of the first stage of the project confirms the viability of our approach which combines the dimensional view of emotions with the appraisal of interaction state.

In the second stage of the project we develop a facial feature tracking application for real-time emotion recognition in a video-stream. Affective awareness of the second version of the agent is based on the output from the facial feature tracking application and the appraisal of the interaction state. This agent's response takes the form of affect-oriented messages designed to interrupt the state of negative flow. The evaluation of the affect-aware agent against an unemotional affect-unaware agent provides positive results, thus confirming the superiority of the affect-aware agent. Although the uptake of the agent was not unanimous, the agent established and maintained good rapport with the users in a role of a caring tutor.

The results of the pilot study and the final evaluation validate our choices in the design of affective interaction. In both experiments, the participants appreciated the

addition of audible feedback messages, describing it as an enhancement which helped them save time and maintain their focus. Finally, we offer directions for future research on affective support which can be conducted within the framework developed in the course of this project.

Contents

List of Figures

List of Tables

Acknowledgements

Thank-you to my supervisor, Associate Professor Antonija Mitrović, for the guidance she readily provided at every stage of the project. Her optimism and enthusiasm made this project run smooth and almost on time. I also thank her for providing the funding and equipment required for this research. Thank-you to my associate supervisor, Associate Professor Lucy Johnston, for her unbiased consideration of this research and for her feedback on the ideas, data analysis and multiple drafts.

I want to say a big thank to my fellow geek-buddies Jay Holland and Nancy Milik for their friendship, miscellaneous tips and tricks, sanity checks and episodical recreational insanity. I also thank Nancy for feedback on the draft of this thesis and for her help in taming LaTeX. EER-Tutor is a result of many people's work out: I want to thank Brent Martin for WETAS code-base, and Pramudi Suraweera for KERMIT's constraint base and other bits and pieces. Thank you to the rest of the ICTG crowd, Amali, Moffat, Amanda, Nilufar and Devon, for the good spirit and company.

Thank-you to Dejan Mitrović for ideas and other help at the initial stages of this project. Thank-you to Associate Professor Andy Cockburn for the advice on data analysis. And of course, I thank all the academic and general staff at the CSSE Department, especially Pete Glassenbury, Blair McMaster, Adrian White, Joff Horlor, Phil Holland, Gillian Clinton and Alex Tobeck, for maintaining the best environment for research.

The most special thank-you goes to Mandala White and to her parents Ken and Ruthanne White.

Finally, thank-you to the people at Google Scholar: you bring the world of science, research and exploration to my desktop. Can it get any better than that?

This research was supported by an ICTG Scholarship and the Battersby-Trimble Scholarship.

CHAPTER 1

Introduction

Learning is a fundamental component of everyone's life. Successful learning is essential for mastering the unlimited variety of skills required for achieving satisfaction in professional and recreational activities on a day-to-day basis. Cognitive Science identifies a number of distinct forms of learning, such as associative learning, spatial learning and skill learning. Skill learning covers both the acquisition of motor skills, such as learning how to serve a tennis ball or how to ride a bicycle, and cognitive skills, such as skills for reading, solving problems within a particular domain, recognising particular patterns, and so on. For example, chess masters have acquired the skill of recognising specific configurations of chess pieces, a skill that helps them both in remembering the arrangement of the game pieces and allows them to think about the game in terms of strategy-defined, goal-oriented patterns of pieces, rather than needing to focus on individual pieces. This research is concerned with cognitive procedural learning, which is described as the acquisition of *procedural knowledge*, the knowledge of how to carry out some procedure, as opposed to *declarative knowledge*, the knowledge that a proposition is correct.

Research in Cognitive Psychology and Education shows that learning outcomes depend on a number of variables associated with the learning process, environment and especially the individual learners—they vary in their amount of prior knowledge, motivation, learning style, natural pace and working memory capacity. Consequently, a uniform predefined instructional sequence can not provide the optimal learning environment for all learners, because it fails to satisfy their changing needs [Holt et al., 1994].

Educational research results published over two decades ago by Bloom [1984] single out one-to-one tutoring as the most effective model of instruction. It has been shown that under one-to-one tutoring conditions the average student performs about two standard deviations above the mean of a conventional class with 30 students per teacher. Bloom calls this phenomenon *the 2 Sigma problem*. Success of one-to-one tutoring is based on the ability of tutors to tailor their feedback to the needs of individual learners. By interacting with students and observing their learning progress, tutors choose instructional actions on the basis of the state of the students' knowledge at each step.

Theoretically, anyone has the potential of achieving high performance in learning with the help of individualised tutoring, but one-to-one tutoring is rarely available on a wide scale; there is no easy solution to this dilemma. Individual tutoring often is simply not a plausible option from either the economic or practical perspectives. Being aware of this problem, Bloom [1984] comments:

> An important task of research and instruction is to seek ways of accomplishing this [high level of learning] under more realistic conditions than the one-to-one tutoring, which is too costly for most societies to bear on a large scale.

1.1 Computer-based Systems for Learning

Research in applications of Artificial Intelligence (AI) in Education offers a way of bridging the *2 Sigma* gap with the help of computer-based instructional environments.

The purpose of incorporating AI into instructional environments is to make them responsive to the needs of the individual learner.

Although the use of computers in Education began in the 1950s with Computer-Aided Instruction (CAI) systems [Ayscough, 1977; Last, 1979], the statically defined content and order of instruction in these systems did not offer any significant advantage over the self-directed study performed by reading a text book. In the 1960s, a new generation of CAI systems appeared, often referred to as *branching programs* [Hullfish and Pottebaum, 1971]. Branching was based on the comparison of students' answers to predefined solutions. A lack of adaptiveness and domain expertise did not allow these systems to provide individualised learning support at the optimal level. The first attempts of making CAI system *intelligent* started in the 1970s. Our research focuses on a class of such systems known as Intelligent Tutoring Systems (ITSs).

The area of ITS is an interdisciplinary research field that draws on knowledge from Artificial Intelligence (AI), Human-Computer Interaction (HCI), Linguistics, Education and Cognitive Psychology; the latter carries a pivotal significance in that ITSs are designed to support learning, which takes place in the human mind. The central component of Cognitive Psychology is the Cognitive Architecture [Anderson et al., 1995], which focuses on the organisation and workings of a human mind. The goal of ITS research is to develop educational environments capable of delivering individualised instruction comparable to one-to-one tutoring. Modern-day ITSs successfully capture the one-to-one aspect of individualised tutoring by adjusting presentation and feedback according to the individual learner's cognitive state.

1.2 Affective Processes in Learning Context

Research outcomes suggest a strong interaction between cognitive and affective processes (otherwise known as emotions) in the human mind. For example, there is a reliable correlation between the quality of one's emotional state and memory capacity [Goleman, 1995]. Research also suggests that in the educational context, stress, anxiety, and frustration experienced by a learner can severely degrade learning outcomes. Goleman [1995] elaborates on the matter in his book, *Emotional Intelligence*:

> The extent to which emotional upsets can interfere with mental life is no news to teachers. Students who are anxious, angry, or depressed don't learn; people who are caught in these states do not take in information efficiently or deal with it well.

In a bid for deeper understanding and better facilitation of the learning process, education researchers during the last two decades have been trying to ascertain further how learning is influenced by affective processes which are an irrepressible aspect of human nature. The semantic component of social interaction, most frequently expressed as speech, is often accompanied by the affective component of social interaction, which is considered equally or sometimes even more important then the semantic component [Bickmore, 2004]. Although people in general are not always aware of how exactly their language, posture, facial expression and eye gaze convey their emotions, these underpin people's interactions and navigation in the social world [Mishra and Hershey, 2004; Knapp, 1978; Kendon, 1983; Klingzing and Tisher, 1986].

Along with keeping track of the students' cognitive states, human tutors are capable of factoring the students' affective state into their tutoring strategies [du Boulay and Luckin, 2001; Kort and Reilly, 2002]. Like real-life tutors, ITSs can become more effective by recognising the affective state of their users. Incorporating analysis of affective state into synthesis of feedback and instruction flow can elevate the interaction with the learner to a new level and make a difference not only in the learner's perception of the interaction, but in the learning outcomes as well. However, present-day ITS research is facing a wide range of interaction design and technical problems that arise during the development of affect-aware ITSs.

Researchers have been grappling with the question of what appropriate behaviour is within an interactive learning environment. Some real-life scenarios, such as problem solving in learning contexts, can often leave people feeling stressed, anxious and frustrated. Despite the best efforts of software and hardware designers, when transferred to computer-mediated learning environments, these scenarios can leave the users doubly exasperated. Since etiquette is highly context-dependent, what may be appropriate in one situation, may be inappropriate in another. A software program offering help may be exhibiting generally appropriate HCI etiquette, but inappropriate educational etiquette.

1.3 Affective Gap in Intelligent Tutoring Systems

Du Boulay and Luckin [2001] state that ITSs initially evolved in a rather lopsided way. They claim that although on one hand ITS developers put much effort into highly detailed models of particular domains, on the other hand not nearly enough effort was put into the research and implementation of the communicative teaching techniques such as explaining, persuading, arguing, demonstrating, describing and so on. Du Boulay and Luckin explain this phenomenon by saying that at the dawn of ITS development the first systems were not grounded in such general communicative competence because it was beyond the state of the art. In the same vein, Kort and Reilly [2002] call for a re-engineering of the current state of educational pedagogy by shifting the focus of research towards expert teachers "who are adept at recognising the emotional state of learners, and, based upon their observations, take some action to scaffold learning in a positive manner".

Computers, however, since their early days have been implicitly designed without awareness of the affective communication channel; computers respond to people as if they were computers too, making people adjust to the computer protocol and interact on a sub-human level. This contradicts the main assumption guiding HCI research: "People should not have to change radically to *fit in with the system*—the system should be designed to match their requirements" [Preece et al., 1994].

This inherent lack of affective fit between technology and its users is particularly significant in the area of ITSs: a failure to recognise affective processes might impose a serious limitation on interaction types which are fundamentally social in nature. This suggests that without taking into account the interaction between the cognitive and affective processes ubiquitous in human activities, educational systems might never approach their full potential.

Speaking in a broader sense, frustrating interaction with a computer system can often leave a user feeling negatively disposed toward the system and its designers; such negative experiences could alter perceptions of trust, cooperation and good faith on the part of the user [Klein et al., 2002]. On the other hand, enabling computers to recognise and adapt to users' affective states can, in theory, improve the quality of interaction.

Recent research on affect recognition in computer-mediated environments opens new perspectives, although very little research has explored the ways a computer can be used to address the emotional state of its user in the learning context [Klein et al., 2002].

1.4 Pedagogical Agents

As described in the work of Reeves and Nass [1996], people tend to view electronic media in a social way, as if they were other people. Reeves and Nass's research supports the premise that computers are not merely neutral tools. Their study also emphasises the importance of the social and affective relationships that can develop between a computer and a learner [Mishra et al., 2001]. A better understanding of these relationships is essential to building smarter tools for learning.

Recent research suggests the use of pedagogical agents in ITSs as a medium for delivering feedback to the users [Johnson et al., 2000; Lester et al., 1999b]. A learner who enjoys interacting with a pedagogical agent has a more positive perception of the overall learning experience and consequently opts to spend more time in the learning environment [Johnson et al., 2000]. Pedagogical agents are software characters capable of expressing human-like behaviours and emotions. Research on pedagogical agents draws on Allport's [Allport, 1935] classic definition of social psychology:

> The scientific investigation of how the thoughts, feelings and behaviours of individuals are influenced by the actual, imagined or implied presence of others.

Supporting this definition, pedagogical agents are known to enhance the social view of interaction with ITS [Lester et al., 1999b].

Affective pedagogical agent technologies are still very much in their infancy, and little is known about their effectiveness in learning environments [Johnson et al., 2000]. In spite of positive results, experimental studies on pedagogical agents carried out to date have had a number of limitations. There has been little research on ITSs with affective pedagogical agents which integrate in their behaviour the cognitive state and affect-recognition techniques based on Psychophysiology or Computer Vision.

Our hypothesis is that ITSs enhanced with affective pedagogical agents aware of users' emotional states will offer better learning support and that users' interaction with an agent will reduce negative feelings, providing a more enjoyable experience of interaction overall. Klein et al. [2002] offer a slightly different view on affective computing, saying that since computers are not capable of genuine empathetic understanding of feelings, affective response produced by a computer is only a form of artificial empathy. However, this empathy is sincere in the sense that its designers' goal is to try to enable the user to feel empathised with.

Heylen et al. [2005] suggest that although ITS research may be far removed from pedagogical agents capable of reading learners' minds, researchers can start designing agents on the basis of findings from Social Psychology, Psychophysiology, Computer Vision, Cognitive Science, Linguistics, Natural Language Processing and Artificial Intelligence. Such agents' behaviour may not be based on the same empathetic capabilities of humans, but these research efforts are crucial for bridging the affective gap in generic HCI, and in particular in ITS; this in turn may prove to be instrumental in the efforts of closing the 2 Sigma gap.

1.5 Thesis Contributions

Our research is a step towards an ideal ITS attuned to both cognitive and affective processes. We attempt to bring together affect-recognition research and ITSs, enhanced with affective pedagogical agents. Our contributions to the area of ITS and pedagogical agents are two-fold.

First, we introduce a design approach for an affective pedagogical agent whose emotional behaviour is modelled on the basis of the dimensional view of emotions. The agent infers users' emotions by monitoring the users' cognitive state and responding to the users' emotions with its affective facial expressions. This version of the agent, however, does not involve affect-recognition—in this thesis we refer to this agent version as the *affect-inferring* agent. The evaluation of the dimensional approach to modelling emotions for the agent and ITS users based on cognitive-state appraisal proves its viability: the agent secured a strong rapport with the users during the first evaluation study.

Second, we develop a facial feature tracking application for real-time affect recognition from a video stream; in this application the dimensional model of emotions is also the foundation of affective state inference. Synthesising affect-recognition and cognitive states of the system users to guide the agent's behaviour, we develop a version of the agent referred to in this work as the *affect-aware* agent. The positive outcomes of the affect-aware agent's evaluation demonstrate that these agents do have an impact on users' interaction with an ITS.

Appendices C and D include the publications produced by this research to date.

1.6 Guide to the Thesis

This section outlines the contents of the remaining chapters of this thesis. Chapter 2 provides an overview of the background research relevant to our project, focusing on ITSs, pedagogical agents and study of emotions. Chapter 3 describes our approach to design and implementation of the affect-inferring pedagogical agent for EER-Tutor, the test-bed ITS for our research. This chapter also details the outcome of a pilot study of the agent-enhanced version of EER-Tutor.

The affect-aware agent is described in Chapter 4; the chapter provides the description of a facial feature tracking algorithm along with the affective state detection and response logic guiding the agent's affective behaviour. Chapter 5 discusses the design and results of the experimental study with the affect-aware agent in EER-Tutor. Chapter 6 contains the summary of our work and presents conclusions drawn on the basis of our research.

CHAPTER 2

Related Research

Research on ITS and affective pedagogical agents draws on a variety of other scholarly disciplines. This chapter sets the background for the discussion on ITSs and affect-aware pedagogical agents. Section 2.1 provides high-level information on ITSs; it describes typical ITS architecture and common approaches to ITS implementation. Section 2.1.1 introduces one of the implementation approaches, Constraint-based Modelling, in greater detail. This approach lies at the foundation of the ITS which we used as the test-bed of our research. This ITS for conceptual database modelling, EER-Tutor, is described in Section 2.1.2. The early implementations of pedagogical agents did not explicitly focus on affect-oriented content and presentation; in the recent research, however, the affective component of interaction both from the user-side and the agent-side, has become more prominent. Even though the boundary between basic pedagogical agents and affective pedagogical agents is not always clear, we introduce these two types of agent in separate Sections. Section 2.2 presents the research on pedagogical agents along with a few examples of ITSs with agents; Section 2.3 discusses affective pedagogical agents and introduces a number of affective agents' examples. Section 2.4 is dedicated to the study of emotions and applications of emotions research in the context of learning and ITSs. Section 2.4.1 presents the two fundamental approaches to emotion description and classification, while Section 2.4.2 talks about the ways of measuring human emotions. Facial feature tracking, one of the approaches to emotion recognition, is described in Section 2.5. Section 2.6 provides the motivation and reasons for affect integration into ITS, followed by the discussion of affective etiquette in the educational context in Section 2.7.

2.1 Intelligent Tutoring Systems

Intelligent Tutoring Systems are task-oriented problem-solving environments designed for learning support in specific instructional domains. They shift the focus from knowledge transfer to knowledge construction by actively adapting the instructional process to the needs of the individual user. Kort and Reilly [2002] describe learning as the process of extraction of meaningful insights from the vast body of data: "Studying is like panning gold where the answers are the nuggets buried in a ton of otherwise uninteresting gravel". These "nuggets" represent the individual pieces of the "jigsaw puzzle of knowledge" Kort and Reilly [2002]. Like one-to-one tutoring, ITSs come to the rescue of learners by offering help with discovery and organisation of the "big picture" of the structured knowledge body.

There has been a number of ITSs developed for well-known domains such as Mathematics, Physics, programming, database design and foreign language acquisition. Examples include the following systems: Andes Physics Tutor is a problem solving environment for introductory college-level Physics [VanLehn et al., 2005]; Algebra Cognitive Tutor is an ITS that teaches problem-solving skills in a high-school Algebra course [Anderson et al., 1995]; AutoTutor supports problem-solving in college Physics and other domains [Graesser et al., 1999]; and STEVE is an environment for teaching hierarchical, multi-step procedures, such as how to start a large air compressor [Johnson et al., 1998].

ITSs are known to improve learning performance by 0.3–1.0 standard deviations. For example, Lisp-Tutor, an ITS for teaching Lisp programming language, claims improvement of one standard deviation [Anderson et al., 1995]. In addition to this, students who use the tutor take 30% less time to master the domain. SQL-Tutor, an ITS for teaching Structured Query Language (SQL) for databases, improves performance by 0.65 standard deviations in just two hours of interaction with the system [Mitrović et al., 2001]. Atlas, a tutoring system for teaching Physics, improves performance by 0.9 standard deviations [Freedman, 1999]. An ITS named AutoTutor comes in two versions: the computer literacy version is designed to help users learn basic computer topics covered in an introductory course (hardware, operating systems, and the Internet), while the conceptual physics version is designed to help users learn Newtonian physics; AutoTutor achieves learning gains of approximately 0.8 standard deviations [Graesser et al., 2005]. Between 20 and 25 hours of interaction with SHERLOCK, a tutor for technical troubleshooting in avionics, is equivalent to four years of on-the-job experience [LaJoie and Lesgold, 1989; Katz et al., 2000].

The capability of individualised instruction in ITS hinges on the interaction of the target knowledge domain model, expert model, and the model of the learner's knowledge, usually described as student model. The interaction of these components is usually guided by the pedagogical module; Figure 2.1 presents a typical high-level view of ITS architecture. The following is a brief description of these components.

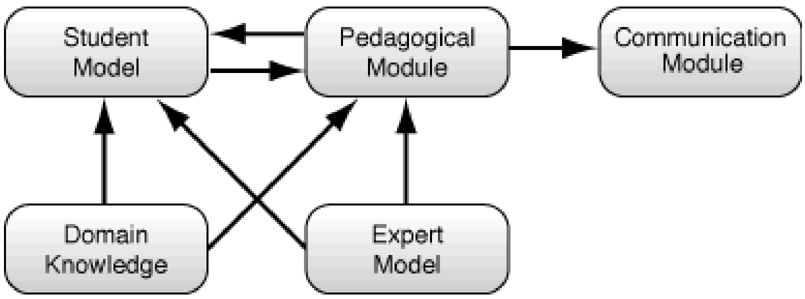


Figure 2.1: Basic ITS architecture. Adapted from Beck et al. [1996].

Domain Model: This component contains the knowledge being taught. Generally, it requires significant knowledge engineering effort to represent a domain in a suitable way; sometimes domain knowledge is represented as facts and procedures, while in other instances the knowledge needs to be expressed as concepts and mental models.

Expert Model: The expert model shares some of the domain knowledge being taught to the learner. However, it is more than just a representation of the data; it is a model of how someone skilled in a particular domain represents the knowledge and uses it for solving problems in the domain. The expert model is used to compare the learner's solution to the expert's solution, pinpointing the parts of the solution that need further attention.

Student Model: The student model performs two functions. It evaluates the user's solution and dynamically maintains a model of the users's knowledge. The units of knowledge in student models are usually characterised by certain correspondence to the units comprising the target domain. Student models are constructed by deducing the user's competence in the target knowledge domain through their interactions with the system and the quality of their solutions. The student models may contain an estimate of the user's long-term knowledge (overall domain mastery) and short-term knowledge (the mistakes made in the last attempt). Maintaining a good approximation of the users's knowledge is essential as all pedagogical decisions depend on the information found in the student model. The pedagogical decisions taken by the system are a direct reflection of the quality of the student model. At each moment, the state of student model and the expert model determine the response of the system. This feature makes ITSs stand above other types of computer-based instructional systems.

ITS research lists a number of student-modelling approaches for short-term and long-term knowledge representations. Model tracing (MT) [Anderson et al., 1995] and constraint-based modelling (CBM) [Ohlsson, 1994; Mitrović et al., 2003] are two important short-term student modelling techniques. MT focuses on representing procedural knowledge, whereas CBM focuses on declarative knowledge. The well-grounded methods for modelling long-term student knowledge include overlays [Holt et al., 1994] and stereotypes [Rich, 1979]. Overlay models represent the knowledge of the user as a subset of the domain expert's knowledge. The initial state of an overlay model assumes that the student has no knowledge. The model is populated as the user interacts with the system. Stereotype modelling also models the users's domain knowledge with respect to the desired knowledge. It overcomes the problem of starting from an empty model by classifying the user into levels of expertise, usually based on the pre-test score.

Pedagogical Module: This component provides a model of the teaching process. The student model and domain model provide input data to this component. The pedagogical module acts as the driving engine of the tutoring system by setting the pace and tone of instruction presentation. Pedagogical strategies vary from traditional goal-oriented methods [Anderson, 1993] to more informal discovery-oriented learning [Lesgold, 1988; Shute et al., 1989]. Tutoring systems that adopt the goal-oriented approach initiate and control user activity. Novices find this method of teaching more suitable as it provides close guidance. However, advanced learners find such systems restrictive and not sufficiently challenging. In contrast, discovery-oriented exploration involves learning from experience by promoting knowledge construction through self-explanation, induction and ontology-based experiences.

Communication Module: This component is the interface—it mediates the learners' interaction with the ITS. Typically, the interface presents the tools and structure for solutions, delivers feedback and critiques of user's attempts.

2.1.1 Constraint-based Modelling

CBM arises from Ohlsson's theory of learning from performance errors [Ohlsson, 1996] and is based on the Cognitive architecture[1]. CBM was proposed as a way of overcoming the intractable nature of Student Modelling. Unlike MT, CBM does not require extensive studies of student bugs for compilation of bug libraries; this is an important trade-off because it helps conquer the intractability of the student model. A CBM model represents domain knowledge as a set of explicit constraints on correct solutions in that domain [Mitrović and Ohlsson, 1999]. At the same time, constraints implicitly represent all incorrect solutions. In this way, constraints partition all possible solutions into correct and incorrect ones.

Each constraint specifies a property of the domain that is shared by all correct solutions. A constraint is an ordered pair (C_r, C_s), where C_r is the relevance condition determining a problem state in which the constraint is relevant, and C_s is the satisfaction condition defining the state in which the constraint is satisfied. If a constraint is relevant in some state then it must also be satisfied in order for the solution to be correct, otherwise the solution contains an error. Thus, the semantics of a constraint are: *if the C_r condition is true, then the C_s must also be true, otherwise something has gone wrong*. The following example is taken from a well-known problem, the Towers of Hanoi[2]:

$$C_r = <\textit{If disk X is on peg Z and disk Y is on peg Z and X is on top of Y}>,$$
$$C_s = <\textit{then X is smaller than Y}> (\textit{or else there is an error})$$

In this example, the relevance condition, C_r, is the complex clause *disk X is on peg Z and disk Y is on peg Z and X is on top of Y*, and the satisfaction condition C_s, is the clause *X is smaller than Y*.

Generally constraints are divided into two types: syntax and semantic. Constraints of the first type represent syntactic properties of the target knowledge domain; they

[1] Cognitive architecture refers to the design and organisation of the mind [Wilson and Keil, 1999].

[2] In the Tower of Hanoi problem, a stack of disks with holes in the centre are to be moved from one peg to another in accordance with a set of rules: only one disk is to be moved at a time, a larger disk cannot be on top of a smaller one, and at the end, the stack of disks should be on a specified peg.

refer only to the user's solution. Constraints of the second type represent semantic properties of the domain; they operate on the relation between the user's solution and the ideal solution. Of course, the distinction between the two kinds of constraints is not strict, and some constraints inspect both the syntax and the semantics of the user's solution [Mitrović et al., 2001].

CBM tutors evaluate solutions by matching them against the constraint set. First, all relevance components are matched against the problem state. Second, the satisfaction components of constraints matching the problem state in the first step (the relevant constraints) are tested. If a satisfaction pattern matches the state, the constraint is satisfied, otherwise, it is violated. The short-term student model consists of all satisfied and violated constraints. The long-term student model mainly consists of the list of all constraints used by the student and the history of constraint usage.

A number of constraint-based tutors have been developed within the Intelligent Computer Tutoring Group (ICTG) at the University of Canterbury: SQL-Tutor is a tutor for teaching Structured Query Language (SQL) [Mitrović, 2003; Mitrović et al., 2001]; CAPIT is a system that teaches the rules of punctuation and capitalisation in English [Mayo and Mitrović, 2001]; NORMIT [Mitrović, 2002] is an ITS that teaches data normalisation; LBITS is an English language tutor [Martin, 2001]; and COLLECT-UML is an environment for collaborative learning of Object-Oriented software design [Baghaei et al., 2005, 2006].

2.1.2 EER-Tutor

EER-Tutor is a CBM ITS for teaching the skill of Enhanced Entity-relationship (EER) modelling. Originally the idea of applying CBM in the domain of database modelling came to existence in KERMIT [Suraweera and Mitrović, 2002, 2004], which was a stand-alone ITS for basic Entity-Relationship (ER) modelling (displayed in Figure 2.2). Subsequently KERMIT was re-implemented in a Web-based authoring shell for Constraint-based ITSs, WETAS [Martin and Mitrović, 2002a,b]. Shortly after this, ER-Tutor became EER-Tutor, when its constraint-base and problem set were extended to handle Enhanced Entity-Relationship (EER) modelling [Suraweera and Mitrović, 2004; Zakharov et al., 2005]. Now EER-Tutor's constraint-base includes about 160 constraints and its problem set offers 60 ER and EER exercises.

We chose EER-Tutor as the test-bed ITS for our research because EER modelling presents a considerable learning challenge that is likely to influence users' affective state. EER design is a complex task that allows for creativity and variability in the solution space. Although EER modelling does produce an outcome defined in abstract terms, there is no algorithm for obtaining that outcome.

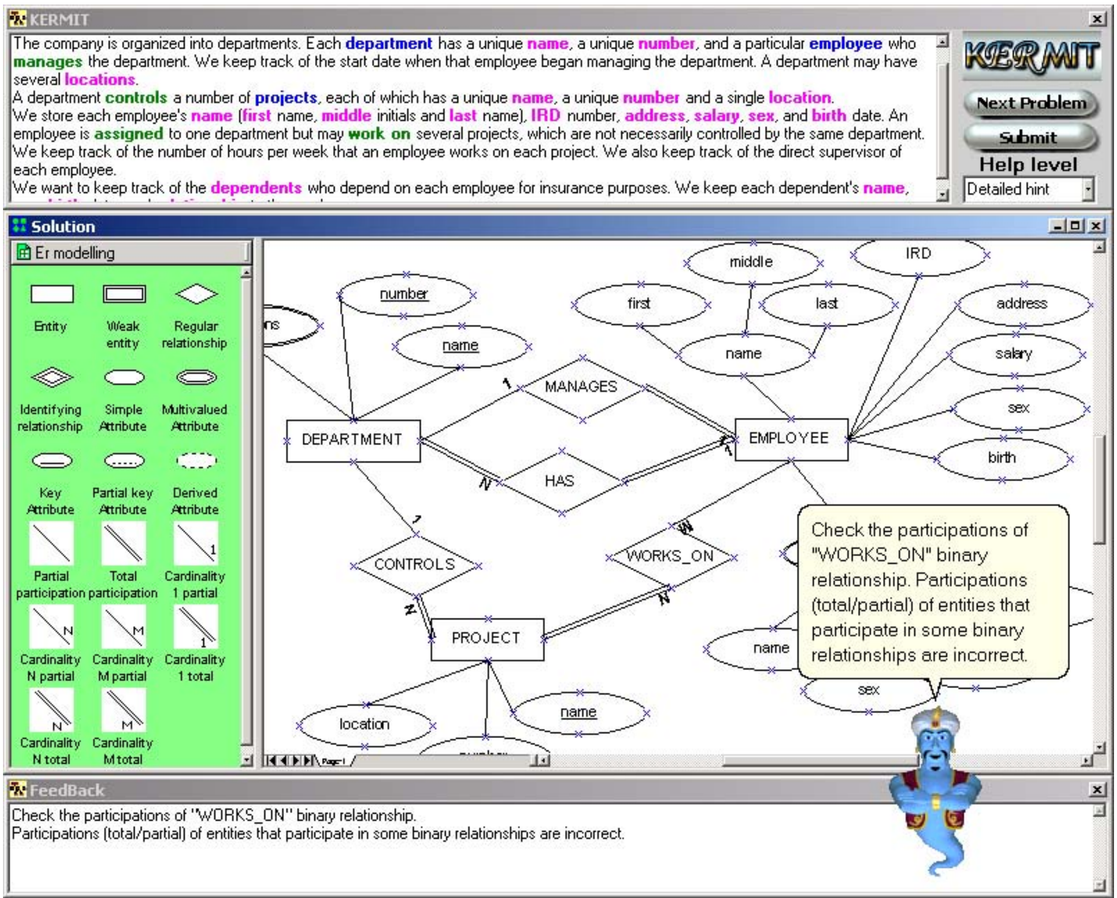


Figure 2.2: KERMIT: an ITS for ER Modelling. Adapted from Suraweera and Mitrović [2002].

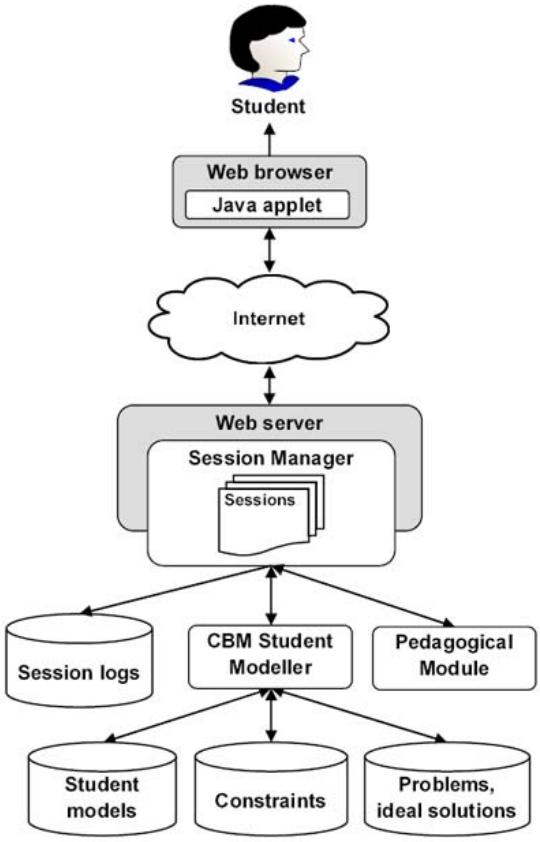


Figure 2.3: Architecture of EER-Tutor.

EER-Tutor's architecture can be traced back to SQLT-Web, the first Web-based CBM tutor [Mitrović, 2003]. The architecture of the system is shown in Figure 2.3. The server-side logic runs on top of an Open Source Web server AllegroServe [Foderaro, 2004]. AllegroServe and the rest of server-side components of EER-Tutor are implemented in Common Lisp [Steele Jr., 1990].

Communication between the interface and the server is built entirely on HTTP requests and responses exchanged by the Web-browser and Web-server. The session manager provides multi-user support for an arbitrary number of concurrent sessions and interacts with the rest of the server-side components. The individual session is established when a user signs on to the system. The session manager records users' actions and system responses in log files. The pedagogical module passes submitted solutions

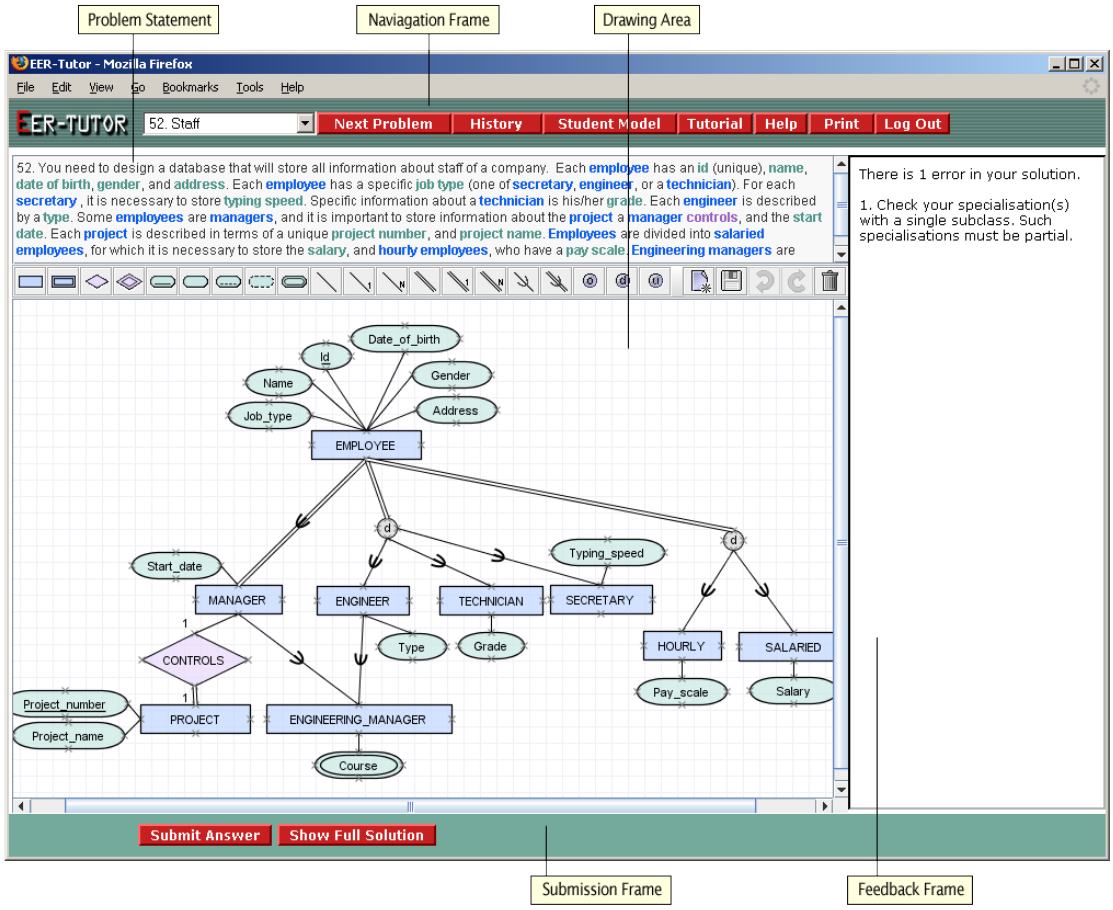


Figure 2.4: EER-Tutor's interface.

to the student modeler, which diagnoses the solution, updates the student model and sends the results of the diagnosis back to the pedagogical module. After this stage, the pedagogical module generates feedback messages.

