Facial Signs of Affect During Tutoring Sessions*

Dirk Heylen[1], Mattijs Ghijsen[2], Anton Nijholt[1], and Rieks op den Akker[1]

[1] University of Twente
[2] University of Amsterdam, The Netherlands

Abstract. An emotionally intelligent tutoring system should be able to taking into account relevant aspects of the mental state of the student when providing feedback. The student's facial expressions, put in context, could provide cues with respect to this state. We discuss the analysis of the facial expression displayed by students interacting with an Intelligent Tutoring System and our attempts to relate expression, situation and mental state building on Scherer's component process model of emotion appraisal.

1 Introduction

INES is an Intelligent Tutoring System that the Human Media Interaction group of the University of Twente is developing as a test-bed for research on multimodal interaction, intelligent agent technology and affective computing ([1] and [2]). In previous work we have investigated strategies for the tutoring agent to give appropriate emotionally intelligent feedback. Ultimately, one would want the system to choose it actions fitting the personality of the students, their reactions to what is happening, their motivation and other aspects of their mental state, in the hope that this will optimize the learning process. In the work described in [3] we discussed how, in the system developed so far, the choice of teaching strategy, the kind of feedback and the form in which this was realized - the kind of dialogue act and stylistic features of the utterance - were co-determined by an hypothesized emotion model of the student. The model changed dynamically on the basis of the level of student activity and the way the student performed the exercise (the number and type of errors). This model and the resulting system was evaluated by having the system generate responses in a number of scenario's and letting people judge the appropriateness of the sessions for different settings of the system.

The design of a tutoring system is a fertile ground for further studies on affective interaction. An obvious issue on the agenda is how we can detect more accurately what the student is actually experiencing during the exercise. The idea that facial expressions may provide cues for this is something that readily springs to mind ([4],[5]), but which is not without its problems. The (marked) facial expressions displayed during interactions might only be few in number

* Part of this work was carried out in the context of the EU Project HUMAINE (IST-2002-507422).

J. Tao, T. Tan, and R.W. Picard (Eds.): ACII 2005, LNCS 3784, pp. 24–31, 2005.

and it is may not be easy to recognize expressions automatically. Moreover, as is well-known, the expressions that are being displayed cannot be taken as simple read outs of the mental state. Associating an expression displayed with an emotion experienced is not a trivial problem (also from a methodological point of view). The remark in [6] that "None of the methods I describe claim to recognize the underlying emotion, but only the expression on the user's face." (p. 175) is telling in this respect. Besides, many facial expressions are determined by other factors than the emotional state: they may be adaptors, have a conversational function, or be expressive of a meta-cognitive state as "thinking face" described in [7], for instance. And, finally, even if we were somewhat confident that the cues we interpret tell us something about the mental state this may not be something useful for optimizing the tutoring process. We therefore started a pilot experiment in which we looked for answers to the following questions: what kinds of facial expressions occur on student's faces during interactions with the INES system and how can we make sense of them. We collected video material from students interacting with the system and explored a method to describe the affective information.

We collected over one hour of video material of students interacting with the INES system. In this paper we will discuss our analysis of the number and kind of facial expressions displayed and the situations in which they occurred. The question we are interest in is what the expressions tell us about the mental state and what events in the context triggered the state/expression. This information can be used by the tutoring system to adapt its strategy. There are various ways one could go about, none completely reliable. One way to proceed to get information about the mental state is to ask the students post-hoc. Another method is to rely on the experimenter's interpretative skills: how do people who view the video interpret the expression in context. For our pilot we relied on this second approach, which is not without its limitations but becomes more reliable when judgements of different people agree. A third method would be to look up the displays in the dictionary of facial expressions and determine their meaning in that way. Of course, the problem is that no dictionaries of this kind exist that can be used for this purpose. Moreover, it is clear that if such a dictionary existed it would list multiple meanings for each expression and map multiple expressions to the same meaning. Nevertheless, emotion theorists, communication researchers and facial expression specialists have made inventories of associations between expression and mental state. These attempts at mapping may be used heuristically or as a first model of what happens. With this in mind we have compiled a kind of facial expression dictionary[1] from the hints in the literature and made an attempt to evaluate this on our data. We have looked at Scherer's component process approach to find a way to come to grips with the relation between facial expressions, the situation they occur in and the mental state of the student ([8] and [9], [10]). The aim is to derive a table associating elements

[1] We are reluctant to use the word dictionary as the term expresses a rather naive view on the relation between facial expression and emotion. On the other hand, for the system to be implemented simplifications will have to made anyway.

of the tutoring situation, the facial expressions that occur in that situation and the mental state one might assume to hold that is consistent with the data and that might be of use for the tutoring system. In the next section we describe how we collected and analyzed the data.

