

Demonstration of the Emote Wizard of Oz Interface for Empathic Robotic Tutors

Shweta Bhargava¹, Srinivasan Janarthanam¹, Helen Hastie¹, Amol Deshmukh¹,
Ruth Aylett¹, Lee Corrigan², Ginevra Castellano²

¹School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh

²School of Electronic, Electrical and Computer Engineering, University of Birmingham

sb426, sc445, h.hastie, a.deshmukh, r.s.aylett@hw.ac.uk,
ljc228, g.castellano@bham.ac.uk

Abstract

We present a Wizard of Oz (WoZ) environment that was designed to build an artificial embodied intelligent tutoring system (ITS) that is capable of empathic conversations with school pupils aged between 10-13. We describe the components and the data that we plan to collect using the environment.

1 Introduction

We present a Wizard of Oz (WoZ) environment that was built as a part of the EC FP7 EMOTE project¹. The objective of this work is to collect multimodal interaction data to build an artificial embodied intelligent tutoring system (ITS) that is capable of empathic conversations with school pupils aged between 10-13. Specifically, the EMOTE (EMbodied-perceptive Tutors for Empathy-based learning) project aims to design and evaluate a new generation of robotic tutors that have perceptive and expressive capabilities to engage in empathic interactions with learners in schools and home environments. The project will carry out interdisciplinary research on affect recognition, learner models, adaptive behaviour and embodiment for human-robot interaction in learning environments, grounded in psychological theories of emotion in social interaction and pedagogical models for learning facilitation. An overview of the project can be found in (Deshmukh et al., 2013).

Wizard of Oz is an effective technique in Human Computer Interaction (HCI) studies where an interactive agent, which is not yet fully autonomous, is remotely controlled by a human wiz-

ard. However the participants who are interacting with the agent are not told that the agent is being remotely controlled. The wizard may be tasked to control one or many parts of the agent such as speech recognition and understanding, affect recognition, dialogue management, utterance and gesture generation and so on. Studies have shown that users “go easy” on computers during interaction and therefore interaction with “wizzarded” system are at the level of complexity that can be learned and emulated (Pearson et al., 2006).

The WoZ environment presented in this paper will be used to collect data to inform the algorithms for affect recognition and empathic dialogue management. The WoZ environment is designed to collect data on how human tutors aided with a robotic interface adapt to learners’ emotions and cognitive states in tutorial tasks. In this study, the wizard plays the same role as that of affect recognition and dialogue management modules in the actual final system.

2 Previous work

Wizard-of-Oz (WoZ) frameworks have been used in several studies since (Fraser and Gilbert, 1991) in order to collect human-computer dialogue data to help design dialogue systems. WoZ systems have been used to collect data to learn (e.g. (Strauss et al., 2007)) and evaluate dialogue management policies (e.g. (Cuayáhuitl and Kruijff-Korbayova, 2012)).

3 The EMOTE Wizard of Oz environment

The WoZ environment consists of the wizard’s desk, the interactive touch table, sensors, and the robotic embodiment as shown in Figure 1. The

¹<http://emote-project.eu/>

wizard will be seated in a different room away from the learner.

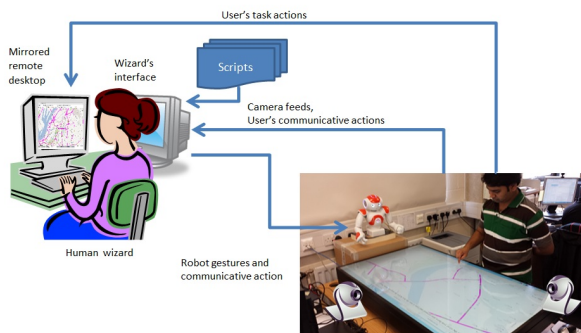


Figure 1: Wizard of Oz environment

3.1 Wizard's desk

The wizard's desk consists of two display screens. The touch table display at the user end will be mirrored on to one of the displays at the wizard's desk using which the wizard can observe the learner's activities related to the educational application. Another display will contain the Wizard Interface, a software application that allows the wizard to interact with the learner (see Figure 2). The Wizard Interface consists of four panels: task control, information, learner response and operations. In the task control panel, the wizard will be able to choose a task plan for the learner and access the tool and curriculum scripts (XML file). The tool script contains information on how to use the tools that are at the disposal of the learner. For instance, to create a marker on the map, one has to click on the appropriate tool and click on the map and so on. The curriculum script contains information on the skills that the learner needs to exercise or develop during his interaction with the system. For instance, in order to identify the right direction, the system will present the mnemonic phrase "Naughty Elephants Squirt Water" in various forms such as a hint, question, pumping move, etc. to provide support to the learner. The information panel contains the video feed from two cameras (see Section 3.4). This will allow the wizard to determine the affective state of the learner. The learner's response to the agent's utterances (such as answering questions in the curriculum scripts) will also be displayed in the learner response panel. Finally, the operations panel provides options for the Wizard to respond to the learner based on the tools

and curriculum scripts. These responses are either customised or predefined. The customised responses facilitate the wizard to execute robot movements on lower level (individual head, arm movements) and predefined responses contain a list for combined predefined speech, sound and behaviours.