The client-side of EER-Tutor (its interface) consists of a set of dynamic HTML. The main page of the interface contains an embedded Java applet as shown in a screenshot of the interface in Figure 2.4. The *Navigation Frame* provides a set of controls for the system; the *Problem Statement* contains the text for problem №52, which is one of the largest problems in EER-Tutor's curriculum. The applet provides a set of drawing tools (positioned under the *Problem Statement*) for creating diagrams as solutions to the problems presented by the system; the *Drawing Area* shows a nearly-complete solution created by a user. The *Submission Frame* controls let the user submit solutions to be analysed by the system (the *Show Full Solution* button lets users view the ideal solution to the current problem when they choose to do so). The *Feedback Frame* is where EER-Tutor displays the feedback messages in response to solution submissions.

Like the other ITSs, the interface of EER-Tutor acts as a mediator between the user and the system, enabling the user to conduct dialogues with the system. EER-Tutor provides on-demand feedback in text format to its users immediately after each attempt submission; the feedback is based entirely on the short-term student model. When the user first chooses a new problem to work on, EER-Tutor presents the user with the text of the problem and a blank *Drawing Area*. The user starts building up the solution by creating some diagram components in the *Drawing Area* with the help of the drawing tools; each drawing tool corresponds to an EER diagram component. For example, problem №5 sets the following task:

> *Some students live in student halls. Each hall has a name (unique) and an address. Each student has a number (unique) and a name. Assume that there are students living in every hall.*

After reading the text of the problem, the user might choose to create a regular entity type[3] by using the appropriate drawing tool; then the user labels the newly-created entity by choosing the appropriate word or phrase in the problem text. In this case the user labels the entity `STUDENT`. In the next step, the user creates a key attribute[3] and labels it `Number`. In this manner the user keeps adding the elements to the solution until solution is complete. However, the user might choose to obtain feedback on incomplete solutions—EER-Tutor lists the errors indicating the missing diagram components or other problems with the solution. The user is allowed to submit solutions as many times as desired at any stage; when the solution is correct and complete the system responds with a congratulatory message. Depending on the complexity of the problem, it can take the average user between a few minutes up to an hour to develop a complete solution.

[3]"Regular entity type", "key attribute" and "relationship" are the terms from the domain of ER modelling [Elmasri and Navathe, 2003]. Table 2.1 provides a summary of ER notation.

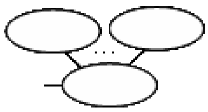
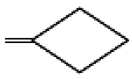
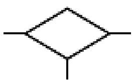
Symbol	Meaning
	Regular Entity
	Weak Entity
	Relationship
	Identifying relationship
	Simple Attribute
	Multivalued Attribute
	Composite Attribute
	Derived Attribute
	Key Attribute
	Partial Key Attribute
	1:1 Cardinality Ratio
	1:N Cardinality Ratio
	M:N Cardinality Ratio
	Total Participation
	N-ary Relationship

Table 2.1: Summary of ER diagram notation. Adapted from Elmasri and Navathe [2003].

Suppose the user submits an attempt containing an incorrectly specified participation of an entity type in a relationship[3] as shown in Figure 2.5. The user erroneously uses a partial participation connector due to lack of experience in extracting the modelling requirements from the problem statement. In particular, the phrase *Assume that there are students living in every hall* implies that every entity in the total set of `HALL` entities must be related to `N STUDENT` entities, which implies total participation of the `HALL` entity type in the `LIVE_IN` relationship.

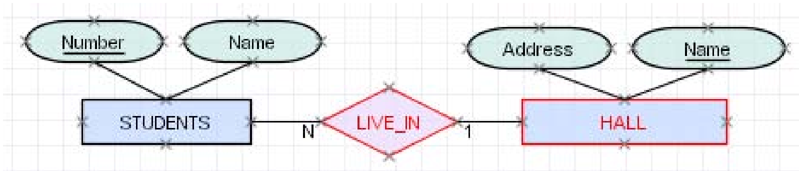


Figure 2.5: Incorrect solution to problem №5 in EER-Tutor.

In response to this solution, EER-Tutor will produce the following error message, associated with the violated semantic constraint which is aimed to validate the correctness of the participation of the entity types in a relationship:

> *The participation constraint specifies whether each entity of a specific type has to appear in the relationship type (total participation) or not (partial participation). The participation of the connector between the highlighted constructs is incorrect.*

Referring to the same erroneous solution in Figure 2.5, the above message starts with the general concept which, most likely, has not been internalised by the user. This is the step aimed at specialising the corresponding rule in the procedural memory of the learner, so that next time when a similar situation arises, the user will be able to differentiate correctly between choosing partial or total participation. The second sentence of the message ties the concept to the situation at hand, simultaneously pointing out the error and allocating the blame.

2.2 Pedagogical Agents

Animated pedagogical agents in computer-based environments draw upon human-to-human social communication patterns by embodying observable human characteristics such as the use of gestures, speech and facial expressions. Pedagogical agents are often capable of general conversational and tutoring functions; they make conversation (greeting, introduction and closing), set tasks, answer questions, provide instruction support, feedback, explanation, evaluation, motivation, and so on [Heylen et al., 2005].

Several studies show that pedagogical agents improve users' learning, engagement and motivation [Johnson et al., 2000; Mitrović and Suraweera, 2000; Mishra and Hershey, 2004]. These findings are consistent with the view of one-to-one tutoring as a kind of interpersonal interaction; tutoring situations are essentially social encounters, where the user's goal is to learn something and the tutor's job is to assist the user in achieving this goal [Heylen et al., 2005]. Whatever the social/cognitive factors are that explain the gap between the effectiveness of learning from a human tutor and an ITS, pedagogical agents are lending themselves to making the gap smaller.

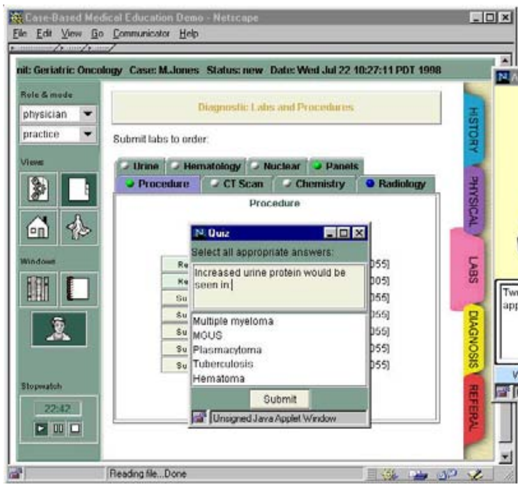
Research carried out independently of pedagogical agents cites evidence that electronic media has the power to appeal to the social aspects of human nature [Reeves and Nass, 1996]; coming from a slightly different angle Mishra et al. [2001] essentially state the same observation as Topffer's law: "All interfaces, however badly developed, have personality." Capitalising on the social aspects of human nature, research on pedagogical agents aims to optimise learning environments to bring them even closer to the one-to-one human tutoring ideal.

The users' perception of agents, as well as their effectiveness, depend on multiple characteristics of agents, such as believability, gender, ethnicity and instructional role [Baylor and Kim, 2004; Lester et al., 1999b]. Johnson et al. [2000] state that agent's believability is a product of two forces: (a) the visual qualities (realism and aesthetics) of the agent and (b) the properties of the behaviour control system that drives the agent in response to evolving interactions with the user; the latter poses considerably more complex design issues than the former. Johnson et al. [2000] describe two general approaches used in behaviour control: behaviour-space-based and generative. The behaviour-space approach relies an a number of canned animations sequenced as building blocks of complex behaviours. The generative approach relies on 2D or 3D graphical model, segmented into movable parts and controlled by algorithms that dynamically generate the transitions of the movable parts from the current position to the desired one.

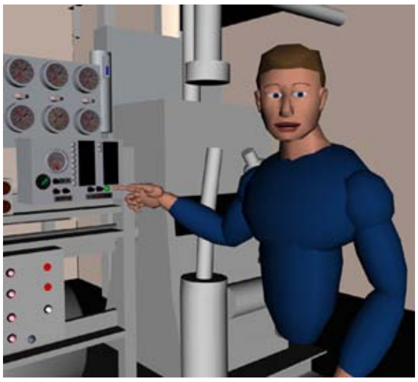
Research indicates that human-like agents produce the impression of stronger credibility and are more beneficial for learning then pet-like or animal-like agents [Parise et al., 1996; Moreno et al., 2002]; further research cites evidence that agents taking on the motivator and tutor (mentor) roles versus the expert role result in stronger learning outcomes [Baylor, 2005; Baylor and Kim, 2004; Prendinger et al., 2005]. This is described more fully in Section 2.7. The work of Prendinger and Ishizuka [2001] promotes social role awareness as a desirable capability of animated agents; their incorporation of “social filters” into mental models of animated agents qualifies the agent’s expression of its emotional state by the social context, thereby enhancing the agent’s believability as a conversational partner or virtual teammate.

Figure 2.6 gives a few examples of pedagogical agents tested in teaching environments for various knowledge domains. ADELE (Agent for Distance Education: Light Edition) is designed to support solving of exercises delivered over the World Wide Web [Shaw et al., 1999]. In the application of a case-based clinical diagnosis, ADELE can highlight interesting aspects of the case, as well as monitor and provide feedback as the student works through a case. STEVE (Soar[4] Training Expert for Virtual Environments) teaches users how to perform procedural tasks, such as operating or repairing complex devices [Johnson et al., 2000; Rickel and Johnson, 1997]. Herman the Bug is a pedagogical agent introduced in Design-A-Plant, a learning environment for the domain

[4]Soar (State, Operator And Result) is a general cognitive architecture for developing systems that exhibit intelligent behaviour. See `http://www.isi.edu/soar/soar-homepage.html`



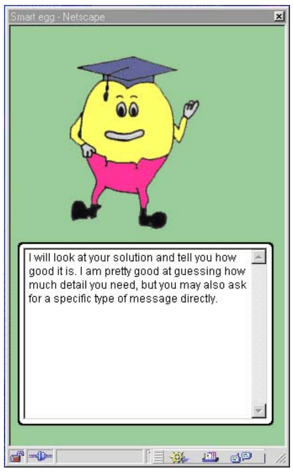
(a) ADELE at the task.



(b) STEVE from Soar.

(c) Herman the Bug.

(d) Cosmo—Internet Protocol Advisor.



(e) Smart-Egg agent.

Figure 2.6: Examples of animated pedagogical agents. Images (a), (b), (c) and (d) adapted from Johnson et al. [2000]; images (e) adapted from Mitrović and Suraweera [2000].

of botanical anatomy and physiology [Lester et al., 1999a]. The agent exhibits life-like behaviour as it provides advice to students solving problems. Cosmo inhabits the Internet Protocol Adviser, which is a learning environment for the domain of Internet packet routing [Lester et al., 1999b]. He provides advice to learners as they decide how to ship packets through the network to the specified destination. Smart Egg [Mitrović and Suraweera, 2000], developed with the ADELE toolkit, assists students in learning Structured Query Language (SQL) with the help of SQL-Tutor [Mitrović, 2003]. The agent explains the system's functions, provides feedback on users' actions and informs users about additional ways of getting help or background information.

ADELE, Herman the Bug, Cosmo and Smart Egg are examples of agents implementing the behaviour-space approach, while STEVE is an example of the generative behaviour approach. Johnson et al. [2000] describe these agent examples as the "first generation" of animated pedagogical agent implementations—all of their communicative capabilities were very limited.

2.3 Affective Pedagogical Agents

Pushing the sophistication of pedagogical agents a level higher, ITS researches say that for pedagogical agents to be believable, they need to monitor the emotional state of the learners in addition to monitoring their cognitive state [Heylen et al., 2005]; this is because of the implicit importance of affect in the everyday relationships between teachers and learners [Cooper, 2002, 2003]. Affect in pedagogical agents brings to light a host of new complex issues to be considered in the design of affective interactions.

The presence of an affective agent in an ITS implies that there are two sets of emotions and goals (somewhat interrelated but independent at the same time): those of the agent and those of the user. Obviously, it is naïve to assume that the emotions of the learner will simply mirror those expressed by the agent. If the agent is always happy, perhaps the users may get put off because the agent's behaviour may be perceived as phony and inadequate. In addition, people are known to be irritated and distracted by the inappropriate *helpful* interventions and bouncing of Microsoft's Office assistants, particularly the Microsoft Paper Clip [Schaumburg, 2001; Serenko, 2007].

Interaction with a human tutor may contain social acts with emotion-changing potential [Heylen et al., 2005]; for example the tutor often has to offer critiques or praise of the user's actions. Human tutors are capable of taking into consideration the effect of their feedback. If the pedagogical agents are to mimic the human tutors' affective behaviour, the agents' designer need to endow them with *social* and *emotional* intelligence. Social intelligence is the ability to understand the emotions of interaction partners; emotional intelligence is the ability to understand one's own emotions. Goleman [1995] elaborates on the subject of social and emotional intelligence by saying that rather then dealing with raw affect, social and emotional intelligence involves the knowledge of how to play in an interpersonal, social game while being sensitive to the feelings of others. Based on such knowledge, affective pedagogical agents should be able to deliberately choose an affective-behavioural strategy suitable for achieving the desired effect; the agents should possess the knowledge of how to link the cognitive and affective experience of the learner in an attempt to meet the learner's needs.

Brna et al. [2001] state that pedagogical agents need to have knowledge of users' emotional and cognitive traits beyond the current session; patience, tolerance, appreciation, consideration of how well the learner accepts emotional support, understanding of how malleable the learner's emotional state is—all these factors may impact the course of interaction on a case-by-case basis. Consequently, affective pedagogical agents need to embody a higher order of emotional behaviour; they have to maintain the history and status of their own emotional state and that of the learners, and they have to have the capability of self-regulation of emotional state and support for the learner's emotional state.

Heylen et al. [2005] summarise the major variables and factors participating in the interplay of cognitive and emotional states (Table 2.2). The mental state row lists some of the fundamental psychological states that influence the learning process. The row of emotional axes lists the affective polarities influenced by the mental state. The values for the emotion axes are described in greater detail in Section 2.7.

Social emotions are closely connected with interpersonal attitudes, including dominance and affiliation, antagonism, and trust; these define how the student and tutor relate to each other on an interpersonal level. Attribution, for instance, is one of the central topics in social psychology and it plays a crucial part in the tutoring situation; it concerns the way people explain their own behaviour and that of others. In the learning context attributions typically describe the reasons for the students' success or failure; tutors and students may disagree about each other's attributions. *Perceivers* tend to overestimate the influence of personal or dispositional factors and underestimate the situational factors; *actors*, on the other hand, tend to attribute the outcomes of their actions more to situational factors.

Attributions may affect such mental states as motivation and attention. These kinds of effects have to be considered by the emotionally and socially intelligent tutor. In addition to the assessment of current affective state of each other, beliefs the participants have about one another are crucial to meaningful and effective communication in real-life social interaction; similarly, affective pedagogical agents also need to rely on

Variables	Values
Mental state	Learning Success, Attention, Collaboration, Motivation, Self-Presentation, Self-Esteem, Attribution
Emotion axes	Anxiety — Confidence, Boredom — Fascination, Frustration — Euphoria, Dispirited — Encouraged, Terror — Enchantment
Social emotions	Embarrassment, Pride, Dislike, Joy for Other, Gratitude
Interpersonal factors	Dominance, Affiliation, Trust

Table 2.2: Variables influencing affective and cognitive states. Adapted from Heylen et al. [2005].

the knowledge about the users with whom they interact. In addition, the potential of emotion-changing events may vary with a particular person's personality type.

Burleson [2006] presents a view which brings together the interactions between motivation, task progress and learning styles based on the work of Dweck [2000]. Experimental results suggest that one's beliefs about their own intelligence profoundly affect their motivation, learning, and behavioural strategies, especially in response to their perception of failure; Dweck's research has identified two predominant groups of learners: *incrementalists*, who believe their own intelligence can be enhanced, and *trait-based* learners, who believe their intelligence is largely fixed. When *incrementalists* fail at a task, their intrinsic motivation tends to increase, and they tend to believe that they can improve their future performance; however, when trait-based learners fail, they exhibit avoidance and decreased intrinsic motivation, believing instead that their previous performance more accurately defines their ability.

In order to manage the explosive complexity of possible combinations of personal and social traits, researchers tend to focus on small subsets of the parameters of interaction and agent characteristics. The topic of affect expression in agents' action has received much attention in recent years. Affect may be communicated through speech, tone of voice, facial expressions and gestures. Previous experimental studies describe multiple methods for learners' affective state recognition. Heylen et al. [2005] suggest that the main elements that should be considered for determining the emotional state of the learner are the level of the user's activity, the difficulty of the task, and the long and short-term emotional and cognitive history of the learner. A detailed description of affect-recognition strategies is given in Section 2.4. The remainder of this section contains brief descriptions of a few instances of affective agents and affect-aware systems.

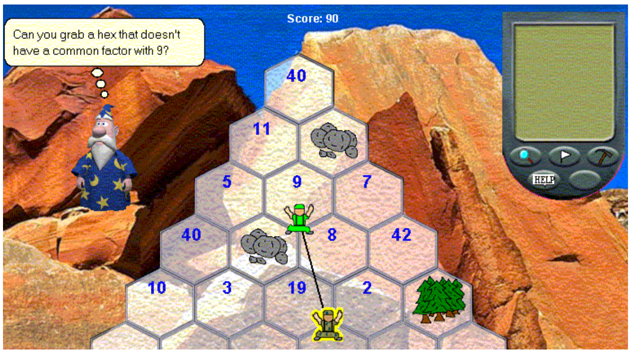
The work of Conati and Maclaren [2004] and Conati [2002] describes the Prime Climb factorisation tutor (Figure 2.7a) designed to integrate the model of student learning with a probabilistic model of student emotions; the Prime Climb agent is implemented with the Microsoft Agent Package. Each of the two players involved in the game has a pedagogical agent who provides on-demand and unsolicited hints. The

Prime Climb agent's tailored hints aim at helping students learn number factorisation. The knowledge of user's affect is required for improving the accuracy of the agent's interventions; the calculations of the student's affect are based on a probabilistic student model and appraisal of the learner's actions, where Dynamic Bayesian Networks (DBNs) are used to handle the uncertainty in the student model assessment.

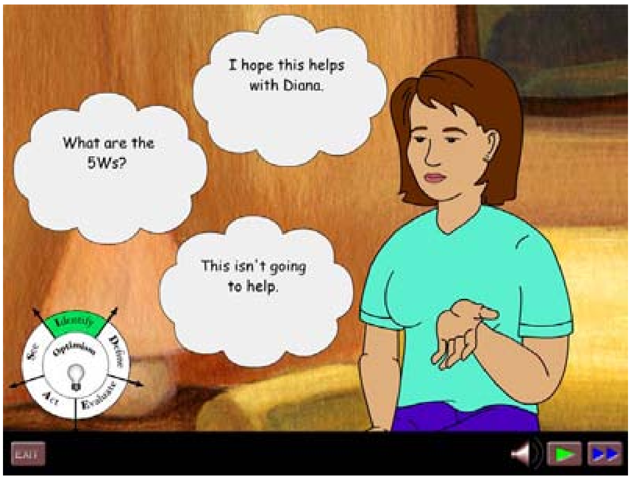
Carmen's Bright Ideas is a multimedia environment designed to teach problem solving skills to mothers of pediatric cancer patients. Learners make decisions and take actions on behalf of a character in the story, and see the consequences of their decisions; the story characters are realised by autonomous agents. This project on Interactive Pedagogical Drama for health interventions [Marsella, 2003; Marsella et al., 2003] uses affective elements of textual dialogue to inform and adapt their agents with the goal of altering the user's affective states through changes in their perspective rather than in the task.

The affective aspects of tutoring dialogues are explored in INES—an ITS designed to help students practice nursing tasks using a haptic device and a virtual environment (Figure 2.7c). INES pays special attention to affective control in the tutoring process by selecting appropriate feedback; it takes into account the elements of the student's character, the detrimental consequences of the student's errors, and the emotional effects of errors and feedback [Heylen et al., 2005; Poel et al., 2004; Heylen et al., 2004].

Recent work by Baker et al. [2006] introduces an agent designed to detect and respond affectively to users gaming an ITS; the system gives a gaming student supplementary exercises focused on exactly the material the student bypassed by gaming, and responds with negative emotions through an animated agent. Scooter the Tutor (Figure 2.8a) is designed to benefit students by serving as a continual reminder that the student should not game. Scooter is also intended to invoke social norms in students by expressing negative emotion when students game—Scooter's display of anger is a natural social behaviour in this context; Scooter's pet-like appearance makes the anger appear less challenging, but at the same time it conveys the message about Scooter's strong attitude towards gaming.



(a) Merlin in Prime Climb.

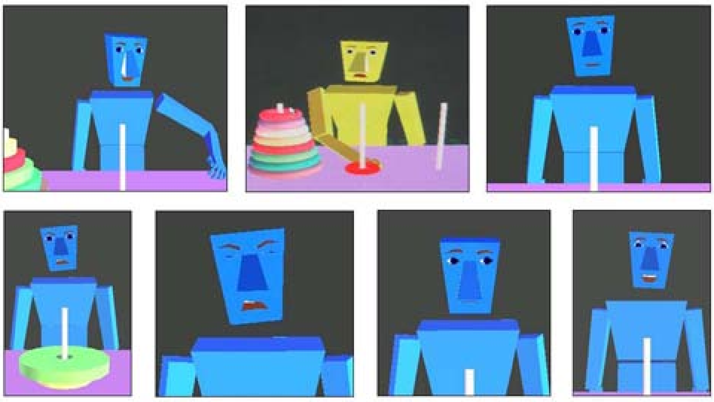


(b) Interaction with Carmen through though balloons.

(c) A view of INES with a haptic device in the foreground.

Figure 2.7: Examples of affective pedagogical agents and environments. Image (a) adapted from Conati [2002]; image (b) adapted from Marsella [2003]; image (c) adapted from Heylen et al. [2004].

(a) Scooter the Tutor in action.



(b) The Learning Companion's range of affective expressions.

(c) The main screen of Tactical Language Training System.

Figure 2.8: More examples of affective pedagogical agents and environments. Image (a) adapted from Baker et al. [2006]; image (b) adapted from Burleson [2006]; image (c) adapted from Johnson et al. [2005].

The work of Burleson and Picard [2004] and Burleson [2006] describes Learning Companion, an agent capable of maintaining a dialogue with the learner and displaying a wide range of emotional expressions. Leaning Companion is a part of a multi-modal system which can pick up the clues of the learners' affective state in real time through a set of physiological sensors and video stream. When a student smiles, Learning Assistant can choose to smile back; this responsiveness is possible because of the agent's behaviour engine and dynamically scripted repertoire of actions.

The Centre for Advanced Research in Technology for Education (CARTE) at the University of Southern California takes the idea of pedagogical agents even further in the Tactical Language Training (TLT) system [Johnson et al., 2004, 2005], which is designed to rapidly teach students how to speak in a foreign language and behave in a foreign culture. TLT combines an intelligent tutoring system with a 3D game where learners get to practice their skills in simulated social situations. In the system, the learners interact with a tutoring agent that listens to their pronunciation and gives appropriate feedback, while in the game they have conversations with characters using speech and gestures. Figure 2.8c shows the player's character in the centre introducing himself to an Iraqi man in a café; the player is accompanied by an aide character (middle left), who can offer suggestions on what to do if the player gets stuck. A virtual tutor evaluates the learner's speech and gives feedback on errors, while providing encouragement and attempting to help the learner overcome his negative affect.

The work of Graesser et al. [2005] reports on the development of a version of AutoTutor with a pedagogical agent Marco that perceives and responds to learner emotions. This version of AutoTutor is augmented with sensors and signal processing algorithms that detect the affective states of learners. Emotions are classified on the basis of dialogue patterns, the content covered, facial expressions, body posture, mouse haptic pressure, and keyboard pressure. The AutoTutor project is developed for the improvement of the methodologies designed for basic affective states classification (such as confusion, frustration, boredom, interest, excitement, and insight) through the analysis of patterns of facial expressions, body movements, and dialogue activity during interaction with AutoTutor.

Gratch et al. [2002] emphasise the difference between the communication-driven and the simulation-driven approaches to agents' emotional displays. In the communication-driven approach, a pedagogical agent chooses its emotional expressions on the basis of its desired impact on the user; this approach is adopted by most tutoring applications. The second approach is concerned with simulating true emotions. Cosmo from Internet Protocol Advisor (Figure 2.6d in Section 2.2) is an example of the first approach; the multi-agent Mission Rehearsal Exercise environment is an example of the second approach [Marsella and Gratch, 2001].

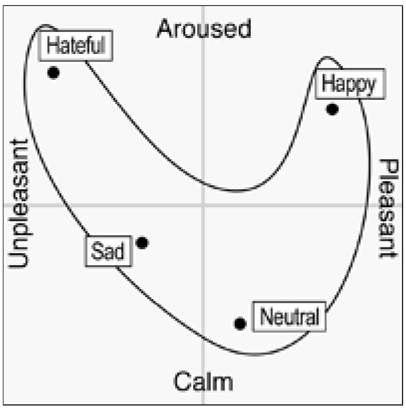
2.4 Study of Emotions

This section provides an overview of the psychological and psychophysiological perspective on emotions and the ways that emotions are detected, measured and categorised. Psychology describes emotions as psychological processes that function in the management of goals. An emotion is typically elicited by evaluating an event as relevant to a goal; the emotive response is positive when the goal is approaching and negative when progress towards the goal is impeded [Wilson and Keil, 1999]. In this way, emotions are viewed as action dispositions [Bradley and Lang, 2000]. The core of an emotion is readiness to act in a certain way [Frijda, 1986]; it is an urgency, or prioritisation of some goals and plans over others.

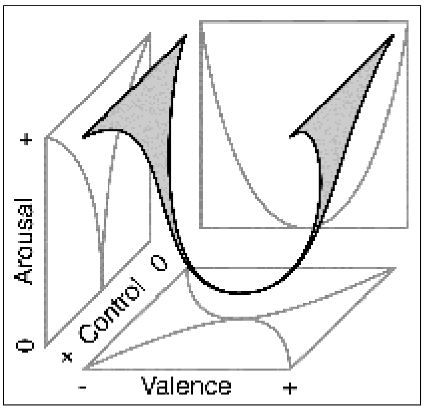
2.4.1 Dimensional versus Categorical Views of Emotions

There are two major theoretical approaches to the study of emotion: dimensional and categorical. Theorists who use the categorical approach to emotion attempt to define specific categories or types of emotions [Izard, 1977; Plutchik, 1980; Ortony and Turner, 1990]. Research in this area suggests that there are a number of basic emotions (estimates range from three to more than 20) which are combined to produce all the emotional states which people experience. Among the emotions most often designated as basic emotions are disgust, anger, happiness, grief, and fear. In addition, intermediate emotional states are accounted for through the blending of various basic emotions.

The dimensional approach [Bradley, 1994; Bradley and Lang, 2000; Lang et al., 1995] conceptualises emotion as having two or perhaps three basic underlying dimensions along which the entire range of human emotions can be arranged. The most common dimensions are valence (which ranges from happy to sad) and arousal (which ranges from calm to excited). The third less-often mentioned dimension is dominance (ranging from in control to out of control). Figure 2.9a shows a 2D emotion space with some common emotion labels mapped on it; Figure 2.9b shows the shape of a 3D emotion space. Research using the dimensional approach has shown that emotions elicited by pictures, television, radio, computers and sounds can be mapped onto an emotional space defined by the arousal and valence axes, where the viewers' levels of arousal and valence consistently predict emotional, cognitive, and physiological responses to emotional stimuli.



(a) Mapping of emotions onto a two-dimensional emotion space.



(b) Three-dimensional emotion space.

Figure 2.9: Two and three-dimensional emotion spaces. Adapted from Dietz and Lang [1999].

In our research, we adopt the dimensional approach: the continuous nature of valence and arousal in this approach (versus the discrete states in the categorical approach) underpins the choices which determine the implementation of modelling the agent's emotions in EER-Tutor described in Section 3.3 and the algorithm for detecting and modelling user emotions is presented in Sections 4.3 and 4.4. The dimensional approach eliminates the need for classifying the emotional states as belonging to certain categories; potentially, this resolves a number questions in emotion modelling and emotional response in the form of facial expressions. For example, within the categorical approach the decision of where to draw the boundary between the neutral and happy state would involve the definition of some type of threshold for the observed parameters which measure affective response; with the dimensional approach, such questions do not arise—the label associated with a particular emotional display carries a lot less significance then the observed parameters of the emotional state.

2.4.2 Emotion Observation and Measurement

Emotions produce multiple responses and thus is it common to group them into three broad categories [Lang, 1993]:

1. Overt acts of behavioural sequences, including the defining survival actions or their variants (fight, flight, sexual approach, threat displays), as well as the modulation by emotion of these behaviours (as in stress-induced performance deficits or in mood-dependent association).

2. Emotional language, including expressive communication (threat or distress cries, sounds of contentment and sexual passion, verbal attack), and evaluative reports (descriptions of feelings, attitudes and self-ratings).

3. Physiological reactions, changes in the somatic muscles (regulating voluntary movement) and in the viscera (internal organs of the body, especially the heart,

liver, or intestine), that are the logistic support of overt acts, of associated affective displays (facial muscle patterns, blushing), or of preparation for these responses.

Physiological and behavioural reactions to affective stimuli significantly correlate with judgements of affective valence and/or arousal; research literature describes several of these relationships [Greenwald et al., 1989; Lang et al., 1993; Bradley et al., 1992; Bradley and Lang, 2000]. In particular, there is a high dimensional correlation between valence reports and electro-myographic activity of *corrugator supercilii*, *zygomaticus major* and *zygomaticus minor* muscles (shown in Figure 2.10). *Corrugator* muscles are located under the skin of the medial half of the eyebrow; they pull eyebrows together medially and cause the furrowed brow of the worried look, a facial action which is an index of distress [Ekman and Friesen, 1986; Fridlund and Izard, 1983]; thus, significant firing of motor units in the *corrugator* muscles are expected when an affective stimuli is judged to be unpleasant. *Zygomaticus major* elevates and draws the corner of the mouth laterally; *zygomaticus minor* elevates the upper lip; both of these muscles are involved in the smile response; *zygomatic* muscle activity increases for pleasant stimuli, peaking for stimuli high in affective valence.

Psychophysiological research literature describes affective state indexing based on a number of physiological responses obtained through physiological sensors:

Surface Electromyography (EMG) measures muscle activity by detecting and amplifying the electrical impulses that are generated by muscle fibers when they contract. The amplitude of the resulting electrical signal is proportional to the strength of contraction. For affective valence indexing, an EMG sensor is placed on the skin surface on the *corrugator* muscle.

