What makes human so different? Analysis of human-humanoid robot interaction with a super Wizard of Oz platform

Guillaume Gibert, Maxime Petit, Florian Lance, Grégoire Pointeau and Peter Ford Dominey

Abstract— **Humanoid robots are more and more realistic but these systems still fail to be as friendly and natural as humans in interaction. Behavioural models of interaction controlling these robots cannot capture and replicate the extreme complexity of human communication yet. To determine the real limitations and key factors one must impose on a behavioural model to maintain a human-robot interaction as natural and efficient as a human-human interaction, we have built a super Wizard of Oz setup. This platform consists of a FaceLab sensor able to track a confederate's rigid and non-rigid face motion and gaze in real time and of an iCub robot able to replicate the confederate's movements. An evaluation of the platform was performed: the confederate's movements were accurately tracked and replicated by the robot with less than 200 ms delay. Thanks to binaural microphones placed in the robot ears and a video camera situated behind the robot, the confederate perceives the scene being in the robot's place. By manipulating specific movements without modifying the rest of the dynamics, the platform can be used to determine the acceptable limits for the human partner for various parametric manipulations. For instance, we have started investigating the effect of damping head movements in dyadic interaction.**

I. Introduction

Humanoid robots use advanced technologies and theories and are getting more and more complex. Yet, these systems still fail to be as friendly and natural as a 'real' human in interaction. The main reason of this failure may be due to the extreme complexity of human multimodal communication. The common approach consists of building behavioural models of interaction from theories and/or observation of real data. In fact, the implementation of these models commands the limitations of those systems as they cannot replicate the richness of human behaviours. Recently, a certain number of teams have started investigating the role of certain nonverbal behaviours in dyadic interaction by turning around the problem. They used an enhanced Wizard of Oz (WoOZ) setup that was mirroring some facial and eye movements on a robot face rather than proposing a set of pre-defined responses as in the traditional WoOZ [1]. For instance, Hiroshi Ishiguro and colleagues [2, 3] developed an android robot Geminoid HI-1 which closely resembles its scientific originator. It can be remotely controlled by teleoperation: a confederate's lips and head movements are cloned on the robot face. The authors have started investigating how real human feel when interacting with this 'almost' human [3]. The main issue is that the confederate perceives the scene as a 3rd person. Another system composed of an eye and head tracker, a robot head, a pair of camera motion devices (robot eyes) and a teleoperation link that connected the motion tracker to the motion devices has been proposed [4]. The confederate watches and reacts to the video stream of the person interacting with the robot. This system is more accurate than the previous one for cloning rigid head motion and gaze. On the other hand, the use of a non-realistic head robot does not allow replicating non rigid facial movements (lips, jaw, eyebrows, etc.). Recent results [5] have shown that it is possible for a human and a rat to interact and to fulfill the tasks of a game using similar technique. The human was represented in a rat arena by a small robot that was slaved to the human's movements, whereas the tracked rat was represented to the human in the virtual reality by a humanoid avatar. Another system Telesar V enables a user to bind with a dexterous robot and experience what the robot can feel from its fingertip when manipulating and touching objects remotely [6]. In this system, the operator can feel the robot's body as his own body through visual, auditory, kinesthetic sensation. However, the user must wear a virtual reality head mounted display and motion capture gloves.

*Research supported by the ANR SWoOZ project (11 PDOC 019 01) and the FP7 EFAA project (270490).

G. Gibert is with the Stem Cell and Brain Research Institute, INSERM U846, Bron, France and with the Université Lyon 1, 69003 Lyon, France (corresponding author, phone: +33 (0)4-72-91-34-41; fax: +33 (0)4-72-91-34-61; e-mail: guillaume.gibert@ inserm.fr).

M. Petit, F. Lance, G. Pointeau & P. F. Dominey are with the Stem Cell and Brain Research Institute, INSERM U846, Bron, France and with the Université Lyon 1, 69003 Lyon, France (e-mail: {firstname.name}@ inserm.fr).