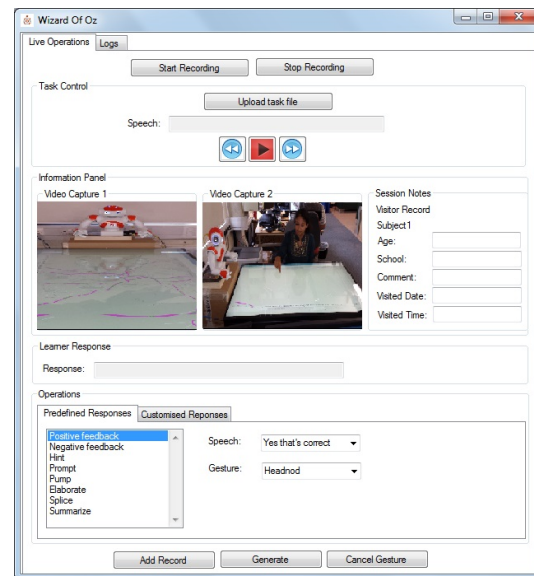


Figure 2: Wizard's Interface

3.2 Touch table

The interactive touch table is a 55 inch Multitaction table capable of sensing multiple touch events simultaneously. The educational application is displayed on the table surface. A map based application has been developed to teach learners basic and advanced map reading skills (see Figure 3). The touch interface allows the learner to use touch to click, drag and zoom the map. The application has two panels of GUI objects such as buttons and text boxes namely, the tools panel and the interaction panel. The tools panel consists of tools that the learner can use to manipulate the map, while using the interaction panel the learner can interact with the tutor. Some of the tools that are currently available are to get grid references for a position on the map, dropping markers on the map, change map types, etc. For instance, if the tutor asks a yes/no question, the learner can respond by pressing the yes or the no button. The learner can answer the tutor's questions by typing into the text box in the interaction panel.

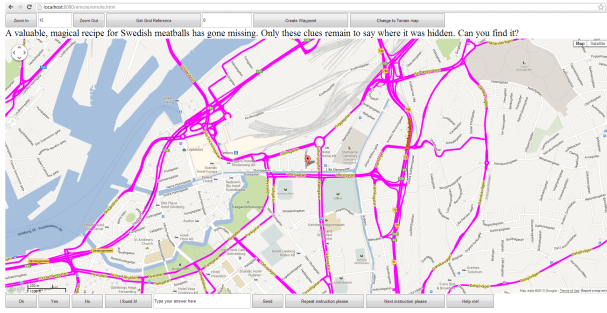


Figure 3: Map reading skills application

3.3 Robotic embodiment

The robotic embodiment is a Nao robot (torso version) that sits on the side of the touch table. It is capable of head, arm and body gestures in addition to synthesised speech. The robot receives the text and gestures selected by the wizard through the Wizard Interface. Tutor's utterances will be synthesized into speech using the in-built text to speech (TTS) engine while the gestures are realised using appropriate head, arm and body motions. To increase naturalness, the robot will also have idle movement in-between wizard selections.

3.4 Sensors

The environment has an array of sensors such as two video cameras and a Kinect sensor. A Kinect sensor and a video camera are placed in front the learner. Another camera is placed in front of the robot (as shown in Figure 1).

4 Data collection

In this section, we discuss the data that we aim to collect using the WoZ environment. We intend to collect these data during experiments where human tutors play the wizard's role and the learners from in the 10-13 year age-range will play the role of learners. The task for the learner is to carry out an expedition using the map application that he or she is provided with. In order to solve the steps of the expedition, the learner will have to exercise his/her map reading skills. Map reading skills such as compass directions, contour lines, grid lines, etc. will have to be exercised using appropriate map tools provided in the application. The tutor's role is to observe the learner responses (both verbal and physical) and respond to them appropriately using the interaction panel in the Wizard Interface application.

Simultaneous video feeds from two cameras and the Kinect sensor will be recorded during the tutor-learner interaction. These data will be further used for affect recognition tasks based on learner's head, arm and body gestures. The interaction between the tutor and the learner in terms of tutor dialogue actions, utterances and learner responses in terms of button presses will also be logged.

5 Demo

We propose to demonstrate the WoZ environment set up using two laptops: learner desktop with the map application and another with the wizard's interface. The learner desktop will also display a simulated Nao robot. We will also exhibit the logs that we collect from the pilot studies with a Geography teacher acting as the wizard tutor and school pupils as tutees.

Acknowledgements

This work was partially supported by the European Commission (EC) and was funded by the EU FP7 ICT-317923 project EMOTE. The authors are solely responsible for the content of this publication. It does not represent the opinion of the EC, and the EC is not responsible for any use that might be made of data appearing therein.

References

- H. Cuayáhuitl and I Kruijff-Korbayova. 2012. An Interactive Humanoid Robot Exhibiting Flexible Sub-Dialogues. In *Proceedings of the NAACL-HTL, Montreal, Canada*.
- A. Deshmukh, G. Castellano, A. Kappas, W. Barendregt, F. Nabais, A. Paiva, T. Ribeiro, I. Leite, and R. Aylett. 2013. Towards empathic artificial tutors. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*.
- N. Fraser and G. N. Gilbert. 1991. Simulating speech systems. *Computer Speech and Language*, 5:81–99.
- J. Pearson, J. Hu, H. P. Branigan, M. J. Pickering, and C. Nass. 2006. Adaptive language behavior in HCI: how expectations and beliefs about a system affect users' word choice. In *Proceedings of the SIGCHI conference on Human Factors in computing systems, Montreal*.
- P. M. Strauss, H. Hoffmann, and S. Scherer. 2007. Evaluation and user acceptance of a dialogue system using Wizard-of-Oz recordings. In *Proceedings of 3rd IET International Conference on Intelligent Environments*.