A single EMG sensor placed on the *corrugator* muscle is capable of providing only half of the data required to estimate the emotional valence, because the *corrugator* muscle activity indicates negative affective valence. In order to make predictions of positive affective valence, it is necessary to measure activity of

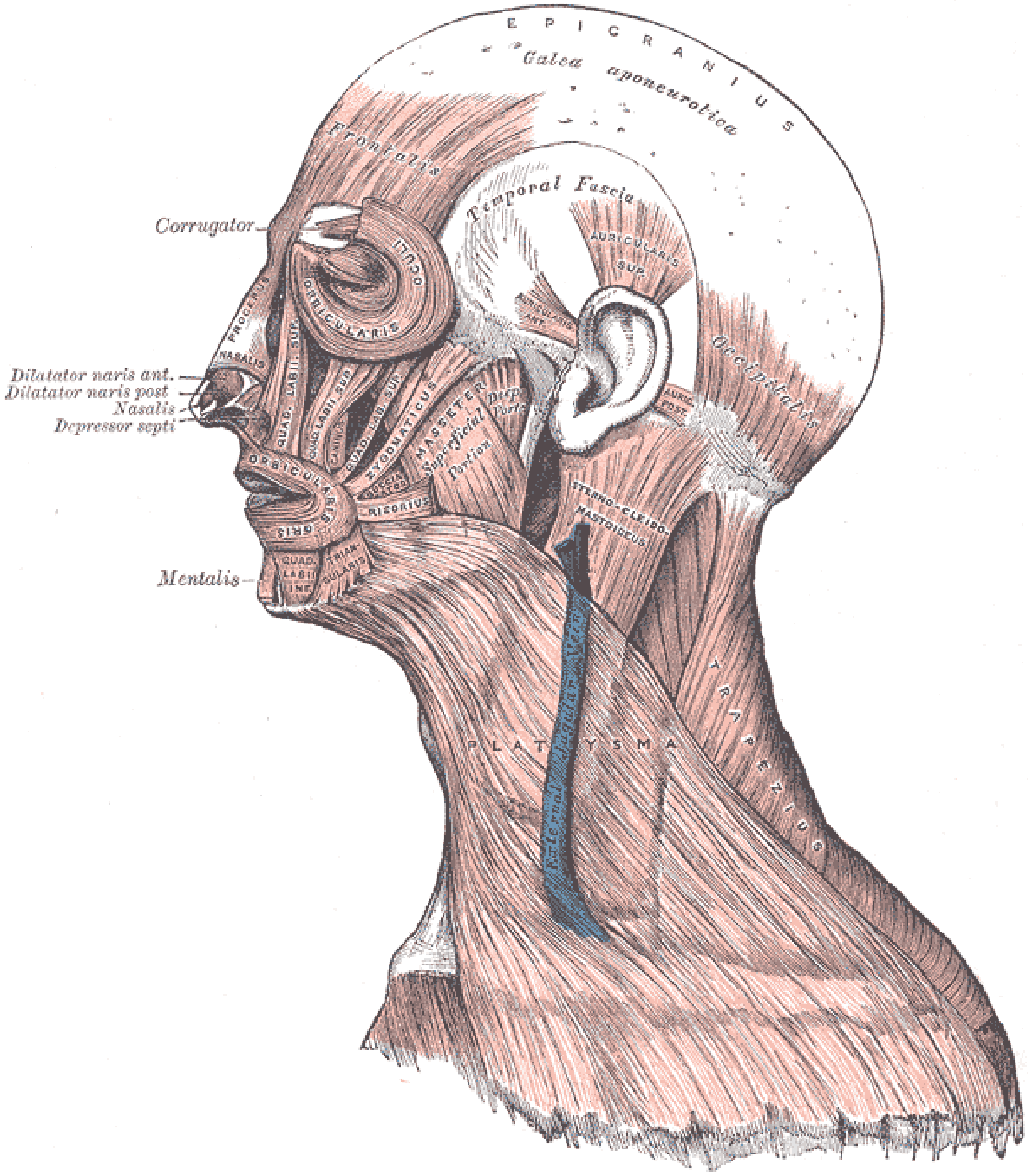


Figure 2.10: Muscles of facial expressions. Adapted from Gray [1918].

the *zygomatic* muscles. Literature on physiological data processing also indicates that the EMG signal is not always reliable unless the top layer of dead skin cells is removed with abrasive cream before the application of the sensor electrodes [Blascovich and Seery, 2005]. This type of preparation for experiments may be considered as intrusive by the subjects. Figure 2.11a shows an EMG sensor.

Galvanic Skin Response (GSR) measures skin's ability to conduct electricity. "Skin conductance", "electro-dermal response" and "galvanic skin response" are different terms for similar physiological measures. GSR is described as a reliable measure of arousal. The amount of skin conductance activity increases linearly as ratings of arousal increase, regardless of emotional valence. Low electrical voltage is applied through two electrodes in order to establish an electric circuit where the subject becomes a variable resistor. The variation in conductance, which is the inverse of the resistance, is the indicator of GSR. Figure 2.11b shows a GSR sensor; GSR sensor is usually strapped to two fingers of one hand.

Blood Volume Pulse (BVP) measurement relies on photoplethysmography[5]. BVP sensor shown in Figure 2.11c is to be worn on a fingertip. Experimental studies measuring BVP signal in the absence of physical exertion document significant heart rate deceleration for unpleasant stimuli as well as greater peak acceleration for pleasant materials [Bradley and Lang, 2000]. The Electro Cardiogram (ECG) sensor can also be used to monitor heart rate.

Respiration (RSP) measures breathing frequency and amplitude. The RSP sensor shown in Figure 2.11d is sensitive to stretch. When strapped around a chest or abdomen, it converts the expansion and contraction of the rib cage or abdominal area, to a rise and fall of the signal. RSP sensor is commonly used in conjunction with other sensors to aid in the detection of affective states, as demonstrated by the work of Healey and Picard [2000].

Figure 2.11e shows BodyMedia SenseWear armband—a device for collecting GSR, body temperature and heart rate (through a chest strap which works with the armband).

[5] Photoplethysmography is based on bouncing infra-red light against the skin surface and measuring the amount of reflected light; this varies with the amount of blood in the blood vessels under the skin. After each heart beat, when there is more blood in the skin, blood reflects red light and absorbs other colours. Between pulses, the amount of blood decreases and less red light is reflected.

Nasoz et al. [2003] used such device to analyse physiological signals associated with emotions in order to recognise affective states of users; such module was used in a multi-modal system for affect detection. The current level of physiological sensors technology simplifies the process of data collection; the physical contact of the user with an I/O device (such as a keyboard, a mouse, a hand-held device, or a touch pad) may be enough to provide the measurements of the user's heart rate, blood pressure, respiration, temperature, and skin conductivity.

(a) EMG sensor.

(b) GSR sensor.

(c) BVP sensor.

(d) RSP sensor.

(e) BodyMedia SenseWear armband.

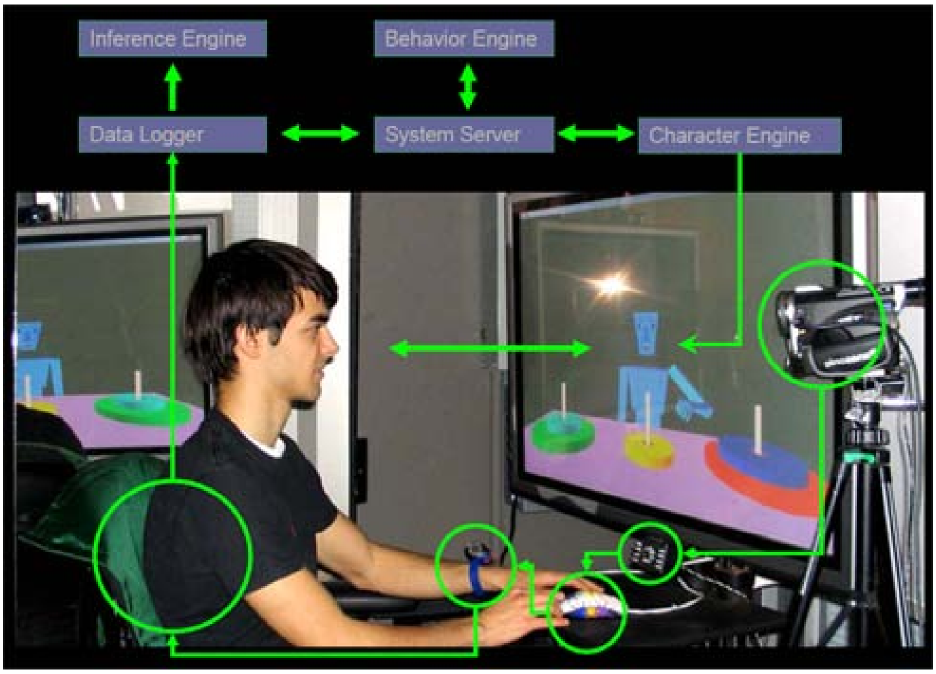
Figure 2.11: Physiological sensors. Images (a), (b), (c) and (d) adapted from Thought Technology Ltd. [2000]. Image (e) adapted from Teller and Stivoric [2004].

Bosma and André [2004] report results of a small-scale study with the use of EMG, GSR, ECG and RSP sensors (Figure 2.12a). The report describes a system that uses affective state recognition to disambiguate dialog acts; that is, the authors were interested in determining the level of the user's commitment to the agent in a problem-solving environment enabled with natural-language processing. Figure 2.12b shows the research platform for the Learning Companion agent presented in the work of Burleson and Picard [2004] and Burleson [2006] (Figure 2.8b). Apart from the GSR sensor, the system uses a sensor to detect the intensity of the user's grip on the mouse (which has been shown to correlate with frustration) and a posture sensor in the chair (used to classify motivational states such as engagement, boredom, and break-taking). In addition, a facial feature tracking application detects facial expressions and movements such as head nod/shake, mouth fidgets, smiles, blink events, and pupil dilations. These sensors have been developed over the past several years and validated in a variety of experiments by the Affective Computing Group at the Massachusetts Institute of Technology.

The work of Nasoz et al. [2004] and Paleari et al. [2005] describes VALERIE (Virtual Agent for Learning Environment Reacting and Interacting Emotionally)—a project built within the MAUI (Multimodal Affective User Interface) framework [Lisetti and Nasoz, 2002]. MAUI framework is designed for determining affective states on the basis of inputs which register autonomic nervous system signals, facial expressions and voice; Figure 2.13 shows the architecture of the MAUI framework which integrates the kinesthetic, auditory and visual modalities. The experimentation with the framework reports the use of three Machine-Learning algorithms: (1) k-Nearest Neighbour (KNN) algorithm [Mitchell, 1997], (2) Discriminant Function Analysis (DFA) [Nicol and Pexman, 1999], and (3) Marquardt Back-propagation (MBP) algorithm [Hagan and Menhaj, 1994] for recognising sadness, anger, surprise, fear, frustration and amusement, with the MBP algorithm performing better than both DFA and KNN for all emotion classes except for surprise.

Healey and Picard [2000] use input from EMG, ECG, RSP and GSR sensors to detect stress in car drivers. Conati et al. [2003], however, point out in their work that the level of noise in physiological signals increases in the environments where subjects have high mobility, because the sensors are sensitive to motion artifacts. Bradley and Lang [2000] state that physiological signals, like EMG, BVP or RSP are not easy to interpret; results of experimental studies often appear contradictory or inconclusive.

(a) Affect recognition for dialogue disambiguation.



(b) Research platform for affect recognition in Learning Companion's project.

Figure 2.12: Physiological sensors in learning environments. Image (a) adapted from Bosma and André [2004]; image (b) adapted from Burleson [2006].

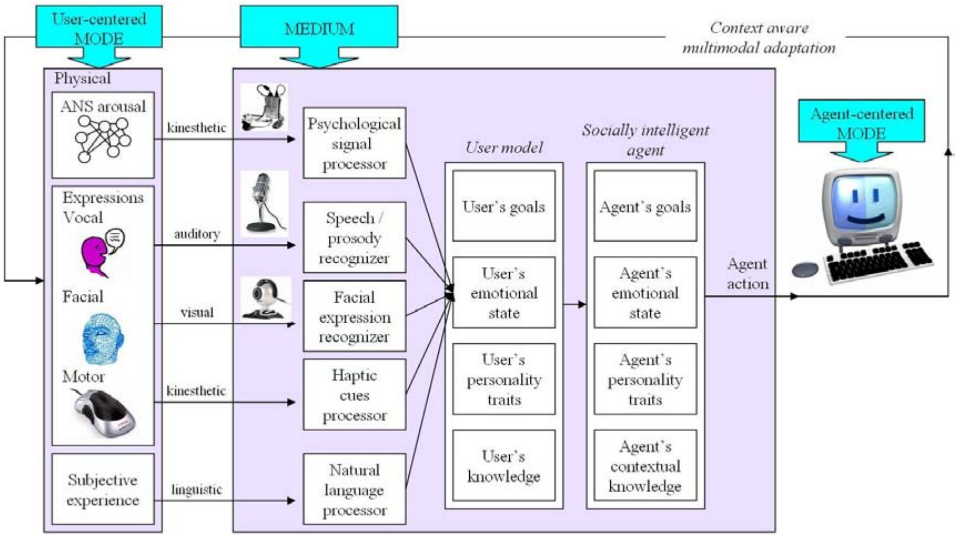


Figure 2.13: MAUI framework architecture. Image adapted from Paleari et al. [2005].

The comprehension of emotion response diversity continues to remain the central challenge to research in emotion theory; for practical purposes of affective state indexing, researchers are recommended to focus on small subsets of physiological measures.

Higher-order Psychological Constructs

Social Psychology theorists state that a perfect physiological index would have a completely transitive one-to-one relationship with the psychological construct [Cacioppo et al., 2007]; unfortunately, such completely transitive relationships are relatively uncommon. Instead, many-to-many relationships between psychological constructs and indexes based on physiological responses appear to be the rule. The work of Blascovich and Seery [2005] suggests that the likelihood of observing one-to-one relationships between the psychological constructs and physiological indexes can be increased by careful limitations of the context and by constructing indexes based on overlapping multiple responses over time. Several physiological models provide theoretical linkages for establishing physiological indexes of these higher-level social psychological constructs.

An example of such model is the model of challenge and threat; Blascovich and Seery [2005] present a biopsychosocial model[6] describing the interaction of affective and cognitive processes; the theory claims that stress is a lay concept that smothers the distinction between important variables and gradations in affective and cognitive processing. As an alternative to the concept of stress, the theory suggests existence of two motivational states: challenge and threat. These states are considered in goal-oriented situations with minimal metabolical demand and maximised requirements for cognitive responses, such as learning, taking tests, decision making processes, and so on. Challenge is characterised by one's resources balancing the difficulty of the task, while threat is characterised by the person's resources being outweighed by complexity of the task. Observable transitions between the states of challenge and threat depend on neural and endocrine responses, such as blood vessel contractility[7], cardiac output[8] and the resulting total peripheral resistance of the vascular system.

The theory claims that the balance of challenge and threat underpins the interactions between cognitive and affective variables. Challenge is regarded as a desirable factor that increases one's cognitive performance, while threat is the factor responsible for hindered performance. It has been shown that subjects under a state of challenge outperform subjects under a state of threat with very few exceptions. The argument of this theory potentially opens a new horizon for further ITS research and experimentation. However, reliable indexing of challenge and threat requires elaborate physiological sensors instrumentation for recording cardiovascular responses. In the course of our work, we explored the feasibility of directing our focus towards detection of higher-level mental states such as challenge and threat, but we concluded that the challenges posed by this approach could not have been resolved in the time frame allocated for a Master's thesis.

[6]The biopsychosocial model proposes that biological, psychological and social factors all play a significant role in human functioning, including in mental processes.

[7]Speed and strength of blood vessel contraction.

[8]Blood volume pumped through the heart.

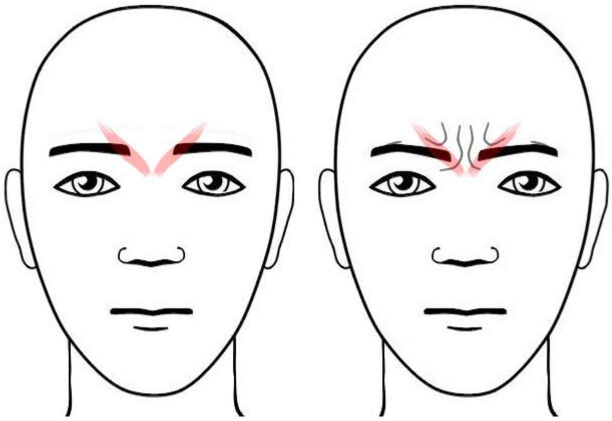
2.5 Facial Feature Tracking

As an alternative to EMG signal, affective signal processing may rely on Computer Vision techniques for facial feature tracking. With facial feature tracking, positive affective valence can be read through variations of the distance between the corner of the mouth and the outer corner of an eye on one side of a face; negative affective valence can be read through variations of distance between the inner corners of eyebrows. This approach is described in further detail in Section 4.2. Facial Action Coding System (FACS) [Ekman and Friesen, 1978] labels the action units associated with the *Corrugator supercilii* and *Zygomaticus major* as units №4—*Brow Lowerer* (Figure 2.14a) and №12—*Lip Corner Puller* (Figure 2.14b) respectively[9].

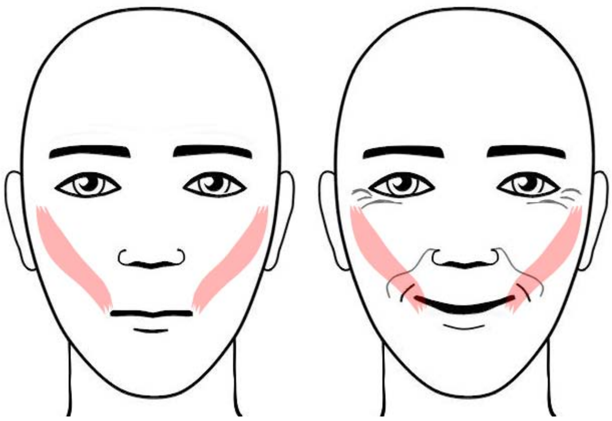
Computer Vision, and in particular, facial feature tracking, is not the focus of our research; however background information is included because it is relevant to the implementation of the affective pedagogical agent for our research. There are two general approaches to facial recognition and feature tracking: holistic and feature-based. The first approach treats the face as a complete unit, while the other approach searches for specific features in the image to identify a face. Prior research in Computer Vision reports numerous successful techniques for face recognition and facial feature tracking:

Principal Component Analysis (PCA) method can be used for face detection, feature detection and emotion recognition as described in the work of Calder et al. [2001]. PCA is based on information theory concepts; it seeks a computational model that best describes a face by extracting the most relevant information contained in that face. The Eigenfaces approach is a PCA method, in which a small set of characteristic pictures are used to describe the variation between face images. The goal is to find the eigenvectors (eigenfaces) of the covariance matrix of

[9] Incidentally, Bosma and André [2004] state in their work that they used EMG signal to detect frowning with the sensors placed as in Figure 2.12a; this placement of the sensors, however, does not allow measurement of the activity of the "frown muscle", *Corrugator supercilii*. Instead it measures the activity of *Frontalis (pars lateralis)* muscle (facial action unit №2, Outer Brow Raiser), which is described as the attention muscle, associated with the feelings of astonishment, surprise, admiration, fright or terror



(a) Action unit №4.



(b) Action unit №12.

Figure 2.14: Facial action units associated with changes in affective valence. Images adapted from Flores [2005].

the distribution, spanned by training a set of face images. Later, every face image is represented by a linear combination of these eigenvectors. Recognition is performed by projecting a new image onto the subspace spanned by the eigenfaces and then classifying the face by comparing its position in the face space with the positions of known individuals.

Active Appearance Model (AAM) is a technique for matching a statistical model of object shape and appearance to a new image. The models are built during a training phase. The training supervisor provides a set of images, together with coordinates of landmarks that appear in these images. The algorithm uses the difference between the current estimate of appearance and the target image to drive an optimisation process. By taking advantage of the least squares techniques, it can match to new images very efficiently. The approach is used for matching and tracking faces as described in the work of Cootes et al. [2001].

Gabor wavelets method represents a face template as a linear combination of continuous 2D odd-Gabor wavelet functions (filters). The weights and position, scale and orientation parameters of each wavelet are determined during the network training phase so that the maximum amount of image information is preserved for a given number of wavelets. The use of Gabor filters in image analysis is biologically motivated as they model the response of the receptive fields of the orientation-selective cells in the human visual cortex [Daugman, 1985]. Feris et al. [2002] describe the use of multi-level networks of Gabor wavelets for estimation of facial feature positions.

Hybrid approach is described in the work of Bourel et al. [2000]; detection of facial features is done through application of the common image processing techniques, such as edge detection, adaptive thresholding and so on. There are a number of similar hybrid techniques published in recent years.

Some of the methods described above have been used in the commercial toolkits for facial feature tracking and emotion detection. For example, the Logitech Video Effects package uses a Gabor-based facial feature detection software development kit (SDK) created by Neven Vision. Considering the prohibitive price of Neven Vision SDK and similar commercial products, we developed an in-house facial feature tracking application based on the functionality available in the Open Computer Vision (OpenCV) library provided by Intel[10] [Bradski, 2000]. The implementation details of our approach to facial feature tracking are provided in Section 4.2.

[10] See the Open Source Computer Vision Library—`http://www.intel.com/technology/computing/opencv/`

2.6 Interaction of Affective and Cognitive States

A number of studies expose complex interaction between cognitive and affective states [Goleman, 1995]. People engulfed by negative feelings display diminished abilities with respect to attention [Niedenthal and Kitayama, 1994], memory retention [Kahneman, 1973], learning [Lewis and Williams, 1989], creative thinking [Isen et al., 1987] and polite social interaction [Goleman, 1995]. Uncontrolled emotions can have profound long-term effects such as heightened levels of stress, development of severe depression [Gross and Muñoz, 1995], addiction [Cooper et al., 1995], and other health problems [Goleman, 1995].

Human intelligence, however, imbibes emotion-regulating skills and strategies that can mediate emotional states to varying degrees [Gross and Muñoz, 1995]. Emotion regulation has been identified as a primary component of emotional intelligence. Emotional intelligence implies the ability to recognise and express emotions, coupled with the ability to regulate these emotions, harness them for constructive purposes, and skillfully handle the emotions of others[11]. Emotional self-regulation is described as an aptitude and skill for modulating and managing one's emotional state.

In emotion management, people tend to rely on a variety of methods, such as interaction with media and/or other people, engagement in sports or work, meditation or praying, use positive thinking, and consumption of foods and other substances such as alcohol, tobacco, and drugs. Myers [1998] describes two generic varieties of support for emotion regulation applicable in human-human interaction (HHI): passive support and active support. Passive support is used by people to manipulate moods, without necessarily addressing or discussing the emotions themselves. Media, activities, food and other substances fall into this category. In contrast, active support occurs when people discuss or otherwise address their emotions directly, as a means of managing them; talking to a parent or friend about what is upsetting, and how that makes the person feel,

[11]The skills of emotional intelligence have been argued to be a better predictor than IQ for measuring aspects of success in life (Goleman 1995, citing Vaillant 1995; Felsman and Vaillant 1987).

is an example of active emotional support. Social interaction can fall into either category. Active listening is perhaps the best known example of active support—it may be described as providing sincere, non-judgemental feedback to an emotionally distressed individual, with a focus on providing feedback of the emotional content [Nugent and Halvorson, 1995]. Active listening, when practised effectively, demonstrates elements of both empathy[12] and sympathy[13].

Since humans are predisposed to respond socially to computers, the same tactics for emotional support and response that people use in real-life communication are applicable in HCI in general, and more specifically, in computer-mediated educational systems capable of recognising emotions. For example, recent research proves that computers can relieve feelings of frustration by providing support for the users' natural ability to regulate their emotions [Klein et al., 2002]. ITS researchers during the last decade have been trying to determine what affective response an ITS should offer to guide the user through solving complex problems and what can be done to ensure high knowledge transfer rates and to support the user's determination in the face of stress, anxiety and frustration in the process.

As discussed in Section 2.7, human tutors can pick up many clues from non-verbal communication; with varying degrees of depth, the tutor is always aware of the learner's affective and cognitive state. This awareness allows a human tutor to choose an appropriate affective response at each step. Unfortunately, there is no *cookbook* defining all the rules for HHI that HCI developers can simply implement [Bickmore, 2004]; however, many of the most fundamental rules have been defined in work by socio-linguists and social psychologists. One simple rule of thumb suggested in the work of Mishra and Hershey [2004] is to apply what has been found appropriate for HHI to the design of HCI. However, the feedback design in current computer-based educational systems is often based on simplistic and erroneous framework where praise is assumed to affect

[12] Empathy is commonly defined as one's ability to recognise, perceive and directly feel the emotion of another.

[13] Sympathy is an emotional affinity in which whatever affects one correspondingly affects the other, and its synonym is pity.

behaviour positively, irrespective of context [Mishra and Hershey, 2004]. There have been studies showing that being praised for success in a task perceived as easy, may have a negative effect on a student's self-confidence; while being blamed for failing in a task perceived as difficult, may lead to a positive effect [Henderlong and Lepper, 2002].

2.7 Affective Etiquette in Learning

HCI does acknowledge the need to avoid negative affective states such as frustration. To avoid a user being swamped by negative affect, frequently mentioned in HCI solutions include either (a) trying to determine and fix the problem causing the negative feelings, and/or (b) preemptively trying to prevent the problem from happening in the first place [Norman, 1988]. However, there are some fundamental differences between general HCI etiquette and educational HCI etiquette, because it is unrealistic to expect the user to stop making errors during learning. Kay's [1991] view on educational systems is as follows:

> "difficulty should be sought out, as a spur to delving more deeply into an interesting area. An education system that tries to make everything easy and pleasurable will prevent much important learning from happening".

Learning can be frustrating and difficult because it involves exposing learners' errors in thinking and gaps in their knowledge, while forcing them to comprehend difficult subject matters [Mishra and Hershey, 2004]. Etiquette considerations in educational HCI are complicated by the fact that learning from a computer is not just about ease of use—the theory of learning from performance errors suggests that errors are an inseparable component of a learning process [Ohlsson, 1996]. Error correction, in fact, has a critical significance for the improvement of future performance. Thus the two HCI strategies mentioned above have a limited application in educational systems, because learners have to have freedom to make errors to be able to learn from them.

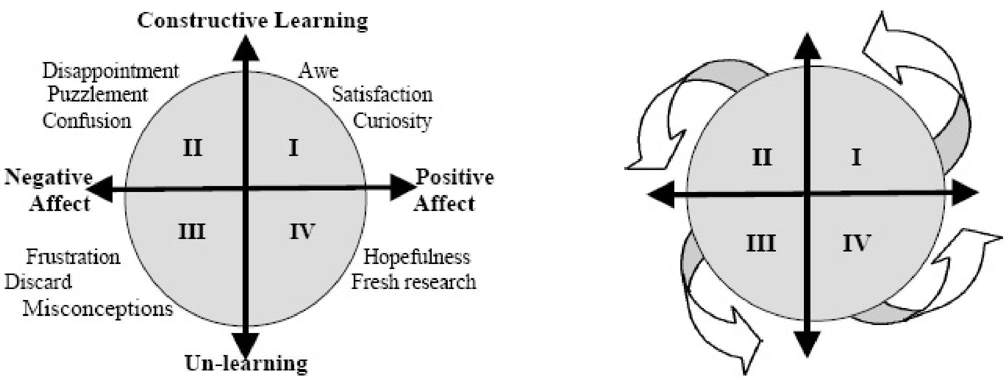
Motivation-supporting Techniques

In their chapter, *Motivation and Failure in Educational Systems Design*, Schank and Neaman [2001] elaborate on how simulated scenarios in the learning-by-doing environments benefit their users; these environments accelerate the pace of learning through exposure to difficult circumstances that may arise less frequently than in real world situations. This inevitably accelerates the rate of failure and, if motivation is sustained, the rate of learning, as "novices are exposed to rare, but critical, experiences". Schank and Neaman acknowledge that the fear of failure is a significant barrier to learning and believe this can be addressed in several ways: minimising discouragement by lessening humiliation, developing the understanding that consequences of failure will be minimal, and providing motivation that outweighs or distracts the learner from the unpleasant aspects of failure. The work of Schank and Neaman shows that by providing learners access to experts at the time of failure, it is possible to sustain learners' motivation as long as they care about what they are doing.

A learner's affective state often determines whether the learner is motivated or challenged; these are key predispositions for certain actions. Heylen et al. [2005] describe how curiosity and puzzlement may lead to investigative actions; frustration also may lead to action, even though it is a more negative affect. The tutor can choose to consider taking certain actions to bring about a change in the emotional state. Lepper et al. [1993] identify four main tactics in motivating learners: challenge the learner, empower the learner with confidence, raise the learner's curiosity, and make the learner feel in control. These tactics can be achieved by various means; for example, the learner can be challenged by being offered appropriately difficult tasks, or by having the difficulty emphasised, or by taking part in some kind of competition. Confidence can be boosted by directly or indirectly maximising the appearance of success. Curiosity is typically raised in Socratic methods when the student is asked to ponder many questions. The tutor can decide to leave the initiative to the student or offer options that suggest that the student can make choices, thereby influencing the student's feeling of being in control.

Kort's Theory of Emotions in Learning

Kort et al. [2001] propose a model of constructive cognitive progress that relates learning and emotion in an evolving cycle of affective states. The model interweaves the emotion axes, such as Anxiety—Confidence, Frustration—Euphoria, Humiliation—Pride, and so on, with the cognitive dynamics of the learning process. Figure 2.15a shows these emotion axes superimposed on the four quadrants defined by the axes of the cognitive directions. In Figure 2.15a the positive valence states are on the right, and the negative valence states are on the left. The upward direction of the vertical axis symbolises the construction of knowledge, and the downward direction symbolises the discarding of misconceptions.



(a) Constructive learning and emotion axes. (b) Flow of emotions in the four-quadrant model.

Figure 2.15: Flow of emotions in learning. Adapted from Kort et al. [2001].

The proposed cyclic trajectory shown in Figure 2.15b ideally begins in Quadrant I and II associated with anticipation, expectation, and exploration, a stage where intervention is to be discouraged as long as the learner stays in the top half of the space, focusing on constructing or testing knowledge. If the learner progresses to disappointment or discouragement (Quadrants III and IV) and stays there too long, then intervention may be productive. Kort et al. [2001] argue that this cycle, including its negative states, is natural to the learning process, and that learners can develop skills to keep moving through the cycle, thus propelling themselves out of the failure mode and into a more hopeful state conducive to continued exploration and learning.

Burleson and Picard [2004] suggest that affect recognition can be used to control and time intervention; an agent devoid of affective awareness can act only on the basis of the problem state. Affective awareness, however, can prevent the agent from intervening inappropriately and from "robbing the learner" of the opportunity to discover the solution on their own.

Flow Theory

Initially proposed by Csikzentmihalyi in 1990, flow is described as a mental state of operation in which the person is fully immersed in what he or she is doing; this state is characterised by a feeling of energised focus, full involvement, and success in the process. The study of flow has shown that conscious awareness of "flow zone" tends to diminish happiness and flow [Csikzentmihalyi, 1990]. Based on these findings, Burleson and Picard [2004] suggest that conscious awareness of frustration, feeling of an impasse and other similar negative influences may diminish these states[14]. If this is the case, affective self-awareness should assist users in mitigating the detrimental influences of negative affective states on their learning. To provide and support such affective self-awareness, the learning system or the affective agent needs to bring to the learner's attention observations about their current emotional state; this can be done either by the means of direct speech communication, textual message or specialised visual interface clues and gadgets. Whatever form it takes, such support can be classified as active emotion support, because it involves some form of communication involving the learner's feelings.

Bringing together the emotional self-regulation, Kort's theory of emotions in learning, and the theory of flow, Burleson and Picard [2004] argue that the task-manipulation

[14] Burleson and Picard [2004], however, alert researches to the phenomenon of negative asymmetry: the staying power of negative affect tends to outweigh the more transient experience of positive affect [Giuseppe and Brass, 2003]; this phenomenon might require special care and consideration in appropriate contexts.

approach sometimes misses important opportunities to help users develop skills to deal with failure and frustration. Instead, Burleson and Picard [2004] advocate and implement an approach that uses affective agents in the role of peer learning companions to help learners develop meta-cognitive skills such as affective self-awareness for dealing with failure and frustration. In their work they make the distinction between, on the one hand, adjusting the environment to facilitate flow, and on the other, empowering the user through self-awareness to participate in self-regulated motivational strategies. In our research we adopt a similar approach by developing an affective agent in the role of a tutor capable of offering active affective support.

CHAPTER 3

Pedagogical Agent for EER-Tutor

The test-bed application, EER-Tutor, and the direction of our research determined the two main requirements defining our choices regarding the implementation of the pedagogical agent for our project. First, since EER-tutor is a Web-based application and its interface is a set of dynamic HTML pages as described in Section 2.1.2, the agent needed to be integrated with the Web-based environment as a part of EER-Tutor's interface. Second, the decision to adopt the dimensional view of emotions in our project as discussed in Section 2.4.1, implied that the agent should have the capability to display emotions in dimensional space rather then categorical space.

People Putty[1] Software Development Kit (SDK) and Microsoft Agent[2] SDK have been used by a number of ITS researchers for developing animated pedagogical agents. People Putty SDK is an easily available commercial tool for building dynamically-controlled 3D virtual characters displayed by Haptek player[3]—a Web-browser plugin. Microsoft Agent SDK is a well-known free solution for building animated characters on the basis of sequences of 2D or 3D images by converting them into animation frames; Microsoft Internet Explorer exposes an Application Programmer Interface (API) for

[1]See `http://www.haptek.com/products/peopleputty/`—Haptek's People Putty SDK.
[2]See `http://www.microsoft.com/msagent/`—Microsoft Agent home page.
[3]See `http://www.haptek.com/`—Haptek Inc., People Putty is a product of Haptek.

integrating Microsoft Agent characters. Both of these packages represent complete solutions which can be used for developing and integrating animated pedagogical agents into ITSs[4]; only People Putty SDK, however, allows for a continuous control (versus discrete) of the agents' emotional appearance.

Out of the possible instructional roles we chose the tutor (mentor) role, because of its positive influence on learning as described in Section 2.2. At the same time, from the affective interaction point of view, passive affective support seems intuitively congruent with the role of tutor; as discussed in Section 2.6, under passive affective support the affective state is not addressed directly. Consequently, we tried to create an agent with a persona of a tutor interested in the learner's progress, and we attempted to provide the agent with the capability of acknowledging the leaner's emotions indirectly through its own empathetic emotional appearance, while trying to keep the learner focused on the task at hand. Thus we designed an agent which expresses solidarity with the user—it cheers with the user's success, is empathetic when the user experiences difficulties and keeps company to the user in neutral situations. Similar agent behaviour was earlier adopted in the work of Lester et al. [2000].

Our approach relies on a simplified model of emotions described by a single dimension—affective valence, ranging from negative to positive. A student's affective state for this version of the agent is inferred on the basis of the student's cognitive state, but the agent's implementation does not explicitly store the student's affective state. Rather, the inferred value of the student's affective valence is used to adjust the value of the agent's valence. This approach creates an effect where the agent mirrors the student's emotions, as if empathising with the student.

[4] Apart from these solutions, there is a variety of sophisticated commercial applications for producing 3D models or frame sequences; however this is not enough for creating an interactive animated character. A dedicated animation engine is required for combining non-interactive frame sequences or "driving" a 3D model to produce interactive behaviours. Along with the animation engine and presentation layer (implemented as a browser plugin), such an agent can be integrated into an ITS. Where resources permit, ITS research groups employ some of the commercial solutions even including 3D game engines; also some researches develop in-house pedagogical agents solutions to suit their needs.

