Foundation of an affective tutoring system: learning how human tutors adapt to student emotion

Samuel Alexander* and
Abdolhossein Sarrafzadeh

Institute of Information and Mathematical Sciences,
Massey University,
Private Bag 102 904, NSMC, New Zealand
Fax: +64-94-418-181
E-mail: S.T.Alexander@massey.ac.nz
E-mail: H.A.Sarrafzadeh@massey.ac.nz
*Corresponding author

Stephen Hill

School of Psychology,
Massey University,
Private Bag 11-222, Palmerston North, New Zealand
Fax: +64-63-505-673
E-mail: S.R.Hill@massey.ac.nz

Abstract: The developing field of Affective Tutoring Systems (ATSs) has created a need to understand *how* such tutoring systems should adapt to the emotional state of students. To this end, an observational study of human tutors was conducted to learn how human tutors adapt to the affective state of students. This knowledge can then be used to implement the tutoring strategies of an ATS and is hence a critical foundation in its development. This paper presents the methodology and results of this observational study of human tutors and looks ahead to the future development of an animated pedagogical agent capable of detecting, expressing and *adapting* to emotion.

Keywords: Affective Tutoring Systems (ATSs); animated pedagogical agents; facial expressions; human tutors.

Reference to this paper should be made as follows: Alexander, S., Sarrafzadeh, A. and Hill, S. (2008) 'Foundation of an affective tutoring system: learning how human tutors adapt to student emotion', *Int. J. Intelligent Systems Technologies and Applications*, Vol. 4, Nos. 3/4, pp.355–367.

Biographical notes: Samuel Alexander received his BSc (Hons) in 2003 from Massey University in Palmerston North, New Zealand and is currently a Doctoral Candidate at Massey University in Auckland, New Zealand. His research interests are in the areas of intelligent tutoring systems and affective computing.

Abdolhossein Sarrafzadeh is a Senior Lecturer in Computer Science and the Leader of the Next Generation Intelligent Tutoring Systems Project at Massey University in Auckland, New Zealand. His research interests include affect-sensitive intelligent tutoring systems, vision-based emotion detection, human–computer interaction and knowledge-based systems. He is author of over 60 research publications and his current research focus is primarily in the area of affect-sensitive systems.

Stephen Hill is a Lecturer in Cognitive and Experimental Psychology at Massey University, Palmerston North, New Zealand. His PhD (University of Canterbury, 2001) critically evaluated the emerging embodied approach to cognitive science. His other research interests include philosophy of mind, evolutionary psychology, adaptive person perception and attention.

1 Introduction

Many researchers now feel strongly that Intelligent Tutoring Systems (ITSs) would be significantly enhanced if computers could adapt according to the emotions of students (Alexander, 2004; Kort et al., 2001; Picard, 1997). This idea has spawned the developing field of Affective Tutoring Systems (ATSs): ATSs are ITSs that are able to adapt to the affective state of students in the same ways that effective human tutors do (de Vicente, 2003; Sarrafzadeh et al., 2004). ATSs have a very short history: it seems that the term 'ATS' was first used only several years ago (Alexander et al., 2003; de Vicente, 2003), although the popular concept of an ITS adapting to perceived emotion can be traced back at least as far as Rosalind Picard's (1997) book *Affective Computing*. However, so far as the author is currently aware, no ATS has yet been implemented, although several groups are working towards this goal (Alexander, 2004; Kort et al., 2001; Litman and Forbes, 2003).

One of the main barriers in implementing an ATS is knowing how to adapt to the emotions of students, assuming that these emotions can be reliably identified. Therefore, to fill a gap in the psychology and education literature, it was decided to conduct an observational study of how human tutors adapt their tutoring based on the affective state of students. The ways in which human tutors adapt their tutoring according to their empathy with students can then be used as a platform on which to build the tutoring strategies of an ATS.

