14

Social User Interfaces

A. Nijholt, D. Heylen, B. de Ruyter, and P. Saini

"Knowing others leads to wisdom, knowing the self leads to enlightment."

Lao Tzu

14.1 Introduction

Current technological and research developments pertaining to Ambient Intelligence, Ubiquitous Computing or Pervasive Computing share an impetus towards embedding computation in our social and physical environments making it an inseparable part of our daily lives. One consequence of embedding technology in this way in our everyday life is that today's user–system interaction paradigm will change substantially. Interaction is expected to be continuous, prolonged, and tied in with the physical spaces that surround us. It will involve a disparate range of interaction devices, affect social interactions and will often bridge physical and virtual worlds. Additionally, technology is expected to become more intelligent and to adapt itself to our needs and the dynamics of the environments we live in. To make this possible, AmI systems will need to be equipped with sensorial and reasoning capabilities. One consequence of confronting users with, for example, home dialog systems that have perceptive and reasoning capabilities is that additional expectations are created. Since people already have a tendency to attribute human like properties to interactive systems (Reeves and Nass 1996), it is expected that implementing human like properties in such home dialog systems will have an important impact on the user–system interaction. Similar observations can be made for other environments, e.g., smart office environments, (virtual) collaborative work or meeting environments and smart educational environments.

In this chapter we look at user–system interaction from this particular viewpoint. How can we include aspects of human social interaction in the interface and what difference will it make? We will investigate this in the paradigm of Ambient Intelligence. That is, we assume that there are sensors that perceive the user or inhabitant of a smart environment and that the information that is obtained from the sensors is being processed, fused, and interpreted in order to give useful, acceptable, and effective responses, taking into account characteristics of the user (interests, personality, moods,

emotions, background, culture). Hence, in this chapter we look at smart environments to see how social interaction can be improved rather than simply the efficiency of the interaction.

Obviously, in order to do experiments, and in order to be able to build prototypes showing our ideas, we have to confine ourselves to rather stripped-down versions of smart environments, but not in a way that we loose the principles that we want to research. Rather than looking at traditional human–computer interaction we want to look at environments in which smart and mobile objects, human and virtual inhabitants share the environment and have sensorial capabilities that allow them to communicate with each other and to provide support to each other. Another principle is that we want to explore the role of verbal and nonverbal cues in social interaction in smart environments. That is, we need to take into account multimodal interactions between users and the smart environment. On the one hand, the user may be allowed and may decide to interact with the environment using a diversity of modalities and combinations of modalities in order to get his or her intentions clear; on the other hand, the environment can be designed in such a way that it understands and anticipates what the user wants, by tracking and interpreting not only those things the user decides to inform the system about, but also by interpreting the user's actions and whatever else it can determine from sensors that allow their input, e.g., facial expressions, body posture, biometrical information, etc.) to be interpreted from social and emotional points of view. Even when the user does not make an attempt to interact with the environment, the environment can gather useful information about the user.

In this chapter we introduce case studies where users interact with so-called *dialogical robots* (physical robots, embodied agents, smart objects) and where some level of affect and social intelligence in the interaction was introduced. The first case study concerns a home dialog system that has been extended with verbal and nonverbal interaction cues in a Wizard of Oz experiment in order to measure the impact of social intelligence characteristics. The second case study describes the design and the development of a prototype of a state-of-the-art multimodal dialog system with underlying models of affect. Both systems have also in common that the interaction with the dialogical robot is meant to provide real-time support to other actions of the user in the environment.

The remainder of this chapter is organized as follows. In the next section we present a short overview of state-of-the-art human–agent interaction from a multimodal, social, and emotional interaction point of view. Agents may be human, embodied virtual agents, fully virtual with no visual representation or physical robots. In Sect. 14.3 we introduce the iCat robot system and discuss how, in a Wizard of Oz experiment, it has been used to show how adding social intelligence to a human–robot interaction makes a difference in the appreciation of the human partner of the topics that are discussed. Sect. 14.4 is about the Intelligent Nursing Education Environment System (INES) and

our attempts to make the system aware of what a student experiences during the interaction with the system. These experiences need to be translated to an internal state of the system from which actions can be planned in order to adapt the system and its inhabitants to a new situation. Finally, in Sect. 14.5 we draw some conclusions and discuss future research.