2 INES

The current demo version of INES implements an exercise which teaches students how to give a subcutaneous injection in the arm of a virtual patient. Students can interact with the patient using speech. The objects in the virtual world that are needed to give the injection are manipulated using a haptic device (a Phantom). The tutoring agent provides instructions and feedback through spoken natural language output and haptic feedback through the haptic device. The tutor is also represented by a talking head.

Fig. 1. Camera positions

We used two webcams (as is shown in Figure 1) in the experiment. One webcam was placed on top of the screen of the system where the student is working with the INES system. The other webcam was placed behind the student and captures the actions with the phantom and the screen. All subjects received an explanation of the exercise in advance (as would be the case with an actual nursing student). They carried out the exercise three times in succession.

2.1 The Exercise

An exercise consists of the following steps:

1. The student has to ask (speech) the patient if she wants to place her right arm on the table and roll up the sleeve of the right arm.
2. Next, the student needs to disinfect the region on the upper arm where he or she wants to inject the medicine. This is done by using the haptic device. When the student moves the haptic device a 3D graphical representation of a pair of tweezers holding a ball of sterile cotton will also be moving accordingly.

3. Next, the student has to insert the needle into the skin of the patient. This is again done by using the haptic device and a 3D representation of a syringe. The tip of the needle has to be positioned right under the skin. Because of the force feedback from the device the student can feel the depth of the needle in the skin layer and the force that needs to be used to get the needle at the required depth.
4. The final step is to inject the medicine and withdraw the needle. This again is done by use of the haptic device.

The actual exercise a nurse has to perform, consists of many more steps before and after this sequence ([11]) but the steps shown above are the only ones were used in the experiment [2]. Table 1 shows a transcript of a part of the exercise.

Table 1. Transcript of an exercise

```
Student: Put your right arm on the table.
Patient: I didn't understand what you said, could you repeat the sentence?
Patient: Where do I have to put my arm?
<student raises eyebrows and pulls down mouth corners>
Student: Put your right arm on the table.
Patient: OK
<patient places her arm on the table>
Student: Could you pull up your sleeve?
Patient: Which sleeve should I pull up?
Student: Of your right arm.
Patient: OK
<patient pulls up sleeve>
<student starts disinfecting the arm of the patient>
<student smiles>
<student has finished disinfecting the arm of the patient>
<student smiles>
<student starts injecting the medicine>
<student has completed the task>
<student smiles>
```

3 Expressions

Each time a subject performed the exercise it was recorded in a take. By watching the takes, the marked facial expressions were put in a table, together with the situation in which they occurred. Table 2 summarizes the result of all takes.

A variety of facial expressions occur but on the whole most of the time the students remain largely 'expressionless'. The table shows, for instance, that when the patient asks for clarification ("Which arm should I put on the table") this is accompanied by several movements: eyebrows are raised, a frown shows, and the head is tilted backwards. They can be interpreted as an instance of slight surprise. The student does not expect the patient to say something except for "I did not understand what you said". When the patient pulls up her sleeve

similar displays occur: smiles, eyebrow raises, a head nod, and a pulling back of the head. The head nod in this particular instance functions as a kind of acknowledgement. The table shows that seemingly autonomous actions by the patient are greeted with quite a few expressions.

Table 2. The occurrence of facial expressions in different situations

Situation / Facial Expression	smile	brows up	mouth down	frown	nod	tilt	head back
P. does not understand S.							
P. asks for clarification		2		1			1
P. repeatedly does not understand S.	2						
P. does not respond to what S. says	1	1	1				
P. says something unexpected							
S. asks P. to do something		1					
S. gives P. more information	1						
P. pulls up her sleeve	4	2			1		1
P. puts arm on the table	1	2			2	1	
S. disinfects the arm	7	1					
S. finished disinfecting	2						
S. injects the medicine	3						
S. finished injecting the medicine	1	2	1				
OTHER	1	2	1				

A second class of situations that brings forth expressions on the face contains the cases where the student manipulates the haptic device. When disinfecting and injecting a lot of smiles occur and some other expressions.