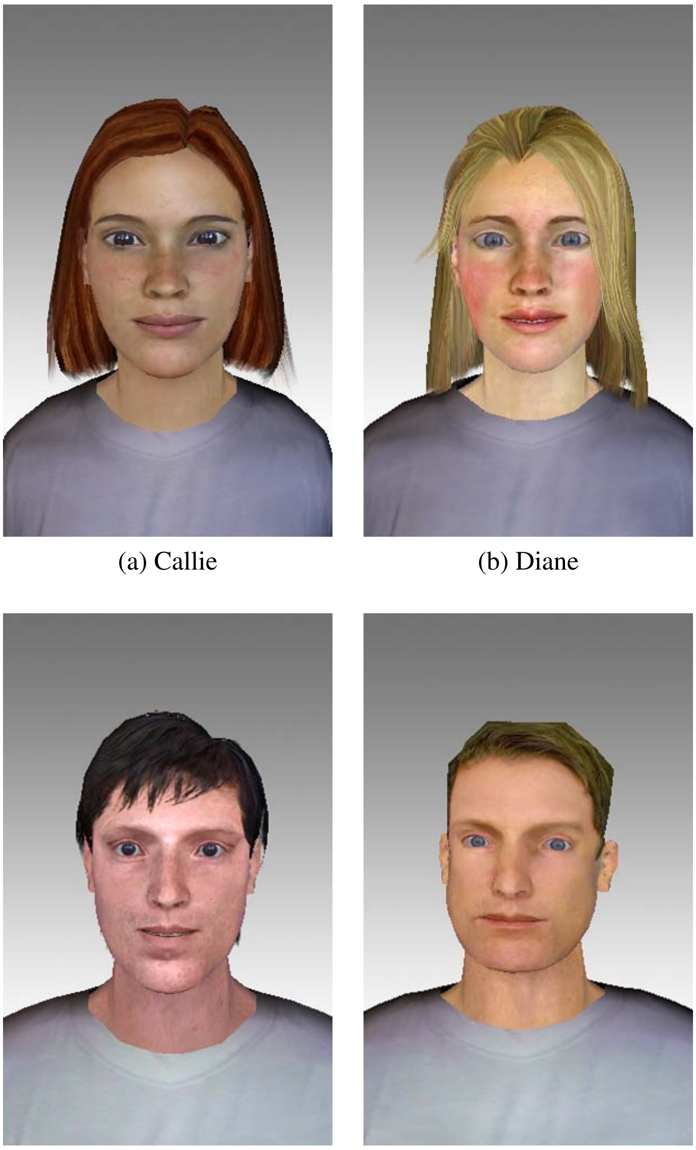
This chapter elaborates on the decisions and details relevant to the affect-inferring agent development and reports the outcome of this part of the project. The development stage included three steps: generation of Haptek characters, integration of these characters into EER-Tutor, and creation of the agent's behaviour logic. Section 3.1 describes the first step—development of the agent. Section 3.2 explains the second step—how the agent is integrated and controlled in EER-Tutor. The third step, the design of the agent's behaviour, is covered in Section 3.3. Sections 3.4 and 3.5 describe the pilot study of the agent-enhanced version of EER-Tutor and present the pilot study results.

3.1 Agent Implementation

Haptek characters produced in People Putty look like 3D human figures with a high level of visual realism. In order to accommodate the variations in the preferences of the target audience[5], we created four characters of both genders shown in Figure 3.1; two female characters: Callie and Diane, and two male characters: David and Mark. In agreement with our decision to have the agents assume the role of tutor, the Haptek characters were designed to appear as young people approximately 20 to 30 years of age; this is consistent with the stereotype of a tutoring group generally represented by postgraduate students.

People Putty SDK allows for fine-grain control over the agent's features and behaviour which is consistent with our adoption of the dimensional view of emotions. People Putty exposes a two-level API for controlling its characters' emotional appearance through a set of parameters associated with the characters' facial features. At a higher level, the SDK provides an interface for creating moods by controlling the agents' emotional appearance through altering the values for particular emotional dimensions (such as anger or happiness), including emotion intensity and decay; Figure 3.2a shows the mood-building interface. This level of control can be used for producing "canned"

[5]The work of Baylor and Kim [2004] provides experimental evidence indicating that users' performance may vary depending on the user's stereotypes and attitudes formed by the users towards the agents.



(a) Callie (b) Diane

(c) David (d) Mark

Figure 3.1: Haptek characters developed in People Putty.

behaviours where a pre-set emotional appearance is appropriate; for example, this approach is suitable for a greeting sequence or a sign out sequence.

On the lower level, the emotional appearance can be controlled through a set of fine-grain parameters defining the position and shape of the eyebrows, the corners of the mouth, the overall position of the head, and so on; in People Putty's context these parameters are called switches. We have chosen a subset of switches to control the characters' appearance changes along the affective valence dimension ranging from sad to happy.

The negative valence switches are `expMouthSad` (corresponding to the FACS action unit №15 Lip Corner Depressor, activated by *Depressor anguli oris* muscle), `expBrowsSad` (action unit №1—Inner Brow Raiser, activated by *Frontalis, pars medialis*) and `energyLow` (the switch which makes the character act in a more subdued manner); when set to values close to 1, these switches make the character look as shown in Figure 3.2b.

The positive valence switches are `expMouthHappy` (action unit №12—Lip Corner Puller, activated by *Zygomaticus major*), `expEyesTrust` (action unit №6—Cheek Raiser, activated by *Orbicularis oculi, pars orbitalis*)[6] and `energyHigh` (the switch which makes the character more active, as if adding more enthusiasm to the character's behaviour); when set to values close to 1, these switches make the character look as shown in Figure 3.2c.

Haptek's implementation makes use of Microsoft Speech API (SAPI) for synthesised speech output; Haptek characters use SAPI Level 4 or Level 5 compatible Text-to-Speech (TTS) engines to produce verbal narrations along with realistic lip-sync movements. For the experimental studies we acquired two reportedly high-quality TTS Cepstral[7] voices—one male and one female.

Haptek Hypertext scripting interface supplies a simple control and query system for character manipulation and internal information access [Haptek Inc., 2002]; in Web-based Haptek deployment scenarios this interface is accessible though a JavaScript wrapper which must be included in the HTML pages hosting Haptek characters. The JavaScript wrapper acts as a thin layer between the Haptek Player and the rest of the

[6] As stated in the work of Miles and Johnston [2007], activation of *orbicularis oculi* is the key distinguishing element between *enjoyment* and *not-enjoyment* smiles; from the perceiver's point of view, smiles that feature contraction of *orbicularis oculi* have been associated with higher ratings of positive mood by others [Scherer and Ceschi, 2000].

[7] See `http://www.cepstral.com/`—Cepstral Text-to-Speech engine.

(a) Higher level: Mood creation interface.

(b) Lower level: Effect of negative switches.

(c) Lower level: Effect of positive switches.

Figure 3.2: Two levels of controlling Haptek characters.

system. The values of the Haptek character's switches can be changed dynamically through the JavaScript Haptek wrapper; the TTS functionality and other Haptek environment controls are available through the same JavaScript wrapper.

3.2 Agent's Integration into EER-Tutor

Like the Web-based architecture of EER-Tutor described in Section 2.1.2, the agent's implementation is split between the client and server. The client-side of the agent's implementation is responsible only for displaying the agent and retrieving behaviour updates from the server; all control logic for the agent is implemented on the server side of EER-Tutor. In Figure 3.3 the agent-related components are shown with a double line. The client side of the application has a Haptek player window embedded in the main interface. On the client side, the agent appears to the users as a "talking head" with an upper body as shown in Figures 3.4a and 3.4b, where David is present on the right-hand of EER-Tutor's workspace. Above the agent window we placed the agent-control toolbar with five buttons; *Home* restores the state of the agent's window, making the agent face forward[8]; *Refresh* is a work-around for the Haptek player rendering quirks—in rare cases the Haptek characters might be displayed incorrectly; *Avatar*[9] takes users to the page where they can choose another agent; *Repeat* allows the user to have the agent read out the last set of feedback messages, *Quiet* makes the agent stop verbalising feedback messages until the user chooses to reverse this state.

Corresponding to each user logged into the system at any given time, the server-side application maintains an object encapsulating the agent state variables (Haptek character's name, the verbal feedback buffer, and so on) along with the values for the

[8]Haptek player allows the figures to be turned around and moved closer to or further away from their original position.

[9]For the sake of simplifying the big picture for EER-Tutor's users, we chose the term *avatar* in favour of the term *pedagogical agent*, even though *avatar* does not entirely fit the meaning of *pedagogical agent*. The term *avatar* in the Internet context is used to refer to 2D or 3D representations of users in the Internet forums, instant-messaging programs, blogs and network games [Lessig, 1999]. In our context, however, we refer to the pedagogical agents as *avatars* implying that they represent the EER-Tutor in person.

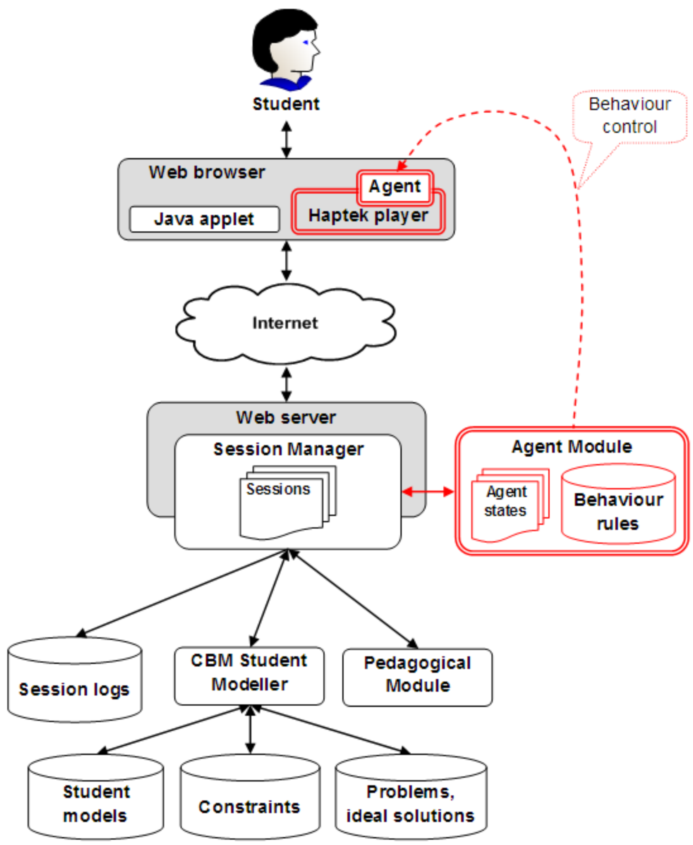


Figure 3.3: Agent's integration into EER-Tutor.

Haptek switches described in Section 3.1, which define the agent's emotional state. The agent's state changes are driven by user actions; that is, user actions (represented as interface events), when processed by the server, result in updates of the session history objects associated with every user and the corresponding agent state objects, which define the agent's behaviour.

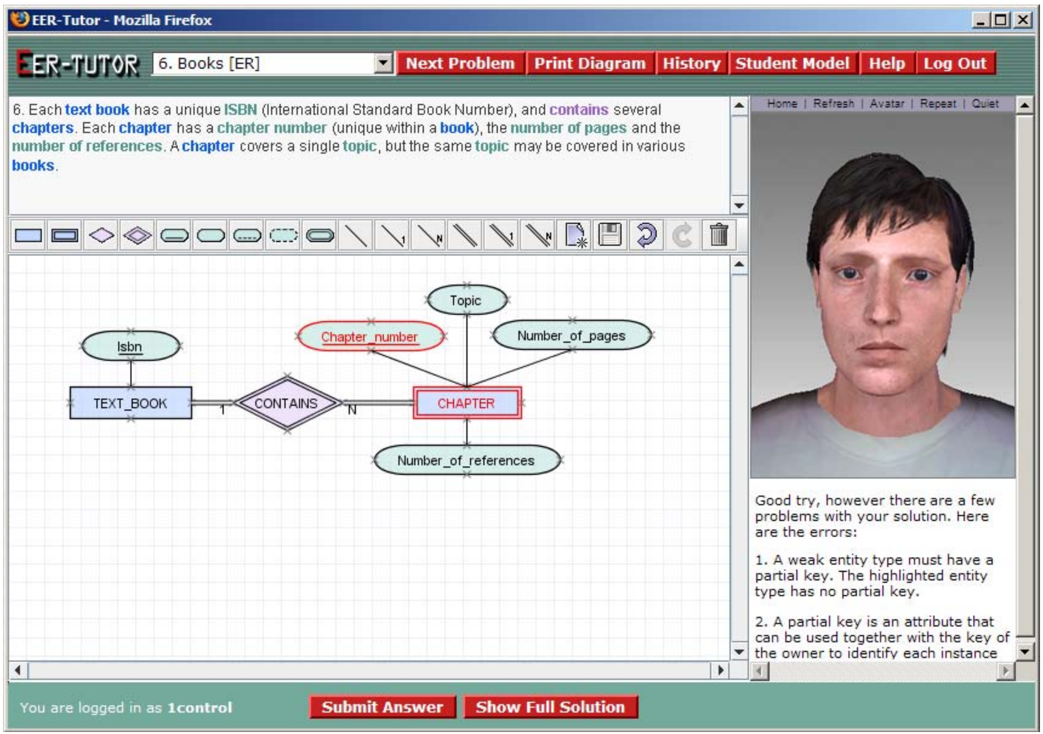
Every interface event (for example, a new problem selection request or solution submission) results in a corresponding request type being sent to the server. After the server processes the request and responds to the client, the agent control script queries the server for an update of the agent's verbal feedback and affective appearance appropriate in the context of the most recent event.

The logic defining the feedback and the changes of the agent's affective state is implemented as a set of rules, triggered by certain session history states. These rules define the agent's behaviour; the rules are described in detail in Section 3.3. The mechanism of the agent state updates revolves around matching the current session state against this set of rules.

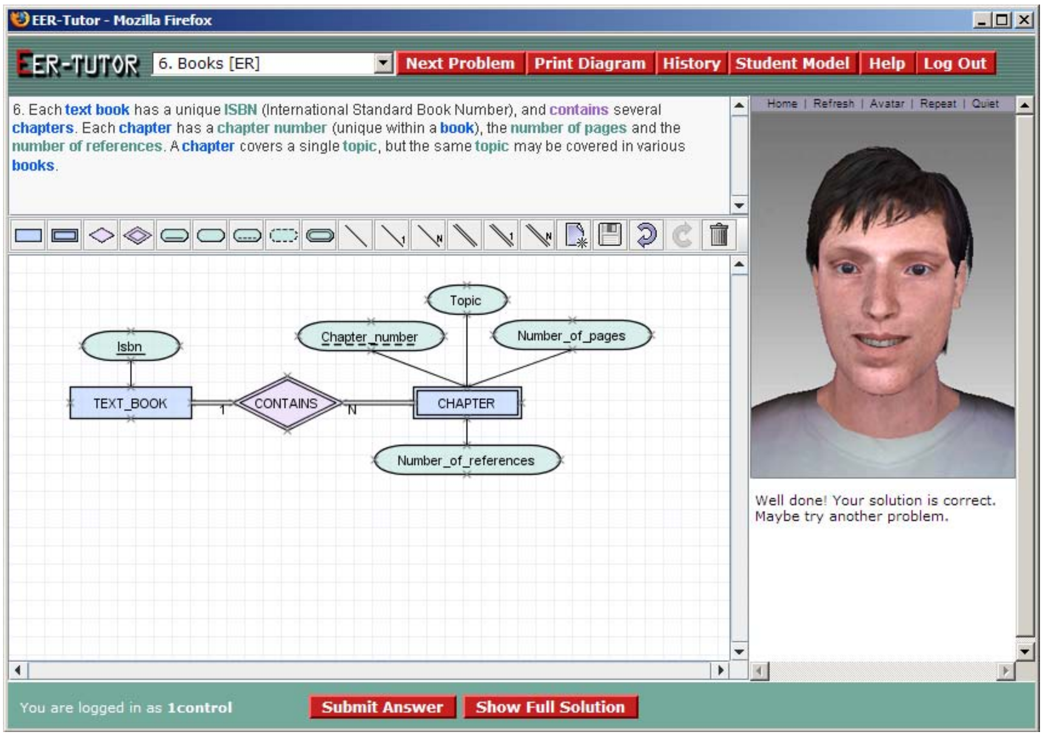
In this way, even though the Haptek character on the client-side is completely unaware of the user's actions and student model, to the users it appears that the agent is actively observing their actions. For example, when a user clicks the *Submit Answer* button, the Web browser initiates a request encapsulating the details of the interface event (in this case, the student's solution submission). The server updates the session history with the results of the request and upon the behaviour update request the server matches the agent's behaviour rules against the current session history state. Based on the specific rule that fires during processing, the action associated with the rule will trigger the agent's emotional state change and feedback buffer update. Thus the agent's appearance and feedback gets updated to match the current state of the agent's parameters.

The agent control mechanism relies on Asynchronous JavaScript, Sockets and eXtensible Markup Language (AJAX) [Crane et al., 2005]. We extended the Haptek's JavaScript wrapper by adding agent-control code for communication with the server. Most interface events result in immediate behaviour updates retrieved from the server through AJAX requests; this communication is asynchronous in the sense that the server does not need to keep track of all its clients; instead the clients pull behaviour updates from the server through the AJAX interface. The server sends back XML-based responses containing the values of the agent's switches and feedback messages appropriate in the current context of the session. In Figure 3.3 this flow of control is shown with a dotted line connecting the agent module on the server-side to the Haptek character on the client side.

The screenshot in Figure 3.4a shows the state of the interface immediately after an incomplete solution submission, which comes in a series of attempts with errors: the agent displays some-what sad features. The user mistakenly defined the *Chap-*



(a) After incorrect submission.



(b) After correct submission.

Figure 3.4: EER-Tutor screenshots with the agent.

ter_number attribute as a key attribute, instead of making it a partial key attribute. The erroneous diagram component, the *Chapter_number* key attribute, is shown in red. The agent looks slightly unhappy, according to the rule matching the current session state. When appropriate, the error messages presented by the agent refer the user to the errors in the diagram; in our example the first error message says: *A weak entity type must have a partial key. The highlighted entity type has no partial key.* This approach to engaging the user in active information processing is supported by Mayer's first principle of multi-media design, *Multimedia effect for transfer*, which states that providing the learner with a narration results in better learning when it is supplemented by corresponding visual aids [Mayer, 2001].

Along with feedback in audible form, the user is also presented with the most recent feedback messages in textual form. Even though the sixth principle of multi-media design, *Redundancy effect for transfer*, states that such redundancy negatively affects learning [Mayer, 2001], we consider that the EER-Tutor context justifies this redundancy because the complexity of the domain knowledge and abundance of information may be difficult to process without textual representation. For example, when the user makes several mistakes, the list of corresponding messages may help the user to stay focused on the task instead of wasting effort on remembering multiple feedback messages. The screenshot in Figure 3.4b shows the state of the interface immediately after a submission of a correct solution: the agent looks happy.

In this way the agent interacts with the user through visual displays and textual/verbal messages. Visual behaviour carries the load of affective communication, while textual/verbal feedback carries the cognitive load. The textual/verbal feedback for this version of the agent is based around the feedback messages defined in the original sans-agent version of EER-Tutor. However, some of the behaviour-space rules add "a personal touch" to some of EER-Tutor's original feedback messages as described in the following section.

The agent's affective state is programmed to respond over time to trends of affective changes, instead of rapidly fluctuating from positive to negative valence even when frequent alternating positive and negative valence updates arrive as a result of rule activation. This is done to make the agent's behaviour appear more human-like. The human emotional state generally can appear to have certain amount of inertia. Similarly, the agent's affective valence does not vary rapidly. For example, if a submission with a few errors happens to come in a series of unsuccessful attempts, the agent will have an unhappy/concerned look as the result of its affective valence being dampened by the repeated activation of this rule. However, if this is a single unsuccessful attempt, the change in the agent's appearance will be rather subtle. Some events, however, produce more noticeable changes then the others. For example, correct solution submission on the first attempt is bound to make the agent happy.

Along with event-triggered updates, the agent-control code continuously queries the server (repeating at the 1.5s intervals) for affective appearance updates even in the absence of interface events. While user actions may colour the agent's affective appearance in a certain way, the agent is capable of emotional self-regulation, as if mimicking the human behaviour as described in Chapter 2; in terms of dimensional emotion theory this tendency can be expressed as a tendency of affective valence value to gravitate over time to the neutral state[10]. The agent's affective state tends to gradually return to neutral, so if the agent momentarily may appear very upset or happy, after a minute or so (depending on the strength of the latest affective response) the agent inevitably "settles down", provided there are no new affect-bearing interface events.

3.3 Affect Inference

The cognitive view of emotions states that the valence of one's emotional response depends on the desirability of the situation [Ortony et al., 1988]. The validity of this approach, based on the appraisal of the learner's actions, has been demonstrated in the

[10] A similar approach for modelling emotions is described in the work of Corradini et al. [2005]: the agent's emotional state in its emotional space is controlled by emotion increments/decrements associated with behaviour templates. In the absence of emotion-eliciting user input, the agent's emotional state nudges back to its default state.

work of Conati and Maclaren [2004] and Conati [2002] as described in Section 2.3. Similarly, in our work we assume that continuous lack of cognitive progress will be accompanied by a negative affective state, because the user will be dissatisfied with the state of the current task; conversely, a satisfactory progress will result in a positive affective state.

Since EER-Tutor maintains student models, the short-term student models can be used to index the users' affective states. In our agent's design, however, the affective logic does not directly rely on the student models. Instead, the logic relies on session history, which includes a wider variety of information other then just solution submissions. The rationale behind this approach is twofold. First, most actions, such as sign on, problem selection and so on, may require a response or acknowledgement on the part of the agent; in many cases such responses would not have to carry much affective load, although failing to respond in a similar HHI situation could be interpreted as lack of attention and tact. For example, skipping to the instructional part without acknowledging the student's presence at the beginning of a tutorial session can be interpreted as rudeness. Second, seemingly neutral actions under some circumstances might be associated with certain changes in the users' affective state and thus should be addressed by the agent. For example, repeated submissions of the same solution to a problem might indicate a negative affective state, such as frustration. In general, we believe that any sequence of rapidly repeated actions should not go unnoticed, because it might indicate negative affective states such as irritation or frustration.

The logic of the agent's affective and verbal behaviour is explicitly programmed into a set of fifteen rules which pair pedagogically and emotionally significant session states with the suitable responses on the part of the agent. Examples of a few rules are provided below. First we describe the rules' structure and categories. A rule consists of three parts, defined as follows:

1. *State pattern:* a pattern describing a certain event or session state which warrants a response from the agent.

2. *Feedback messages:* a set of equivalent verbal/textual responses. When the rule is activated, one of the responses is chosen randomly to avoid repetitiveness. Most rules defined for EER-Tutor agent have between three and seven alternative messages.

3. *Affect adjust value:* a command to modify the agent's affective appearance; these changes are expressed as numeric values. Positive values nudge the agent's affective valence towards the positive end of affective valence dimension; negative values move the affective values towards the negative end.

Listing 3.1 provides the source code for the rule №5, defining the agent's response to a correct solution achieved on the first attempt. EER-Tutor builds session histories by pushing event descriptions in the form of property lists[11] onto the front of each individual session history record. Listing 3.2 shows a sample session history record; the most recent event shown in lines 1–6 would match the rule pattern specified in the lines 6–9 of Listing 3.1.

Based on the categories of session states that the rules are associated with, the rule set is divided into three categories:

1. *Activity transition category* (AT): this category unites the rules applicable to session states indicating the user switching between various types of actions. For example, agent selection, problem selection, request to display the open student model or help information—all these situations are covered by rules in this category. First time sign on, subsequent sign on and sign out rules also belong to this category. Most rules in this category define neutral affective reactions; the sign on rules, however, make the agent look quite happy. This category consists of seven rules.

2. *Problem submission category* (PS): these rules carry the most work in the agent's affective behaviour, because the majority of the user activity in EER-Tutor is cen-

[11] In Common Lisp, a property list or a plist is a list where every other element, starting with the first, is a symbol that describes what the next element in the list is [Seibel, 2005].

```
;;; Correct submission on the first attempt.
;;; NO affective interjection REQUIRED here.
(def-agent-rule 5
    "Correct submission on first attempt agent rule."
   ;; 1 - Session history pattern.
  (let ((last-entry (get-entry :submission 0 *history*)))
    (and last-entry
         (zerop (errors last-entry))
         (= 1 (attempt-number last-entry))))
  ;; 2 - List of equivalent feedback messages.
  (("Great job! See if you can solve another problem just as fast.")
   ("Brilliant! You've solved this problem in one attempt.")
   ("Outstanding! It took you only one attempt to solve this problem.")
   ("Congratulations! You've solved this problem in one attempt.")
   ("Great effort! Do you want to try another problem now?"))
  ;; 3 - Affect-adjust value.
  ;; Move affective valence to the positive end.
  0.9)
```

Listing 3.1: Source code defining the rule №5.

```
((:submission (:event-time 3392076824)
              (:problem-number 6)
              (:attempt-number 1)
              (:errors 0)
              (:violated-constraints nil)
              (:empty-attempt nil))
 (:new-problem (:event-time 3392076820)
               (:problem-number 6))
 (:new-problem (:event-time 3392076809)
               (:problem-number 5))
 (:agent-selection (:event-time 3392076801)
                   (:agent-symbol :female-character-1))
 (:agent-selection (:event-time 3392076792)
                   (:agent-symbol :male-character-1))
 (:login (:event-time 3392076788)))
```

Listing 3.2: An example of a session history record.

tred around submitting solutions to the curriculum problems. All rules in this category carry some affective reaction—some of them make agent look happy, others make it look sad. This category consists of eight rules.

3. *Repeating sequence category* (RS): these rules catch repeated sets of identical actions. These rules move the agent's valence value towards the negative end of the affective axis. This category has only two rules.

Following are a few rule descriptions to demonstrate the effect intended by these rules; Table 3.1 provides the complete list of rules, classified by category and accompanied by the descriptions of the identifying session states, examples of feedback messages and affective adjust values.

№1: First-time sign on: agent introduces itself and welcomes the user with an extended message describing EER-Tutor's workspace; the agent's affective state is set such that the agent smiles (AT category instance).

№2: Subsequent sign on: agent welcomes the user and while smiling suggests the user to choose a problem to work on (AT category instance).

№4: Repeated submissions of identical erroneous attempts: the agent tells the user that the solution looks the same since the last attempt, lists the errors and suggests that the user tries to resolve some of these errors (RS category instance).

№6: Correct solution on the N-th attempt: the agent congratulates the user for solving the problem and while smiling suggests that the user tries another problem. Figure 3.4b shows the state of the interface and the agent as determined by this rule—the agent looks happy (PS category instance).

№7: Submission with a single error: the agent starts with an encouraging phrase (as if it is enthusiastic because the user is getting closer to correct solution) and presents the feedback message; affective valence is nudged slightly to the positive end (PS category instance).

№8: Submission with a few errors: the agent starts with a friendly introductory phrase and reads the errors; the agent's affective state is nudged slightly towards

№	Cat.	State Description	Feedback Message Example	Aff-adj.[a]
1	AT	First-time sign on event	“Welcome to EER-Tutor. I’m your companion avatar for this session to help you learn about EER modelling.”	0.8
2	AT	Subsequent sign on event	“Hello and welcome back to EER-Tutor. Please choose a problem to work on.” …	0.8
3	PS	Empty solution submission	“Your diagram is empty. I can’t give feedback on empty diagrams. Why don’t you try adding some components to your solution?” …	-0.4 (-0.2)
4	RS	Three repeated submissions with no changes to solution	“You are not making changes to your solution. There are still some errors:” + [the list of error messages] …	-0.5 (-0.3)
5	PS	Correct submission on the first attempt	“Great job—you solved this problem in one attempt! See if you can solve another problem just as fast.” …	0.9
6	PS	Correct submission on the N-th attempt	“Well done! Your solution is correct. Maybe try another problem.” …	0.7
7	PS	Submission with one error	“Almost there, there’s only one mistake:” + [the error message] …	0.6 (0.3)
8	PS	Submission with more the one error	“There are a few mistakes in your solution. Here are the errors:” + [the list of error messages] …	-0.4 (-0.2)
9	RS	Five repeated requests for new problems	“You keep flicking through problems without attempting them. Maybe try solving a problem.” …	-0.4 (-0.2)
10	AT	Next problem request	“Okay, let’s try this problem.” …	0.5
11	AT	Agent selection request	“Hello there. Ready to practice?” …	0.5
12	PS	Complete solution request	“Maybe next time try to identify mistakes yourself.” …	-0.5
13	AT	Display history request	No verbal feedback[b]	0.5
14	AT	Display open student model request	No verbal feedback	0.5
15	AT	Sign out request	“Bye, then. I hope you’ve enjoyed learning with EER-Tutor.”	0.0

Table 3.1: Rules determining the agent’s behaviour.

[a] The values shown in parentheses in this column are for the evaluation study discussed in Chapter 5.

[b] The rules №13 and 14 shift the agent’s affective state value towards the positive side of the spectrum.

the negative side. Figure 3.4a shows the state of the interface and the agent as determined by this rule (PS category instance).

№11: Selection of a new agent: the agent introduces itself and smiles (AT category instance).

There are also rules defining affective behaviour for repeated submissions of identical solutions and repeated abandoning of problems which we dubbed *problem surfing*. The precedence of session states is implicitly encoded in the order of the rules; during the rule matching process, the first match results in the corresponding action on behalf of the agent. All four Haptek characters are guided by the same set of rules; in other words, there are no differences between the agents' personalities.

3.4 Pilot Study

In July 2006 we conducted a pilot study of the affect-inferring agent. This study was designed to validate the appropriateness of the agent's personality in the process of the users' interaction with the agent-enhanced version of EER-Tutor. We recruited 20 volunteers (16 male and 4 female) among the postgraduate students at the Department of Computer Science and Software Engineering at the University of Canterbury. All participants were familiar with the EER modelling concepts. The disparity between the users' levels of EER expertise was not an important factor, because the study aimed to collect qualitative data only; we did not evaluate multiple experimental conditions.

The experiment was carried out as a series of individual sessions. At the start of a session, a participant was provided with a verbal description of the task. Participants were required to spend 45 to 60 minutes exploring the system's functionality and solving problems while paying attention to the agent's emotions and feedback messages. The participants were asked to choose an agent before they were able to navigate to the workspace. During the session, the participants were free to choose a different agent at any time. At the end, the participants were required to fill out questionnaires designed to assess their experience of learning with the agent and their perception of the agent's persona; the questionnaire is included in Appendix B. The outcome of the study was evaluated on the basis of the questionnaires. The results of the pilot study are given in the following section.

3.5 Pilot Study Results

The participants showed a preference for female agents as the first choice (14 versus 6) irrespective of their sex. Among male participants only 38% chose a male agent, while all female participants chose a female agent at the start of the session. Discussion on social attraction towards media cites evidence in favour of both similarity-attraction and complimentary-attraction [Linek and Scheiter, 2006], although the lack of scale in the pilot study does not allow us to support either side.

On the basis of questionnaire responses, the male agents won in their effectiveness and overall impression on the participants. The participants who chose a male agent (a total of 6) reported they enjoyed using EER-Tutor better, giving it a rating of 4.1 ($\sigma = 0.4$) out of 5. The participants who chose the female agent (a total of 14) rated their experience at 3.7 ($\sigma = 0.7$) out of 5. The average learning success (rated on the scale of 1 to 5) reported by participants with male agents is higher then the female agents' group: 3.5 ($\sigma = 0.5$) versus 3.1 ($\sigma = 0.7$). This learning estimate difference is consistent with the actual learning outcomes, measured by the number of constraints learned during the session: 3.3 ($\sigma = 4.8$) for male agents versus 2.1 ($\sigma = 2.3$) for female agents. Table 3.2 summarises the results, although the small number of participants does not allow us to treat these results as being statistically reliable.

We attribute the apparent association between higher ratings and better learning outcomes for the male agents to the difference in the quality of male and female Cepstral TTS voices used in the pilot study. The male TTS voice sounds more human-like then the female voice; even though the participants were not specifically asked to rate the quality of TTS voices, 50% of the participants who used female agents stated that they were frustrated by or disliked that voice. No such comments were made about the male voices. Our interpretation of the voice quality effect is supported by the theory of cognitive load [Sweller et al., 1998], which states that processing unnatural-sounding or machine-like voices imposes higher cognitive demands on learners than natural voices do; when not consumed by voice processing, cognitive resources are free to be used for deeper processing of instructional material.

The general response to the agents was positive—75% rated the agents as a useful feature. Of the 25% of participants who thought the agent's presence was unnecessary, half rated audible narration as useful. Overall, the participants were enthusiastic about narration—50% stated that narration was the most helpful feature, because it made it possible for them to keep their eyes on the diagram and begin correcting errors while listening to the narration; participants commented that this helped save time and enabled them solve problems faster. The prevailing positive uptake of the verbal feedback feature is very encouraging, because these comments were not explicitly solicited from users by the questionnaires.

Both male and female agents' responses and appearance were rated as adequate by around 90% of both male and female participants. Free-form feedback indicates that the agents' persona was appreciated for not being "in-your-face". Some users commented that they liked the "feeling of company" and "human presence" created by the agents. Some users made comments suggesting that the agents should be more active and versatile in their behaviour; others commented on being distracted by the agents' movements, such as blinking and breathing, in the background. For example, some users wanted the agents to be more dynamic, even suggesting that we "have them fly over the workspace" or "make emotional changes more apparent", while others wanted the agents to be less obtrusive.

The pilot study highlighted the need for the agents' behaviour to appear more natural. Many comments suggested that we enable the user to have greater control over the agents' behaviour and voice properties. Some participants said they would have liked to be able to control the speed of the narration to make it faster or slower. Some users (around 25%) did not pay attention to the agents' affective appearance, but a small proportion (around 15%) commented on the agents "flipping out" and getting "too emotional" too soon. One participant commented that he or she felt "emotionally blackmailed and manipulated by the agent".

Group	Estimated learning	Actual learning
Male agent	3.5 ($\sigma = 0.5$)	3.3 ($\sigma = 4.8$)
Female agent	3.1 ($\sigma = 0.7$)	2.1 ($\sigma = 2.3$)

Table 3.2: Estimated learning (measured on a Likert scale of 1 to 5) and actual learning (measured by the number of learned constraints) for male and female agents.