14.2 Social Interfaces and Multimodal Interaction

In our investigations we have looked at the design of a social interface between humans and dialogical robots in a smart environment. The users must be able to address the environment and this environment needs to be aware of the interaction that is going on. In this context, we investigate how verbal and nonverbal interaction with embodied agents and situated devices can take place in smart environments. We consider especially interactions that are also guided by affect and maybe even by an implicit desire to have some development of a social relationship between user and system. This will require us to build systems that have social intelligence.

The notion of social intelligence is not a well-defined one. It is useful to take as a starting point the ability of an agent to relate to other actors or agents in a society, understand them, and interact effectively with them. This ability can be contrasted with other kinds of intelligence, for example, the ability to solve complex logical problems and the ability to monitor one's own and others' emotions and to use that information to guide one's thinking and actions (Nishida et al. 2005). Being able to relate with other agents in a society also means being able to develop personal relationships. Social intelligence in this sense means knowing how to judge and behave in everyday social situations. However, as the same authors point out, social intelligence can also deal with social structures and social conventions. The questions in this case are how do these structures and how do these conventions constrain, guide or stimulate the way individual agents interact with each other. When designing environments that can be shared among participants this latter point of view should put emphasis on facilitating common ground building and community development. Although we certainly do not exclude these latter points of view in the design of our systems, i.e., we do not exclude the building of community knowledge in the home and educational dialog environments that we consider here, the main emphasis of this chapter concerns the first notion of social intelligence, i.e., the ability to relate not only in an intellectual way but also on an affective and interpersonal level to the agents we interact with.

Social intelligence requires an understanding of social situations in which people meet each other. When people meet face to face, they receive input from all the sensory modalities. They may see, feel, hear, smell, and touch each other. They assess social situations by the input to their sensors that comes from facial expressions, body posture, gestures, paralinguistic features of the

speech such as prosodic signals, and from bodily contact or the distance that people keep (proxemics). They are aware of many of the cues that they provide to the other in response. Their actions are not just motivated as a response but they are chosen to make a specific impression on the other. Being able to perceive and interpret these cues and respond to them appropriately is an important part of showing social intelligence. In short, getting along with people requires intelligence, and applying this intelligence requires understanding of all the cues that are signaled not only verbally but also nonverbally.

When a computer system, such as a communication robot or an embodied conversational agent, lacks social intelligence, there will typically be situations where it will not understand its human conversational partner and will make wrong assumptions about how to proceed in an interaction. Moreover, it will display nonverbal behaviors that do not fit the social situation. It may show a lack of social awareness by displaying emotions, body postures, facial expressions, and gaze behavior that disrupts the interaction rather than taking care of a smooth, socially appropriate communication between system and conversational participants. Among the basic behaviors that have clear implications for the social, interpersonal relation – particularly when they are not timed and executed properly – are gaze behavior and keeping the appropriate distance. These cues can be very subtle. Establishing eye contact, or avoiding it, is determined by several factors. For instance, when the speaking process demands a high cognitive load, people tend to look away from the listener to avoid distraction from the feedback provided by the listener through the visual backchannel. However, by looking at a speaker, a listener not only may show that he is attending to what is being said, but also mark his interest. Conversely, when not looking at the speaker, this might be interpreted in all kinds of ways: the listener is distracted, is not showing any interest, and is starting to think about a reply. The way the eye contact is broken and where the listener is looking at alternatively, may distinguish between these interpretations. In the interactions with dialogical robots, and other forms of Ambient Intelligence, people will expect similar cues that tell a user when and how to engage with the system, whether or not the system is attending to the user.

Another way in which social interaction between humans is regulated is stated in the form of equilibrium theory. *Equilibrium* theory, proposed by Argyle and Argyle and Dean (1965), is a hypothesis about the way different nonverbal behaviors, such as gaze, distance, and touch, compensate for one another in order to signal the "social distance" between people. For instance, if the social distance between people is high (for instance people who do not know each other waiting for a bus) but they are forced to stand close together, they will try not to look at each other as both of these behaviors mark close relationships. In an interesting experiment, Bailenson et al. (2001) have investigated the equilibrium theory in a virtual setting. The question here is how a system can mediate the intimacy level with users in an ambient environment.

In the following sections we discuss experiments with systems that we have designed to show socially intelligent behavior.

The case studies that we will introduce here deal with the role of affect and social intelligence in a home environment and in an educational environment. Although the studies were not conducted from this point of view from the start, it turned out to be useful to look at them this way, allowing observations on modeling, developing, and maintaining social relations with artifacts and the environment. Maybe more importantly, it proved interesting to investigate how the user in his or her interaction behavior is influenced by the introduction of affect and social intelligence in the interface.

