What Is It Like to Be a Bot? Toward More Immediate Wizard-of-Oz Control in Social Human–Robot Interaction

Sam Thellman
Dept. of Computer and Information Science,
Linköping University
Linköping, Sweden
sam.thellman@liu.se

Jacob Lundberg
Computer Science and Engineering Program,
Linköping University
Linköping, Sweden
jaclu010@student.liu.se

Mattias Arvola
Dept. of Computer and Information Science,
Linköping University
Linköping, Sweden
mattias.arvola@liu.se

Tom Ziemke
Dept. of Computer and Information Science,
Linköping University
Linköping, Sweden
tom.ziemke@liu.se

ABSTRACT
Several Wizard-of-Oz techniques have been developed to make robots appear autonomous and more social in human–robot interaction. Many of the existing solutions use control interfaces that introduce significant time delays and hamper the robot operator's ability to produce socially appropriate responses in real time interactions. We present work in progress on a novel wizard control interface designed to overcome these limitations: a motion tracking-based system which allows the wizard to act as if he or she is the robot. The wizard sees the other through the robot's perspective, and uses his or her own bodily movements to control it. We discuss potential applications and extensions of this system, and conclude by discussing possible methodological advantages and disadvantages.

Author Keywords
Social robotics; social interaction; Wizard of Oz; HRI methodology; HTC Vive; humanoid; Pepper robot.

ACM Classification Keywords
• **Human-centered computing~Interaction techniques**
• *Human-centered computing~User studies*

INTRODUCTION
The problem of how to simulate robot autonomy and responsiveness in social human–robot interaction (HRI) is relevant to exploring prospective robotic systems that are not yet technically realizable [7, 9]. It is also a substantial problem in research that utilizes interaction with robotic systems to study human social cognition [12]. The "Wizard of Oz" (WoZ) approach to this problem is to use some technical solution that allows a human operator (the "wizard") to monitor and control some part(s) of the behavior of the robot system, usually without the awareness of the person interacting with the robot [3, 9].

Several wizard control interfaces have been developed to simulate autonomy in HRI [e.g., 4, 5, 10]. However, these systems often use: (A) monitoring mechanisms which restrict the wizard's ability to discern subtle human behavior crucial to the production of an appropriate response, such as position-fixed cameras and cameras that do not capture the full-body and facial expressions of the human interaction partner, and; (B) control mechanisms which restrict the wizard's ability to regulate the behavior of the robot freely and timely. These limitations often introduce significant time delays to wizard responses and makes it difficult for the wizard to respond in socially appropriate ways to human social behavior. Furthermore, Martelaro [7] points out that the fact that many of the existing wizard control interfaces are based on custom-made GUI software and hardware controllers leads to (C) a standardization issue: control mechanisms map differently (or not at all) across different robot platforms that each have different restrictions as to what the robot can and cannot do. High-level behavior control interfaces have been proposed to address this issue [6]. However, while these solutions enable researchers to work with different types of robots they do not bring the simulation of the robot's behavior closer to the appearance of autonomy in social interaction.

We created a novel WoZ system to overcome the above limitations (A) and (B), and to alleviate (C), using a virtual-reality headset and two handheld controllers (HTC Vive) (Figure 1). Our approach was based on the assumption that the best way to simulate autonomy in social human–robot interaction is to, as closely as possible, let a human operator monitor the interaction and control the robot as if he or she was *in the position of the robot.* The goal was, more specifically, to enable the wizard to perceive the world from the perspective of the robot's "eyes", including the ability to shift perspective by moving the robot's head, and to be able to freely move any of the robot's movable physical parts through the movement of one's own corresponding limbs. A second assumption was that human motion is translatable into a wide range of humanoid robot motions, and that the standardization issue (C) therefore can be somewhat remedied, while at the same time improving the quality of the resulting simulation. Inspired by Nagel's

HAI 2017, October 17–20, 2017, Bielefeld, Germany.

ACM ISBN 978-1-4503-5113-3/17/10...$15.00.
Include DOI string here from eform confirmation