II. Motivations and Objectives

The proposed system is designed to respond to the following question: **Which are the aspects of human-likeness that are more relevant for interaction?**

To determine the real limitations and key factors one must impose to a behavioural model to maintain a human-robot interaction as natural and efficient as a human-human interaction, we have built and used a super-WoOZ setup. The originality of this project relies on the full control of robot's behaviours by a human becoming a puppeteer and on the real time manipulation of specific behaviours. In our setup the user only wears headphones and can easily interact without any additional sensor. Parametric manipulations can be applied to one or several behaviours (for instance, gaze, head movements, mouth movements, blinks, and acoustics) at a time. Manipulations could be varied, for example adding a constant delay while preserving the head movements or adding a vertical and/or horizontal offset to gaze trajectories. The question I would like to ask to the invited speakers is:

III. METHODS

A. Procedure

The *super WoOZ* platform consists of:

- a sensor (FaceLab5, Seeing Machines) to determine the confederate's face rigid and non-rigid motion and gaze position at any time;
- a software program to apply online manipulation to specific parameters;
- a humanoid robot (iCub [7]).

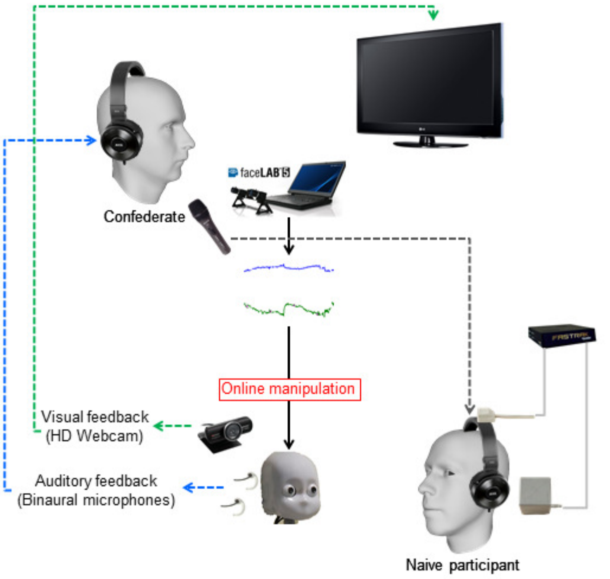

Figure 1: Schematic representation of the super Wizard of Oz setup. A FaceLab sensor captures the confederate's rigid head motion and gaze position continuously; then, an online manipulation can be applied to any parameter; finally, the iCub robot replicates the original or manipulated confederate's movements.

A schematic representation of the setup can be seen on Figure 1. The FaceLab sensor (Seeing Machines) provides accurate measurement of the confederate's eye gaze (05-1° of rotational error) and head rigid motion (±1mm of translational error and ±1° of rotational error) On the robot side, the iCub robot (http://www.icub.org/), a 1 metre high humanoid robot with eyes, head, arms and hands fully controllable in real-time. Between the sensor and the robot, a software program captures the gaze and face rigid and non-rigid motion data from the FaceLab sensor and applies (or not) modifications to the signals before sending it to the robot which mimics the confederate's gaze and head/face motion. A perfect correspondence exists between the robot's and the confederate's movements for the eye and head rotations. The Facelab sensor provides the landmark positions on the face and more specifically on the lips and the eyebrows. This information is used to estimate the movements of the lip/jaw and the eyebrows which are then simulated on the robot face using a set of diodes. Three different eyebrow shapes were displayed on the robot face depending on the confederate's face. Each eyebrow was considered separately. The robot mouth could be either open or close depending on the level of opening of the confederate's mouth (see Figure 2). In addition, the eye closure was estimated using the FaceLab sensor and replicated on the robot eyelids. The naïve participant wears headphones to hear the confederate's voice captured by a microphone remotely. A motion tracking sensor is attached to the headphones to track his head rigid movements.