One case study is about human–robot interaction and the other about human–virtual human interaction. Obviously, for both human–human interaction studies are relevant. Also, the two studies certainly have to take into account the context in which the interactions are taking place. In fact, the possibility to do this and to learn from this has been the main reason to discuss them in this chapter.

In both cases we are designing and modeling interactions that take place in an environment where many different sensors gather information. Multimodal models of interaction are able to fuse such information and a multimodal dialog manager is able to provide an interpretation – on a semantic and a pragmatic level. From that interpretation the system should be able to decide about an effective next action that aims at supporting its human partners in the environment.

The two systems we will look at in this chapter have been built and investigated in our research environments. They are the iCat system on human–robot interaction, and the INES on human–embodied agent interaction. In both cases there is face-to-face interaction between a human conversational partner and a synthetic partner (robot or embodied agent). However, more importantly, the interaction between human and dialogical robot is guided by what is happening in the environment which itself can provide information about the interaction.

Our first case study is devoted to the so-called iCat system (de Ruyter et al. 2005). This study is an experiment on simulated human–robot interaction in a home environment. In a Wizard of Oz simulation it is shown how positive feelings about the interaction and the topics that are discussed can be induced in the human partner by making the iCat robot react in a socially intelligent way.

The second case study is devoted to the role of affect in a smart educational environment. We discuss a specific environment called the INES where multimodal interaction between a student, a patient and a tutor is modeled (Heylen et al. 2005). Again, we have a situation where there is direct interaction between a user (a student) and a virtual human (the patient), where there is an environment, embodied by a virtual tutor, that knows what is going on and tries to understand the multimodal interactions between patient and student.

14.3 The Impact of Affect and Social Intelligence: the iCat Case

In the continuous strive for natural interaction, research into computational and robotic characters is demonstrating an increasing interest into the topic of social intelligence. Nevertheless, the benefits that it might offer to users are to this point only hypothesized. Rather than examining the effect of singular factors conducive to social intelligence, this case study examines what broader benefits could be brought upon the interaction experience by a more socially complex and coherent home dialog system that is perceived as more socially intelligent. Such a holistic examination that would show the relevance and importance of social intelligence in the domain of human–computer interaction has not been attempted before.

In this case study we report on a controlled experiment into the effects of perceived social intelligence in a home dialog system addressing the following research questions:

– Will test participants be able to perceive the level of social intelligence implemented in the home dialog system?
– What is the effect of bringing the concept of social intelligence into a home dialog system on the perception of quality of the interactive systems (other than the home dialog system) in the environment?
– Will the participant's acceptance for home dialog systems increase if the concept of social intelligence is implemented into these systems?

In the following sections, we present, a home dialog system using a robotic interface, the iCat. We describe an experiment addressing the research questions raised above. The results of this experiment are discussed leading to some conclusions in the final section.

The home dialog system used in our study takes the form of an "interactive Cat," or just iCat. The iCat is a research platform for studying social robotic user-interfaces. It is a 38 cm tall user-interface robot and is implemented as a desktop robot since it lacks mobility facilities; see Fig. 14.1. The robot's head is equipped with 13 standard R/C servos that control different parts of its face, such as the eyebrows, eyes, eyelids, mouth, and head position. With this setup we are able to generate many different facial expressions that are needed to create an emotionally expressive character.