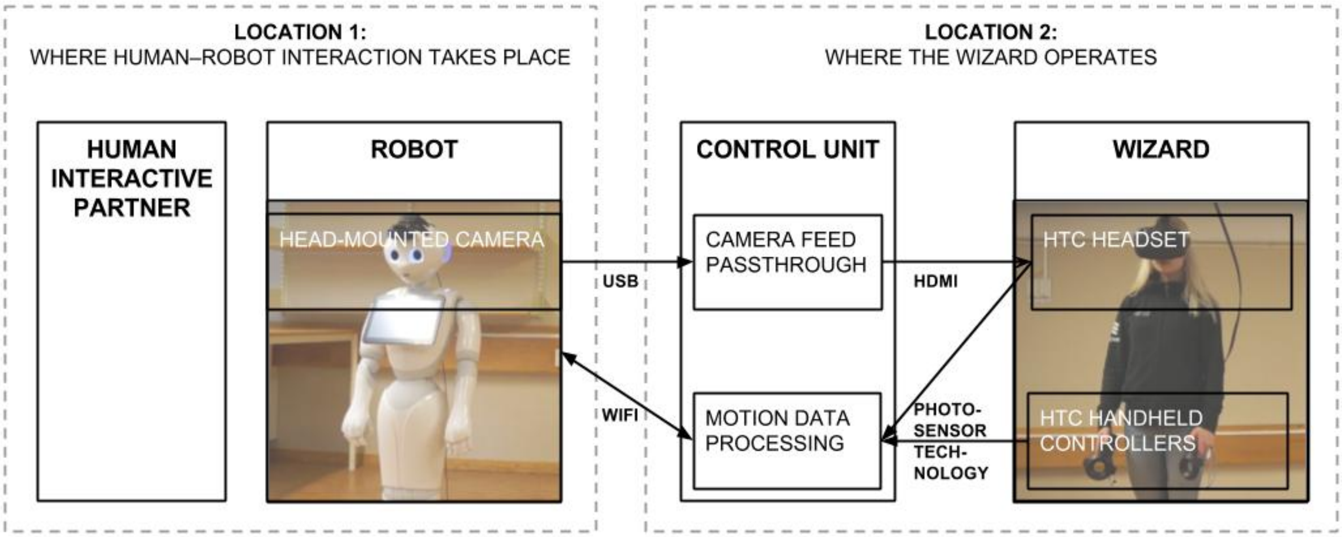

Figure 1. System overview

seminal paper "What is it like to be a bat?" [8], we wanted to create a wizard experience – or approximation – of what it is like to *be* the robot, and not only to control it.

TECHNICAL DESCRIPTION

The developed system tracks the wizard's head and arm movements using an HTC Vive headset and two HTC Vive handheld controllers (Figure 2). These data are processed on a control unit (a PC) which receives wizard motion data and passes commands to a Pepper humanoid robot which mirrors the movements of the wizard upon execution. Arm movements are processed using an inverse kinematics solver which interfaces with the robot manufacturer's Cartesian space API. As a result, the wizard can control 2 degrees of freedom (DOF) in the head of the robot (yaw and pitch), 5 DOF in the arms (shoulder pitch and roll, elbow yaw and roll, and wrist yaw), and 1 DOF in the robot's wheels (body rotation/direction) using his or her own body movements. The wizard can control the robot's planar movement any direction (2 DOF; forwards, backwards, left, right or anywhere in between) using the touchpad on one of the HTC Vive controllers[1]. The robot's hands can be opened and closed using the trigger button on either controller (1 DOF). The surroundings of the robot are captured using one high-resolution camera (1920x1080 pixel resolution running at 30 frames per second) mounted just above the robot's eyes. The camera feed is relayed through the control unit to the wizard's headset which fully encloses his or her field of view. The perspective of the camera feed shifts as the wizard moves the robot around in the room or turns its head. See Figure 1 for a diagrammatic overview of the system and Figure 3 for an illustration of mimicked wizard movements. Further details on the implementation of the design can be found in a technical report [2].

[1] We hope to implement room-scale movement to replace this mechanism, which would allow the wizard to move the robot in any chosen direction.

INITIAL FORMATIVE TESTS

This section reports on two initial formative tests that took place during the development of the system. At this point in development, the camera feed, and the head, arm and body movement was operational; movement of wrists, hands and body rotation using gestures was not yet implemented. Both tests focused on the experiences of the human operator or "wizard" controlling the robot. There was, however, no designated interaction partner with whom the wizard was tasked to interact. The purpose of these tests was to get an initial assessment of how well the system performed technically from the perspective of a naïve user, and was not primarily directed at the system's ability to convincingly simulate autonomous behavior and responsive social interaction.

Method

Four persons were asked to use all the system's available functionality in a controlled testing environment and to subsequently answer three questions: "Where do you feel that your consciousness resides; in your body or in the body of the robot?", "When do you feel that the experience is

Figure 2. HTC Vive headset (left) and HTC Vive handheld controllers (right).