CHAPTER 4

Agent's Affective Awareness

Needless to say, inference of the affective state on the basis of the user's cognitive state and in particular, session state, described in the previous chapter is not ideal. On one hand, in the EER-Tutor's context, events resulting in the modification of the session history sometimes may occur rare and far apart. Finer-grain awareness of the learner's affective state in the absence of the affect-indexing interface events is required for precise intervention timing; if the learner progresses to disappointment or discouragement and stays there too long, then intervention may be productive. On the other hand, the assumption that continuous lack of cognitive progress is accompanied by a negative affective state will inevitably fail in some situations. As discussed in Section 2.7, depending on the individual's personal characteristics, failure sometimes may lead to curiosity, exploration and investigative actions.

The remainder of this chapter describes the affect-aware agent version for EER-Tutor. Section 4.1 substantiates the implementation choices made in our approach to facial feature tracking. Affective response implementation consists of three layers: frame-by-frame facial feature detection, affective state calculation in the context of the current session based on the analysis of feature positions, and finally, affective response generation. Section 4.2 presents the first level, the facial feature tracking algorithm.

The second level, the logic of affective state inference on the basis of detected features is described in Section 4.3, followed by Section 4.4 which describes the third level—integration of affective states into the agent's behaviour and response.

4.1 Feature Tracking Approach

Among the facial feature tracking techniques listed in Section 2.5, OpenCV directly supports PCA; however, PCA has a number of weaknesses, some of which would be difficult to overcome in an experiment requiring real-time facial feature tracking:

- PCA is translation-dependent—recognition accuracy is negatively affected by variations of the target object location in the image.
- The algorithm is scale-dependent—target features are hard to detect at various scales.
- PCA is background-dependent—changes of the background degrade recognition accuracy.
- PCA is also lighting-dependent—light intensity changes adversely affect recognition accuracy.

Background and lighting conditions were relatively easy to control in the experimental setup during our evaluation. Arrangements to control face position and proximity to the camera, however, are not acceptable in our context, because they would constrict people's natural tendency for free head movement; an attempt to restrict participants' head movements would make them feel uncomfortable and unnatural. Therefore instead of PCA, we chose the hybrid approach based on geometric constraints and lower-level image processing functionality available though the OpenCV library. We developed an algorithm for feature tracking which utilises a combination of common image processing techniques, such as thresholding, contour-tracing and detection of points with high eigenvalues [Shi and Tomasi, 1994] provided in the OpenCV library. The algorithm also relies on our implementation of integral projections [Mateos, 2003] developed for this project and described in the following section.

4.2 Feature Tracking Algorithm

Our feature detection algorithm includes five steps. In this section, we describe the steps and the associated image processing techniques. Figure 4.1 shows facial features labelled according to the corresponding algorithm steps; the cut-in image in the lower-right corner displays intermediate processing results. For better feature tracking accuracy we modified the OpenCV library source code to retrieve frames from the camera in higher resolution, 640×480 px, not in 320×240 px, which is used as the default resolution. Throughout the algorithm, the focus of attention is shifted among a number of regions of interest, determined by the algorithm on the basis of the anthropomorphic constraints describing human face geometry [Tian et al., 2005; Farkas, 1994]. The steps involved are as follows:

1. Face region extraction—the algorithm retrieves a frame from the video stream, converts it to grayscale colour space and convolves it with a 3×3 Gaussian kernel to eliminate noise in the image [Gonzalez and Woods, 1993]; these operations are provided by the OpenCV library. Then the algorithm finds all rectangular regions containing faces; this is achieved by running Haar object classifier [Viola and Jones, 2001; Wilson and Fernandez, 2006], `cvHaarDetectObjects`, with a cascade trained to detect human faces (also provided with the OpenCV library). On completion of this stage, the largest rectangle indicates the location of a face closest to the camera. In our experiment only one participant must be present in front of the computer workstation during the session. If there were no faces detected in the frame, which may happen when the user's is looking away from

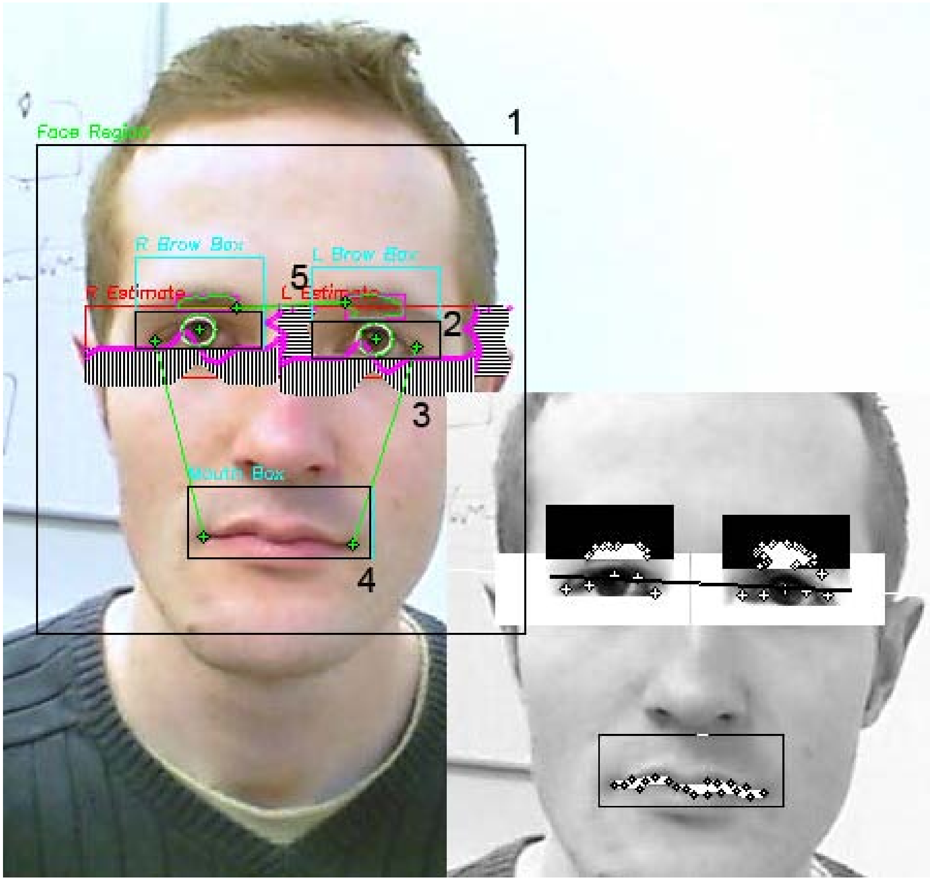
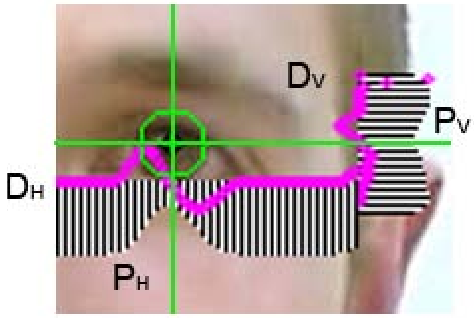


Figure 4.1: Detected features labelled according to algorithm steps.

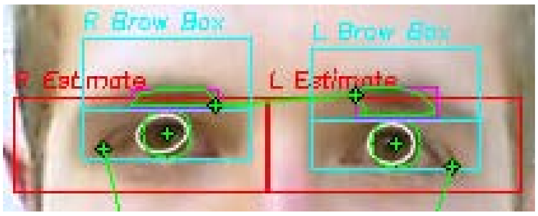
the screen or looks down at the keyboard, then the algorithm steps onto the next frame; otherwise the algorithm advances to the following step.

2. Iris detection—feature extraction starts with iris detection. Based on anthropometric face proportions our implementation code defines two regions containing the eyes. Then, for each eye region the algorithm calculates the vertical and horizontal integral projections[1] and convolves each of them with a one-dimensional Gaussian kernel of size 3 to smooth the resulting projection. We developed the code for integral projection processing for this step; a visualisation of the vertical and horizontal projections is shown in Figure 4.2a—the vertical and horizontal in-

[1] An integral projection is a one-dimensional pattern, obtained through the sum of a given set of pixels along a given direction.



(a) Integral projections and derivatives.



(b) Verification of detected iris parameters with ellipse mapping.

Figure 4.2: Pupil detection with integral projections.

tegral projections are labelled P_V and P_H respectively; the projections are shown as black-and-white histograms. Both projections in the example have the global minima corresponding to the region of lowest luminosity representing the pupil area. Then, to obtain the minima coordinates, we calculate the derivatives D_V and D_H from the projections; in Figure 4.2a the derivatives are shown as continuous graphs. The point where the vertical derivative intersects the base projection axis indicates the vertical position of the pupil; the point of intersection of the horizontal derivative with the base projection axis indicates the horizontal position of the pupil.

When both pupil positions are known, the algorithm verifies the results of pupil detection by mapping ellipses to the thresholded (binary) images of the eye regions; sample code for ellipse mapping is provided with the OpenCV library. An iris radius should be approximately equal to $\frac{1}{9}$th of the distance between the centres of the irises; if the centres and shorter axes dimensions of the resulting ellipses do not match the detected iris parameters ($\pm 20\%$ is the acceptable error margin in our algorithm), then it is assumed that either the frame shows a face during an eye blink or while the user is looking down at the keyboard. This results in the assumption that the irises' centres were not detected correctly, which causes the algorithm to abandon the current current frame and step onto the next frame. Otherwise, if irises detection was successful, the algorithm adjusts the size and location for the eye regions, which results in smaller bounding boxes around the eyes. Figure 4.2b shows both mapped ellipses (shown in white) and irises indicating successful iris detection in the current frame.

3. Outer eye corners detection—during this step, the algorithm detects the features with high eigenvalues in the regions limited by the eyes' bounding boxes. The OpenCV library encapsulates this functionality in the `cvGoodFeaturesToTrack` function. The features returned by this function are first filtered to remove the points found above the line connecting the centres of the irises and then the list of the remaining points is sorted by the points' distance to the centre of the respective irises, starting from the closest point to the farthest; in the intermediate results cut-in of the Figure 4.1 the points in the image identify the regions with strongest eigenvalues, which usually represent corners. The outer-most features in the respective regions indicate the positions of the outer right and left eye corners.

4. Mouth corners detection—this step is similar to the previous step. High eigenvalue features found in the subregion of the frame contained in the mouth's bounding box are sorted by their horizontal position. The two extreme points represent the left and right mouth corners.

5. Inner brow corners detection—our algorithm uses the OpenCV function `cvFindContours` to trace the brows' contours in the thresholded region above the eyes' bounding boxes; as a result the leftmost point in the right eyebrow con-

tour and the rightmost point in the left eyebrow contour are designated at the inner brow corners.

Throughout the algorithm execution a number of consistency checks verify the correctness of the geometric relationships between the extracted features and interrupt the algorithm flow by making it to the next frame if any inconsistency between the feature positions is suspected in the current frame.

Tracking Session Calibration

Feature detection relies on the threshold values for processing eye, brow and mouth regions. To accommodate lighting variations, these threshold values have to be chosen manually during algorithm calibration at the start of each tracking session; the interface of the feature tracking application includes the trackbar widgets for setting the threshold values as shown in Figure 4.3. The thresholds are adjusted one-by-one; optimal values are detected visually, resulting in the corresponding set of features displayed in the application window. For best results during calibration, the subject is recommended to look at the centre of the screen.

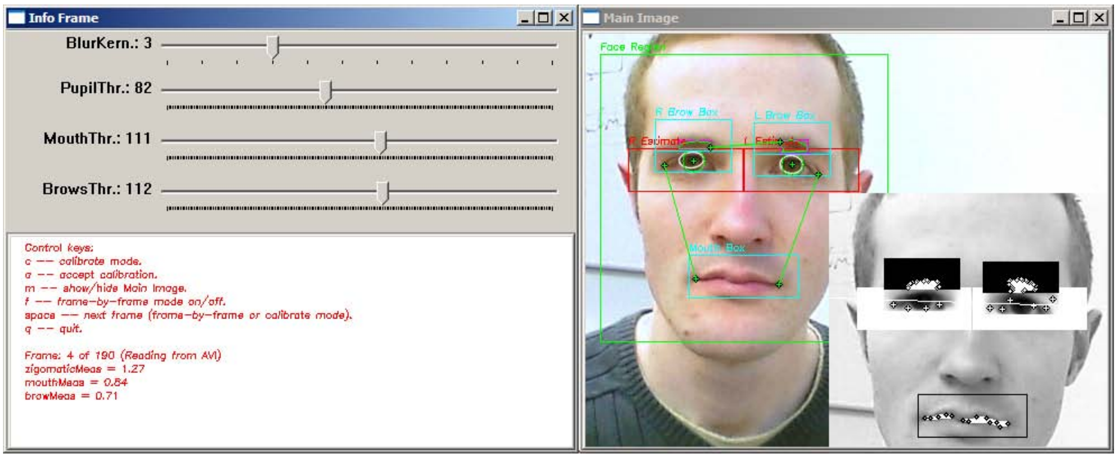


Figure 4.3: Feature tracking calibration interface.

4.3 Affective State Detection

Our implementation of affective state detection is based on the idea of facial animation parameter normalisation described in the work of Pandzic and Forchheimer [2002]; the feature displacement is measured in this technique on the basis of a set of facial parameters for a neutral expression. At the start of every tracking session, during the calibration stage described in the previous section, the algorithm stores three ratios, RE_0, RM_0, and RB_0, which are used later in the session to calculate the changes in facial expressions. The first ratio, RE_0, (4.1) is the ratio of the average distances between the outer eyes and mouth corners $E_R \rightarrow M_R$ and $E_L \rightarrow M_L$, to the distance between pupils, $P_R \rightarrow P_L$; these distances are shown in Figure 4.4.

$$RE_0 = \frac{E_R M_R + E_L M_L}{2 P_R P_L} \tag{4.1}$$

The second ratio, RM_0[2], (4.2) is calculated as the distance between outer mouth corners, $M_R \rightarrow M_L$, divided by the distance between pupils.

$$RM_0 = \frac{M_R M_L}{P_R P_L} \tag{4.2}$$

And the third ratio, RB_0, (4.3) is the distance between eyebrows, $B_R \rightarrow B_L$, again divided by the distance between pupils.

$$RB_0 = \frac{B_R B_L}{P_R P_L} \tag{4.3}$$

During the session, the subjects may move closer to or further away from the camera, but for frames with neutral expressions, when the user is looking at the screen[3] the values of the running ratios for frame k, RE_k, RM_k and RB_k, calculated in the same way as RE_0, RM_0 and RB_0, remain relatively close to the RE_0, RM_0 and RB_0 values

[2]This ratio, RM_0, is redundant, because theoretically the RE_0 ratio is enough for indexing positive affective valence, but we've chosen to use RM_0 to compliment the calculation accuracy.

[3]The frames where the user is looking away or down will be skipped by the tracking algorithm.

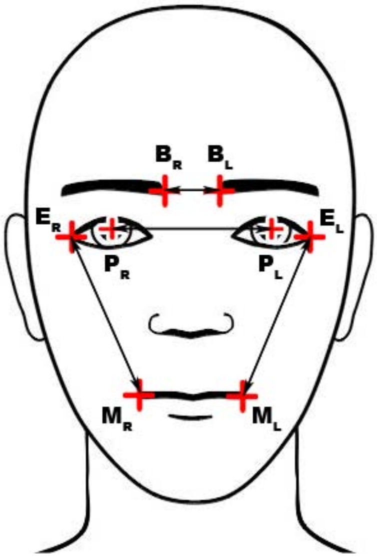
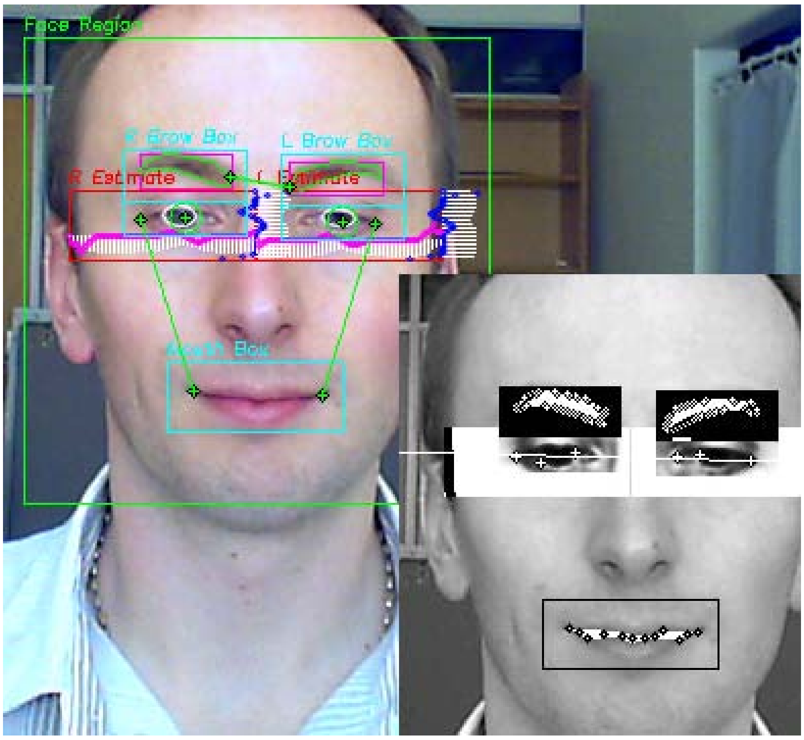


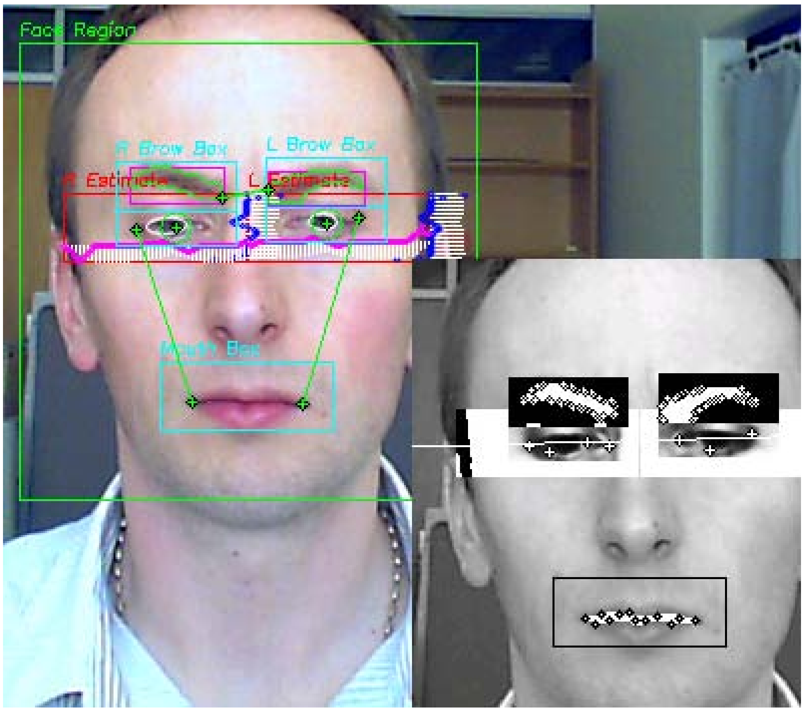
Figure 4.4: Facial points.

respectively. On the other hand, affective facial expressions, such as smiling or frowning, make the RE_k, RM_k and RB_k values differ from the RE_0, RM_0 and RB_0 values. Analysing the differences between the ratios saved during the calibration step and the ratios calculated for each frame is the key to detection of affective state variations. For example, the positive value of $RE_0 - RE_k$ (eye corner to mouth corner ratio difference) and the positive value of $RM_k - RM_0$ (mouth corner ratio difference) indicate positive affective valence accompanied by a smile. The negative value of the $RB_0 - RB_k$ (brow ratio distance) indicates negative affective valence accompanied by a furrowed brows look on the subject's face.

Figure 4.5 shows two frames which in the context of the session would "vote" for positive and negative affective states respectively. Table 4.1 shows the values of the ratio differences for both frames. The absolute values for the two ratios associated with positive affective valence are largest in the column for Figure 4.5a, and conversely, the absolute value of the ratio associated with frowning is the largest in the column of values for Figure 4.5b.



(a) Positive valence



(b) Negative valence

Figure 4.5: Positive and negative valence frames

Ratios difference	Positive valence (4.5a)	Negative valence (4.5b)
$RE_0 - RE_k$	0.09	-0.03
$RM_k - RM_0$	0.14	0.02
$RB_0 - RB_k$	-0.02	-0.07

Table 4.1: Values of tracking ratios for Figure 4.5.

The key feature of our implementation is that each frame is analysed in the context of the session. The algorithm does not attempt to determine affective state in every individual frame; rather our algorithm makes its decisions on the basis of the observed changes throughout the session. As described earlier in Chapter 3, EER-Tutor is a web-based application which has its functionality divided between client and server-side. Similarly, the logic of affect recognition is also divided between the client and server-side[4], as shown in Figure 4.6. The affect-related components integrated into EER-Tutor's architecture are shown with a double line.

Feature-tracking algorithm, described above, runs on the client-side; the application maintains three buffers holding the values for the three ratio differences for each frame where facial features were found. Every three seconds[5], the client-side code calculates the running average for each of the three ratios and posts these values via HTTP to the server which keeps track of the data for all concurrent EER-Tutor sessions. We wrote a simple wrapper for WinSock[6] and used it as a static library to be compiled with facial feature tracking code to enable network communication. In Figure 4.6 this flow of data from feature tracking application to the affect-calculating module is shown symbolically with a dotted line. The decision-making logic, which tells apart negative and positive affective states, runs on the server. The server maintains affective state data for each user in the affective state object with three data channels; each data channel corresponds to

[4]The algorithm does not have to be split between server and client side, but we chose to do so for the purposes of our experiment; both agent behaviour updates and feedback message construction logic runs on the server side. To factor affect-awareness into agent behaviour and feedback it was a natural decision to implement the final stage of affective state calculation on the server.

[5]This interval ensures detection of relatively short displays of emotions; the work of Levenson [2003] provides evidence that the average duration of an emotion display lasts up to five seconds.

[6]See `http://msdn2.microsoft.com/en-us/library/ms740673.aspx`—Windows Sockets 2

one of the three ratio differences described earlier. A data channel object maintains the maximum and minimum values recorded during the session, and is capable of returning the normalised current value. With every consecutive request received from the client feature tracking application, the data channels in the affective state object are updated to reflect the changes in the observed affective state, registering transitions between negative, neutral and positive affective states.

The agent module monitors the affective state object associated with the given session, with the purpose of generating affective feedback in response to interface events; also, the affective state changes may result in the agent's affective interjections. The following section provides the detailed description of the agent's affective logic.

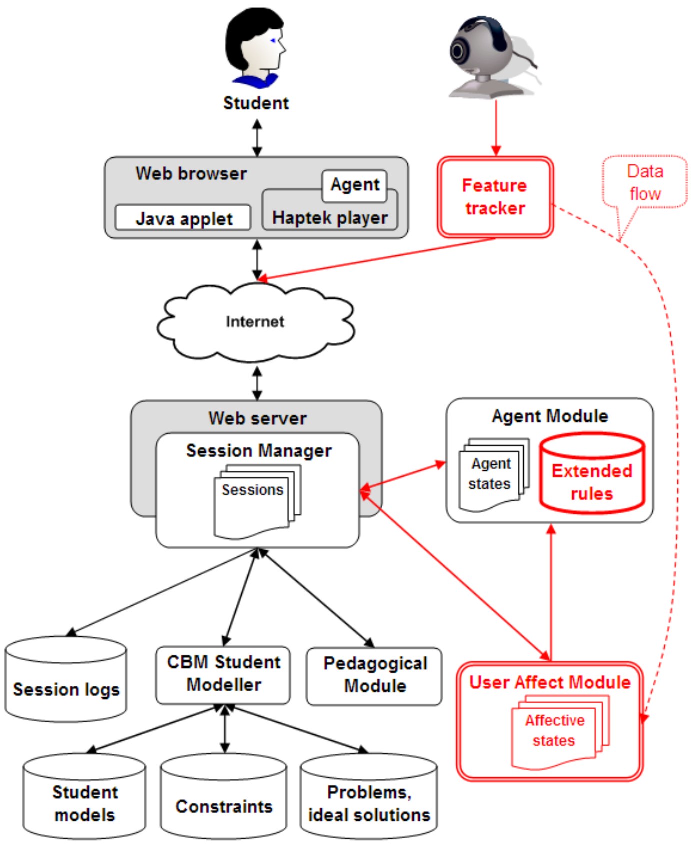


Figure 4.6: Integrated affective awareness components.

The averaging of the ratio differences values on the client side dampens the interference of inaccurate feature tracking calculations; in addition, the affective state valence value changes in such a way that it appears somewhat inert. In the event of frequent variations between the positive and negative ratios, the valence of the affective state object changes as if it is guided by the general trend of data received from the client. The server-side code also takes care of the situations when the subject might be looking away from the camera or down at the keyboard of the workstation while typing. To make the system more robust in absence of feature tracking data, the affective valence value gravitates towards the neutral position, as does the affective valence in human emotional processes.

4.4 Agents' Affect-aware Persona

The affect-aware version of the agent inherits most of its behaviour from the version of the agent based on session history appraisal described in Section 3.3, that is, the affective appearance of the agent is controlled by the same basic set of rules. However, some of the original rules (described in detail further in this section) have been extended to use the added affective awareness to make the agent address the user's feelings to provide active affective support. As discussed in Section 2.6, active support is characterised by direct discussion of emotions as a means of managing them. The affect-aware agent version differs from the previous version in two ways: first, this version of the agent differs by its capability to differentiate between positive and negative affective states and address only negative affective states through additional feedback messages. Second, apart from responding to user-generated interface events, the agent has the capability of intervening with affect-oriented messages when the system observes a degradation of the user's affective state.

The rationale for this approach is based simultaneously on the flow theory and on the model of cyclic flow of emotions in learning, both of which are discussed in Section 2.7. As described in the work of Csikzentmihalyi [1990], the state of positive flow may be disrupted by making the subject aware of the flow; thus the agent does not need to interfere if there is no negative affect. When the user is happy with the state of the session, it is unlikely the agent's affective feedback will improve anything, even if the agent is beaming with happiness and enthusiasm; if anything, such an interference may break the mood or unnecessarily distract the user. On the other hand, making the subject aware of their negative flow state may distract them from their negative feelings and move them along towards their goal [Klein et al., 2002].

Kort et al. [2001], however, emphasise that the cyclic model of emotions in learning views transient negative affect as an integral part of the learning process. This approach has been exemplified in the work of Burleson and Picard [2004] who describe the affect-aware learning assistant which focuses its attention on the affective elements of progress. If, for example, the majority of symptoms indicate that there is no frustration on the part of the user even when the user has not exhibited progress in terms of the task state, the system assumes that curiosity, exploration, or familiarity are at play and the Learning Companion allows the user to further explore the task without explicit intervention. If on the other hand, there are signs believed to be associated with stress, the agent might intervene.

Certainly, the affect-oriented messages triggered by negative affective states run the risk of making a bad situation worse, because a user afflicted by negative feelings might consider any interruptions irritating. With this in mind, we tried to design our agent's affective feedback and interjections to be as unobtrusive as possible; the agent only provides affect-oriented content if the subject's facial feature tracking data indicates the dominance of the negative affective state. In our implementation, the interval to be taken into consideration is a configurable parameter; for the evaluation study it was set to two minutes. Thus the agent, while responding to interface events, such as solution submission, may add an affect-oriented message to its feedback relative to the situation only if the negative affect has been observed during the last two minutes and if the user did not receive affective feedback during that time. The same logic is applied to the agent's affective interjections in the absence of interface events.

To make the agent responsive to observed affective valence changes during the session, we extended the functionality of the agent-control module. When the server receives a request for the agent's behaviour update, the agent-control module queries the user-affect module to retrieve the affective valence value; this value determines the agent's affective verbal response, if any. A subset of rules in the original set described in Section 3.3 has been modified with the additions of affect-oriented messages addressing negative affective states. We identified five rules, № 3, 4, 7, 8, 9, which require consideration of the affective state. In other words, all the rules defining the agent's behaviour still apply to the original set of session states described in Section 3.3, but the subset of five rules also runs the affective valence test to determine if the affective verbal message needs to be appended to the rule's base feedback message. Again, each rule has a pool of equivalent affect-oriented messages. This set of rules represents one third of the complete set, but three out of five of these rules belong to the problem submission rule category; these are the rules catching all submissions with errors, which makes them the most frequently used rules in EER-Tutor. Table 4.2 lists these rules along with affect-oriented add-on message examples; we tried to compile the messages in such a way as if the agent takes on a part of the blame and somewhat delicately reminds the user that the negative feelings are not there forever. Some messages also acknowledge the complexity of the task.

The following are examples of feedback messages used by the agent for the unsolicited affective interventions which are triggered by the observed changes in the affective state during the absence of interface events. Overall these messages are similar to the add-on affect-oriented messages—they are intended to address the user's negative feelings and express empathy:

- "I'm sorry if you are feeling frustrated—it's just that some of the problems demand lots of work."
- "You don't appear happy—I'm sorry if you are not enjoying this experience."

- "I apologise if you feel negative about this practice session—some of the solutions are quite complex."
- "I'm sorry if you are not enjoying this practice session—if you keep going though, you will be better prepared for future assessment."
- "Some of these problems require a lot of thinking and time—I'm sorry if you are not enjoying this session."

The evaluation of the affect-aware agent with facial feature tracking and experimental results are described in the following chapter.

№	State Description	Affect-oriented Add-on Message Examples
3	Empty solution submission	“You seem frustrated at the moment; I’m sorry if you are not enjoying the session.”
4	Three repeated submissions with no changes to solution	“If you are feeling negative at the moment, the best way I can help is by listing the errors.” “It appears you are somewhat frustrated with this particular problem, but you will feel better when you resolve these errors.” “I’m sorry to see you are not feeling happy, but these error messages are your clues to help you find the right solution.” ...
7	Submission with one error	“Just a little more effort and you will get there—it will make you feel great.” “I hope you are not feeling too negative now—actually you are very close to completion.” “You should be able to get over and done with this problem very soon—don’t feel bad about it.” ...
8	Submission with more then one error	“I’m sorry if it appears this problem is too hard—-some problems are harder than others.” “You might be feeling frustrated now, but you can definitely improve your skills if you keep practicing.” “I hope you keep pushing on, even though it might be difficult at the moment.” ...
9	Five repeated requests for new problems	“I’m sorry if you feel unhappy. I hope you can find a problem to work on.” “It seems you are somewhat frustrated. Would it help if you started working on one of the simpler problems?” “You don’t look very pleased with this experience. I’m sorry if this session is not going well for you at the moment.” ...

Table 4.2: Rules with affect-oriented add-on messages.

CHAPTER 5

Evaluation Study

The study was designed with the purpose of evaluating the pedagogical agents' strategies for affect-recognition and affective response described in the previous chapter. During March–April 2007, we ran the evaluation study of the affect-aware version of the pedagogical agent. Section 5.1 describes the experimental environment and experiment design, followed by the results of the study in Section 5.2. The limitations of the study are listed in Section 5.4 and finally, Section 5.3 discusses the outcomes of the evaluation and offers some reflections on our work.

5.1 Experiment Description

The study was designed as the experimental versus control condition comparison of two agents. The experimental group had access to the full-featured affect-aware version of the agent (as described in Section 4.4) whose affective awareness was based on facial feature tracking. The control condition, however, had the plain affect-unaware version of the agent. This version of the agent was guided by the original set of rules (as in the pilot study), but the affect-adjust part of the rules was disabled; this was so the agent did not generate affective facial or verbal reactions, but always remained neutral. The task-oriented feedback for both conditions was identical. From the user's point of view the difference between the two conditions was in the capability of the affect-aware agent to change its facial expression and generate affect-oriented feedback.

The participants for the experiment were recruited on a voluntary basis from the COSC226 class which is an introductory course on relational databases taught in the second-year curriculum of the Department of Computer Science and Software Engineering at the University of Canterbury. All users were familiar with EER modelling and EER-Tutor because the COSC226 class was introduced to the sans-agent version of EER-Tutor in the laboratory sessions a week prior to the evaluation study.

The participants were randomly allocated to the control and experimental conditions. The study was conducted as a series of individual sessions, one session per participant, in the Departmental HCI laboratory for single-user experiments. Figure 5.1 shows the experimental environment. A webcam for facial feature tracking was mounted on top of the monitor and aimed at the participant's face. We used the *Logitech QuickCam Pro 5000* webcam, operating at the frame rate of 15 fps., resolution of 640×480 px. For improving the accuracy of facial feature tracking, we ran the sessions in a controlled lighting environment—two 1000W video-studio lights were pointed away from the participant towards a white screen, which worked as a source of diffused white light. Participants wore head-phones to hear the agent's feedback.

The participants were expected to take a 45-minute session with EER-Tutor, while solving problems of their choice from EER-Tutor's curriculum at their own pace. Before each session the experiment administrator provided a verbal description of the task. To ensure informed consent for data collection, the experiment administrator provided a high-level description of the purpose of the webcam and a description of the nature of the data collected during the session; the descriptions did not contain references to the measurement of affective responses and were identical for both conditions. When the participant was ready to sign onto EER-Tutor, the experiment administrator started facial feature tracking session on the experimental workstation by first calibrating the tracking parameters as discussed in Section 4.2 and then removing the facial feature ap-

(a) A participant in front of the workstation with a webcam mounted on the monitor.

(b) Controlled lighting environment setup.

Figure 5.1: Experiment environment setup.

plication window from the screen, leaving only EER-Tutor's interface available for the participant during the session. For the duration of the session the participants were left alone in the room. At the end of the session, the experiment administrator prompted the participant to sign out of EER-Tutor anf fill out the questionnaire. Finally the participant was debriefed on the nature of the experiment and the underlying research.

5.2 Evaluation Results

A total of 27 participants (24 male, 3 female) took part in the experiment—13 participants (11 male, 2 female) were allocated to the control condition and 14 (13 male, 1 female) to the experimental condition. The average age in the control and experimental conditions was 22 ($\sigma = 6.5$) and 24 ($\sigma = 7.1$) years respectively. We did not expect to observe significant difference between the conditions from the point of view of objective learning performance, because the sessions were short; therefore the questionnaires were the main source of data for the between-condition comparison. Questionnaire items 1–3 recorded responses on the Likert scale, with the ranges of 1–3, 1–5 and 1-5 respectively. Responses to questions 7–9 required Yes/Maybe/No answers[1]. We used Mann-Whitney U Test for Likert scales, (one independent variable, two independent groups) to compare the responses. Table 5.1 summarises the results of the evaluation based on the questionnaire given in Appendix B.