A camera installed in the iCat's head is used for different computer vision capabilities, such as recognizing objects and faces. The iCat's foot contains two microphones to record the sounds it hears, perform speech recognition, and to determine the direction of the sound source. By determining the direction of a sound source, the iCat can exhibit turn-to-speaker behavior. Also, a loudspeaker is installed to play sounds and to generate speech. Furthermore, iCat is connected to a home network to control in-home devices, e.g., light, VCR, TV, radio, and to access the Internet. Finally, touch sensors and multicolor LEDs are installed in the feet and ears to sense whether the user touches
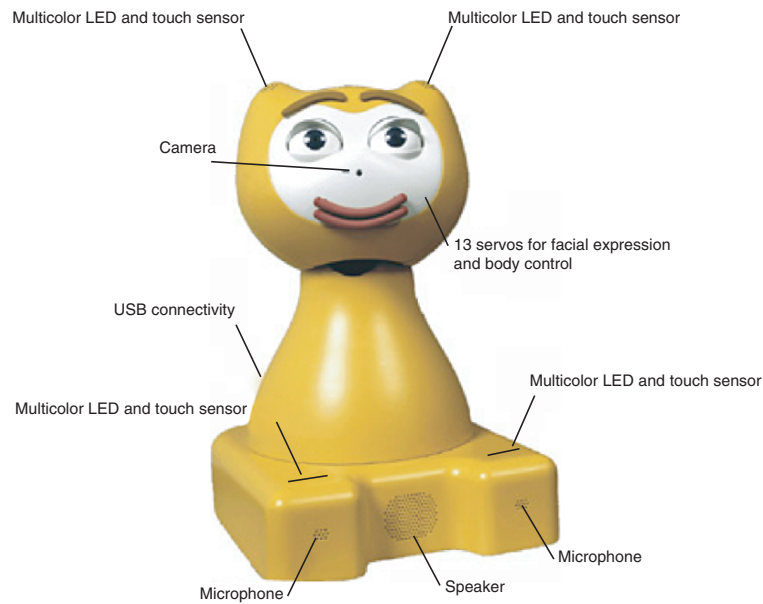
Multicolor LED and touch sensor

Multicolor LED and touch sensor

Camera

13 servos for facial expression
and body control

USB connectivity

Multicolor LED and touch sensor

Multicolor LED and touch sensor

Microphone

Microphone

Speaker

Fig. 14.1. Hardware setup of the home dialog system "iCat"

the robot and to communicate further information encoded by colored light. For instance, the operation mode of the iCat such as sleeping, awake, busy, and listening, is encoded by the color of the LEDs in the ears.

14.3.1 Experiment

A total of 37 paid subjects participated in the experiment (15 women and 22 men). These participants were selected to have at least some basic experience with e-mail and Internet and were externally recruited and randomly assigned to one of two experimental conditions.

The experiment took place at the HomeLab (de Ruyter and Aarts 2004). Participants would be left with the iCat in the living room of the Home-Lab, while the experimenter would observe and control the experiment from the observation station of the HomeLab. We adopted a one-factor between-subjects design in which social intelligence was manipulated. There were two conditions.

Condition 1: Social Intelligence

During this condition the robot would talk (using synthesized speech) with lip synchronization, blink its eyes throughout the session, and display facial expressions while exhibiting the following selected social intelligence aspects:

– *Listening attentively*: by looking at the participant when she or he is talking and occasional nodding of the head.
– *Being able to use nonverbal cues the other displays*: responding verbally to repeated wrong actions of the participant by offering help.
– *Assessing well the relevance of information to a problem at hand*: by stating what is going wrong, before offering the correct procedure.
– *Being nice and pleasant to interact with*: by staying polite, mimicking facial expressions (smile when participant smiles for example), being helpful.
– *Not ignoring affective signals from the user*: by responding verbally or by displaying appropriate facial expression to obvious frustration, confusion, or contentment.
– *Displaying interest in the immediate environment*: the immediate environment being the participant and the equipment used in tasks, by carefully monitoring the person and the progress of the tasks.
– *Knowing the rules of etiquette*: by not interrupting the participant when she or he is talking.
– *Remembering little personal details about people*: addressing the participant by name, remembering login information, and passwords if asked.
– *Admitting mistakes*: by apologizing when something has gone wrong, but also when no help can be provided upon participant's request.
– *Being expressive*: by showing facial expressions while talking, if appropriate.
– *Thinking before speaking and doing*: by showing signs of thinking (with facial expression) before answering questions or fulfilling the participant's request.

The behaviors for the social intelligence condition were available as pre-programmed blocks for the experimenter who would observe (the socially intelligent condition) and listen to the participant and would type in responses for the iCat to utter. Further, the experimenter would initiate these pre-programmed social behaviors at appropriate moments in a Wizard of Oz fashion.

Condition 2: Social Neutrality

The iCat did not display any facial expressions and did not blink its eyes. It talked and used lip-synchronization, but the aspects of social intelligence listed above did not drive the talking. It responded verbally only to explicit questions from the participant. The only self-initiated help was when the participants really got stuck and could not continue without help.

We underline that contrary to studies listed in the previous section, we did not seek to assess the impact of each of the low level behaviors listed here. Rather, we hoped that their combination would lead the iCat to be perceived as socially intelligent and it is the impact of this perception that we aimed to assess, see Fig. 14.2.