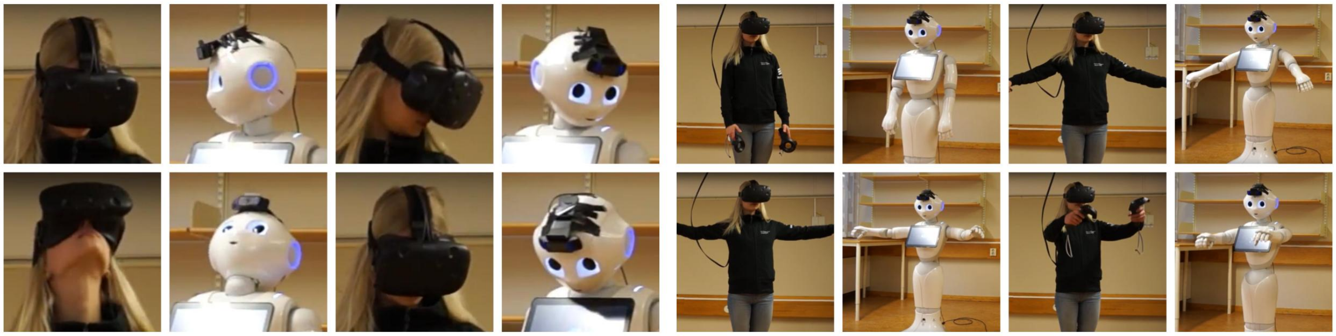

Figure 3. Robot head and arm movements generated by corresponding human wizard movement.

problematic?", and "What do you think of when you see yourself as a robot?". When the test begun, the robot was turned toward a mirror so that the user could get a view of the robot body that they were now in control of (Figure 4).

Thirteen participants at the 2017 Swedish annual conference on Cognitive Science and Information Technology (KVIT) were also asked to freely state their opinions about the system after using it (Figure 4).

Results

Where do you feel that your consciousness resides; in your body or in the body of the robot?
A majority of the participants claimed that their frame of reference shifted from the perspective of their own body to that of the robot, as long as they were standing still and only moved their head and arms.

When do you feel that the experience is problematic?
A majority of the participants experienced some lag in the arm-movements, and that the arms of the robot did not always behave as one's own arms. Some of the participants felt that moving around using the controllers confused their perception of the position of the robot's body, and some experienced the field of view provided through the headset as too limited and stated that this was somewhat disruptive.

What do you think of when you see yourself as a robot?
A majority of the participants stated that they were partly or fully "convinced" by the mirror impression that the robotic body was their own.

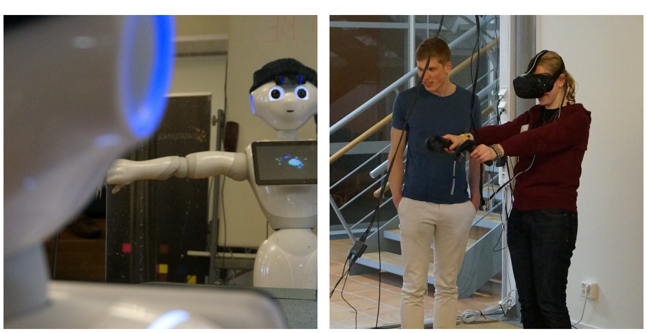

Figure 4. Robot looking at mirror image (left) and a person using the system at KVIT conference (right).

Comments during the KVIT conference
Six out of the thirteen persons used the words "fun" or "cool" to describe the experience. Four people thought that the experience was "convincing". Three persons pointed toward the arm movements as problematic. Three persons thought that the experience was "confusing" or "weird" in the beginning. Three persons stated that head movements were responsive. Two persons complained about the limited field of view.

CONCLUSIONS

Our initial testing suggests several advantages with the system. Firstly, participants reported having no issues controlling the robot using the system (barring a few glitches in the translation of arm movements). We also observed that the participants could use the system directly with minimal instructions. We believe that these two advantages stem from the fact that our system allows the wizard to produce robot behavior using his or her instinctive embodied ways of acting, and that the wizard therefore knows intuitively how to control the robot. This is not a quality of most existing WoZ solutions, which utilize traditional input methods such as keyboard, mouse, or joystick to control the robot. Secondly, most participants reported that they experienced a shift from the perspective of their own physical self to the perspective of the robot when using the system, i.e. they experienced what is was like to be the robot. We attribute this experience to the combination of filling the wizard's field of view with the robot's camera feed and giving the wizard the ability to shift the perspective freely by moving his or her head. We also noted issues with the system, including experiences of motion sickness, lag in arm movements, and a limited field of view in the headset. These preliminary results from our initial user study need to be re-examined in more detail in the future.