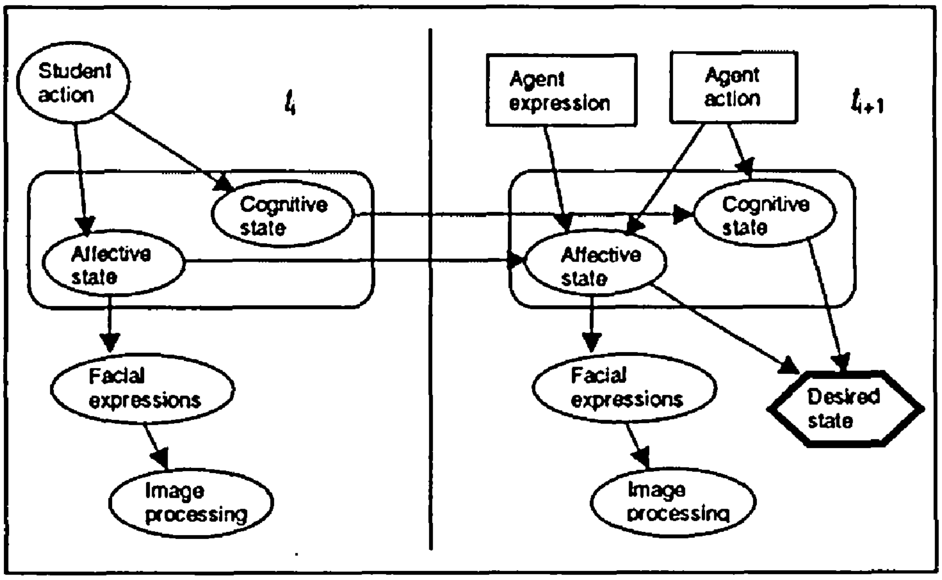
The main aim of this paper is to present the methodology and results of this study of human tutor empathy and to briefly discuss future work. These results include analysis of student and tutor emotional expressions and the probabilities of a tutor's response given student cognitive and affective state information over a sequence of interactions. Future work will be to map these results directly to the tutoring strategies module of an ATS that is currently being developed. The next section provides a background and rationale to the study.

2 Background

The overall aim of this research is to develop an ATS that is capable of recognising both the cognitive and affective state of students and of usefully adapting to this information. As well as recognising expressed emotions through automated image processing, the tutoring system will also be able to show emotions through an animated pedagogical agent. The aim of the ATS is summarised in Figure 1 in a model adapted from

Conati (2002): the left hand side of the diagram at time t_i represents the cognitive and affective state of the student immediately following a student action, where the affective state is identified by detecting the student's facial expression. At time t_{i+1} the system's animated agent then responds to the cognitive and affective state of the student by adapting both its tutoring and its own facial expression, with the intention of mapping the student's cognitive and affective states to a particular desired state.

Figure 1 Model of the ATS



2.1 *Animated pedagogical agents*

Though few, if any existing ITSs can recognise emotions, many ITSs have been developed that can show emotions through an animated pedagogical agent (Johnson et al., 2000; Prendinger and Ishizuka, 2004). Animated pedagogical agents are 'lifelike autonomous characters that cohabit learning environments with students to create rich, face-to-face learning interactions' (Johnson et al., 2000). Animated agents carry a persona effect, which is that the presence of a lifelike character can strongly influence students to perceive their learning experiences positively (van Mulken et al., 1998). The persona effect has been shown to increase learner motivation, especially in technical domains, although its overall benefits remain unclear (van Mulken et al., 1998). Assuming that the affective state of students can be reliably identified, animated agents are able to show timely empathy towards students through their own facial expressions and gestures.

2.2 *Identifying affective state*

There are several different ways that computers can attempt to identify the affective state of users. These can be divided into two main groups: methods that aim to detect emotions based upon their physical effects and methods that aim to predict emotions based upon understanding their causes. For example, methods that detect the physical effects of emotions are concerned with the affective state information carried through mediums such as facial expressions, gestures, speech and physiological responses like heart-rate and skin conductance (Picard, 1997).

