Multimodal Interaction in a Haptic Environment

Anton Nijholt Sander Kole Job Zwiers
University of Twente, CTIT
PO Box 217, 7500 AE Enschede, the Netherlands
E-mail: anijholt@cs.utwente.nl

Abstract

In this paper we investigate the introduction of haptics in a multimodal tutoring environment. In this environment a haptic device is used to control a virtual piece of sterile cotton and a virtual injection needle. Speech input and output is provided to interact with a virtual tutor, available as a talking head, and a virtual patient. We introduce the haptic tasks and how different agents in the multi-agent system are made responsible for them. Notes are provided about the way we introduce an affective model in the tutor agent.

1. Introduction

The Intelligent Nursing Education Environment System (INES) is an application that allows students to use multimodal interaction with a virtual reality (VR) environment and an embodied tutor to learn procedural tasks, e.g. to give a virtual patient a subcutaneous injection. 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.

This education environment has been built using our multi-agent platform. The virtual tutor receives input from different agents, for example, from error agents that keep track of what the student is doing with the haptic device. When appropriate the tutor displays its approval by its facial expressions and it interacts with the student using speech recognition and speech synthesis. In our prototype system we are experimenting with different kinds of pedagogical strategies, each of them requiring different kinds of interaction (verbal and non-verbal) with the virtual tutor. The tutor is aware of the history of the interactions, in particular the errors made by the student, and it uses this information to introduce affect in the interaction. The student can also communicate with the patient. For example, asking her to move her arm or to roll up a sleeve. Any communication with patient or tutor is related to the handling of the haptic device. This device is represented as a piece of sterile cotton or an injection needle in the virtual world that is displayed on the screen.

In Figure 1 we show the current INES configuration, consisting of a student interacting with the system using speech, keyboard and haptic device, a virtual tutor interacting with the student, a virtual 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.

There are more projects on haptics and graphics for medical training. They aim at allowing students to train on virtual patients rather than having to practice on a vulnerable human patient. Among them are projects on injection simulators, where trainees feel realistic forces when attempting to position the needle and inject a fluid and where different feedback can be obtained for different vein sizes and patient profiles, e.g. [1]. Unlike these projects, in our research we aim at modeling the interaction between student, system, tutor and patient. Hence, our research is on introducing a haptic modality in a multimodal interaction environment, rather than aiming at realistic rendering of haptics, graphics and animations.

Section 2 of this paper introduces the INES architecture. In section 3 we zoom in on the multimodal interaction and section 4 is devoted to the role of haptics. Notes on affect are in section 5. Future research is mentioned in section 6.

Figure 1. The INES system

Proceedings of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems
0-7695-2310-2/05 $20.00 © 2005 **IEEE**

2. The INES agent architecture

Our agent architecture allows different contents for the different modules and interaction modalities (haptics, speech, keyboard, and their integration) to connect the student to the different components of the system. Our agent platform allows introducing agents, defining their properties and defining their communication capabilities [4]. Several types of agents are distinguished: sensor agents (proactive) and processing agents (reactive). For example, in the chosen domain of subcutaneous injection a proactive 'sterile agent' looks every second if there are objects that are not sterile while they should be. On the other hand, there is, e.g., a reactive 'feedback agent' that acts when it receives a message from another agent; this feedback agent then will determine whether or not feedback should be given. Other agents help to observe the performing of tasks by a student and to provide explanation. From a global point of view the main three agents in the system are the TaskAnalyzerAgent, checking the correct order of the steps in the tasks and subtasks, the PhantomAnalyzerAgent, checking the speed, position, force and angle of the needle and the location of the objects during the haptic simulation, and the SpeechAnalyzerAgent, taking care of communicating and analyzing the speech input. ErrorAgents receive information (in XML format) from these AnalyzerAgents and are able to determine when something is wrong, what is wrong (angle, excessive force, exercise takes too much time, etc.), how harmful it can be, and to communicate this to the TutorAgent that can take appropriate actions such as informing the student, do nothing (let the student make an error), offer an other exercise or demonstrate a subtask. The 'harmfulness' of the error influences the tutor's emotion (joy and distress) and whether or not the tutor should respond with an appropriate affective act.