Fig. 14.2. Snapshot from the test sessions showing respectively: the participant (LIV1), an overview of the living room (LIV2 and LIV3), and a close-up of the iCat (LIV4)

14.3.2 Tasks

Participants were asked to perform two tasks: (a) program a DVD recorder to record three broadcast shows for the upcoming week and (b) complete an online auction. The auction task involved registering for the service and buying several items on a list. For registration as a new user, the site required a valid web-accessible e-mail. Participants could give iCat their e-mail details (login and password) if they wanted iCat to monitor their selected items. In asking the help of the iCat, participants would also need to entrust it with their password.

14.3.3 Measures

A multiple set of measures was designed to test both the direct effects of iCat's behaviors and the potential implicit spillover effects like satisfaction with the DVD recorder.

Social Behaviors Questionnaire (SBQ). In the absence of existing validated instruments to assess social intelligence in interactive systems, a questionnaire was developed for the purpose of this study. Its purpose was to verify whether we succeeded in creating two separate conditions that the participants would

rate differently in terms of social intelligence. The questionnaire (described in a separate publication) was built up of five-point scales rating the agreement of subjects to statements such as the following:

– The robotic cat takes others' interests into account.
– The robotic cat does not see the consequences of things.
– The robotic cat says inappropriate things.
– The robotic cat is not interested in others' problems.
– The robotic cat tells the truth.

User Satisfaction Questionnaire (USQ). The USQ is an instrument developed previously in-house for assessing user satisfaction with consumer products (de Ruyter and Hollemans 1887). The USQ was used to assess the satisfaction with a DVD recorder that participants had to operate during the experiment.

The Unified Theory of Acceptance and the Use of Technology (UTAUT). This questionnaire measures technology acceptance (Venkatesh et al. 2003). We used the UTAUT with some adaptations for the home domain, to measure the extent to which participants would use iCat at home after the experiment.

In a post-experimental questionnaire participants could indicate on a five-point scale what they thought about their own performance during the experiment.

Finally we noted the number of times participants asked the robot general questions and the number of times they asked questions about the experimental tasks. We also noted the number of times that participants looked at the robot during the entire session.

14.3.4 Results

The results from the SBQ verify the distinctness of the experimental conditions that we wanted to create: participants rated the socially intelligent iCat as more social than the neutral one, which validates the design of behaviors exhibiting social intelligence.

The USQ also had a differential effect between the two conditions. Since the USQ was developed to test satisfaction with a consumer product after thorough interaction with that product and the DVD recorder task only consisted of exploring one function in a time frame of 10 min, the significant difference found between the two experimental conditions is remarkable.

The UTAUT was applied to the iCat and, as such, it shows the explicit positive effect of the social intelligence manipulation.

There was no significant effect regarding perceived auction performance; most participants thought they did pretty well in both conditions. The task of buying items was for most of them not a hard one. As such, many of them felt they did very well. Participants would have liked to delegate more chores to iCat and to have asked it more questions. Most participants asked iCat to monitor their items for bids from others (83% of participants).

The only participants not very satisfied with how well they had performed were those who in their daily lives did not spend much time on the web or on the computer.

Overall the impressions were that participants were more "social" with the socially intelligent iCat: they were much more inclined to laugh, ask questions, and ask for elaborations, than with the neutral iCat. They were more curious about the reasons the social robot said the things it said than when they were interacting with the neutral robot. For example, when asked which LDC monitor was a good one (to buy in the auction task) they were happy that iCat could help by naming a product. But they were curious how it knew this and why it was the best. They were also more inclined to ask about iCat's opinion on the other LCD monitors. They asked these questions politely and using full sentences. In the case of the neutral iCat, they were more inclined to take the suggestion of the best LCD monitor for what it was and not continue asking further. In cases that they did ask more, it was usually in shorter and to-the-point command like sentences than in the social condition.

Participants in the socially intelligent condition liked the fact that the robot was expressive in terms of facial expressions, that it nodded and shook its head in response to their talking. Overall they agreed that it was only natural for the iCat to use its potential this way. However, participants in the neutral condition also liked iCat with its more neutral behavior. After all, it is a robot and it should not try or pretend to be anything other than that. Moving and facial expressions would only look like a poor attempt to seem alive and it would likely annoy and distract from whatever you are doing. This finding shows how hard it can be to imagine something you have not experienced. Neither group of participants could imagine iCat being the opposite of what they experienced.