On the other hand, a model that predicts a student's affective state based upon the causes of emotions is presented by Conati (2002). This model takes into account factors such as the personality and goals of the student, as well as a history of the interaction between the student and the tutoring system. The current research uses automated facial expression analysis to detect the affective state of the student, chiefly because a facial expression analysis system is being developed in-house at Massey University (Fan et al., 2003). However, as Conati (2002) points out, the best solution for identifying student emotion will most likely be reached by combining both types of methods: those that detect affective states and those that predict them. This remains an avenue yet to be explored.

2.3 Rationale for the observational study

However, even if an ATS could perfectly identify the affective state of students, it would still need to know what to do with this information before it could adapt its tutoring in a genuinely useful manner. As good human tutors *can* effectively adapt to the emotions of students, the most obvious way to learn about how to adapt to the affective state of students is to study human tutors.

The ways in which human tutors adapt to the affective state of students have yet to be explained. Therefore, the aim of this study is to take a step towards filling this gap in the literature to make it possible to design an ATS that can sensibly make use of affective state information. The results of this study will be directly applied in the development of the tutoring strategies module of an ATS.

Secondly, if the affect-based adaptations of the animated agent are based on human tutors then this should help to increase the believability of the agent. If by definition the agent is pretending to act like a human tutor, then the more human-like it appears to act (through appropriate emotional awareness and expressions), the more believable the agent will be. It is vital that agents are believable as this is the foundation of the persona effect (van Mulken et al., 1998). Therefore given a student's cognitive and affective state, the ATS will not only present the most appropriate tutoring material for these student states, it will do so via an especially believable animated agent that maximises the persona effect.

Also, a recent study by Ghijsen et al. (2005) has shed some light on the types and frequencies of facial expressions exhibited by students during interactions with a computer tutoring system. By comparing the types of facial expressions displayed in front of a computer with the results of the current study, it will be possible to tentatively gauge the similarities and differences in facial expressions displayed by students tutored by computers and students tutored by human beings. This will be very briefly discussed further below.

3 Methodology

3.1 Videoing students

The observational study of human tutors involved videoing several tutors as they tutored students individually. There were three tutors altogether – two of the tutors were teachers at the school where the research was undertaken and the third was a professional tutor.

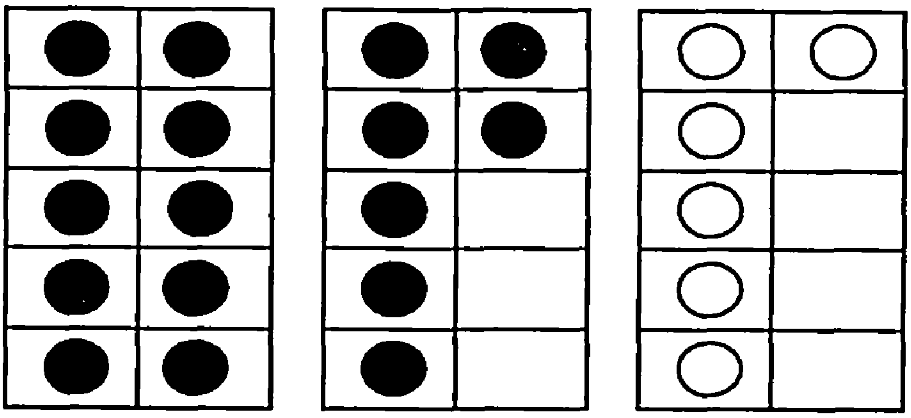
There were nine student participants, all of whom were 8- or 9-year-old students at a school in Auckland, New Zealand. Five participants were tutored by the professional tutor; the other two tutors tutored two students each.