As a next step, we will conduct a more comprehensive testing of the system, including assessments of persons interacting with the controlled robot and comparisons with more traditional input devices for WoZ systems. We will also work on technical improvements to the system, including the addition of a verbal communication channel between the wizard and the interactive partner, room-scale movement which would allow the wizard to move the robot in any direction by bodily movement in physical space. We will also implement a higher quality wireless camera to improve the wizard's monitoring experience, and we will investigate the possibility to add haptic devices that can provide the wizard with kinesthetic awareness of the movements of the robot's joints. Another venue of exploration is haptic feedback to close the sensory-motor loop, and add operator immersion. We hope that these technical additions will mitigate some of the issues experienced by users of the system during our initial formative testing, and that the planned tests will lead to further improvements.

We also would like to note here that eye-contact plays an important role in social interaction and fulfils several functions such as gathering feedback on the other person's reactions [1]. Eye-contact is not possible to achieve with most of the robotic platforms commercially available today, including the one used in the system presented here. However, we would like to see the implementation of eye-tracking tracking technology that tracks the gazes of the wizard and represents them in the eyes of the robot in a similar future system, as we believe that this feature is an important step toward achieving a truly realistic simulation of autonomy and responsiveness in social human–robot interaction.

In conclusion, the initial formative testing of the system indicated that the system succeeded in putting the human operator in the perspective of the robot (i.e., to create the impression of what it is like to be a bot). We believe that this unique quality has distinct advantages compared to existing WoZ systems. Specifically, we believe our approach has potential to decrease wizard response times and to generate more socially appropriate behavior in wizard controlled human-robot interaction. Our next step is to investigate this by testing with people interacting with a robot controlled using the system. Our initial testing also suggested that the interface to the controlled robot became transparent to the operator in the sense that new users knew how to use it intuitively. With this in mind, we also believe that our system has potential to alleviate the standardization issue that arises in WoZ controlled human–robot interaction in part due to the large variety of control mechanisms used [7]. However, the quality of the simulation generated using the system presented here will depend on the unique morphology of each specific robot, and the technical problem of mapping human movement to the physiology of the robot remains for each specific type of robot. Methods to mitigate this issue have been proposed, but need further validation [11].

ACKNOWLEDGEMENTS
The technical work on the implementation of the system presented in this paper was carried out by Robin Christensen, Ylva Lundegård, Emil Segerbäck, Partik Sletmo, Teo Tiefenbacher, Jon Vik and David Wajngot. We also thank the people that participated in the initial formative testing sessions, including attendees at the 2017 KVIT conference.

REFERENCES
1. Argyle, M., & Dean, J. (1965). Eye-contact, distance and affiliation. *Sociometry*, 289-304.
2. Christensen, R., Lundberg, J., Lundegård, Y., Segerbäck, E., Sletmo P., Tiefenbacher, T., Vik, J., & Wajngot, D. (in press). Hur är det att vara en robot? [What is it like to be a bot?]. Linköping: Linköping University.
3. Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of Oz studies—why and how. *Knowledge-based systems*, *6*(4), 258-266.
4. Hoffman, G., Kubat, R., & Breazeal, C. (2008). A hybrid control system for puppeteering a live robotic stage actor. In *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on* (pp. 354-359). IEEE.
5. Huttenrauch, H., Green, A., Norman, M., Oestreicher, L., & Eklundh, K. S. (2004). Involving users in the design of a mobile office robot. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *34*(2), 113-124.
6. Lu, D. V., & Smart, W. D. (2011). Polonius: A Wizard of Oz interface for HRI experiments. In *Proceedings of the 6th international conference on Human-robot interaction* (pp. 197-198). ACM.
7. Martelaro, N. (2016). Wizard-of-Oz Interfaces as a Step Towards Autonomous HRI. In *2016 AAAI Spring Symposium Series*.
8. Nagel, T. (1974). What is it like to be a bat?. *The philosophical review*, *83*(4), 435-450.
9. Riek, L. D. (2012). Wizard of Oz studies in HRI: a systematic review and new reporting guidelines. *Journal of Human-Robot Interaction*, *1*(1).
10. Robins, B., Dautenhahn, K., Te Boekhorst, R., & Nehaniv, C. L. (2008). Behaviour delay and robot expressiveness in child-robot interactions: a user study on interaction kinesics. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction* (pp. 17-24). ACM.
11. Van de Perre, G., De Beir, A., Cao, H.-L., Gómez Esteban, P., Lefeber, D., & Vanderborght, B. (2016). Reaching and pointing gestures calculated by a generic gesture system for social robots. *Robotics and Autonomous Systems, 83*, 32-43.
12. Wykowska, A., Chaminade, T., & Cheng, G. (2016). Embodied artificial agents for understanding human social cognition. *Phil. Trans. R. Soc. B*, *371*(1693).