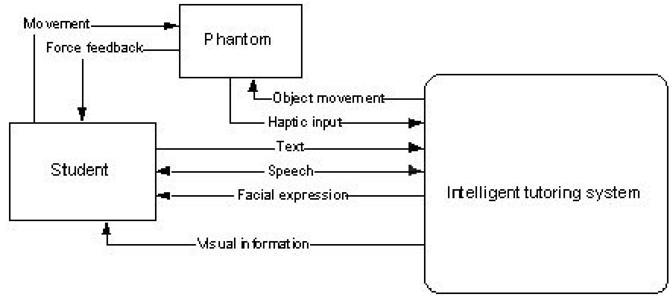

Figure 2. Context diagram of INES

After comparing signals coming from the haptic device with a 'correct model' a message can be constructed containing relevant information for deciding whether to give instruction and the kind of instruction to be given by the tutor. ErrorAgents that take care of the comparison with the 'correct model' are alerted in critical zones around objects in the virtual environment.

3. Multi-modal interaction with INES

In INES students interact through an interface that makes information available about the patient, the current task and its subtasks. The student can ask for demonstration and explanation. Text-to-speech synthesis is

Subtask:	**Conditions indicating the student is working on this specific subtask:**	**Checks needed:**
1. Positioning of the virtual patient through speech	• Speech detection.	• Check if the correct utterances are done by the student.
2. Disinfecting the injection area	• Selection of the sterile cotton using a pair of tweezers which is used as avatar for the PHANToM and as disinfecting equipment	• Check if the virtual patient was correctly positioned. • Check if the correct area is being disinfected.
3. Actual injecting of the needle in the injection area	• Selection of the syringe which is used as avatar for the PHANToM and for the injection.	• Check injection angle and speed. • Check whether needle touches only the correct upper arm. • Check whether injection is done in disinfected area.
4. Administering of the medication through the injected syringe	• Pressing the button available on the PHANToM.	• Check whether the needle is within the injection area. • Check whether the injection depth is correct (skin layer).
5. Removal of the needle from the injection area	• The student has just pressed the button on the PHANToM and medication is administered.	• Check whether a ten seconds waiting period has elapsed before the needle is retracted.

Table 1. Haptic subtasks with their conditions and needed checks

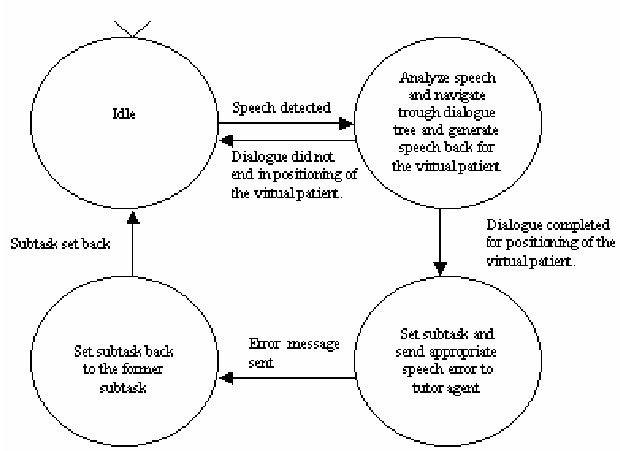

Figure 3. SpeechAnalyzerAgent

used to give feedback for a student. Part of the screen contains a view on the task in progress. In a first version of the system it showed videos for the different subtasks. In the current version we have a 3D VR environment where virtual humans play the roles of patient and, available on a second monitor, tutor and the nursing student can use speech input and a haptic device to perform an injection. In Figure 2 we display a context diagram of our system.