Each participant was tutored for about 20 min, which generated a total of approximately 3 hr of video footage. The participants were selected solely on the basis of their mathematics ability, so that they would find the tutoring exercise that was used in the study to be challenging, but achievable. This exercise is discussed in Section 3.2.

3.2 Tutoring exercise

The domain that was chosen for the observational study was the concept of part-whole addition. The study used an existing exercise developed by the New Zealand Numeracy Project (2003) that encourages students to add numbers by transforming the initial equation to make the first addend up to the next 10. For example, 17 + 6 would become 17 + 3 (to make 20) + 3 = 23. Students learn this reasoning by manipulating tens frames and counters, as shown in Figure 2: in this example the student should move three counters from the tens frame on the right across to make the middle tens frame up to ten.

Figure 2 Tens frames and counters in the maths exercise



3.3 Coding the videos

To analyse the videos, a coding scheme was developed expanding on previous work by Person et al. (2003). This scheme was used to extract data from each tutoring video to describe the behaviours, facial expressions and expression intensities of students and tutors.

Each tutoring video was divided into several hundred clips, with each clip being either a student or tutor turn in the tutoring dialogue – student and tutor turns describe the behaviour of the actor in any given clip. The tutor and student turns that were used for the coding scheme are given in Tables 1 and 2, respectively. The facial expressions that were used for the coding scheme are given in Table 3. The intensity of each expression was rated as either 'low' or 'high' – neutral expressions were assigned a low intensity by default. The coding scheme was applied to each of the clips, thus generating the raw data of the study.

Table 1 Frequency of tutor turns used in the coding scheme

Code	*Tutor Turn*	*Frequency*	*% age*
1	Pose initial problem	8	–
2	Pose harder problem	63	3
3	Pose easier problem	6	–
4	Pose similar problem	71	4
5	Ask new question	336	18
6	Ask about error	5	–
7	Request clarification	55	3
8	Pump for additional information	454	24
9	Assess knowledge of topic	0	–
10	Global assessment	0	–
11	Positive immediate feedback	276	14
12	Positive delayed feedback	1	–
13	Neutral immediate feedback	285	15
14	Neutral delayed feedback	2	–
15	Negative immediate feedback	3	–
16	Negative delayed feedback	1	–
17	Reminding example	7	–
18	Hint	64	3
19	Answer own question	14	1
20	Answer student question	10	1
21	Rearticulate/discuss problem	38	2
22	Rearticulate/discuss question	44	2
23	Rearticulate/discuss solution	123	6
24	Comment about tutor ability	1	–
25	Comment about student ability	15	1
26	Comment about problem	6	–
27	Complaint	0	–
28	Other	22	1
	Total	1910	100

Table 2 Frequency of student turns used in the coding scheme

Code	*Student turn*	*Frequency*	*% age*
1	Correct answer/action	820	74
2	Partial answer	23	2
3	Error-ridden answer	106	10
4	No answer	70	6
5	Related question	14	1
6	Unrelated question	0	–
7	Valid statement	20	2
8	Invalid statement	5	–
9	Reminding example	0	–
10	Meta-comment	7	1
11	Acknowledgement	20	2
12	Complaint	0	–
13	Thinking aloud	5	–
14	Thinking silently	0	–
15	Other	16	1
	Total	1106	100

Table 3 Frequencies of student and tutor expressions

Expression	*Frequency (%)*	
	Student	*Tutor*
Neutral (low)	62	86
Neutral (high)	–	–
Happy (low)	23	9
Happy (high)	6	3
Confused (low)	3	–
Confused (high)	–	–
Frustrated (low)	–	–
Frustrated (high)	–	–
Disappointed (low)	–	–
Disappointed (high)	–	–
Bored (low)	–	–
Bored (high)	–	–
Surprised (low)	–	–
Surprised (high)	–	–
Apprehensive (low)	4	1
Apprehensive (high)	–	–
Disgusted (low)	–	–
Disgusted (high)	–	–

4 Results

The nine tutoring videos were divided into over 3000 sequential clips of student and tutor turns. Each of these clips was coded to describe the actor's behaviour, facial expression and intensity of facial expression.