14.4 Modeling Affect in a Dialogical Robot

INES is an application designed at the University of Twente that allows students to learn procedural medical tasks. The procedures of subcutaneous injection have been implemented in our system as a first example. This task requires the execution of several subtasks, for example, taking care that the instruments are sterilized, that there is communication with the patient, and that the injection is done in a correct way. Besides as a research project on multimodal interaction, the INES application serves as a basis for research on affective computing.

14.4.1 Multimodal Interaction

In Fig. 14.3 we show the current INES configuration, consisting of a student interacting multimodally with the system using speech, keyboard, and a haptic device. It also shows a virtual tutor interacting with the student, a virtual

Fig. 14.3. Student interacting with the Intelligent Tutoring System

patient that can be addressed by the student and objects in the virtual world, e.g., the patient and the needle that can be manipulated using haptics and speech. The virtual tutor has been implemented as an embodied agent (a talking face). When it is considered useful it displays its (dis)approval by its facial expressions. It interacts with the student mainly by using speech recognition and speech synthesis.

The student can also communicate with the patient. For example, for this particular application, asking him or her to move his or her arm or to roll up a sleeve. Communication with patient or tutor is mostly related to the handling of the haptic device. This device is, for this particular application, represented as an injection needle in the virtual world that is displayed on the screen.

The input for the virtual tutor in the current INES environment consists of keystrokes, mouse movements and clicks, movements and force using the haptic device, and speech. There is a limited build-up of the interaction history. A more complete interaction history will be obtained by embedding the characteristics of a generic multimodal interaction architecture in the INES environment; see also Hofs et al. (2003). This earlier work contained models of multimodal interaction – in that particular case tuned to speech and pointing – that can be used for haptic input combined with speech, mouse, and keyboard input, and that allows the embedding of multimodal input in a discourse model that also keeps track of the history of the interaction. However, even now, the multimodal input, embedded in its situational and dynamic context, allows the tutor agent to make assumptions and from that compute possibly emerging emotions of the student that is performing a (sub-) task in the INES environment. Being able to respond in an appropriate way corresponding to the student's emotion (sympathizing) will make the tutoring process more effective (Kort et al. 2001).

This nursing education environment has been built using our own multiagent platform. The virtual tutor receives input from different agents, for example, from a collection of error agents that keep track of what the student is doing with the haptic device. The agents that currently have been implemented in the environment track the activities of the student, notice

the errors that are made, interact with the student, and change the teaching environment. In particular the so-called ErrorAgents know about the direct performance of the student: does the student use the right angle of the needle when trying to give an injection, what is the speed and what is the force that is used, has the needle been sterilized, does the student take too much time, does he or she have many questions or is asking too often about explanations, etc.

14.4.2 Affective Interactions

In our work on INES, we have also started to build a tutor agent that tries to be sensitive to the mental state of the student that interacts with it. Tutoring situations are essentially a social encounter, the goal of which is for a student to learn some task or acquire knowledge with the tutor acting in all kinds of ways to assist the student with this goal. The actions of a tutor are also not just restricted to pure instructions but they should also create the right emotional conditions for a student to act. The fact that the tutoring situation is a social encounter means that influencing the emotional state proceeds through social acts with emotion changing potential. For instance, the tutor has the status to judge (criticize or praise) the student for his actions.

The emotional state of the student contributes a lot to whether a student is motivated or challenged, which are key conditions for certain actions. Curiosity and puzzlement may lead to investigate problems. But also frustration 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. Learners can be motivated by challenging them, giving them confidence, raising their curiosity, and making them feel in control. These goals can be achieved by means of various tactics. The student can be challenged by selecting appropriately difficult tasks, or by having the difficulty emphasized or by having some kind of competition setup. Confidence can be boosted by maximizing success directly (praising) or indirectly ("it was quite a difficult task, you managed to do"). 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 the student can make choices and thereby influence the student's feeling of being in control.

In our prototype system we are experimenting with different kinds of pedagogical strategies: Socratic methods, active student learning, using deep explanatory reasoning, etc. Each of these requires different kinds of interaction (verbal and nonverbal) with the virtual tutor. The tutor in our system is aware of the history of the interactions, in particular the activity level of the student, the number and kinds of errors made by the student, and it uses this information to introduce affect in the interaction. The system not only figures out what teaching strategy fits the current context best, but also what kind of speech act is appropriate, which wording should be chosen – choosing between

a reprimanding or more encouraging turn of phrase – and which facial expression should be displayed. For instance, the tutor may ask the student a series of questions first, before moving on to having the student make the injection. Or it might decide to give a demonstration first.

