Facial Affect Displays during Tutoring Sessions

Mattijs Ghijsen
University of Amsterdam
Kruislaan 419
1089 VJ Amsterdam, The Netherlands
mattijs@science.uva.nl

Dirk Heylen
University of Twente
P.O. Box 217
7500 AE Enschede, The Netherlands
d.k.j.heylen@utwente.nl

Anton Nijholt
University of Twente
P.O. Box 217
7500 AE Enschede, The Netherlands
anijholt@ewi.utwente.nl

Rieks op den Akker
University of Twente
P.O. Box 217
7500 AE Enschede, The Netherlands
infrieks@ewi.utwente.nl

ABSTRACT

An emotionally intelligent tutoring system should be able to provide feedback to students, taking into account relevant aspects of the mental state of the student. Facial expressions, put in context, might provide some 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 the component process model of affective states.

Keywords

Affective interaction, Facial Expressions, Human Computer Interaction

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 multi-modal interaction, intelligent agent technology and affective computing ([4] and [6]). In previous work we have investigated strategies for the tutoring agent to give appropriate, "emotionally intelligent" feedback. Ultimately, one would want to obtain a system that chooses it actions to fit the personality of the student, his reactions to what is happening, his motivation and other aspects of his 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). The exact ways in which the model changed its parameter settings on the basis of this input were determined by basing ourselves on ideas taken from the literature on tutoring and emotion research in general, mixed with common sense insights. 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 ([7]. We decided to explore this option despite the fact that we were well aware of the many problems involved. The (marked) facial expressions displayed during interactions might only be few in number and it is may not be easy to recognize expressions automatically. The expressions displayed cannot be considered as simple read outs of the mental state; associating an expression displayed with an emotion experienced[1] is a huge problem (also from a methodological point of view). 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 metacognitive state as "thinking face" described in [2], 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. The questions that we have tried to answer so far were the fol-

[1]The remark in [8] 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.

lowing: what kinds of facial expressions occur on student's faces during interactions with the INES system and how can we make sense of them. In a pilot experiment we collected video material from students interacting with the system and explored a method to describe the affective information.

We have 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. 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. 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. Despite this, 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 crude facial expression dictionary[2] 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 ([9] and [10], [11]). The aim is to derive a table associating elements 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.

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 natural language output (speech) and haptic feedback through the haptic device. The tutor is also represented by a talking head.

Figure 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.

The Exercise

The exercise a student had to conduct in the experiment was to administer medication to a patient by injection. An exercise consists of the following steps:

1. The student has to ask the patient if she wants to place her right arm on the table. The student also has to ask the patient to roll up the sleeve of the right arm. The student can ask these questions by talking to the patient.

2. Next, the student needs to disinfect a 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. The student now has to move the sterile cotton over the patients upper arm to disinfect it.

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

[2]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.

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 ([5]) but the steps shown above are the only ones were used in the experiment [6]. Table 1 shows a transcript of a part of the exercise[3]. The original exercise was in Dutch but for this illustrative purpose it has been translated to English.

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 describes the first take of subject 2. Table 3 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.

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.

[3]Tables can be found at the end of the paper.

INTERPRETATION

In [10], 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 4 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 [11].

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 5. 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 4. 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. This is presented in Table 2. These indicate 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 mental state. 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[4]. Such an analysis is presented in Table 6. 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.

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 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. Eventually this should provide a basis for use in the a tutoring system or so it is hoped.

DISCUSSION

The goal of our pilot experiment was to find out the frequency and the kind of facial expressions we can expect students to display during a tutoring session and to explore a method to relate facial expression, situation and mental state.

One of the problems with determining the mental state from facial expressions is that there is no one-to-one correspondence. Likewise, the same situation can give rise to different appraisals. It is not the situation as such but the way the person experiences it that is important. The situational factors, and certainly the coarse situation type that we have used, do not completely determine the mental state of course. By looking at more data such as the one we have collected and by analyzing it more thoroughly one could, however, find out what is normal to expect and explain what happens.

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. The inference from facial expression to mental state is one that is hard to make in general. In particular contexts, though, it might be easier. 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. Other aspects of the mental state, attention, concentration, may be more interesting to take into account. The problem with interpreting facial expressions is that many factors contribute to the expression. Expressions may also serve a conversational function, for instance.

It may also be useful to look at other visual cues in this respect. While looking at the data, we noticed that some typical movements of the head can be observed during different stages of the exercise. The head almost freezes, for instance, when full concentration is needed to give the injection with the haptic device. Investigating what these and other head movements may tell us, is currently one of the items on our research agenda.