Whenever sequences of tutor turns or student turns appeared in the data, they were compressed into a single turn so that there was a 1:1 ratio between student and tutor turns. This led to the creation of 6 new kinds of student turn and 112 new kinds of tutor turn, such as 'tutor gives positive immediate feedback and pumps for additional information' – a composite of the existing tutor turns 11 and 8 (see Table 1). This has made the data much easier to work with.

4.1 Frequencies of student and tutor turns

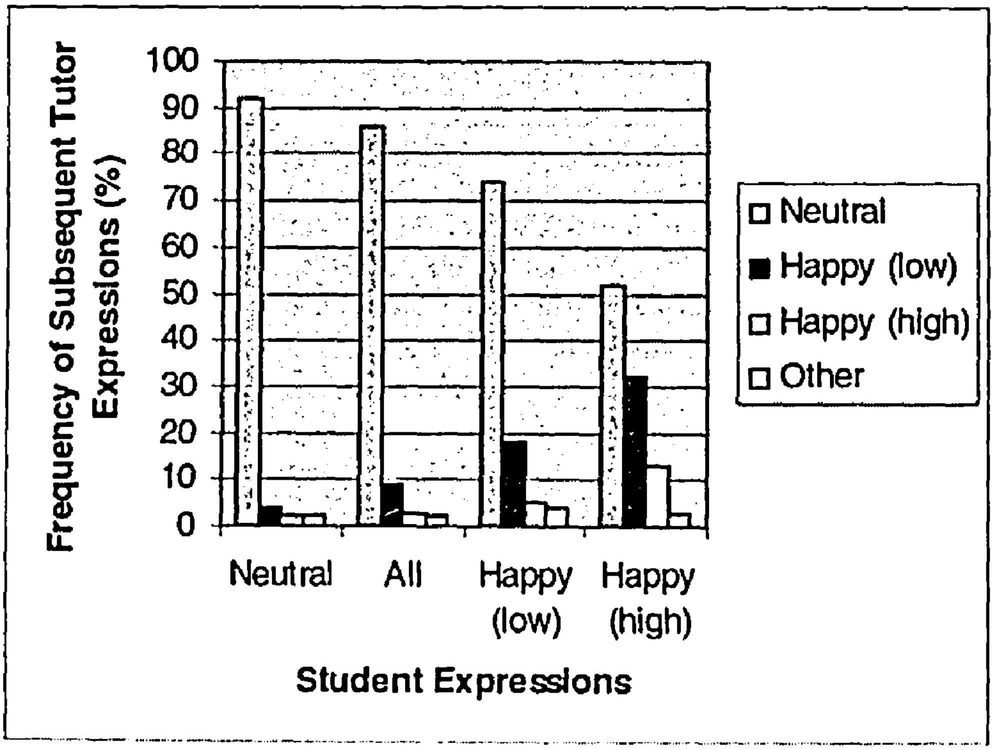
Tables 1 and 2 give the respective frequencies of student and tutor turns in the coding scheme. Perhaps unsurprisingly, almost all student turns were related to answering questions, with these turns occurring a total of 92% of the time. The occurrences of tutor turns were more widely spread across the coding scheme, although 'ask new question', 'pump for additional information', 'positive immediate feedback' and 'neutral immediate feedback' between them totalled 71% of all tutor turns. Negative feedback was almost never used by tutors.