Multimodal interaction is made possible through haptic simulation, speech and visualization (changes in the environment). We designed a Java 3D virtual world (inhabited by the virtual patient) and a corresponding C++ PHANToM haptic world. Updates between the worlds travel in both directions and are done through a driver and with a frequency of 30 Hz. For example, object force feedback information and object 3D representation information (including the camera view and the virtual object that visualizes the PHANToM in the virtual world) from the virtual world are made available to the haptic world. Haptic actions update the virtual world. Certain actions cannot be performed by the haptic device. For instance, one of the subtasks that has to be performed is the positioning of the patient in order to be able to give the injection. The student can use speech input to perform this subtask. An agent has been introduced in the platform that takes care of the communication with the recognizer. The recognizer employs a grammar in which possible utterances have been specified.

There are interactions between student and patient and between student and tutor. Depending on the progress and the errors made by the student, the tutor chooses dialogue acts, i.e., responses to previous student activity. A list of dialogue functions for the tutoring process has been constructed using the SWBD-DAMSL dialogue acts. Each of the functions can only be executed when certain preconditions are satisfied [5]. Apart from the choice of dialogue act, taking into account previous dialogue acts, there is the choice of facial expressions in the tutor's face.

4. Haptics in the interface

In Table 1 we give an overview of the haptic tasks that have to be performed by the student. Conditions are shown for every subtask. These conditions cannot possibly be monitored all by the agent responsible for the analysis of the subtask the student is working on. This would require too much overhead for this agent. Other agents monitoring the multi-modal interaction components will send messages when these conditions are met. The agent responsible for the task analysis will then deactivate the currently running error agents and start the error agents belonging to the changed subtask.

For each subtask error agents are required to monitor the student's actions and send messages to the tutor agent about these actions. Each check that has to be done is listed in the table. From these checks the appropriate error agents are evoked. The tasks of these agents are listed below with the help of (deterministic) finite automata (DFA) descriptions. In Figure 3 we show the DFA for the SpeechAnalyzerAgent. It is responsible for analyzing the spoken interaction by the student with the virtual patient. Similarly not displayed here, we designed a DFA for the SterileCottonAgent. It detects errors on the disinfecting subtask of a student's exercise. The student is able to choose a piece of sterile