Question 1 responses compared participants' level of expertise in EER modelling; as in the male and female ratio and age data, there were no significant difference between the control and experimental groups for the question, suggesting that both groups represented equal sampling. The main positive outcome of the evaluation is determined by the responses to questions 8 and 9, which rank the adequacy and appropriateness of agent's behaviour in the role of a caning tutor, and the usefulness of the agent in EER-Tutor. In question 8, 64% of the experimental condition participants answered "Yes" versus only 30% in the control condition; for question 9 the "Yes" responses

[1]For the analyis, the answers were encoded as $Yes = 2$, $Maybe = 1$, $No = 0$.

were 43% and 15% respectively. The affect-aware agent's behaviour adequacy and appropriateness were rated higher then the neutral agent's behaviour, providing a reliable result (Mann-Whitney U Test, $U = 57$, $N_C = 13$, $N_E = 14$,$p < 0.05$). Similarly, the usefulness of the affect-aware agent was rated higher in comparison the usefulness of the neutral agent, also showing a reliable result (Mann-Whitney U Test, $U = 56$, $N_C = 13$, $N_E = 14$,$p < 0.05$).

Questionnaire Item №	$\bar{x}$		σ		U	P	Diff.
	contr.	exper.	contr.	exper.			
Q1. Estimated Expertise	1.62	1.64	0.50	0.50	88	0.44	
Q2. Estimated Learning	3.23	3.57	0.93	1.09	71	0.16	
Q3. Enjoyed agent	3.23	2.93	0.73	1.27	76	0.22	
Q7. Noticeable agent's emotions	0.62	0.43	0.96	0.76	85	0.33	
Q8. Adequate agent's behaviour	0.92	1.50	0.86	0.76	57	0.04	✓
Q9. Agent useful	0.54	1.14	0.78	0.86	56	0.04	✓

Table 5.1: Evaluation results.

Questions 2 and 3 did not show a difference between the participants' perception of learning and enjoyment levels with either version of the agent. Questions 4 (first chosen agent) and 5 (choice justification) responses cannot be analysed in the same way as other responses, because of an inconsistency between the control and experimental conditions code implementation; for the control condition, the participants were not presented with the agent-selection page upon sign on; instead, one of the male agents (David) was displayed by default. Consequently, question 6 free-form responses indicate that most users did not know it was possible to choose a different agent or did not feel the need to change the agent. Question 7, ranking the participants' perceptions of the agents' emotional expressiveness, did not reveal significant difference between the two conditions—30% in the control condition answered "Yes" versus 15% in the experimental condition. This result is somewhat unexpected, because the affective facial expressions were generated only for the experimental condition; section 5.3 discusses this outcome in more detail.

Table 5.2 contains the responses to the questionnaire items discussed in this section. In addition, the Aff-R column provides the counts of affect-oriented verbal feedback delivered by the agent, while the Aff-I column gives the number of affective interjections made by the agent; this data was extracted from session logs. The results in Aff-R and Aff-I columns show that 64% of the experimental condition participants received affect-oriented feedback as a result of negative affective state observed through facial feature tracking. The absence of affective response for some participants is largely explained by the unresolved issues with our facial feature tracking implementation, which did not work as expected for some participants who wore facial hair or glasses during the experiment. For participant №21 facial feature tracking did not work at all, because this participant was of African ethnicity. Section 5.4 lists the limitations of the study associated with facial feature tracking.

As in the pilot study, verbalised feedback was appreciated by the participants; even though the questionnaire did not elicit comments on verbalised feedback, 33% of participants (approximately equal proportions for each condition) stated that verbal feedback was a useful addition to EER-Tutor because it helped the users to remain focused on the diagram and work faster. A number of participants claimed verbalised feedback saved them time. Only 11% stated that verbalised feedback was unnecessary or distracting. The *Quiet* button in agent-control toolbar described in section 3.2 allowed the users to make the agent stop verbalising feedback; only 20% used this feature to stop verbal feedback, but all these participants turned the verbal feedback on again within one to four minutes.

The participants interest in the verbal feedback is explained by the research presented in the recently published book, *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*, by Nass and Brave [2005]. The authors of the book provide evidence that the awareness of non-human origin of speech is not enough for the "brain to overcome the historically appropriate activation of social relationships by voice." These findings have a vast potential for the future development of affective pedagogical agents in ITSs.

№	Gender	Age	Q1	Q2	Q3	Q7	Q8	Q9	Aff-R	Aff-I
1	m	17	2	4	3	No	Yes	Maybe		
2	m	22	1	2	3	No	No	No		
3	f	20	2	5	3	No	No	No		
4	m	20	1	2	3	No	Maybe	No		
5	m	19	2	2	3	No	Maybe	No		
6	m	19	2	4	3	No	Yes	No		
7	m	19	2	4	4	Yes	Yes	Yes		
8	m	18	1	3	3	No	Maybe	No		
9	m	38	1	3	5	Yes	Maybe	Yes		
10	m	21	2	3	3	Yes	Yes	Maybe		
11	m	24	2	4	3	Yes	No	No		
12	m	18	2	3	4	No	No	Maybe		
13	f	34	1	3	2	No	No	No		
$\bar{x}$		22.2	1.6	3.2	3.2					
σ		6.4	0.5	0.9	0.7					
14	m	34	2	3	4	No	Yes	Yes		
15	m	19	2	3	1	No	Yes	No		
16	m	19	1	3	2	No	Maybe	Maybe		5
17	m	40	1	3	4	Maybe	Yes	Maybe		5
18	m	34	1	5	3	No	Yes	Yes		3
19	m	28	1	4	3	No	Yes	Yes	2	3
20	m	20	2	4	3	No	Yes	Yes	6	4
21	m	18	2	5	5	No	Yes	Yes		
22	m	21	1	4	5	Maybe	Yes	Maybe	2	3
23	m	26	2	3	3	No	Yes	Yes	1	2
24	f	22	2	3	2	Yes	Maybe	Maybe		
25	m	17	2	5	3	Yes	Maybe	No		1
26	m	23	2	1	1	No	No	No	1	
27	m	20	2	4	2	No	No	No		
$\bar{x}$		24.7	1.6	3.6	2.9					
σ		7.3	0.5	1.1	1.3					

Table 5.2: Questionnaire responses for control (1–13) and experimental (14-27) conditions.

Free-form questionnaire responses suggest that the participants received the affect-aware version with interest and approval; for example one participant commented: "I liked when the avatar told me I shouldn't worry because of feeling uncomfortable about the question I was working on." There were two other comments of a similar nature. Three participants, however, stated they felt annoyed when the agent misdiagnosed their affective state; one user commented: "The avatar kept asking me if I was feeling negative when I wasn't."

Another comment suggests that we, as interaction designers, were not clear enough about the emotions and the intention the pedagogical agent was to communicate to the user: "I needed encouragement when I wasn't doing very well, but instead got a sad face". This user clearly did not find the agent's empathy encouraging. In this situation, verbal expression of empathy combined with a positive facial expression could have had a better effect.

We also observed in some cases that the agent did not match the users' expectations in terms of its ability to provide situation-specific unsolicited hints and assistance when the participants were struggling with the task. For example, one user commented: "When it was obvious I was lost, the avatar didn't offer any tips or appropriate questions." It appears that the agent's presence raised the level of users' expectations associated with the EER-Tutor's ability to guide users and provide hints. This functionality is implicit in EER-Tutor, because at any time the list of errors can be obtained by submitting an incomplete solution; in the agent's presence, however, some users wanted the agent to take it upon itself to provide context-specific hints. Without such ability the agent was perceived as "... a helper that was not very helpful." Some users also reported that the agent was a distracting influence because of its "fidgeting".

In general, approval of the pedagogical agent's presence in EER-Tutor dominates the questionnaire responses. The following are some examples of positive free-form feedback from the questionnaires:

- "Avatars made EER-Tutor seem more polished."

- "I was told to stop surfing problems and complete one—it made me laugh and kept me interested."
- "I liked it that the avatar wasn't being overbearing—only interacting with me when needed, when I submitted a solution."
- "It is better to hear 'well done' coming from a human-like computer-generated avatar, rather than displayed as a string of characters on the screen."
- "Their presence helped me think better when I was solving problems."
- "Avatars gave the system a human feel."

5.3 Study Results Discussion

The results of this evaluation study follow in a line of natural progression from the pilot study; the outcomes of both studies lend their support to the use of affective pedagogical agents in ITSs. Even though it appears from both the pilot study and the final evaluation there were two camps of participants, 'we-want-more-interaction' and 'agent-is-distracting', the former camp was dominant.

A peculiar experimental outcome is the negative result for question 7. Neither experimental nor control conditions differed in their perception of the agent's emotional responses—both groups reported almost no awareness of the agent's affective facial expressions. There are a number of explanations for this, each of which could have equally contributed to the outcome:

- The pilot study results prompted us to reduce the absolute values of the affect-adjust coefficients because a number of users indicated in their responses that the agent was getting too upset too easy. The reason is that the pilot study participants' task included free system exploration—the participants were allowed to "play" with the system to test the robustness of the interface and the agent; this resulted in a very sad-looking agent, because some participants took it upon themselves to test how easy or hard it was to make the agent lose its emotional balance by such actions as repetitive identical submissions or problem surfing; the Affect-adjust column in Table 3.1 shows the new values in parentheses. In retrospect, we think that the affect-adjust values should not have been changed for the evaluation; it might be that for the evaluation study affective expressions in some cases were too subtle to be noticed.

- The participants were not explicitly instructed before the experiment to observe emotional changes; consequently, while focusing on problem-solving, some participants may have not paid attention to the agent's facial expressions.

- Bickmore [2004] states that several studies of mediated human-human interaction have found that the additional nonverbal cues provided by video-mediated communication do not affect performance in task-oriented interactions, but in interactions of a more relational nature, such as getting acquainted, video is superior [Whittaker and O'Conaill, 1997]. Consequently, it appears, in the context of the study, participants were less conscious of affective facial expressions, but more aware of affect-oriented messages expressed verbally. Responses to the free-form questions indicate that verbalised feedback was more appreciated then visual feedback. Some participants also interpreted non-affective messages as affective; for example one user stated in the questionnaire that the agent appeared to get angry because of problem-surfing.

Both the pilot study and final evaluation indicate a range of preferences associated with pedagogical agents and affective communication. Affective interaction is individually driven, and it is reasonable to suggest that in task-oriented environments, affective communication carries less importance for a certain proportion of learners. Also, some learners might not display emotions in front of a computer, or some users might display emotions differently; even though people do tend to treat computers and other digital media socially, it does not necessarily mean that the HHI and HCI responses are equivalent.

The active affective support which we attempted to implement in the affect-aware version of the agent theoretically takes the interaction between the agent and learner to a new level—it brings the pedagogical agent closer to the learner. This opens a whole new horizon of social and ethical interaction design issues associated with the human nature traits. For example, some people are naturally more private about their feelings; such individuals might respond to the invasion of their private emotional space by a perceptive affective agent with a range of reactions from withdrawal to fear. Others might resent being reminded about their feelings when they are focusing on a cognitive task; in such situations people might unconsciously refuse to acknowledge their feelings all together. Although the interplay of affective and cognitive processes always underpins learning outcomes, affective interaction sometimes might need to remain in the background; whatever the case, an ITS should let the user decide on the level of affective feedback, if any, thus leaving the user in control. This is recommended in the work of Kay [2000, 2001] on learner-controlled student modelling and scrutability.

Affective state in computer-mediated learning environments and in HCI in general is known to be negatively influenced by the mismatch between the user's needs and the available functionality; inadequate interface implementations, system's limitations, lack of flexibility, occurrences of errors and crashes—all these factors contribute to the affective state. In most cases, it is difficult or virtually impossible to filter out the noise generated by this kinds of problems. These considerations add another level of complexity to modelling such ill-defined domains as the human-like emotional behaviour.

It is fair to say that not all computers need to be aware of their user's emotions—most machines need only remain rigid straightforward tools [Picard, 2000]. At the same time, the inevitable and undisputed truth is that humans are affective beings, guided by a complex system of emotions, drives, and needs. Some aspects of affect-recognition in HCI may remain ill-defined or hidden, just as in some HHI scenarios one can never be completely and utterly sure of the perceived experience. Affective agents may improve the learner's experience in variety of ways, and these will be perceived differently by every individual learner; agents may ease frustration, make the process more adaptive, interesting, intriguing, appealing. If affect-recognition and affective agents can attract more learners and improve learning outcomes, bringing ITS research a step closer to bridging the affective gap and resolving the *2 Sigma* problem, this step is worth taking.

5.4 Study Limitations

We see the short duration of the evaluation study as the main limitation of our research. It was unlikely that the experiment revealed significant performance difference after a 45-minute session when the differences between the conditions were as subtle as they were in our study. Another limitation lies in the interpretation of furrowed brows: along with displeasure and mental anguish, this expression can sometimes indicate concentration, attention, or the execution of physical or intellectual effort [Farkas, 1994].

In our implementation the agent had no way of distinguishing between the concentration states and the negative affective states, which may explain some irritation reported by users in Section 5.2. Concentration may have been misdiagnosed by the system as a negative state, but as discussed in Section 2.7, concentration along with short-term negative affect constitutes an important stage of learning. Interfering with this stage would have surely resulted in irritation with the agent, rather then the desired effect of breaking the negative affective state. In retrospect, we could have used the user's cognitive progress to help the agent differentiate between concentration and long-term negative affect. Research provides evidence of correlation between the strength of task-evoked pupillary responses and cognitive load [Blascovich and Seery, 2005]; the use of this measure could reliably improve the differentiation between concentration and negative affective state. We believe that a more elaborate implementation of facial feature tracking could benefit our research by enabling detection of higher-level mental states, such as concentration, curiosity, boredom and so on, through the multi-level analysis of action patterns as described in the work of Kaliouby and Robinson [2004].

As stated in Section 5.2, our facial feature tracking implementation did not work for a participant of African ethnicity; we were unable to calibrate the system, because the iris position detection step in the facial feature tracking algorithm failed. The session ran without facial feature tracking. In addition, the accuracy of the agent's affective state judgement has been adversely affected by facial hair and glasses; similar problems are common for facial recognition and feature tracking.

CHAPTER 6

Conclusions

This thesis presented our research on recognition and support of affective states in ITSs. In our work, we adopted the dimensional approach to modelling emotions; this approach is reflected in our implementation decisions and affective interaction design. This final chapter completes the thesis by first summarising our work in Section 6.1. Section 6.2 lists our research contributions. Section 6.3 offers directions for future work on affective awareness and affective pedagogical agents with a view towards enhancing learning with ITSs.

6.1 Research Summary

We described an approach to modelling an affective pedagogical agent's persona as a set of rules corresponding to the interaction session states which require responses from a pedagogical, affective or social point of view. Each rule defines an emotional response, adjusting the agent's emotional state in accordance with the situation matching the rule. The agent's facial expressions depend on the variations of the agent's emotional state, which changes throughout the session, in response to the users's actions. We defined this version of the agent as the affect-inferring agent, because the agent's affective state is based on the appraisal of the learner's actions; the appraisal model is implicitly programmed into the agent's behaviour rules.

Our approach to modelling affective persona is characterised by simplicity, flexibility and scalability. The agent's persona can be developed incrementally by adding new rules; the larger the number of rules, the more versatile and interactive the persona is. With this approach it is possible to develop different types of personas defined by their own sets of rules.

We created a set of rules defining a tutor-like persona, mimicking a somewhat informal interaction in a learning environment with passive emotional support. Four animated characters were designed to take on the tutor persona in our evaluation studies. We integrated the agent into the test-bed application for our research. The results of a pilot study indicated a good outlook on the viability of the suggested approach—our agent's persona secured positive ratings during the pilot study.

From the pilot study we learned that the quality of the agent's voices was critical for maintaining the agent's rapport with the users and not interfering with the flow of the learning process. The pilot study also revealed the breadth of user preferences when it came to interacting with affective pedagogical agents and suggested that finding an ideal persona to "click" with all users was unrealistic. Needless to say, affective agents imbibed with social and emotional intelligence, need to project respect towards the human individuality observed in the style of interaction and learning. Some learners saw the agent as a distraction, while others wished the agent was more interactive and entertaining.

In the second stage of the project, we enhanced the agent's persona with affective awareness based on facial feature tracking to give the agent a more realistic level of emotional intelligence; this version of the agent is identified in our work as the affect-aware agent. We wrote an application which extracts the locations of a set of facial features in real-time from a video-stream and identifies the user's affective state. The affect calculations from feature tracking data and affect modelling were designed with the dimensional emotion model in mind. The relevant rules in the original set were modified to accommodate variations in the observed affective state of the user; the modifications were carried out by adding affect-oriented feedback which was meant to be delivered to the user afflicted by the negative affective state.

In our interaction design we relied both on the Flow theory and the model of the cyclic model of emotions in learning which ties together cognitive and affective processes in the learner's mind. We tried to make affect-oriented feedback and unsolicited affect-oriented interjections as unobtrusive as possible, but at the same time we aimed to have the agent provide active affective support by addressing the user's negative affective states. The evaluation of the affect-aware agent version against the non-affective agent provided positive results: the participants preferred the affect-aware agent as their tutor; at the same time the affect-aware agent was rated as a more valuable addition to EER-Tutor.

The agents' uptake was not unanimous, but the evaluation results advocate the presence of affective pedagogical agents, with the affect-aware agent demonstrating superiority over its non-affective counterpart. The participants' responses indicate the need for making the agent's behaviour more configurable. An encouraging and somewhat unexpected outcome of the study concerned verbal feedback: audible narration was welcomed by most participants irrespective of their expectations and perceptions of the agents' persona.

6.2 Research Contributions

The benefits of affective pedagogical agents in learning have been addressed and explored to a certain depth in the ITS research in recent years. Although we did not analyse the participants' learning performance during the experiment, the evaluation results support the use of affective pedagogical agents in ITSs. Our research, however, focused on modelling the affective agent's emotions and behaviour as well as modelling the users' affective state. The interaction of both affective states and the user's cognitive state were at the centre of our attention in the process of interaction design.

We see proof of the plausibility and practicality of the dimensional emotion modelling approach as the main contribution of our research: the dimensional approach was used in our project for modelling both the agent's emotions and the user's affective states. We feel that the dimensional approach presents a viable approach to emotion modelling, because it eliminates the complexity of border-line cases and unlocks the multiple shades of emotional states which would be described by a single label within the categorical approach. Additionally, we view our facial feature extraction and tracking algorithm along with its affective state detection logic as another important contribution component. This algorithm lends itself easily to the integration with the dimensional emotion modelling.

We view our second significant contribution to be the implementation and evaluation of the rule-based agent behaviour that synthesises cognitive and affective states. At the same time, the evaluation supports the viability of combining affective awareness with active emotional support in pedagogical agents, when the agent directly addresses the emotional state of the user.

In spite of the positive outcomes of our project, we acknowledge that our research barely scratches the surface of the vast domain of modelling emotional and social intelligence in pedagogical agents; this domain's complexity and ill-defined characteristics pose countless challenges to ITS researchers. We hope that our effort makes the field of ITS research richer through contributing our experiences to its shared knowledge repository.

6.3 Future Work

We supported our research with a significant volume of technical work and development, such as implementation of facial feature tracking, implementation of affective state inference, affective state modelling, affective pedagogical agent development and agent's integration into EER-Tutor. All these components and processes provide a framework for future research on affective pedagogical agents in ITSs. The range of ideas for future work in the context described in this project covers potentially straightforward enhancements and complex open-ended questions.

For example, the addition of problem selection strategies to EER-Tutor's pedagogical module can certainly improve the quality of hints provided by affective agents and take the user's perception of the agent's intelligence to a new level; students learn faster if they attempt problems that are most suitable to their cognitive state. Integrating previously explored problem selection strategies [Mitrović and Martin, 2003, 2004] with the agent's behaviour rules is likely to make the agent's behaviour more closely resemble the behaviour expected from an intelligent and attentive tutor.

Our implementation of the pedagogical agent does not distinguish between the various character/personality types, though in real life the pedagogical strategy selection skills of expert teachers hinge on their capability to discriminate between various types of learners [Heylen et al., 2005; du Boulay and Luckin, 2001].

Another option for future research could be based around an investigation of the visual relationships between users and agents; such a project could be based around pupil-movement tracking to explore when, how much and how often a user looks at the agent's face. There is plenty of scope for research in facial feature tracking observation in combination with affective agents; for example, tracking data can be used to infer the influence of affective pedagogical agents on stress levels, motivation and concentration.

Currently EER-Tutor affective agent supports only one-way verbal/textual communication. However, two-way verbal communication is a ubiquitous component of a real-life tutor-learner interaction. The evaluation study free-form responses include suggestions regarding the enhancement of EER-Tutor agent with a Natural Language Processing (NLP) interface; such addition could contribute to the agent's realism. More importantly, NLP would most likely enhance the flow in the system, as was observed through the addition of the verbalised feedback.

References

Allport, G. W. (1935). Attitudes. In Murchison, C. A., editor, *A handbook of Social Psychology*, volume 2, pages 798–844. Clark University.

Anderson, J. R. (1993). *Rules of the Mind*. Lawrence Erlbaum Associates.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., and Pelletier, R. (1995). Cognitive Tutors: Lessons Learned. *The Journal of the Learning Sciences*, 4(2):167–209.

Ayscough, P. B. (1977). CALCHEMistry. *British Journal of Education Technology*, 8:201–3.

Baghaei, N., Mitrović, A., and Irwin, W. (2005). A Constraint-Based Tutor for Learning Object-Oriented Analysis and Design using UML. *International Conference on Computers in Education*, pages 11–18.

Baghaei, N., Mitrović, A., and Irwin, W. (2006). Problem-Solving Support in a Constraint-based Tutor for UML Class Diagrams. *Journal of Technology, Instruction, Cognition and Learning*, 4(1-2).

Baker, R. S. J., Corbett, A. T., Koedinger, K. R., Evenson, S., Roll, I., Wagner, A. Z., Naim, M., Raspat, J., Baker, D. J., and Beck, J. E. (2006). Adapting to When Students Game an Intelligent Tutoring System. In Ikeda, M., Ashley, K. D., and Chan, T.-W., editors, *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*, volume 4053 of *Lecture Notes in Computer Science*, pages 392–401, Jhongli, Taiwan. Springer-Verlag GmbH.

Baylor, A. L. (2005). Preliminary Design Guidelines for Pedagogical Agent Interface Image. In *Proceedings of the 10th International Conference on Intelligent User Interfaces*, pages 249–250, New York, NY, USA. ACM Press.

Baylor, A. L. and Kim, Y. (2004). Pedagogical Agent Design: The Impact of Agent Realism, Gender, Ethnicity, and Instructional Role. In Lester, J. C., Vicari, R. M., and Paraguaçu, F., editors, *Proceedings of the 7th International Conference on Intelligent Tutoring Systems*, volume 3220 of *Lecture Notes in Computer Science*, pages 592–603, Maceió, Alagoas, Brazil. Springer-Verlag GmbH.

Beck, J., Stern, M., and Haugsjaa, E. (1996). Applications of Artificial Intelligence in Education. *Crossroads*, 3(1):11–15.

Bickmore, T. W. (2004). Unspoken Rules of Spoken Interaction. *Communications of the ACM*, 47(4):38–44.

Blascovich, J. and Seery, M. D. (2005). Visceral and Somatic Indexes of Social Psychological Constructs. In Kruglanski, A. W. and Higgins, E. T., editors, *Social Psychology: Handbook of Basic Principles*. Guilford Publications, 2nd edition.

Bloom, B. S. (1984). The 2 Sigma Problem: The Search for Method of Group Instruction as Effective as One-to-One Tutoring. *Educational Researcher*, 13(6):4–16.

Bosma, W. and André, E. (2004). Exploiting Emotions to Disambiguate Dialogue Acts. In Vanderdonckt, J., Nunes, N. J., and Rich, C., editors, *Proceedings of the 9th International Conference on Intelligent User Interfaces*, pages 85–92. ACM Press.

Bourel, F., Chibelushi, C. C., and Low, A. A. (2000). Robust Facial Feature Tracking. In Mirmehdi, M. and Thomas, B. T., editors, *Proceedings of the 11th British Conference on Machine Vision*, volume 1, pages 232–241, Bristol, UK. British Machine Vision Association.

Bradley, M. (1994). Emotional memory: A dimensional analysis. In *Emotions: Essays on Emotion Theory*. Lawrence Erlbaum Associates, Hillsdale, New Jersy.

Bradley, M. M., Greenwald, M. K., Petry, M. C., and Lang, P. J. (1992). Remembering Pictures: Pleasure and Arousal in Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(2):379–390.

Bradley, M. M. and Lang, P. J. (2000). Measuring Emotion: Behavior, Feeling and Physiology. In Lane, R. D. and Nadel, L., editors, *Cognitive Neuroscience of Emotion*, Series in Affective Science, pages 242–276. Oxford University Press, New York.

Bradski, G. (2000). Programmer's Toolchest—The OpenCV Library. *Dr. Dobb's Journal November 2000, Computer Security*.

Brna, P., Cooper, B., and Razmerita, L. (2001). Marching to the Wrong Distant Drum: Pedagogic Agents, Emotion and Student Modelling. Proceedings of the 2nd International Workshop on Attitude, Personality and Emotions in User-Adapted Interaction, held in conjunction with the 8th International Conference on User Modeling.

Burleson, W. (2006). *Affective Learning Companions: Strategies for Empathetic Agents With Real-Time Multimodal Affective Sensing to Foster Meta-Cognitive and Meta-Affective Approaches to Learning, Motivation, and Perseverance*. PhD thesis, Massachusetts Institute of Technology.

Burleson, W. and Picard, R. (2004). Affective Agents: Sustaining Motivation to Learn Through Failure and a State of "Stuck". Proceedings of the Workshop on Social and Emotional Intelligence in Learning Environments, held in conjunction with the 7th International Conference on Intelligent Tutoring Systems.

Cacioppo, J. T., Tassinary, L. G., and Berntson, G. G. (2007). *Handbook of Psychophysiology*. Cambridge University Press.

Calder, A. J., Burton, A. M., Miller, P., Young, A. W., and Akamatsu, S. (2001). A Principal Component Analysis of Facial Expressions. *Vision Research*, 41(9):1179–1208.

Conati, C. (2002). Probabilistic assessment of user's emotions in educational games. *Applied Artificial Intelligence*, 16(7-8):555–575.

Conati, C., Chabbal, R., and Maclaren, H. (2003). A Study on Using Biometric Sensors for Monitoring User Emotions in Educational Games. *Proceedings of User Modeling Workshop on Assessing and Adapting to User Attitudes and Effect: Why, When, and How*.

Conati, C. and Maclaren, H. (2004). Evaluating a Probabilistic Model of Student Affect. In Lester, J. C., Vicari, R. M., and Paraguaçu, F., editors, *Proceedings of the 7th International Conference on Intelligent Tutoring Systems*, volume 3220 of *Lecture Notes in Computer Science*, pages 55–64, Maceió, Alagoas, Brazil. Springer-Verlag GmbH.

Cooper, B. (2002). *Teachers as Moral Models? The Role of Empathy in Teacher/Pupil Relationships*. PhD thesis, Leeds Metropolitan University.

Cooper, B. (2003). Care—Making the Affective Leap: More Than a Concerned Interest in a Learner's Cognitive Abilities. *International Journal of Artificial Intelligence in Education*, 13(1):3–9.

Cooper, M. L., Frone, M. R., Russell, M., and Mudar, P. (1995). Drinking to Regulate Positive and Negative Emotions: A Motivational Model of Alcohol Use. *Journal of Personality and Social Psychology*, 69(5):990–1005.

Cootes, T. F., Edwards, G. J., and Taylor, C. J. (2001). Active Appearance Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685.

Corradini, A., Mehta, M., Bernsen, N.-O., and Charfuelan, M. (2005). Animating an interactive conversational character for an educational game system. In *Proceedings of the 10th International Conference on Intelligent User Interfaces*, pages 183–190, New York, NY, USA. ACM Press.

Crane, D., Pascarello, E., and James, D. (2005). *Ajax in Action*. Manning Publications Co., Greenwich, CT, USA.

Csikzentmihalyi, M. (1990). Flow: The Psychology of Optimal Experience. *New York: HarperPerennial*.

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Optical Society of America, Journal, A: Optics and Image Science*, 2:1160–1169.

Dietz, R. B. and Lang, A. (1999). Æffective agents: Effects of agent affect on arousal, attention, liking and learning. In *Proceedings of the Third International Cognitive Technology Conference*, San Francisco, CA, USA.

du Boulay, B. and Luckin, R. (2001). Modelling Human Teaching Tactics and Strategies for Tutoring Systems. *International Journal of Artificial Intelligence in Education*, 12(3):235–256.

Dweck, C. S. (2000). *Self-Theories: Their Role in Motivation, Personality, and Development*. Psychology Press (UK).

Ekman, P. and Friesen, W. V. (1978). *Facial Action Coding System*. Consulting Psychologists Press.

Ekman, P. and Friesen, W. V. (1986). A New Pan Cultural Facial Expression of Emotion. *Motivation and Emotion*, 10(2).

Elmasri, R. and Navathe, S. B. (2003). *Fundamentals of Database Systems*. Addison-Wesley, 4th edition.

Farkas, L. G. (1994). *Anthropometry of the Head and Face*. Raven Press.

Felsman, J. K. and Vaillant, G. E. (1987). Resilient Children as Adults: A 40-year Study. *The Invulnerable Child*, pages 289–314.

Feris, R. S., Gemmell, J., Toyama, K., and Kruger, V. (2002). Hierarchical Wavelet Networks for Facial Feature Localization. In *Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 118–123, Los Alamitos, CA, USA. IEEE Computer Society.

Flores, V. C. (2005). Artnatomy/Artnatomia. `http://www.artnatomia.net`.

Foderaro, J. (2004). AllegroServe—a Web Application Server. `http://opensource.franz.com/aserve/index.html`.

Freedman, R. (1999). Atlas: A Plan Manager for Mixed-Initiative, Multimodal Dialogue. *International Conference on Artificial Intelligence in Education, Workshop on Mixed-Initiative Intelligence*.

Fridlund, A. J. and Izard, C. E. (1983). Electromyographic Studies of Facial Expressions of Emotions and Patterns of Emotions. In Cacioppo, J. T. and Petty, R. E., editors, *Social Psychophysiology: A Sourcebook*, pages 243–286. Guilford Press, New York.

Frijda, N. H. (1986). *The Emotions*. Studies in Emotion and Social Interaction. Cambridge University Press, New York.

Giuseppe, L. and Brass, D. (2003). Exploring the Social Ledger: Negative Relationships and Negative Asymmetry in Social Networks in Organizations. *Academy of Management Review, Special issue: Building Effective Networks, Academy of Management, Chicago, IL. April*, 4.

Goleman, D. (1995). *Emotional Intelligence*. Bantam Books, New York.

Gonzalez, R. C. and Woods, R. E. (1993). *Digital Image Processing*. Addison-Wesley Reading, Mass.

Graesser, A. C., Chipman, P., Haynes, B. C., and Olney, A. (2005). Autotutor: An Intelligent Tutoring System With Mixed-Initiative Dialogue. *IEEE Transactions on Education*, 48(4):612–618.

Graesser, A. C., Wiemer-Hastings, K., Wiemer-Hastings, P., and Kreuz, R. (1999). AutoTutor: A Simulation of a Human Tutor. *Cognitive Systems Research*, 1(1):35–51.

Gratch, J., Rickel, J., André, E., Cassell, J., Petajan, E., and Badler, N. I. (2002). Creating interactive virtual humans: Some assembly required. *IEEE Intelligent Systems*, 17(4):54–63.

Gray, H. (1918). *Anatomy of the Human Body*. Lea & Febiger Philadelphia, 20th edition.

Greenwald, M., Cook, E., and Lang, P. (1989). Affective Judgement and Psychophysiological Response: Dimensional Covariation in the Evaluation of Pictorial Stimuli. *Journal of Psychophysiology*, 3:51–64.

Gross, J. J. and Muñoz, R. F. (1995). Emotion Regulation and Mental Health. *Clinical Psychology: Science and Practice*, 2(2):151–164.

Hagan, M. and Menhaj, M. (1994). Training Feedforward Networks with the Marquardt Algorithm. *Neural Networks, IEEE Transactions on*, 5(6):989–993.

Haptek Inc. (2002). *Haptek Automated Personalities Programmer's Technical White Paper*. Haptek Inc.

Healey, J. and Picard, R. (2000). SmartCar: Detecting Driver Stress. *Proceedings of the 15th International Conference on Pattern Recognition*, pages 218–221.

Henderlong, J. and Lepper, M. R. (2002). The Effects of Praise on Children's Intrinsic Motivation: A Review and Synthesis. *Psychological Bulletin*, 128(5):774–795.

Heylen, D., Nijholt, A., and op den Akker, R. (2005). Affect In Tutoring Dialogues. *Applied Artificial Intelligence*, 19(3–4):287–311.

Heylen, D., Vissers, M., op den Akker, R., and Nijholt, A. (2004). Affective Feedback in a Tutoring System for Procedural Tasks. *Affective Dialogue Systems*, 3068:244–253.

Holt, P., Dubs, S., Jones, M., and Greer, J. (1994). The State of Student Modelling. In Greer, J. E. and McCalla, G. I., editors, *Student Modelling: The Key to Individualized Knowledge-based Instruction*, volume 125 of *Computer Systems and Sciences*, pages 3–35, Berlin. Springer-Verlag GmbH.

Hullfish, W. and Pottebaum, W. (1971). Take on a Digital Assistant: The Computer as a Teaching Aid. *Music Educators Journal*, 57(5):83–87.

Isen, A. M., Daubman, K. A., and Nowicki, G. P. (1987). Positive affect facilitates creative problem solving. *Journal of Personality and Social Psychology*, 52(6):1122–1131.

Izard, C. E. (1977). *Human Emotions*. Emotions, Personality, and Psychotherapy. Plenum Press, New York.