4.2 Frequencies of student and tutor facial expressions

The frequencies of student and tutor facial expressions are both given in Table 3. For both students and tutors, neutral expressions were by far the most common: this was especially the case for tutors, for whom 86% of all expressions were neutral. The second and third most commonly appearing expressions were also the same for both students and tutors, with smiling (low) the second and smiling (high) the third most common expressions. However, smiles were much more common for students than tutors: students smiled for a total of 29% of student turns, whereas tutors smiled for only 12% of tutor turns. Students also appeared apprehensive (low) for 4% of turns and confused (low) for 3% of turns, but the combined occurrences of all other student expressions totalled only 2% of student turns. Similarly for tutors, the combined occurrences of all expressions other than neutral and smiling (including confusion and apprehension) totalled only 2% of tutors turns. Expressions of frustration, disappointment, boredom, surprise and disgust were almost entirely absent for both students and tutors.

Tutor expressions were clearly influenced by the student expressions that immediately preceded them. For instance, confused and apprehensive student expressions almost always resulted in neutral tutor expressions, as did neutral student expressions: all three expressions were followed by a neutral tutor expression over 90% of the time. Figure 3 shows how the likelihood of a tutor smiling was significantly affected by whether or not the student was smiling and the intensity of student smiles. Combining low and high intensity tutor smiles, tutors smiled in only 6% of turns following neutral student expressions – exactly half the average of 12% across all student expressions – but this frequency increased to as much as 45% when students smiled with high intensity. Therefore, the probability of tutor smiles significantly rose as students smiled with increasing intensity.

Figure 3 Frequency of particular tutor expressions following neutral, all, smile (low) or smile (high) student expressions



4.3 Searching for sequences of interactions

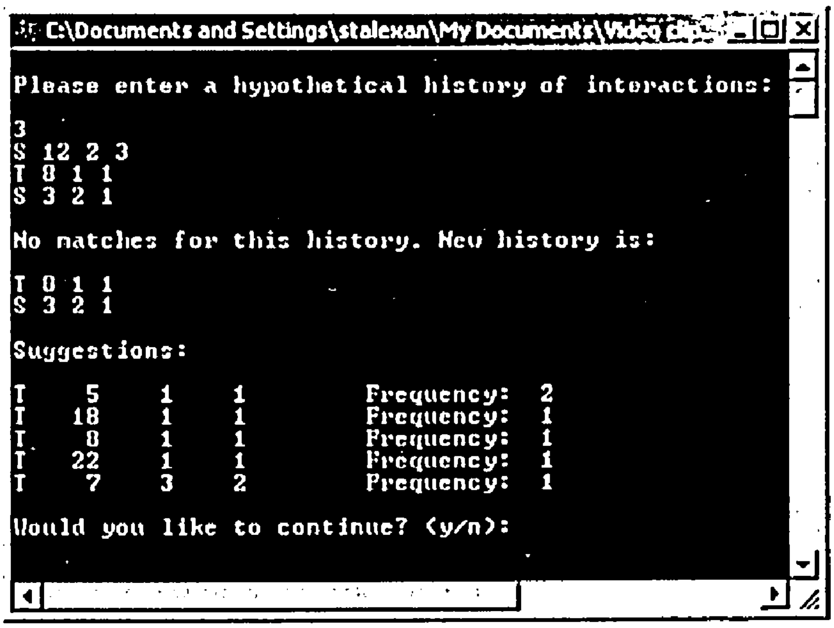
The data contain a wealth of information about the *interaction* between tutors and students during the tutoring process. This information can be used to generate a list of appropriate tutor responses to a student based upon the frequencies in the data of particular tutor responses to particular student turns and expressions. Thus, for any given combination of student turn, facial expression and intensity of facial expression, the following information is readily available:

- the frequencies in the data of all the tutor turns that immediately follow this combination of student states and
- the frequencies in the data of all the tutor facial expressions (and intensities) that immediately follow this combination of student states.

This information is available for any combination of student turn, facial expression and intensity.