[4]One could adopt various methods to attempt this, with more or less reliability ([1])

REFERENCES

1. P. Ekman and M. O'Sullivan. Facial expression: Methods, means, and moues. pages 163–199.
2. M. Goodwin and C. Goodwin. Gesture and coparticipation in the activity of searching for a word. *Semiotica*, 62:51–75, 1986.
3. D. Heylen, M. Vissers, R. op den Akker, and A. Nijholt. Affective feedback in a tutoring system for procedural tasks. In E. André, L. Dybkjaer, W. Minker, and P. Heisterkamp, editors, *Affective Dialogue Systems*, volume 3068/2004 of *Lecture Notes in Computer Science*, pages 244–253, Kloster Irsee, Germany, June 14–16 2004. Tutorial and Research Workshop ADS 2004, Springer-Verlag Heidelberg.
4. M. Hospers, E. Kroezen, A. Nijholt, R. op den Akker, and D. Heylen. *An agent-based intelligent tutoring system for nurse education*, volume Applications of intelligent agents in health care of *Whitestein Series in Software Agent Technologies*, chapter 9, pages 143–159. Birkhauser Publishing Ltd, Basel, Switzerland, 2003.
5. M. Hospers and L. Kroezen. I.N.E.S. intelligent nursing education software. Master's thesis, Departement of Computer Science University of Twente, 2002.
6. S. Kole. 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.
7. C. Lisetti and D. Schiano. Facial expression recognition: where human computer interaction, artificial intelligence, and cognitive science intersect. *Pragmatics and Cognition*, 8(1):185–235, 2000.
8. R. Picard. *Affective Computing*. The MIT Press, Cambridge, Massachusetts, 2000.

9. K. Scherer. *On the Nature and Function of Emotion: A Component Process Approach*, chapter 14, pages 293–318. Approaches to Emotions. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1984.

10. K. Scherer. 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.

11. T. Wehrle, S. Kaiser, S. Schmidt, and K. Scherer. Studying the dynamics of emotional expression using synthesized facial muscle movements. *Journal of Personality and Social Psychology*, 78(1):105–119, 2000.

```
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>
```

Table 1: Transcript of an exercise.

Situation	Facial expression
Patient repeatedly does not understand what the student has said	smile
Patient repeatedly does not understand what the student has said	smile
Patient puts her arm on the table	smile, nod head
Patient pulls up her sleeve	smile, nod head raise eyebrows
Student starts to disinfect the arm	smile
Student is disinfecting the arm	smile
Student is disinfecting the arm	smile
Student is injecting the medicine	smile
Student is injecting the medicine	smile

Table 2: An example of the collected data (Subject 2, Take 1).

Situation Facial Expression	smile	raise brow(s)	mouth corner down	frown	nod	tilt head left	pull head back
Patient does not understand what the student has said.							
Patient asks for clarification		2		1			1
Patient repeatedly does not understand what the student has said	2						
Patient does not respond at all to anything the student says	1	1	1				
Patient says something unexpected							
Student asks the patient to do something		1					
Student give the patient more information	1						
Patient pulls up her sleeve	4	2			1		1
Patient puts her arm on the table	1	2			2	1	
Student is disinfecting the arm	7	1					
Student has finished disinfecting the arm	2						
Student is injecting the medicine	3						
Student has finished injecting the medicine	1	2	1				
OTHER	1	2	1				

Table 3: The occurrence of facial expressions in different situations

Appraisal dimension	Outcome A	Outcome B
Novelty	**High**	**Low**
Suddenness	1 + 2 + 5 + 26/27	–
Familiarity	–	4b + 7
Predictability	–	4b + 7
Intrinsic pleasantness	**High** 5a + 26a	**Low** 4 + 7/43)/44 + 51/52 + (61/62)
Goal significance		
Concern relevance	**High** focusing responses: lower intensity of the cumulation of the two first SECs	**Low** terminating of the responses of the two first SECs
Outcome probability Expectation	**Probable** higher intensity for future responses	**Not probable** lower intensity for future responses
Conduciveness	**Conducive** 6 + 12	**Obstruct** 4e (long) + 7 + 17+ 23/24
Urgency	**Low** de-amplification, low tension	**High** intensification, high tension
Coping potential		
Cause: agent	**Internal personal (self) + external not personal (natural agent) attribution** less intense than external personal attribution	**External personal attribution (other)** intensify existing and future responses, more intense than internal personal or external not personal attribution
Cause: motive	**Not intentional** diminution of intensity of existing and future responses	**Intentional** intensify existing and future responses, more intense than not intentional
Control	**Low** 15 + 25/26 + 41/42/43 + 54 + 61/62 + 64 (1+4)	**High** intensify existing and future responses
Power	**Low** 20 + 26/5 + freezing	**High** 5 + [10 + 23 + 25]/[17 + (23) +24] + 38 + (53/57)
Adjustment	**Low** holding the existing pattern	**High** de-amplification, (12)
Compatibility standards	**Surpassed**	**Violated**
Internal	17 + 24 + (53)	41/42/43 + 54 + 55/56 + 61/62 + 64(gaze avoidance + head down)
External	17 + 24 + (53)	41/42/43 + 54 + 55/56 + 61/62 + 64

Table 4: Predictions of Facial Expression Changes

Situation#	Situation description	Novelty	Pleasantness	Goal Significance	Coping Potential
(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
(7)	Student gives the patient more information.	-	Neutral	Conducive	High
(8)	Patient pulls up her sleeve.	-	High	Conducive	Low
(9)	Patient puts her arm on the table.	-	High	Conducive	Low
(10)	Student is disinfecting the arm.	-	High	Conducive	High
(11)	Student has finished disinfecting the arm.	-	High	Conducive	High
(12)	Student is injecting the medicine.	-	High	Conducive	High
(13)	Student has finished injecting the medicine.	-	High	Conducive	High

Table 5: Situations expressed in terms of SECs.

Facial expression	Novelty	Pleasantness	Goal Significance	Coping Potential
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

Table 6: Results of comparing facial expressions and situations.