Johnson, L., Marsella, S., Mote, N., Si, M., Vilhjalmsson, H., and Wu, S. (2004). Balanced Perception and Action in the Tactical Language Training System. *Proceedings of the InSTIL Symposium on NLP and Speech Technologies in Advanced Language Learning Systems*.

Johnson, W. L., Rickel, J., Stiles, R., and Munro, A. (1998). Integrating Pedagogical Agents Into Virtual Environments. *Presence: Teleoperators and Virtual Environment*, 7(6):523–546.

Johnson, W. L., Rickel, J. W., and 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.

Johnson, W. L., Vilhjalmsson, H., and Marsella, S. (2005). Serious Games for Language Learning: How Much Game, How Much AI. In Looi, C.-K., McCalla, G., Bredeweg, B., and Breuker, J., editors, *Proceedings of the 12th International Conference on Artificial Intelligence in Education*, volume 125 of *Frontiers in Artificial Intelligence and Applications*, Amsterdam, The Netherlands. IOS Press.

Kahneman, D. (1973). *Arousal and Attention*, pages 28–49. Prentice-Hall, Englewood Cliffs, New Jersy.

Kaliouby, R. E. and Robinson, P. (2004). Real-time inference of complex mental states from facial expressions and head gestures. In *Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition*, volume 10, page 154, Los Alamitos, CA, USA. IEEE Computer Society.

Katz, S., Aronis, J., and Creitz, C. (2000). Modeling Pedagogical Interactions With Machine Learning. *Kognitionswissenschaft*, 9(1):45–49.

Kay, A. C. (1991). Computers, Networks and Education. *Scientific American*, 265(3):138–148.

Kay, J. (2000). Stereotypes, Student Models and Scrutability. In Gauthier, G., Frasson, C., and VanLehn, K., editors, *Proceedings of the 5th International Conference on Intelligent Tutoring Systems*, volume 1839 of *Lecture Notes in Computer Science*, pages 19–30, Montréal, Canada. Springer-Verlag GmbH.

Kay, J. (2001). Learner Control. *User Modeling and User-Adapted Interaction*, 11(1):111–127.

Kendon, A. (1983). Gesture and Speech: How They Interact. *Nonverbal Interaction*, pages 13–45.

Klein, J., Moon, Y., and Picard, R. W. (2002). This Computer Responds to User Frustration: Theory, Design, and Results. *Interacting with Computers*, 14:119–140.

Klingzing, H. G. and Tisher, T. P. (1986). Expressive nonverbal behaviors; A review of research on training with consequent recommendations for teacher education. *Advances in Teacher Education*, 2:89–133.

Knapp, M. L. (1978). *Nonverbal Communication in Human Interaction*. Holt, Rinehart and Winston, 2nd edition.

Kort, B. and Reilly, R. (2002). A Pedagogical Model for Teaching Scientific Domain Knowledge. *Frontiers in Education*, 1:(T3A)13–17.

Kort, B., Reilly, R., and Picard, R. W. (2001). An Affective Model of Interplay between Emotions and Learning: Reengineering Educational Pedagogy - Building a Learning Companion. In *Proceedings of the International Conference on Advanced Learning Techniques*, pages 43–48.

LaJoie, S. P. and Lesgold, A. (1989). Apprenticeship Training in the Workplace: Computer-coached Practice Environment as a New Form of Apprenticeship. *Machine-Mediated Learning*, 3:7–28.

Lang, A., Dhillon, P., and Dong, Q. (1995). Arousal, Emotion, and Memory for Television Messages. *Journal of Broadcasting and Electronic Media*, 38:1–15.

Lang, P. J. (1993). The Three System Approach to Emotion. In Birbaumer, N. and Öhman, A., editors, *The Organization of Emotion*, pages 18–30. Hogrefe-Huber, Toronto.

Lang, P. J., Greenwald, M. K., Bradley, M. M., and Hamm, A. O. (1993). Looking at Pictures: Affective, Facial, Visceral, and Behavioral Reactions. *Psychophysiology*, 30:261–273.

Last, R. W. (1979). The Role of Computer-Assisted Learning in Modern Language Teaching. *Assoc. for Literary and Linguistic Computing bulletin*, 7:165–171.

Lepper, M. R., Woolverton, M., Mumme, D. L., and Gurtner, J. L. (1993). Motivational Techniques Of Expert Human Tutors: Lessons For The Design Of Computer-Based Tutors. *Computers as Cognitive Tools*, pages 75–105.

Lesgold, A. (1988). Toward a Theory of Curriculum for Use in Designing Intelligent Instructional Systems. *Learning Issues for Intelligent Tutoring Systems*, pages 114–137.

Lessig, L. (1999). *Code and Other Laws of Cyberspace*. Basic Books.

Lester, J. C., Stone, B. A., and Stelling, G. D. (1999a). Lifelike Pedagogical Agents for Mixed-Initiative Problem Solving in Constructivist Learning Environments. *International Journal of User Modeling and User-Adapted Interaction*, 9(1–2):1–44.

Lester, J. C., Towns, S. G., Callaway, C. B., Voerman, J. L., and FitzGerald, P. J. (2000). Deictic and Emotive Communication in Animated pedagogical Agents. *Embodied Conversational Agents*, pages 123–154.

Lester, J. C., Towns, S. G., and Fitzgerald, P. J. (1999b). Achieving Affective Impact: Visual Emotive Communication in Lifelike Pedagogical Agents. *International Journal of Artificial Intelligence in Education*, 10:278–291.

Levenson, R. W. (2003). Autonomic Specificity and Emotion. In Davidson, R. J., Scherer, K. R., and Goldsmith, H. H., editors, *Handbook of Affective Sciences*, Series in Affective Science. Oxford University Press, New York.

Lewis, V. E. and Williams, R. N. (1989). Mood-congruent vs. Mood-state-dependent Learning: Implications for a View of Emotion. *Mood and Memory: Theory, Research, and Applications*, pages 157–171.

Linek, S. B. and Scheiter, P. G. K. (2006). Speaker/Gender Effect: Impact of the Speaker's Gender on Learning with Narrated Animations. In Sun, R. and Miyake, N., editors, *Proceedings of the 28th Annual Conference of the Cognitive Science Society in cooperation with the 5th International Conference of the Cognitive Science*, pages 483–488, Vancouver, British Columbia, Canada. Cognitive Science Society, Lawrence Erlbaum Associates, Inc.

Lisetti, C. L. and Nasoz, F. (2002). MAUI: A Multimodal Affective User Interface. In *Proceedings of the 10th ACM International Conference on Multimedia*, pages 161–170. ACM Press New York, NY, USA.

Marsella, S. (2003). Interactive pedagogical drama: Carmen's bright ideas assessed. In Rist, T., Aylett, R., Ballin, D., and Rickel, J., editors, *Proceedings of the 4th International Workshop on Intelligent Virtual Agents*, volume 2792 of *Lecture Notes in Computer Science*, pages 1–4, Kloster Irsee, Germany. Springer-Verlag GmbH.

Marsella, S. and Gratch, J. (2001). Modeling the Interplay of Emotions and Plans in Multi-Agent Simulations. In Moore, J. D. and Stenning, K., editors, *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, pages 294–599, Edinburgh, Scotland.

Marsella, S., Johnson, W. L., and LaBore, C. (2003). Interactive Pedagogical Drama for Health Interventions. In Hoppe, U., Verdejo, F., and Kay, J., editors, *Proceedings of the 11th International Conference on Artificial Intelligence in Education*, volume 97 of *Frontiers in Artificial Intelligence and Applications*, pages 341–348. IOS Press.

Martin, B. and Mitrović, A. (2002a). Authoring Web-Based Tutoring Systems with WETAS. *Proceedings of the International Conference on Computers in Education*, pages 183–187.

Martin, B. and Mitrović, A. (2002b). WETAS: A Web-Based Authoring System for Constraint-Based ITS. In de Bra, P., Brusilovsky, P., and Conejo, R., editors, *Proceedings of the 2nd Intertational Conference on Adaptive Hypermedia and Adaptive Web-based Systems*, pages 543–546. Springer-Verlag GmbH.

Martin, B. I. (2001). *Intelligent Tutoring Systems: The Practical Implementation of Constraint-based Modelling*. PhD thesis, University of Canterbury.

Mateos, G. G. (2003). Refining face tracking with integral projections. In Kittler, J. and Nixon, M. S., editors, *Proceedings of the 4th International Conference on Audio- and Video-Based Biometric Person Authentication*, volume 2688 of *Lecture Notes in Computer Science*, pages 360–368. Springer-Verlag GmbH.

Mayer, R. E. (2001). *Multimedia Learning*. Cambridge University Press.

Mayo, M. and Mitrović, A. (2001). Optimising ITS Behavior with Bayesian Networks and Decision Theory. *International Journal of Artificial Intelligence in Education*.

Miles, L. and Johnston, L. (2007). Detecting happiness: Perceiver sensitivity to smile veracity. *Journal of Nonverbal Behavior*, 31:259–275.

Mishra, P. and Hershey, K. A. (2004). Etiquette and the Design of Educational Technology. *Communications of the ACM*, 47(4):45–49.

Mishra, P., Nicholson, M., and Wojcikiewicz, S. (2001). Does My Word Processor Have a Personality? Topffer's Law and Educational Technology. *Journal of Adolescent and Adult Literacy*, 44(7):634–641.

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill Higher Education.

Mitrović, A. (2002). NORMIT: a Web-enabled Tutor for Database Normalization. In Kinshuk, R., Lewis, K., Akahori, R., Kemp, T., Okamoto, L., Henderson, C., and Lee, H., editors, *Proceedings of the International Conference on Computers in Education*, volume 2, pages 1276–1280.

Mitrović, A. (2003). An Intelligent SQL Tutor on the Web. *International Journal of Artificial Intelligence in Education*, 13(2–4):173–197.

Mitrović, A., Koedinger, K. R., and Martin, B. (2003). A Comparative Analysis of Cognitive Tutoring and Constraint-Based Modelling. In Brusilovsky, P., Corbett, A., and de Rosis, F., editors, *Proceedings of the 9th International Conference on User Modeling*, volume 2702 of *Lecture Notes in Artificial Intelligence*, pages 313–322, Berlin. Springer-Verlag GmbH.

Mitrović, A. and Martin, B. (2003). Scaffolding and Fading Problem Selection in SQL-Tutor. In Hoppe, U., Verdejo, F., and Kay, J., editors, *Proceedings of the 11th International Conference on Artificial Intelligence in Education*, volume 97 of *Frontiers in Artificial Intelligence and Applications*, pages 479–481. IOS Press.

Mitrović, A. and Martin, B. (2004). Evaluating Adaptive Problem Selection. In Bra, P. D. and Nejdl, W., editors, *Proceedings of the 3rd International Conference on Adaptive Hypermedia and Adaptive Web-based Systems*, volume 3137 of *Lecture Notes in Computer Science*, pages 185–194. Springer-Verlag GmbH.

Mitrović, A., Mayo, M., Suraweera, P., and Martin, B. (2001). Constraint-based Tutors: a Success Story. In Monostori, L., Váncza, J., and Ali, M., editors, *Proceedings of the 14th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, volume 2070, pages 931–940, Berlin. Springer-Verlag GmbH.

Mitrović, A. and Ohlsson, S. (1999). Evaluation of a Constraint-based Tutor for a Database Language. *International Journal of Artificial Intelligence in Education*, 10:238–256.

Mitrović, A. and Suraweera, P. (2000). Evaluating an Animated Pedagogical Agent. In Gauthier, G., Frasson, C., and VanLehn, K., editors, *Proceedings of the 5th International Conference on Intelligent Tutoring Systems*, volume 1839 of *Lecture Notes in Computer Science*, pages 73–82, Montréal, Canada. Springer-Verlag GmbH.

Moreno, K. N., Klettke, B., Nibbaragandla, K., and Graesser, A. C. (2002). Perceived Characteristics and Pedagogical Efficacy of Animated Conversational Agents. In Cerri, S. A., Gouardéres, G., and Paraguaçu, F., editors, *Proceedings of the 6th International Conference on Intelligent Tutoring Systems*, volume 2363 of *Lecture Notes in Computer Science*, pages 963–971, Biarritz, France and San Sebastian, Spain. Springer-Verlag GmbH.

Myers, D. G. (1998). *Psychology*. Worth Publishers, New York, 5th edition.

Nasoz, F., Alvarez, K., Lisetti, L., and Finkelstein, N. (2004). Emotion recognition from physiological signals using wireless sensors for presence technologies. *Cognition, Technology & Work*, 6(1):4–14.

Nasoz, F., Lisetti, C. L., Alvarez, K., and Finkelstein, N. (2003). Emotion Recognition from Physiological Signals for User Modeling of Affect. In *Proceedings of the 3rd Workshop on Affective and Attitude User Modelling*, Pittsburgh, PA, USA.

Nass, C. and Brave, S. (2005). *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. The MIT Press.

Nicol, A. A. M. and Pexman, P. M. (1999). *Presenting your findings: a practical guide for creating tables*. American Psychological Association.

Niedenthal, P. M. and Kitayama, S. (1994). Motivating the Focus of Attention. In Derryberry, D. and Tucker, D. M., editors, *Heart's Eye: Emotional Influences in Perception and Attention*, pages 167–196. Academic Press, New York.

Norman, D. (1988). *The Psychology of Everyday Things (a.k.a. The Design of Everyday Things)*. Basic Books, New York.

Nugent, W. R. and Halvorson, H. (1995). Testing the Effects of Active Listening. *Research on Social Work Practice*, 5(2):152–175.

Ohlsson, S. (1994). Constraint-Based Student Modelling. In Greer, J. E. and McCalla, G. I., editors, *Student Modelling: The Key to Individualized Knowledge-based Instruction*, volume 125 of *Computer Systems and Sciences*, pages 167–189, Berlin. Springer-Verlag GmbH.

Ohlsson, S. (1996). Learning From Performance Errors. *Psychological Review*, 103(2):241–262.

Ortony, A., Collins, A., and Clore, G. L. (1988). *The Cognitive Structure of Emotions*. Cambridge University Press.

Ortony, A. and Turner, T. J. (1990). What's Basic About Basic Emotions? *Psychological Review*, 97(3).

Paleari, M., Lisetti, C. L., and Lethonen, M. (2005). VALERIE: A virtual agent for a learning environment, reacting and interacting emotionally. In Looi, C.-K., McCalla, G., Bredeweg, B., and Breuker, J., editors, *Proceedings of the Workshop on Motivation and Affect in Educational Software, held in conjunction with the 12th International Conference on Artificial Intelligence in Education*, volume 125 of *Frontiers in Artificial Intelligence and Applications*, Amsterdam, The Netherlands. IOS Press.

Pandzic, I. S. and Forchheimer, R. (2002). *MPEG-4 Facial Animation: The Standard, Implementation and Applications*. John Wiley and Sons.

Parise, S., Kiesler, S., Sproull, L., and Waters, K. (1996). My Partner is a Real Dog: Cooperation with Social Agents. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, pages 399–408, New York, NY, USA. ACM Press.

Picard, R. W. (2000). Toward Computers that Recognize and Respond to User Emotion. *IBM Systems Journal*, 39(3–4):705–719.

Plutchik, R. (1980). *Emotions: A Psychoevolutionary Synthesis*. Harper and Row, New York.

Poel, M., op den Akker, R., Heylen, D., and Nijholt, A. (2004). Emotion based Agent Architectures for Tutoring Systems The INES Architecture. *Cybernetics and Systems*, pages 663–668.

Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., and Carey, T. (1994). *Human-Computer Interaction*. Addison-Wesley, New York.

Prendinger, H. and Ishizuka, M. (2001). Social Role Awareness in Animated Agents. *Proceedings of the 5th International Conference on Autonomous Agents*, pages 270–277.

Prendinger, H., Mori, J., and Ishizuka, M. (2005). Recognizing, Modeling, and Responding to Users' Affective States. In Ardissono, L., Brna, P., and Mitrović, A., editors, *Proceedings of the 10th International Conference on User Modeling*, volume 3538 of *Lecture Notes in Artificial Intelligence*, pages 60–69, Berlin. Springer-Verlag GmbH.

Reeves, B. and Nass, C. (1996). *The Media Equation. How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge Publications.

Rich, E. (1979). User modeling via stereotypes. *Cognitive Science*, 3(4):329–354.

Rickel, J. and Johnson, W. L. (1997). Intelligent Tutoring in Virtual Reality: A Preliminary Report. In du Boulay, W. and Mizoguchi, R., editors, *Proceedings of the 8th International Conference on Artificial Intelligence in Education*, volume 39 of *Frontiers in Artificial Intelligence and Applications*, pages 294–301. IOS Press.

Schank, R. and Neaman, A. (2001). Motivation and failure in educational simulation design. In Forbus, K. D. and Feltovich, P. J., editors, *Smart machines in education*, pages 37–69. AAAI Press.

Schaumburg, H. (2001). Computers as Tools or as Social Actors?—The Users' Perspective on Anthropomorphic Agents. *IJCIS*, 10(1-2):217–234.

Scherer, K. R. and Ceschi, G. (2000). Criteria for Emotion Recognition From Verbal and Nonverbal Expression: Studying Baggage Loss in the Airport. *Personality and Social Psychology Bulletin*, 26(3):327–339.

Seibel, P. (2005). *Practical Common Lisp*. Apress.

Serenko, A. (2007). Are Interface Agents Scapegoats? Attributions of Responsibility in Human-Agent Interaction. *Interacting with Computers*, 19(2):293–303.

Shaw, E., Ganeshan, R., Johnson, W. L., and Millar, D. (1999). Building a Case for Agent-assisted Learning as a Catalyst for Curriculum Reform in Medical Education. In Lajoie, S. and Vivet, M., editors, *Proceedings of the 9th International Conference* *on Artificial Intelligence in Education*, volume 50 of *Frontiers in Artificial Intelligence and Applications*, pages 509–516. IOS Press.

Shi, J. and Tomasi, C. (1994). Good Features to Track. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 593–600.

Shute, V. J., Glaser, R., and Raghavan, K. (1989). *Inference and Discovery in an Exploratory Laboratory*, chapter Learning and Individual Differences: Advances in Theory and Research, pages 279–326. Center for the Study of Learning, Learning Research and Development Center, University of Pittsburgh.

Steele Jr., G. L. (1990). *Common LISP: the Language*. Digital Press, 2nd edition.

Suraweera, P. and Mitrović, A. (2002). KERMIT: A Constraint-Based Tutor for Database Modelling. In Cerri, S. A., Gouardéres, G., and Paraguaçu, F., editors, *Proceedings of the 6th International Conference on Intelligent Tutoring Systems*, volume 2363 of *Lecture Notes in Computer Science*, pages 376–387, Biarritz, France and San Sebastian, Spain. Springer-Verlag GmbH.

Suraweera, P. and Mitrović, A. (2004). An Intelligent Tutoring System for Entity Relationship Modelling. *International Journal of Artificial Intelligence in Education*, 14(3–4):375–417.

Sweller, J., van Merrienboer, J. J. G., and Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3):251–296.

Teller, A. and Stivoric, J. (2004). The bodymedia platform: Continuous body intelligence. In *Proceedings of the 1st Acm Workshop on Continuous Archival and Retrieval of Personal Experiences*, pages 114–115, New York, NY, USA. ACM Press.

Thought Technology Ltd. (2000). ProComp Infiniti Hardware Manual. Web resource, available from `http://www.thoughttechnology.com`.

Tian, Y.-L., Kanade, T., and Cohn, J. (2005). *Handbook of Face Recognition*. Springer-Verlag GmbH.

Vaillant, G. E. (1995). *Adaptation to Life*. Harvard University Press.

VanLehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., Treacy, D., Weinstein, A., and Wintersgill, M. (2005). The Andes Physics Tutoring System: Lessons Learned. *International Journal of Artificial Intelligence in Education*, 15(3):147–204.

Viola, P. A. and Jones, M. J. (2001). Rapid Object Detection using a Boosted Cascade of Simple Features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 511–518. IEEE Computer Society.

Whittaker, S. and O'Conaill, B. (1997). The Role of Vision in Face-To-Face and Mediated Communication. *Video-Mediated Communication*, pages 23–49.

Wilson, P. I. and Fernandez, J. (2006). Facial Feature Detection Using Haar Classifiers. *Journal of Computing Sciences in Colleges*, 21(4):127–133.

Wilson, R. A. and Keil, F. C., editors (1999). *The MIT Encyclopedia of the Cognitive Sciences*. The MIT Press.

Zakharov, K., Mitrović, A., and Ohlsson, S. (2005). Feedback Micro-engineering in EER-Tutor. In Looi, C.-K., McCalla, G., Bredeweg, B., and Breuker, J., editors, *Proceedings of the 12th International Conference on Artificial Intelligence in Education*, volume 125 of *Frontiers in Artificial Intelligence and Applications*, pages 718–725, Amsterdam, The Netherlands. IOS Press.

APPENDIX A

Pilot Study Questionnaire

The following is a copy of the questionnaire that was used in the pilot study described in Section 3.4.

Pilot Study of EER-Tutor with Avatars

Thank you for participating in our pilot study—your feedback is crucial to our future research. This questionnaire is anonymous and if you wish, you may at any time withdraw from participation, including withdrawal of any information you have provided. However, by completing the questionnaire, you indicate your consent for publication of the generalised results of our research findings.

Login name:		Gender:	M / F	Age:	

1. How would you rate your expertise in EER modelling prior to this session? (please circle one)

(a) Novice learner	(b) Confident user	(c) Expert

2. How much did you learn about EER modelling from using EER-Tutor? (please circle one)

Nothing				Very much
1	2	3	4	5

3. How much did you enjoy using EER-Tutor? (please circle one)

Not at all				Very much
1	2	3	4	5

4. Which avatar did you chose when you first logged in? (please circle one)

(a) David	(b) Callie	(c) Mark	(d) Diane

5. Why did you choose this avatar? Please explain.

6. Did you switch to a different avatar in the process of interaction with EER-Tutor? If yes, which ones did you choose and why? (circle more then one if needed and comment)

(a) David	(b) Callie	(c) Mark	(d) Diane	(e) No, I didn't switch

P.T.O.

7. Did you think the avatar's responses to your actions were appropriate? Please comment.

Yes		Maybe		No	

8. Did you notice the emotional changes in the avatar's appearance? Please describe your observations.

Yes		Maybe		No	

9. Did you find the presence of the avatars in EER-Tutor useful? Please elaborate.

Yes		Maybe		No	

10. Is there something you didn't like or found frustrating about the avatars and EER-Tutor in general?

11. What did you like about the avatars and EER-Tutor in general? Please elaborate.

12. Do you have any suggestions for improving EER-Tutor and its avatars?

APPENDIX B

Evaluation Study Questionnaire

The following is a copy of the questionnaire that was used in the evaluation study described in Section 5.1.

Evaluation Study of EER-Tutor with Avatars

Thank you for participating in our evaluation study—your feedback is crucial to our current and future research. This questionnaire is anonymous and if you wish, you may at any time withdraw from participation, including withdrawal of any information you have provided. However, by completing the questionnaire, you indicate your consent for publication of the generalised results of our research findings.

Login name:		Gender:	M / F	Age:	

1. How would you rate your expertise in EER modelling prior to this session? (please circle one answer)

(a) Novice learner	(b) Confident user	(c) Expert

2. How much did you learn about EER Modelling from using EER-Tutor? (please circle one value)

Nothing				Very much
1	2	3	4	5

3. How much did you enjoy the avatars' presence in EER-Tutor? (please circle one value)

Not at all				Very much
1	2	3	4	5

4. Which avatar did you chose when you first logged in? (please circle one answer)

(a) David	(b) Callie	(c) Mark	(d) Diane

5. Why did you choose this avatar? Please explain.

6. Did you switch to a different avatar in the process of interaction with EER-Tutor? If yes, which ones did you choose and why? (circle more then one if needed and please comment)

(a) David	(b) Callie	(c) Mark	(d) Diane	(e) No, I didn't switch

P.T.O.

7. Did you notice any emotional changes in the avatar's appearance? Please state your observations.

Yes		Maybe		No	

8. Did you think the behaviour of the animated character was appropriate and adequate for an avatar? Please elaborate.

Yes		Maybe		No	

9. Did you find the presence of the animated character in the role of an avatar useful? Please elaborate.

Yes		Maybe		No	

10. What did you like about the avatars? Please elaborate.

11. Is there something about the avatars you didn't like or found frustrating? Please elaborate.

APPENDIX C

NZCSRSC 2007 Article

The following article was presented at the 5th New Zealand Computer Science Research Student Conference that was held at the Waikato University in Hamilton, in April 2007.

Intelligent Tutoring Systems Respecting Human Nature

Kon ZAKHAROV[1], Antonija MITROVIC[1], Lucy JOHNSTON[2]

[1]Intelligent Computer Tutoring Group, [2]Department of Psychology
University of Canterbury, Christchurch, New Zealand
{Konstantin.Zakharov, Tanja.Mitrovic, Lucy.Johnston}@canterbury.ac.nz

Abstract: The current level of development in Intelligent Tutoring Systems (ITS) ensures successful cognitive support. However, a number of studies suggest that learning outcomes are significantly influenced by a complex interaction between cognitive and affective states of learners. Little research has been done to investigate the effectiveness of learning with the help of affect-aware ITSs. Recently used approaches to affect recognition rely on facial feature tracking and physiological signal processing, but there is no clear winner among them because of the complexity and ambiguity associated with the task and the low-level data interpretation. The goal of our project is to develop a robust way of affect recognition for creating affect-aware pedagogical agents with the view to improve learners' engagement, motivation and learning outcomes.

1 Introduction

The semantic component of social interaction, most frequently expressed as speech, is often accompanied by the affective (otherwise known as emotional) component of social interaction; this is considered equally or sometimes even more important then the semantic component [1]. Although people are not always aware of how their language, posture, facial expression and eye gaze convey their emotions, these underpin people's interaction and navigation in the social world [2]. People also have a propensity to interact with computers in a social way, as if they were other people [3]. This undermines the idea that computers are merely neutral tools and emphasises the importance of the social relationships that can and will develop between a computer and a user. A better understanding of these relationships is essential for building of smarter tools for learning. For tasks fundamentally social in nature, a failure to include the emotional component in human-computer interaction (HCI) is potentially trimming the bandwidth of the communication channel. By ignoring human nature, computers force their users to act as if they too were machines; this contradicts the main assumption guiding HCI research [4]: "*People should not have to change radically to 'fit in with the system'; the system should be designed to match their requirements*".

The rest of this paper presents a Master of Science thesis project aimed at enabling an Intelligent Tutoring System (ITS) with affective awareness. Section 2 introduces prior research underlying our work. Section 3 outlines our approach and current state of our project. Section 4 concludes the paper with the plans of our further research efforts.

2 Background Research

Researchers have been grappling with the question of what appropriate behaviour is on the part of an interactive learning environment. Since etiquette is highly context-dependent, what may be appropriate in one situation, may be inappropriate in another. Frequently mentioned in HCI solutions to avoid a user being swamped by negative affect include either (a) trying to determine and fix the problem causing the negative feelings, and/or (b) pre-emptively trying to prevent the problem from happening in the first place [5]. Etiquette considerations in educational HCI are complicated by the fact that learning from a computer is not just about ease of use. Learning can be frustrating and difficult because it implies exposing learners' errors in thinking and gaps in their knowledge. Therefore, there are some fundamental differences between general HCI etiquette and etiquette for educational HCI. Cognitive psychology theory of learning from performance errors suggests that errors are an inseparable component of a learning process [6]. Error correction, in fact, has a critical significance for the improvement of future performance.

The foundation of our project is the research on applications of AI in Education. Learners vary in their amount of prior knowledge, motivation, learning style, natural pace and working memory capacity. Consequently, a uniform predefined instructional sequence can not provide the optimal learning environment for all learners. Educational research results published over two decades ago single out one-to-one tutoring as the most effective model of instruction [7]. Its success is based on the ability of human tutors to adjust their feedback based on their interaction with the learner. Real-life human tutors under one-to-one tutoring conditions are aware of the learner's cognitive state; this aspect of tutoring has been successfully captured by ITSs.

2.1 Intelligent Tutoring Systems

ITSs are task-oriented problem-solving environments designed for learning support in specific instructional domains. The capability of individualised instructions in ITSs hinges on the interaction of the target knowledge domain model, usually described as the expert module, and the model of the learner's knowledge, usually described as student model. Student models are maintained and updated by the ITS through interaction with the student; the ITS infers the relevant changes to the mental model of the target domain knowledge. There have been many approaches to student modelling including overlay models, perturbation models, models based on machine learning, and more recently, constraint-based models [8]. ITSs are known to improve learning performance by 0.3-1.0 standard deviations in a variety of target knowledge domains. For example, SQL-Tutor, an ITS for teaching Structured Query Language (SQL) for databases, improves performance by 0.65 standard deviations in just two hours of interaction with the system [9]. Atlas, a tutoring system for teaching Physics, improves performance by 0.9 standard deviations [10]. Between 20 and 25 hours of interaction with SHERLOCK, a tutor for technical troubleshooting in avionics, are equivalent to four years of on-the-job experience [11].

Constraint-based modelling (CBM), the student modelling approach we rely on in our research, arises from Ohlsson's theory of learning from performance errors [6]. A

CBM model represents domain knowledge as a set of explicit constraints on correct solutions in that domain [12]. At the same time, constraints implicitly represent all incorrect solutions. In this way, constraints partition all possible solutions into correct and incorrect solutions. There have been a number of constraint-based tutors developed within the Intelligent Computer Tutoring Group, at the University of Canterbury; one of the examples is EER-Tutor – a system for teaching the skill of Enhanced Entity-Relationship (EER) data modelling [13].

2.2 Affective and Physiological Processes

Emotions are described as psychological states or processes that function in the management of goals. An emotion is typically elicited by evaluating an event as relevant to a goal; it is positive when the goal is approaching and negative when progress towards the goal is impeded. Literature on emotion theory points out that negative affective states characterised by increased levels of adrenaline and other neurochemicals coursing through the body, diminish abilities with respect to attention, memory retention, learning, creative thinking, and polite social interaction [14].

There are two major theoretical approaches to the study of emotion: dimensional and categorical. Theorists who use the categorical approach to emotion attempt to define discrete categories or types of emotion [15]. Research in this area suggests that there are a number of basic emotions (estimates range from three to more than 20) which combine to produce all the emotional states which people experience. The dimensional approach conceptualises emotion as having two or perhaps three basic underlying dimensions along which the entire range of human emotions can be arranged [16]. The most common dimensions are valence (which ranges from happy to sad) and arousal (which ranges from calm to excited). Research using the dimensional approach has shown that emotions elicited by pictures, television, radio, computers and sounds can be mapped onto an emotional space created by the arousal and valence axes. Emotions involve multiple responses and thus is it common to group them into three broad categories: overt acts of behavioural sequences, emotional language and physiological reactions. In our research we focus our attention on the physiological reactions, characterised by changes in the somatic muscles (regulating voluntary mevement) and in the viscera (internal bodily organs, like heart, liver or intestine). Physiological and behavioural reactions to affective stimuli significantly correlate with judgements of affective valence and/or arousal [17]. In respect to valence, there is a high dimensional correlation between valence reports and electromyographic activity of corrugator[1] and zygomatic[2] muscles. There is also significantly greater heart rate deceleration for unpleasant pictures, and relatively greater peak acceleration for pleasant materials. The same literature sources describe electro-dermal activity as a useful measure of arousal.

---

[1] The corrugator muscles are responsible for a lowering and contraction of the brows, a facial action to be an index of distress, associated with unpleasant affective stimuli.

[2] Activity of zygomatic muscle is involved in the smile response. Zygomatic activity increases for pleasant stimuli, is greatest for stimuli high in affective valence.

2.3 Affective Pedagogical Agents

Following in the vein of Topffer's law, "*All interfaces, however badly developed, have personality*" [2], recent research based on the Computers as Social Actors (CASA) paradigm explores ways of broadening the communication channel between people and computers with a view to improving the effectiveness of computer-based educational environments. This research relies on the use of affective pedagogical agents (APAs) which act as a medium for delivering feedback from a computer to its users [18]. APAs are anthropomorphic software characters capable of expressing human-like behaviours and emotions; they are known to enhance the social view/interaction with computers [19]. Pedagogical agents in instructional environments draw upon human-human social communication scripts by embodying observable human characteristics such as the use of gestures and facial expressions. Several studies show that animated agents improve students' learning, engagement and motivation [2, 18]. STEVE, ADELE and Cosmo and are just a few examples of pedagogical agents described in ITS literature. STEVE (Soar Training Expert for Virtual Environments) teaches students how to perform procedural tasks, such as operating or repairing complex devices [18, 20]. ADELE (Agent for Distance Education: Light Edition) is designed to support students solving exercises delivered over the World Wide Web [21]. In the application of a case-based clinical diagnosis, ADELE can highlight interesting aspects of the case, as well as monitor and provide feedback as the student works through a case. Cosmo inhabits the Internet Protocol Adviser, which is a learning environment for the domain of Internet packet routing [19]. He provides advice to learners as they decide how to ship packets through the network to the specified destination.

3 Our Approach and Project Status

The goal of our project is to develop an affect-aware pedagogical agent, integrate it with EER-Tutor and conduct an evaluation study. Our hypothesis is that the effectiveness of computer-mediated learning environments will improve from recognition of the affective state of their users. Incorporating analysis of affective state in the synthesis of feedback can elevate the interaction with the learner to a new level and make a difference not only in the learner's perception of the interaction, but in the learning outcomes as well.

3.1 Pedagogical Agent

We have developed a pedagogical agent and integrated it into EER-Tutor. During the persona design process, we tried to determine what kind of affective displays an affective agent should offer in a learning context in order to support the learner's determination in the face of the inevitable stress, anxiety and frustration involved in learning. One simple rule of thumb suggested by Bickmore [1] is to apply what has been found appropriate for human-to-human interaction (HHI) to the design of educational HCI. Klein et al. [14] identify two types of support for emotion regulation.

First, passive support is used to manipulate moods without necessarily addressing or discussing emotions themselves. Media, activities, food and other substances fall into this category; interactions with people can fall into either category. In contrast, active support occurs when people discuss or otherwise directly address their emotions as a means of managing them.