However, a human tutor's response to a tutoring scenario is certain to be influenced by the history of their interactions with the student throughout the tutoring session. The data also contain information on the way that a human tutor's adaptations can vary according to the history of an interaction. A simple case-based reasoning program has been written that searches the data for any given sequence of interactions and returns the frequencies of tutor responses to the series of student and tutor turns and expressions that is given as input. If the sequence that is input cannot be found in the data, then the least recent turn in the input sequence is deleted and the resulting sequence searched for – this process repeats until either a match in the data is found or the sequence has run out of turns. A sample screenshot of this program is shown in Figure 4: the first column of the input represents whether the actor in the turn was a tutor or a student, the second column represents the student or tutor turn (see Tables 1 and 2), the third column represents the facial expression of the actor (see Table 3) and the fourth column represents whether the intensity of the expression was low or high by the values 1 and 2, respectively.

Figure 4 Screenshot of the case-based reasoning program



5 Discussion

It should be remembered that these results may not have been the same had the study used different aged participants, different tutoring exercises or a different tutoring environment. These factors might need to be considered if these results were to be applied to other domains.

5.1 Expertise of tutors

The entire study has been conducted on the assumption that the three tutors that were videoed are actually worth copying in an ATS. This was felt to be a safe assumption, as both the professional tutor and the two teachers have experience with the New Zealand Numeracy Project, from which the tutoring exercise used in the study was taken. Measuring the expertise of the tutors was felt to be beyond the scope of the current study.

5.2 Relevance of this study to an ATS

A potential criticism of this study is that students may display different expressions when interacting with human beings to the expressions they would display when interacting with a computer tutor. However, a study of student facial expressions when interacting with a computer tutor found that by far the most common student expression was neutral and that easily the next most common expression was a smile (Ghijsen et al., 2005). These are the same as the results that were found in the current study; we might tentatively suggest that the expressions displayed by students are not significantly affected by whether the tutor is human or artificial. This would concur with the Media Equation of Reeves and Nass (1996), which argues that interactions between computers and human beings are inherently social and that the same rules for interactions between human beings also apply to interactions between human beings and computers. However, much more complete studies would be required before firm conclusions could be reached on this particular aspect of the facial expressiveness of students.

5.3 Other methods of identifying affect

Although the observational study considered only the facial expressions of the students and the tutors, the results would be equally relevant to methods where the affective state of students is predicted, based upon the causes of emotions. The factors that are important to predicting affective state include the personality of students and the sequence of their interactions with the tutor: the personality of the particular students that were videoed could easily be added to the data and the sequences of interactions between tutors and students are already implicit in the way the data are represented.

The observational study is also relevant to methods of identifying affect where subtle physiological responses such as heart rate are detected. Human tutors can detect facial expressions, but not many physiological responses; however in both cases affective state information can only be of any use if you know what to do with it, which was the original rationale for studying human tutors assumed to be expert.

However, it may be possible that the human tutors in the study detected affective state information from the acoustic and prosodic elements of student speech; this information was not captured in the data. This was a potential limitation of the current study.

6 Future work

6.1 Consistency of tutors

There were three different tutors that took part in the study; the data also allows a comparison of the ways in which these different tutors adapted to the same student states. By comparing the frequencies of particular tutor responses to particular student states, it will be possible to get some idea of how much variability in approach there was amongst these three tutors. It will also be possible to determine whether the same tutors were consistent in their tutoring across different students.

If statistical analysis finds the data to be consistent across students tutored by the same tutor and consistent between tutors, then this will justify searching *all* the data for sequences of interactions as in the program shown above in Figure 4. If the data are found to be inconsistent, then it should only be carefully searched to avoid confounding the results – the responses of a particular tutor to a particular student may be unrelated to the responses of another tutor to another student. This would be an interesting study in its own right.

6.2 Improving the case-based program

Currently, whenever the case-based reasoning program is searching for a sequence and cannot find a match in the data, it simply keeps shortening the sequence by one turn and repeating the search until a match is eventually found.