Presently, we are not really keeping track of the student's emotional state. Instead, the tutor tries to make reasonable assumptions about the student's mental state based on information obtained from ErrorAgents and InputAgents. For the tutor we distinguish emotions that allow the tutor to feel content or discontent (i.e., joy and distress) and that allow the tutor to feel sympathy for the student (i.e., happy-for and sorry-for). Three agents have been designed that take care of emotions: the EmotionTutor, containing the tutor's emotions, the EmotionStudent, containing the student's emotions, and the EmotionalResponse, containing the algorithms that determine in what emotional way will be responded. Experiments to evaluate the affective behavior of the INES TutorAgent have been reported by Heylen et al. (2005) and Nijholt (2003).

As a next step in the INES research, we have looked at possibilities to measure more accurately what the student is actually experiencing during the exercise that has to be followed and to feed the tutor with this information. Lisetti and Schiano (2000) have shown that facial expressions can reveal something about the cognitive and emotional states of students interacting with a tutoring system. Still, it is not at all clear how a facial expression displayed during an interaction should be associated with a mental state of the student, let alone how to use them to optimize the tutoring process. Nevertheless, facial expressions appear and we can make an attempt to understand what triggered the expression in a particular context, and from that, hopefully, get some information about aspects of the mental state of the student that can be used to provide useful feedback and to adapt the teaching.

In a pilot experiment we collected video material from students interacting with the system; see Fig. 14.4.

We 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 (Scherer 1987; Wehrle et al. 2000).
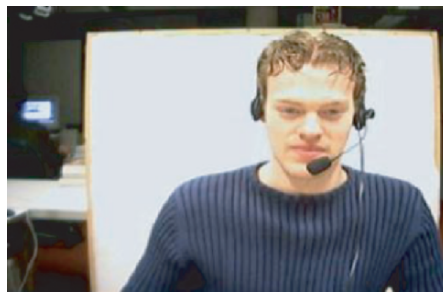

Fig. 14.4. A student working with the INES during the experiment

Scherer tries to present a coherent picture about what elements of a situation trigger what kinds of emotions. His model explains how an organism evaluates stimuli in a series of appraisal checks. The general idea is that the outcome of these checks results in specific facial expressions. This can be used to relate stimulus (situation), facial expression, and appraisal (mental state). This includes dimensions such as novelty (suddenness, familiarity, and predictability), pleasantness, goal significance (relevance, expectation of outcome, etc.), coping potential, and compatibility standards.

In the data we have looked at what facial expressions occur and we analyzed the situations in which they occurred in terms of the stimulus evaluation checks. In this way we can build a database that gives a crude indication of the relation between facial expressions and appraisal. Hence, when our system captures a facial expression in a particular situation it allows us to infer some aspects of the mental state of the user. In our corpus we found the following facial expressions: smile, raise eyebrows, pull down mouth corners, and frowns. Smiles often occurred when the students were manipulating the haptic device, as this seemed pleasurable. This is work in progress, but the aim is to derive a table associating elements of our particular 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 (Heylen et al. 2005).

14.5 Conclusions and Future Research

As technologies in the area of connectivity and computational platforms are moving from ubiquitous, i.e., available throughout the environment, to pervasive, i.e., embedded in out daily lives, and to ambient, i.e., creating intelligent environments, we observe that such systems will have perceptive and reasoning capabilities. This will in its turn lead to raised expectancies by end users.

In addition, the vision of Ambient Intelligence is advocating that technologies should be hidden into the background and that only the system's functionality should be available to the user. This introduces the need for advanced home dialog systems since users will still have to interact with the functionality embedded into the environment.

Combining this ambient and intelligent characteristic of future systems, we see an opportunity for the concept of social intelligence to facilitate user–system interaction. More specifically, we have investigated the effects of applying this concept to interactive systems in both the physical and virtual world. This research has led us to the conclusion that of social intelligence is an essential higher-level context for implementing intelligent and affective systems. Our research has shown that by the application of this concept of social intelligence we cannot only manipulate the user's perception of quality in an interactive system, but we can also significantly increase the user's acceptance of system intelligence embedded into the environment.