We aimed for the agent to be able to acknowledge the leaner's emotions indirectly through its emotional appearance, while trying to keep the learner focused on the task at hand. Thus we designed an agent which would express solidarity with the user at all stages of the interaction – it will cheer with the users' success, be sympathetic with the user facing difficulties and keep company to the user in neutral situations. Similar agent behaviour was earlier adopted and implemented in the work of Lester et al [19].

While in our case, the agent does not yet have a way of determining users' affective state, prior research shows that in a learning context affective state can be indexed on the basis of cognitive state [22]. EER-Tutor maintains long and short-term student models; the state of the student model can be used to index the student's affective states. In our agents' design however, affective logic does not directly rely on the student model. Instead, the logic relies on session history, which includes the history of a wide variety of user actions.

The agents' implementation is split between the server-side and client-side of EER-Tutor. The server-side carries the agents' affective logic and controls the agents' behaviour, while on the client-side the agent appears to the users as a "talking head" with an upper body, embedded in EER-Tutor's work-space. The agent figures have been designed with the help of PeoplePutty toolkit; the web browser displays the agent with Haptek[3] player plug-in. Haptek's character affective appearance is controlled by a number of parameters called switches. Some switches, for example, control lips and eyebrows positions. Haptek characters communicate with the server through AJAX requests. Figure 1 shows the EER-Tutor workspace with a male agent seen on the right-hand side above the feedback pane. The screenshot shows the state of the work-space immediately after a solution submission.

In July 2006, we conducted a formative study aimed at assessing learners' perception, expectations and response to the agents. The general response to the agents was positive – 75% rated the agents as a useful feature. At the same time, half of the participants who thought the agent's presence was unnecessary rated audible narration as useful. Overall, the participants were enthusiastic about narration – 50% stated that narration was the most helpful feature, because it made it possible for them to keep their eyes on the diagram and begin correcting errors while listening to the narration. Participants commented that this helped save time and enabled them solve problems faster.

3.2 Affect Measurement

In our research we have adopted the dimensional approach, since continuous nature of valence and arousal is more suitable for our method of measurement. In order to make EER-Tutor aware of the emotional state of its users we will use a set of sensors from

[3] http://www.haptek.com/ – Haptek Inc., PeoplePutty is a product of Haptek.

Thought Technology[4] to capture the physiological data. We have acquired four sensors for reading physiological signals: blood volume pulse sensor, galvanic skin response sensor, surface electromyography sensor and respiration sensor. In order to convert the analogue signal from the sensors into digital signal we have acquired a data acquisition card (DAQ card) and implemented a prototype module for controlling the DAQ card and sending data to EER-Tutor server.

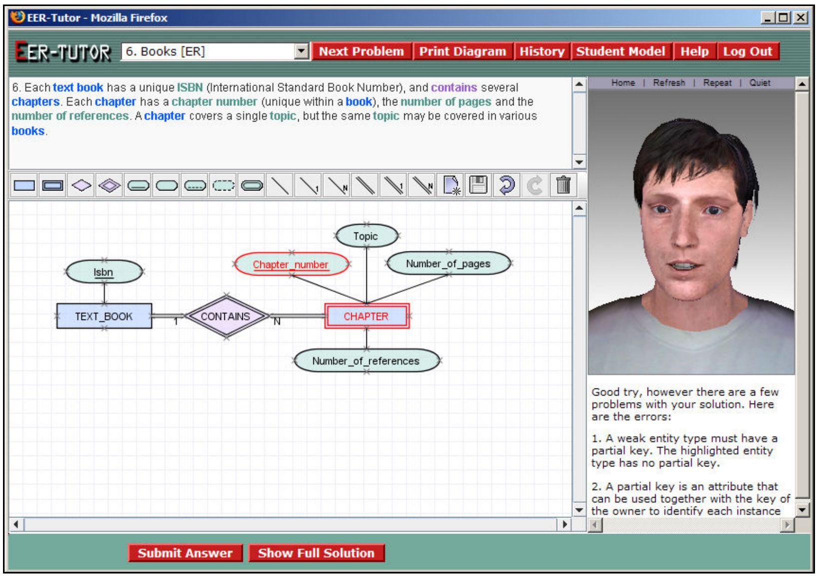


Figure 1. EER-Tutor work space with an agent.

3.3 Facial Feature Tracking

Literature on physiological data processing indicates that the EMG signal is not always reliable unless the top layer of dead skin cells is removed with abrasive cream before the application of the sensor electrodes. This can be perceived as quite intrusive by the experiment participants. As an alternative to EMG we have been working on a facial feature tracking application. Positive affective valence, for example, can be read as the distance between the corner of the mouth and the outer corner of an eye on one side of a person's face. Out of the numerous facial feature tracking approaches and implementation descriptions we have chosen a hybrid approach based on a Haar Classifier for face region detection [23] and feature extraction through common image processing techniques, such as edge detection, adaptive thresholding and integral projections [24]. In our implementation we used the Intel's OpenCV library[5]. Figure 2 shows a video stream frame with the important facial features detected with our prototype application: pupils, eyes and mouth corners. Next we will extend the application

---

[4] http://www.thoughttechnology.com – Thought Technology Ltd.

[5] http://www.intel.com/technology/computing/opencv/ – Open Source Computer Vision Library.

to enable retrieval of eyebrow positions. Before the facial feature application is integrated into EER-Tutor, we will run a brief evaluation of our implementation.

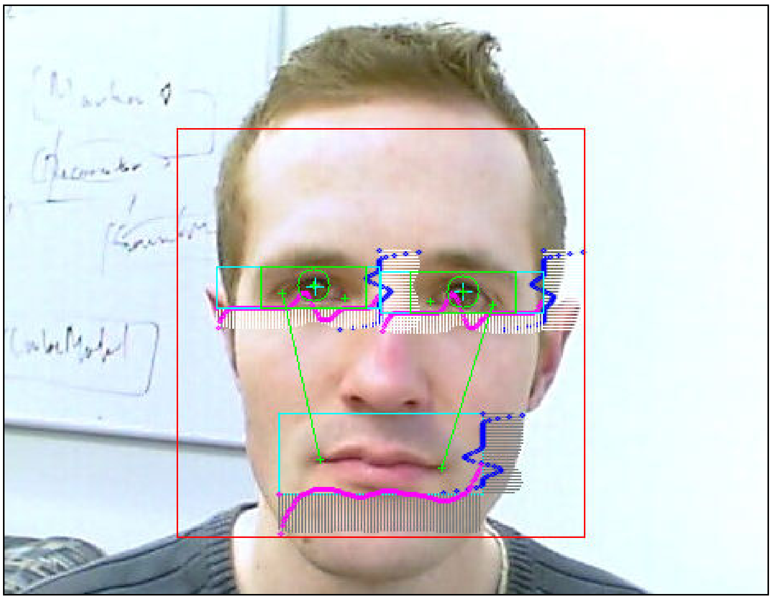


Figure 2. Retrieval of facial features with integral projections.

4 Future Work

In our future work, we will extend our system to identify students' affective states via real-time facial feature tracking and physiological sensors. Incorporating data from the sensory channel and changing the agents' persona rules will give the agents a finer level of affective awareness. This will be followed by a larger evaluation study to asses the agents' impact on users' view of interaction and learning with EER-Tutor.

References

1. Bickmore, T.W.: Unspoken Rules of Spoken Interaction. Communications of the ACM **47** (2004) 38-44
2. Mishra, P., Hershey, K.A.: Etiquette and the Design of Educational Technology. Communications of the ACM **47** (2004) 45-49
3. Reeves, B., Nass, C.: The Media Equation. How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge Publications (1996)
4. Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., Carey, T.: Human-Computer Interaction. Addison-Wesley, New York (1994)
5. Norman, D.: The Psychology of Everyday Things (a.k.a. The Design of Everyday Things). Basic Books, New York (1988)
6. Ohlsson, S.: Learning From Performance Errors. Psychological Review **103** (1996) 241-262

7. Bloom, B.S.: The 2 Sigma Problem: The Search for Method of Group Instruction as Effective as One-to-One Tutoring. Educational Researcher **13** (1984) 4-16
8. Ohlsson, S.: Constraint-Based Student Modelling. In: McCalla, G.I. (ed.): Student Modelling: The Key to Individualized Knowledge-based Instruction, Vol. 125. Springer-Verlag GmbH, Berlin (1994) 167-189
9. Mitrovic, A., Mayo, M., Suraweera, P., Martin, B.: Constraint-based Tutors: a Success Story. In: Ali, M. (ed.): Proceedings of the 14th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, Vol. 2070. Springer-Verlag GmbH, Berlin (2001) 931-940
10. Freedman, R.: Atlas: A Plan Manager for Mixed-Initiative, Multimodal Dialogue. International Journal of Artificial Intelligence in Education, Workshop on Mixed-Initiative Intelligence (1999)
11. LaJoie, S.P., Lesgold, A.: Apprenticeship Training in the Workplace: Computer-coached Practice Environment as a New Form of Apprenticeship. Machine-Mediated Learning **3** (1989) 7-28
12. Mitrovic, A., Ohlsson, S.: Evaluation of a Constraint-based Tutor for a Database Language. International Journal of Artificial Intelligence in Education **10** (1999) 238-256
13. Zakharov, K., Mitrovic, A., Ohlsson, S.: Feedback Micro-engineering in EER-Tutor. In: Breuker, J. (ed.): Proceedings of the 12th International Conference on Artificial Intelligence in Education, Vol. 125. IOS Press (2005) 718-725
14. Klein, J., Moon, Y., Picard, R.W.: This Computer Responds to User Frustration: Theory, Design, and Results. Interacting with Computers **14** (2002) 119-140
15. Ortony, A., Turner, T.J.: What's Basic About Basic Emotions? Psychological Review **97** (1990)
16. Lang, A., Dhillon, P., Dong, Q.: Arousal, Emotion, and Memory for Television Messages. Journal of Broadcasting and Electronic Media **38** (1995) 1-15
17. Bradley, M.M., Lang, P.J.: Measuring Emotion: Behavior, Feeling and Physiology. In: Nadel, L. (ed.): Cognitive Neuroscience of Emotion. Oxford University Press, New York (2000) 242-276
18. Johnson, W.L., Rickel, J.W., Lester, J.C.: Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments. International Journal of Artificial Intelligence in Education **11** (2000) 47-78
19. Lester, J.C., Towns, S.G., Fitzgerald, P.J.: Achieving Affective Impact: Visual Emotive Communication in Lifelike Pedagogical Agents. International Journal of Artificial Intelligence in Education **10** (1999) 278-291
20. Rickel, J., Johnson, W.L.: Intelligent Tutoring in Virtual Reality: A Preliminary Report. Proceedings of the 8th International Conference on Artificial Intelligence in Education. IOS Press (1997) 294-301
21. Shaw, E., Ganeshan, R., Johnson, W.L., Millar, D.: Building a Case for Agent-assisted Learning as a Catalyst for Curriculum Reform in Medical Education. Proceedings of the 9th International Conference on Artificial Intelligence in Education. IOS Press (1999) 509-516
22. Conati, C., Maclaren, H.: Evaluating a Probabilistic Model of Student Affect.: Proceedings of 7th International Conference on Intelligent Tutoring Systems, Vol. 3220/2004. Springer, Brazil (2004) 55-66
23. Viola, P.A., Jones, M.J.: Rapid Object Detection using a Boosted Cascade of Simple Features. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1. IEEE Computer Society, Kauai, HI, USA (2001) 511-518
24. Mateos, G.G.: Refining Face Tracking with Integral Projections. Lecture Notes In Computer Science **2688** (2003) 360-368

APPENDIX D

AIED 2007 Article

The following article has been published in the proceedings of the 13th International Conference on Artificial Intelligence in Education (pages 59–66) held at Marina del Ray Marriott, Los Angeles in July 2007.

Pedagogical Agents Trying on a Caring Mentor Role

Konstantin ZAKHAROV[a], Antonija MITROVIC[a], Lucy JOHNSTON[b]
[a] *Intelligent Computer Tutoring Group,* [b] *Department of Psychology*
University of Canterbury, Christchurch, New Zealand
{konstantin.zakharov, tanja.mitrovic, lucy.johnston}@canterbury.ac.nz

Abstract: We describe the design and evaluation of an affective pedagogical agent persona for Intelligent Tutoring Systems. The goal of our research was to develop an agent embodying a persona of a caring mentor interested in the learner's progress. The agent's behaviour is guided by a set of rules that are triggered by the states of the session history. Four agents were integrated with EER-Tutor for a formative evaluation study. The mentor persona secured strong rapport with the users; the audible narration was seen as a strong feature of the agents.

Introduction

The semantic component of social interaction, most frequently represented as speech, is often underpinned by the affective component, which can be expressed through speech and non-verbal displays such as gestures, posture, facial expression and eye gaze [10]. People have a propensity to transfer the social view of interaction onto their interaction with electronic media. As described by Reeves and Nass [18], people tend to view electronic media in a social way, as if they were other people. However, computers since their early days have been implicitly designed without awareness of the affective communication channel. Computers respond to people as if they too were computers, thus forcing people to adjust to computer protocols and interact on a sub-human level, which contradicts the main assumption guiding Human-Computer Interaction (HCI) research: *"People should not have to change radically to 'fit in with the system' – the system should be designed to match their requirements"* [17].

This inherent lack of affective fit is particularly significant in the area of Intelligent Tutoring Systems (ITS), because research suggests a strong interaction between cognitive and affective processes in human mind. For example, there is a reliable correlation between one's emotional state, memory capacity and motivation [6, 18]. Consequently, a failure to recognise affective processes might impose a serious limitation on interaction types which are fundamentally social in nature, because the affective component is often considered as important as semantic component [3]. Without considering the interaction between the cognitive and affective processes ubiquitous in human interaction, educational systems might never approach their full potential.

HCI does acknowledge the need to avoid negative affective states, such as frustration; common solutions to avoid user frustration include either (a) trying to determine and fix the problem causing the negative feelings, and/or (b) pre-emptively trying to avoid the problem from happening in the first place [7]. However, these approaches have a limited application in ITSs because of the fundamental differences between general HCI and educational protocols. Affective considerations in ITSs are more complex, because learning from a computer is not just about ease of use. It would be unrealistic to expect students to stop making errors during learning. Learning can be frustrating and difficult because it requires exposing learners' errors in thinking and gaps in their knowledge. Theory of Learning from Performance Errors [15] suggests that errors are an inseparable component of learning. Error detection followed by error correction, in fact, is vital for the improvement of future performance.

Recent research suggests using Affective Pedagogical Agents (APAs) in ITSs as a medium for delivering feedback to the users [8, 12]. This research draws on Allport's [1] classic definition of social psychology: *"The scientific investigation of how the thoughts, feelings and behaviours of individuals are influenced by the actual, imagined or implied presence of others"*; in accordance with this statement APAs are known to enhance the social view of interaction with ITSs. This undermines the idea that computers are merely neutral tools and emphasises the importance of the social relationships that can develop between a computer and a learner [14]. A better understanding of these relationships is essential to building smarter tools for learning.

We start by presenting our work on developing an affective persona for agents integrated with EER-Tutor, an ITS for developing Enhanced Entity-Relationship (EER) modelling skills [19, 21]. Section 2 outlines the experiment aimed at assessing learners' perception, expectations and response to the agents, and details the experimental results. Section 3 presents the conclusions and discussion of our research.

1. Developing Affective Pedagogical Agents

When designing the persona, we had to establish what kind of affective response an agent should provide in order to support the learner's determination in the face of the inevitable stress, anxiety and frustration involved in learning. One simple rule of thumb suggested by Bickmore [3] is to apply what has been found appropriate for human-to-human interaction (HHI) to the design of educational HCI. People use a variety of methods to help manage their emotions, such as interacting with media and/or other people, engaging in sports or work, meditating or praying, using positive thinking, and consuming foods and substances such as alcohol. Klein et al. [9] identify two types of support for emotion regulation. First, passive support is used to manipulate moods without necessarily discussing emotions themselves. Media, activities, food and other substances fall into this category. In contrast, active support occurs when people discuss or otherwise directly address their emotions, as a means of managing them.

Out of possible instructional roles we chose the mentor role, because of positive influence on learning demonstrated in previous research [2]. At the same time, from the affective interaction point of view, passive affective support intuitively seems as adequately congruent with the role of a mentor. Consequently, we tried to create an agent with a persona of a mentor interested in learner's progress. We wanted the agent to acknowledge the leaner's emotions indirectly through its emotional appearance, while trying to keep the learner focused on the task at hand. Thus we designed an agent which expresses solidarity with the user – it will cheer with the users' success, be sympathetic when there are difficulties and keep company to the user in neutral situations. Similar agent's behaviour was earlier adopted in the work of Lester et al. [11].

Human mentors can pick up a lot of clues from non-verbal communication; with varying degrees of depth, the mentor is always aware of the learner's affective state and cognitive state. In combination, these two factors allow a human mentor to choose an appropriate affective response at each step. While in our case, the agent does not have a way of determining users' affective state, prior research shows that in a learning context affective state can be indexed on the basis of cognitive state [4]. Cognitive Theory of Emotions states that the valence of one's emotional reaction depends on the desirability of the situation [16]. Thus we can assume that continuous lack of cognitive progress will be accompanied by a negative affective state, because the user will be dissatisfied with the state of the current task. Conversely, good progress will result in a positive affective state. In our approach, we rely on a simplified emotional model described by a single dimension – affective valence.

EER-Tutor maintains student models; the state of student model can be used to index the student's affective states. In our agents, however, affective logic does not directly rely on the student model. Instead, the logic relies on session history, which includes a wide variety of user actions. The rationale behind this approach is twofold. First, most user actions, such as login, problem selection and so on, may require a response or acknowledgement from the agent; in many cases such responses would not have to carry much affective load, although failing to respond in a similar HHI situation could be interpreted as lack of attention and care. Second, under certain circumstances seemingly neutral actions might indicate changes in the users' affective state and thus should be addressed by the agent. For example, repeated submissions of the same solution to a problem might indicate a negative affective state, such as frustration. In general, we assume that any repeating actions should not go unnoticed.

1.1. Agents' Appearance

EER-Tutor [19, 21] is a web-based ITS whose server carries the application load which includes running the constraint-based modelling engine, applying pedagogical logic, maintaining student models and hosting EER-Tutor curriculum. The client is implemented as a set of AllegroServe[1] dynamic HTML pages and a Java applet providing drawing tools for creating solutions; the interface only accepts user input and delivers feedback. Similarly, the agents' implementation is also split between the server and client. The server carries the agents' affective logic and controls the agents' behaviour, while on the client-side the agent appears to the users as a "talking head" with an upper body, embedded in EER-Tutor's work-space. The agent figures have been designed with the help of PeoplePutty[2] toolkit; the web browser displays the agent with Haptek[3] player plug-in. Haptek's character affective appearance is controlled by a number of parameters called switches. Some switches, for example, control lips and eyebrows positions. Haptek characters communicate with the server through AJAX[4] requests. The communication process revolves around retrieving updated values for the switches along with appropriate narration fragments. Figure 1 shows EER-Tutor workspace with a male agent above the feedback pane. The screenshot shows the state immediately after a solution submission. In this case, the submitted solution was incorrect: the user mistakenly defined the *Chapter_number* attribute as a key attribute, instead of making it a partial key attribute. The erroneous diagram component, the *Chapter_number* key

[1] http://allegroserve.sourceforge.net – AllegroServe web application server
[2] http://www.haptek.com/products/peopleputty/ — Haptek's PeoplePutty SDK
[3] http://www.haptek.com/ — PeoplePutty is a product of Haptek
[4] http://www.adaptivepath.com/publications/essays/archives/000385.php

attribute, is shown in red. In general, the error messages presented by the agent refer the user to the errors in the diagram; in our example the first error message says: *A weak entity type must have a partial key. The highlighted entity type has no partial key*. This approach to engaging the user in active information processing is supported by Mayer's first principle of multi-media design, *Multimedia effect for transfer*, which states that providing the learner with a narration results in better learning when it is supplemented by corresponding visual aids [13].

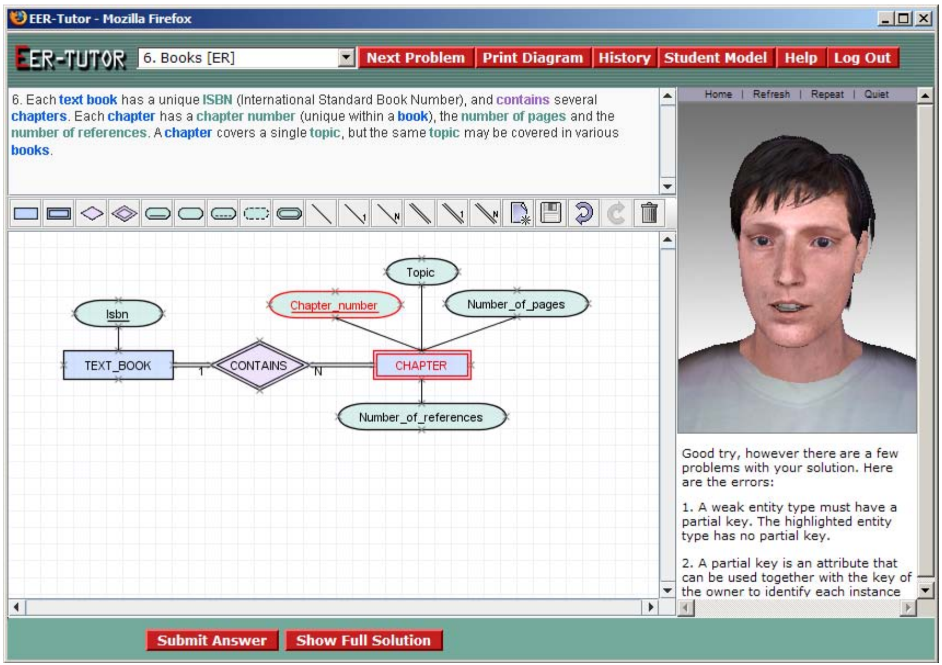


Figure 1. The view of EER-Tutor's work-space with the male agent

Figure 2. The agents designed with the PeoplePutty toolkit.

Along with feedback in audible form, the user is also presented with the most recent feedback messages in the textual form. Even though the sixth principle of multi-media design, *Redundancy effect for transfer*, states that such redundancy negatively affects learning [13], we consider that the EER-Tutor context justifies such redundancy, because the complexity of the domain knowledge and abundance of information may be difficult to process without textual representation. For example, when the user makes several mistakes, the corresponding messages may help the user to stay focused on the task instead of wasting efforts on remembering multiple feedback messages.

Haptek figures rely on Microsoft's Speech Application Programming Interface (SAPI), SAPI4 and SAPI5-compatible Text-to-Speech (TTS) engines to display realistic mouth movements while producing verbal narrations. In order to enhance the agents' realism, we obtained two reportedly higher quality Cepstral[5] voices – one male and one female, instead of using the default TTS voices supplied with Windows XP. Figure 2 shows the four agents we created – two male and two female characters. We chose a generic youthful appearance in order to remain consistency with agent's role of a mentor' since the mentoring group is traditionally represented by younger people.

1.2. Agents' Behaviour

EER-Tutor is driven entirely by user-generated interface events, such as submissions of solutions. Every interface event results in a corresponding request type being sent to the server. When the server processes the request and the client receives the response, the agent control script requests the server for the updates of the agents' emotional display and verbal feedback to be returned in response to the users' most recent action. At this stage the agent's behaviour rules, described later in this Section, are matched against the session history. On completion of this process, the server sends the response defined by the selected rule. In this way, even though the Haptek character is completely unaware of the users' actions and student model, to the user it appears that the agent is actively observing their actions.

The agent control script also runs an uninterrupted loop (repeating at the 1.5s intervals) continuously querying the server for affective appearance and narration updates even in the absence of interface events. While user actions may colour the agent's affective appearance, the agent is capable of emotional self-regulation mimicking human behaviour. The agent's affective state tends to gradually return to neutral, so if the agent momentarily may appear very upset or happy, after a few minutes the agent inevitably "settles down" even in the absence of affect-provoking actions of the user.

The agent's persona is controlled by a set of rules, each of which corresponds to a unique state of the session history and produces a certain reaction from the agent; every rule consists of a pattern (defined in Allegro Prolog[6]) to match a certain history state, a number of equivalent alternative verbal responses and the affective state update command. We have defined just under 20 rules, which make the agent respond to a variety of situations. The rules can be roughly divided into three categories: affectively-neutral rules, rules causing current affective valence to nudge slightly towards the positive end, and rules resulting in a move towards the negative end. The values of positive/negative affect changes are defined individually in each rule. When dominated by positive affect, the agent smiles; negative affect makes the agent appear sad. The following are examples of situations, with descriptions of corresponding responses and affective changes in the agent's appearance:

---

[5] http://www.cepstral.com/ — Cepstral Text-to-Speech engines.
[6] http://www.franz.com/products/prolog/ — Allegro Prolog – integrated extension for Common Lisp.

First-time login – agent introduces itself and welcomes the user with an extended message; the agent's affective state is changed so that the agent smiles.

Selection of a new agent – agent introduces itself and smiles.

Submission with a few errors – agent starts with a friendly introductory phrase and reads the errors; the agent's affective state is nudged slightly towards the negative side. If this attempt happens to come in a series of unsuccessful attempts, the agent will have an unhappy/concerned look as the result of its affective valence being dampened by the repeated triggering of this rule. However, if this is a single unsuccessful attempt, the change in the agent's appearance will be subtle.

Submission with a single error – the agent starts an encouraging phrase and presents the error; affective valence is nudged slightly to the positive direction.

Correct solution – the agent congratulates the user with solving the problem in one attempt and while happily smiling suggests that the user try another problem.

There are also rules defining behaviours for repeated submissions of identical solutions, repeated abandoning of problems, logging out etc. Precedence of session states is implicitly encoded in the order of the rules; during the rule matching process, the first match results in the corresponding action on behalf of the agent. Both male and female agents are guided by the same set of rules.

2. The Experiment

In July 2006, we recruited 20 volunteers (16 male and 4 female) from the third and fourth year students at the University of Canterbury. All participants were familiar with the EER model. The disparity between the users' levels of EER expertise was not an important factor, because the study was aimed at collection of qualitative data only. The experiment was carried out as a series of individual sessions. At the start of a session, a participant was provided with a verbal description of the task. Participants were expected to spend 45 to 60 minutes solving problems. The participants were asked to choose an agent before they were able to navigate to the workspace. During the session, the participants were free to choose a different agent at any time. At the end, the participants were required to fill out a questionnaire, for assessing their experience of learning with the agent and their perception of the agent's persona.

The first-choice preference was given to female agents (14 vs. 6) irrespective of the participants' sex. Among male participants only 38% chose a male agent, while all female participants chose a female agent at the start of the session. On the basis of questionnaire responses, it appears that male agents won in their effectiveness and overall impression on the participants. The six participants who chose a male agent reported they enjoyed EER-Tutor more, giving it a rating of 4.1 ($\sigma = 0.4$) out of 5, compared to the rating of 3.7 ($\sigma = 0.7$) by the participants who chose the female agent. The average learning success (rated on the scale of 1 to 5) reported by participants with male agents is higher then the female agents' group: 3.5 ($\sigma = 0.5$) vs. 3.1 ($\sigma = 0.7$). This learning estimate difference is consistent with the actual learning outcomes, measured by the number of constraints learned during the session: 3.3 ($\sigma = 4.8$) for male agents vs. 2.1 ($\sigma = 2.3$) for female agents. However, the small number of participants does not allow us to treat these results as being statistically reliable.

We attribute the apparent association between higher ratings and better learning outcomes for the male agents to the difference in the quality of male and female Cepstral TTS voices used. The male TTS voice sounds more human-like then the female voice; even though the participants were not specifically asked to rate the quality of TTS voices, 50% of the participants who used female agents stated that they were frustrated by or disliked that voice. No such comments were made about the male voices. Our interpretation of the voice quality effect is supported by the Cognitive Load Theory [20], which states that processing unnatural-sounding or machine-like voices imposes higher cognitive demands on learners than natural voices do.

The general response to the agents was positive – 75% rated the agents as a useful feature. At the same time, half of the participants who thought the agent's presence was unnecessary rated audible narration as useful. Overall, the participants were enthusiastic about narration – 50% stated that narration was the most helpful feature, because it made it possible for them to keep their eyes on the diagram and begin correcting errors while listening to the narration. Participants commented that this helped save time and enabled them solve problems faster.

Both male and female agents' responses and appearance were rated as adequate by around 90% of both male and female participants. Participants' feedback indicates that the agents' persona was appreciated for not being "in-your-face". Some users commented that they liked the "feeling of company" and "human presence" created by the agents. Some users made comments suggesting that the agents should be more active and versatile in their behaviour; others commented on being distracted by the agents' movements, such as blinking and breathing, in the background.

The study highlighted the need for the agents' behaviour to appear more natural. Many comments suggested enabling the user to have greater control over the agents' behaviour and voice properties. For example, some users wanted the agents to be more dynamic, even "have them fly over the workspace" and "make emotional changes more apparent", while others wanted the agents to be less obtrusive; some participants said they would like to be able to control the speed of the narration to make it faster or slower. Some users (around 25%) did not pay attention to the agents' affective appearance, but a small proportion (around 15%) commented on the agents "flipping out" and getting "too emotional" too soon. One participant commented that he or she felt "emotionally blackmailed and manipulated by the agent".

3. Conclusions and Future Work

We described an approach to modelling affective pedagogical agent persona and reported evaluation results. The persona was designed as a set of rules corresponding to the interaction states which require responses from a pedagogical, affective or social point of view. Our approach is characterised by simplicity, flexibility and scalability. The persona can be developed incrementally by adding new rules; the larger the number of rules, the more versatile and interactive the persona is. With this approach it is possible to develop different sets of rules defining different types of personas.

We created a mentor-like persona, mimicking a somewhat informal HHI interaction with passive emotional support. Four animated characters were designed aimed at soliciting perception of the agents' mentor persona. The results indicate a good outlook on the viability of the suggested approach and we will enhance our implementation in future work. Our agents' personae secured positive ratings during the evaluation; however, the lack of scale in the study and the confounding factor of difference between the quality of male and female TTS voices does not allow us to make conclusions about participants' preferences on the agent's sex and its effect on learning. We have learned that the quality of agents' voice is critical for maintaining the agents' rapport with the users and maintaining the flow of the learning process.

This study also reveals the breadth of user preferences when it comes to interacting with affective pedagogical agents and suggests that finding an ideal persona to "click" with all users is unrealistic. Some learners see the agent as a distraction, while others wish the agent to be more interactive and entertaining. Needless to say, affective agents need to respect human individuality in the style of interaction. Even though the study shows the agents were a well-received addition to EER-Tutor, the participants' comments indicate the need for making the agents more flexible. Audible narration was welcomed by most participants irrespective of their expectations of the agents' persona.

In future work, we intend to extend our system to identify students' affective states via real-time facial feature tracking and physiological sensors. Incorporating data from the sensory channel and changing the agents' persona rules will give the agent a finer level of affective awareness.

References

[1] G. W. Allport, *Attitudes*, in C. A. Murchison, ed., *A handbook of Social Psychology*, 1935, pp. 798-794.

[2] A. L. Baylor, Y. Kim, *Pedagogical Agent Design: The Impact of Agent Realism, Gender, Ethnicity, and Instructional Role*, in J. C. Lester, R. M. Vicari and F. Paraguaçu, eds., *Intelligent Tutoring Systems*, 2004, pp. 592-603.

[3] T. W. Bickmore, *Unspoken Rules of Spoken Interaction*, CACM, 47, 2004, 38-44.

[4] C. Conati, H. Maclaren, *Evaluating a Probabilistic Model of Student Affect*, *Intelligent Tutoring Systems*, 2004, pp. 55-66.

[5] C. Gama, *Metacognition in Interactive Learning Environments: The Reflection Assistant Model*, *Intelligent Tutoring Systems*, 2004, pp. 668-677.

[6] D. Goleman, *Emotional Intelligence*, Bantam Books, New York, 1995.

[7] J. Gratch et al., *Creating Interactive Virtual Humans: Some Assembly Required*, IEEE Intelligent Systems, 17, 2002, 54-63.

[8] W. L. Johnson, J. W. Rickel, J. C. Lester, *Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments*, Artificial Intelligence in Education, 11, 2000, 47-78.

[9] J. Klein, Y. Moon, R. W. Picard, *This Computer Responds to User Frustration: Theory, Design, and Results*, Interacting with Computers, 14, 2002, 119-140.

[10] H. Klingzing, T. Tisher, *Expressive nonverbal behaviors; A review of research on training with consequent recommendations for teacher education*, in J. Raths and L. Katz, eds., *Advances in Teacher Education*, Ablex Publishing, 1986.

[11] J. Lester et al., *Deictic and emotive communication in animated pedagogical agents*, Embodied Conversational Agents, 2000, pp. 123-154.

[12] J. Lester, S. Towns, P. Fitzgerald, *Achieving Affective Impact: Visual Emotive Communication in Lifelike Pedagogical Agents*, Artificial Intelligence in Education, 10, 1999, 278-291.

[13] R. E. Mayer, *Multimedia Learning*, Cambridge University Press, 2001.

[14] P. Mishra, M. Nicholson, S. Wojcikiewicz, *Does My Word Processor Have a Personality? Topffer's Law and Educational Technology*, Adolescent and Adult Literacy, 44, 2001, 634-641.

[15] S. Ohlsson, *Learning From Performance Errors*, Psychological Review, 103, 1996, 241-262.

[16] A. Ortony, G. Clore, A. Collins, *The Cognitive Structure of Emotions*, Cambridge University Press, 1988.

[17] J. Preece et al., *Human-Computer Interaction*, Addison-Wesley, New York, 1994.

[18] B. Reeves, C. Nass, *The Media Equation.* Cambridge Publications, 1996.

[19] P. Suraweera, A. Mitrovic, *An Intelligent Tutoring System for Entity Relationship Modelling*, Artificial Intelligence in Education, 14, 2004, 375-417.

[20] J. Sweller, J. van Merrienboer, F. Paas, *Cognitive Architecture and Instructional Design*, Educational Psychology Review, 10, 1998, 251-296.

[21] K. Zakharov, A. Mitrovic, S. Ohlsson, *Feedback Micro-engineering in EER-Tutor*, in J. Breuker, ed., *Artificial Intelligence in Education*, IOS Press, 2005, pp. 718-725.