Most of the time it is not difficult for us, when looking at the video, to imagine what state of mind the student is in when displaying the particular expression. It is almost always obviously what event triggers the occurrence of a marked expression. But in order to use this information in a tutoring system, we wanted to describe the relation between situation, expression and ascribed state more systematically. We looked at the literature for a way to help us explain what is it about the situation and particularly the student's appraisal of it that triggers the facial expression. We chose to explore the use of Scherer's component process analysis as one way to describe the relation between facial expression, type of situation and mental state by stimulus evaluation checks.

4 Interpretation

In [9], Scherer describes a model that explains how an organism evaluates stimuli in a series of appraisal checks. The general idea is that the outcome of these checks result in specific facial expressions. This can be used to relate stimulus (situation), facial expression and appraisal (mental state). Table 3 show the various checks for dimensions such as novelty (suddenness, familiarity, predictability), pleasantness, goal significance (relevance, expectation of outcome, etcetera),

coping potential and compatibility standards. These are related to facial expressions, indicated by Action Unit numbers. The table is adapted from [10].

Table 3. Predictions of Facial Expression Changes

Appraisal dimension	**Outcome A**	**Outcome B**
Novelty	**High**	**Low**
Suddenness	1 + 2 + 5 + 26/27	–
Familiarity	–	4b + 7
Predictability	–	4b + 7
Intrinsic pleasantness	**High**	**Low**
	5a + 26a	4 + 7/43)/44 + 51/52 + (61/62)

If we look at our data, one can think about each of the types of situation that occur and make educated guesses as to what kinds of appraisals these situations are likely to give rise to. For instance, in our case, one could come up with something as in Table 4 (this is part of the table only).

Table 4. Situations expressed in terms of SECs

	Situation description	Novel	Pleasant	Goal	Coping
(1)	Patient does not understand what the student has said.	-	Low	Obstruct	Low
(2)	Patient asks for clarification.	-	Neutral	Neutral	Low
(3)	Patient repeatedly does not understand what the student has said.	-	Low	Obstruct	Low
(4)	Patient does not respond at all to anything the student says.	-	Low	Obstruct	Low
(5)	Patient says something unexpected.	-	Low	Neutral	Low
(6)	Student asks the patient to do something.	-	Neutral	Conducive	High

It should be noted that we have greatly reduced the number of stimulus evaluation checks for the pilot. If the patient pulls up her sleeve, after the student has asked her as part of the exercise, then this is conducive to the goal of the student but the control the student has is low (control being one aspect of coping potential). Novelty is not scored in this table because it is typically not a property of situation types but rather of actual situations. This table gives us then an idea of the *relation between situation and expected appraisal* (**A**).

From the opposite perspective, one can look at the expressions that occur and look up what appraisal dimensions might have led to these expressions, using associations as in Table 3. Here appraisal dimensions are associated with Facial Action Units. Of course the ambiguity of the expression and the nature of the appraisal process operating in real life make these things much more complicated. The theory was not designed to be used as a dictionary. Let alone, that it is

designed to be complete. It is obvious that we are reducing the complexity, hoping to arrive at a reasonable guess about how the student experiences the situation. However, what we get out of this is a crude indication of the *relation between facial expression and appraisal* (**B**).

The data about the actually occurring expressions with the situation when they occur has been tabulated as well. They were derived from the transcripts. This indicates the *relation between facial expression and situation* (**C**). Given **A** and **B**, a system could infer aspects of the mental state of a student when presented with **C**. Either the facial expression will correspond with what can be expected from the theory or the tables will not provide direct information about how to relate the expression and the situation at all. In that case, the system could decide that the student is experiencing something different from what would be expected and can use the association between expression and appraisal (**B**) to make a guess about the most probable mental state or about something deviant that may have happened. All this assumes that the various associations make sense. At the early stages of data collection, however, incompatibilities will lead more likely to adjustments of the 'dictionary' (refining the situation descriptions, changing the analysis in terms of the evaluation checks or changing the associations).