This algorithm could be significantly improved by adding a fuzzy element to the process that considers turns and expressions that are similar to, if not exactly the same as, the precise turns and expressions that are input to be searched for. For instance, 'ask new question' with a low-intensity smile is almost the same as 'ask new question' with a high-intensity smile. Similarly, 'give neutral feedback and ask new question' is almost the same as 'ask new question'. By including similar sequences in the search, it may be possible for the program to make a more balanced estimate of likely tutor responses. A fuzzy approach would also make the data go further, as much more of the data would be relevant to any given search than is the case at present.

6.3 Implementing the ATS

An ATS will be implemented to help 8–9 year old students with exactly the same New Zealand Numeracy Project exercise that was used in the observational study of human tutors. The affective state of the student will be modelled based upon input from an automated facial expression analysis system that is currently being developed in-house (Fan et al., 2003). The ATS will use an animated pedagogical agent to both recognise and display emotions in the manner of a real human tutor.

The improved case-based reasoning program will form the basis for the tutoring strategies module of the ATS. The only inputs required for the module will be a sequence of student and tutor turns and their accompanying facial expressions and intensities: the input of the tutor turns will be self-evident from the history of interactions between the ATS and the student and their accompanying facial expressions and intensities will be detected by the image processing system.

6.4 *Testing the ATS*

Once it is implemented, the ATS will be tested in the same school in which the observational study of human tutors was carried out. The data gathered from these tests should provide valuable information on the effectiveness of including affective state in a tutoring strategies module and also on how empathy affects the persona effect of an animated pedagogical agent.

References

Alexander, S.T.V. (2004) 'Emulating human tutor empathy', *Proceedings of the IIMS Postgraduate Conference*, Albany, New Zealand.

Alexander, S.T.V., Sarrafzadeh, A. and Fan, C. (2003) 'Pay attention! The computer is watching: affective tutoring systems', *Proceedings of E-Learn 2003*, Phoenix, Arizona.

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

de Vicente, A. (2003) 'Towards tutoring systems that detect students' motivation: an investigation', PhD Thesis, School of Informatics, University of Edinburgh.

Fan, C., Johnson, M., Messom, C. and Sarrafzadeh, A. (2003) 'Machine vision for an intelligent tutor', *Proceedings of the International Conference on Computational Intelligence, Robotics and Autonomous Systems*, Singapore.

Ghijsen, M., Heylen, D., Nijholt, A. and op den Akker, R. (2005) 'Facial affect displays during tutoring sessions', *Affective Interactions: The Computer in the Affective Loop (Workshop at IUI 2005)*, San Diego.

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*, Vol. 11, pp.47–78.

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', *IEEE International Conference on Advanced Learning Technologies*, pp.43–48.

Litman, D. and Forbes, K. (2003) 'Recognizing emotions from student speech in tutoring dialogues', *Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, St. Thomas, Virgin Islands.

New Zealand Ministry of Education (2003) *Book 1. The Number Framework*, Numeracy Professional Development Projects, Ministry of Education, Wellington.

Person, N.K., Graesser, A.C. and The Tutoring Research Group (2003) 'Fourteen facts about human tutoring: food for thought for ITS developers', *AI-ED 2003 Workshop Proceedings on Tutorial Dialogue Systems: With a View Toward the Classroom*, Sydney, Australia.

Picard, R.W. (1997) *Affective Computing*, MIT Press, Cambridge, MA.

Prendinger, H. and Ishizuka, M. (Eds) (2004) *Life-Like Characters, Tools, Affective Functions, and Applications Cognitive Technologies Series*, Springer, Berlin Heidelberg.

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

Sarrafzadeh, A., Fan, C., Dadgostar, F., Alexander, S.T.V. and Messom, C. (2004) 'Frown gives game away: affect sensitive tutoring systems for elementary mathematics', *Proceedings of the IEEE Conference on Systems, Man and Cybernetics*, The Hague.

van Mulken, S., André, E. and Muller, J. (1998) 'The persona effect: How substantial is it?' *Proceedings of Human Computer Interaction*, Berlin.