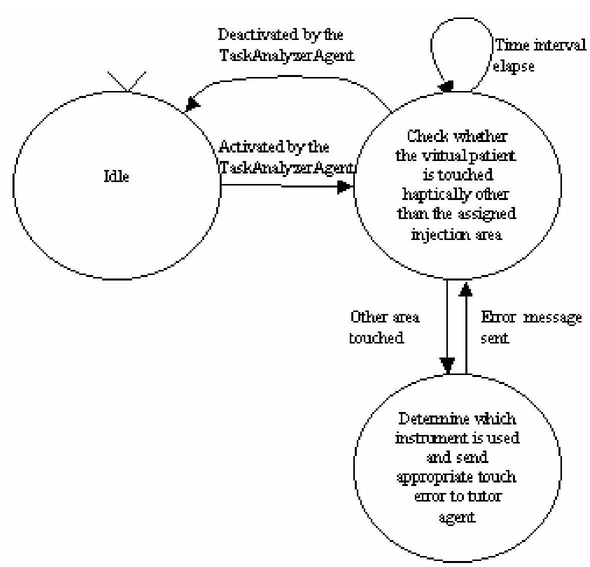

Figure 4. DFA for the TouchAgent

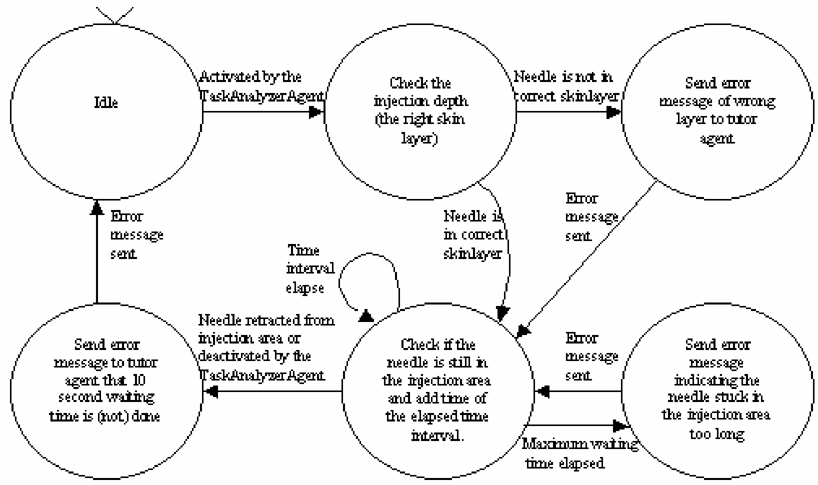

Figure 5. DFA for the PlungerAgent

cotton as a representation of the avatar of the PHANToM. The touching of the virtual patient with the defined medical instruments is guided by the TouchAgent (Figure 4). Its avatar represents a syringe or a pair of tweezers holding the sterile cotton with disinfectant fluid. Finally, in Figure 5 we show the PlungerAgent, responsible for detecting student errors that deal with movements and timing of the use of the needle. In the next section we discuss the way this tutor agent uses affect in its interaction with the student.

5. Affect in the INES tutoring system

The multimodal input, embedded in its situational and dynamic context, allows the tutor agent to make assumptions about possibly emerging emotions of the student. Being able to respond in an appropriate way corresponding to the student's emotion (sympathizing) will make the tutoring process more effective [6].

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 earlier mentioned ErrorAgents know about the direct performance of the student.

We do not really keep track of the student's emotional state. Instead, the tutor makes assumptions based on information obtained from ErrorAgents. 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). The tutor's emotional state needs to be translated into actions: what will be the next dialogue act, how will it be worded and what are the associated displays of nonverbal actions (intonation, facial expressions and body posture). This translation is also dependent on the tutoring strategy followed by the tutor. In our current implementation only rather modest approaches to these problems have been implemented. Three agents have been designed that take care of emotions: the EmotionTutor, containing the tutor's emotions, the Emotion-Student, containing the student's emotions, and the EmotionalResponse, containing the algorithms that determine in what emotional way will be responded. Tests to evaluate the affective behavior of the TutorAgent are reported in [3].

6. Conclusions and future research

INES has become a platform to integrate several of the lines of research that we pursue. The most important one is multi-modal interaction. The agent-based architecture introduced here allows an interaction between (internal) agents, tutor agent and student. The role of haptics and its integration with other modalities has been discussed. Obvious future research includes more realistic haptics and in particular trying to detect the emotions of a student rather than have assumptions about them [2].

References

[1] T. Dang, T.M. Annaswamy & M.A. Srnisvasan. Development and evaluation of an Epidural Injection Simulator with force feedback for medical training. *Medicine Meets VR* 9, Newport Beach, CA, 2000.

[2] M. Ghijsen, D. Heylen, A. Nijholt & R. op den Akker. Facial Affect Displays during Tutoring Sessions. In: Proceedings *IUI 2005*, San Diego, Ca, January 2005.

[3] D. Heylen, A. Nijholt & R. op den Akker. Affect in Tutoring Dialogues. *Applied Artificial Intelligence*. Taylor & Francis, 2005.

[4] Hospers, M., E. Kroezen, A. Nijholt, R. op den Akker and D. Heylen. An agent-based intelligent tutoring system for nurse education. In: J. Nealon and A. Moreno (Eds.) *Applications of Intelligent Agents in Health Care*. Birkhauser Publishing, Switzerland, 2003, pp. 143-159.

[5] D. Jurafsky. Switchboard SWBD-DAMSL Shallow-discourse-function annotation. Technical Report 97-02, University of Colorado, 1997.

[6] R.W. Picard. An affective model of interplay between emotions and learning. Proc. *Intern. Conf. on Advanced Learning Technologies (ICALT)*, 2001.

Proceedings of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems
0-7695-2310-2/05 $20.00 © 2005 **IEEE**