It is also possible to attempt an interpretation of the actually occurring expressions in terms of stimulus evaluation checks, by looking at the expressions and make judgements about the mental state[2]. Such an analysis is presented in Table 5. This table shows how the facial expressions (smile, raise eyebrows, pull down mouth corners and frown) correspond with the characteristics of the situation (expressed as SEC parameters). It shows that smiles often occur in situations that we assume to have a high pleasantness and a conducive goal significance, as one might expect.

Table 5. Results of comparing facial expressions and situations

Facial expression	Novel	Pleasant	Goal	Coping
Smile (Total=22)	3	18	19	14
Raise eyebrows (Total=11)	7	4	2	0
Pull down mouth corners (Total=2)	2	1	1	1
Frown (Total=1)	0	0	0	1

5 Discussion

Looking at the real data one can thus evaluate the predictions of the theory - the adapted dictionary - further. Such investigations can be used as a heuristic procedure to derive more detailed triples (situation, expression, presumed mental state). In case there are mismatches between the conjectured Stimulus Evaluation Checks and the ones that are associated with the facial expressions

[2] One could adopt various methods to attempt this, with more or less reliability ([12]).

one might assume that either the situation specification should be refined or the relations between expression and appraisal might need to be revisited. Clearly, for the simple case we have presented we know that both of these will have to be refined. Our classification of situations, for instance, is too coarse grained. The particular situation with all the contextual features is what gives rise to the actual appraisals. By collecting and analyzing more data in this way one can refine the specification of the associations between expression, appraisal and situation.

Our goal with looking at the facial expressions during tutoring sessions is to get information about aspects of the mental state of the student that may be useful to know for the system in order to adapt its teaching strategy. It remains to be seen whether the facial expressions are a useful cue, whether valid inferences can be made and whether they can lead to useful actions of the system. One of the issues related to this is whether the stimulus evaluation checks provide the information that is appropriate for the system to react to.

References

1. Hospers, M., Kroezen, E., Nijholt, A., op den Akker, R., Heylen, D.: 9. In: An agent-based intelligent tutoring system for nurse education. Volume Applications of intelligent agents in health care. Birkhauser, Basel (2003) 143–159
2. Kole, S.: Tactile, spoken, and visual interaction within an intelligent tutoring system. Master's thesis, Department of Electical Engineering, Mathematics, and Computer Science, University of Twente, Enschede (2004)
3. Heylen, D., Vissers, M., op den Akker, R., Nijholt, A.: Affective feedback in a tutoring system for procedural tasks. In André, E., Dybkjaer, L., Minker, W., Heisterkamp, P., eds.: Affective Dialogue Systems, Heidelberg, Springer (2004) 244–253
4. D'Mello, S.K., Craig, S.D., Gholson, B., Franklin, S., Picard, R., Graesser, A.: Integrating affect sensors in an intelligent tutoring system. In: Affective Interactions, Workshop at IUT (2005)
5. Lisetti, C., Schiano, D.: Facial expression recognition: where human computer interaction, artificial intelligence, and cognitive science intersect. Pragmatics and Cognition **8** (2000) 185–235
6. Picard, R.: Affective Computing. The MIT Press, Cambridge, Massachusetts (2000)
7. Goodwin, M., Goodwin, C.: Gesture and coparticipation in the activity of searching for a word. Semiotica **62** (1986) 51–75
8. Scherer, K.: 14. Approaches to Emotions. In: On the Nature and Function of Emotion: A Component Process Approach. Lawrence Erlbaum Associates, Hillsdale, New Jersey (1984) 293–318
9. Scherer, K.: Toward a dynamic theory of emotion: The component process model of affective states. Copy retrieved [02-26-2004] from: http://www.unige.ch/fapse/emotion/publications/list.html (1987)
10. Wehrle, T., Kaiser, S., Schmidt, S., Scherer, K.: Studying the dynamics of emotional expression using synthesized facial muscle movements. Journal of Personality and Social Psychology **78** (2000) 105–119
11. Hospers, M., Kroezen, L.: I.N.E.S. intelligent nursing education software. Master's thesis, Departement of Computer Science University of Twente (2002)
12. Ekman, P., O'Sullivan, M.: Facial expression: Methods, means, and moues. (In Feldman, R.S., Rimé, B., eds.: Fundamentals of nonverbal behavior) 163–199