B. Auditory-visual Feedback

To bind the user into the robot, the user must sense the scene as if he was the robot. Auditory and visual feedbacks were provided to the experimenter using binaural microphones (MS-TFB-2, The Sound Professionals, Inc.) and a High Definition (HD) webcam (Creative Live Cam Socialize HD). The binaural microphones create a 3D stereo sound sensation to the listener of actually being in the place of the 'person' (i.e. the robot) wearing them. The HD video stream was displayed on a screen allowing the experimenter to sense the scene as the robot would. The confederate can interact freely and naturally with his human partner.

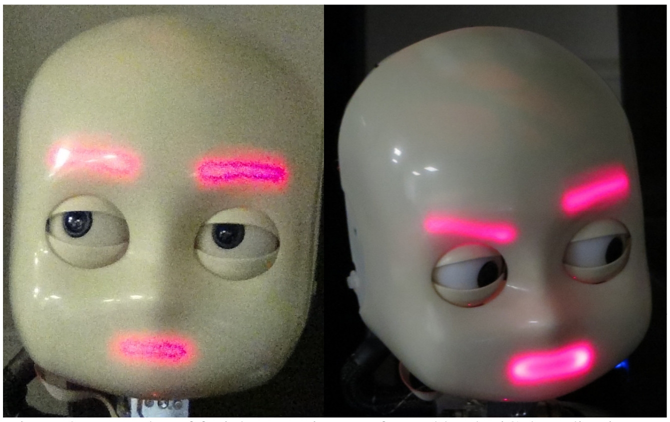

Figure 2: Examples of facial expressions performed by the iCub replicating the confederate's facial movements. The robot mouth can replicate two different shapes either closed or open depending on the confederate's mouth movements. The eyebrows can replicate three different shapes depending on the confederate's eyebrow movements.

IV. PLATFORM EVALUATION

In the SWoOZ setup, the robot mimics the confederate's head and gaze trajectories. The FaceLab sensor estimates the rigid head motion and gaze trajectories of anybody positioned in front of the sensor after a quick (less than 5 seconds) calibration procedure. This sensor can work in two modes: a real-time one with 50 ms delay and a precision mode with 2500 ms delay. Once the data are estimated, they are sent to the iCub robot using YARP commands (in velocity control mode) [8]. The robot mimics the estimated confederate's movements when receiving the YARP commands. It is necessary to assess the delay and the quality of transfer. To perform this evaluation a specific setup was created. The confederate's rigid head motions (and gaze) were estimated with the FaceLab sensor. They were saved on hard drive and sent to the iCub robot. The robot's head motions were estimated using a motion tracking system (Fastrak, Polhemus). This motion capture system tracked the position and orientation of a small sensor with respect to a transmitter at 120 Hz with low-latency (< 4ms). The sensor was positioned on the left side of the robot head. The confederate was seated next to the iCub robot. During the recording, the confederate was asked to perform two series of six head rotations around each axis (yaw, pitch and roll) and also movements towards the left/right, up/down, toward/away from the sensor. The setup allowed recording the head movements using one separate sensor for the confederate and the robot.

The rendering time was estimating using the series of head rotations performed by the experimenter. Examples of head rotations (pitch, yaw, and roll) recorded by the Facelab (experimenter) and the Fastrak (robot) are displayed in Figure 3. A zero-crossing procedure was applied to the head rotation signals. The time difference between the zero-crossing instants for the user and robot head movements was computed. The mean delay between the Facelab and Fastrak head rotations was 136 ms with a standard deviation of 28 ms. The minimum value was 100 ms, the maximum value was 200 ms and the mode was 150 ms.

The movement accuracy was also estimated with the head rotations data. The Facelab data was delayed by 150 ms (mode of the delay values). The difference signals were then computed between the rotations around the three axes of the robot and the experimenter. The average rotational error was very low (< 10e-3 rad) with maximum absolute values less than 0.14 rad. The rotational errors are mainly due to gravity.

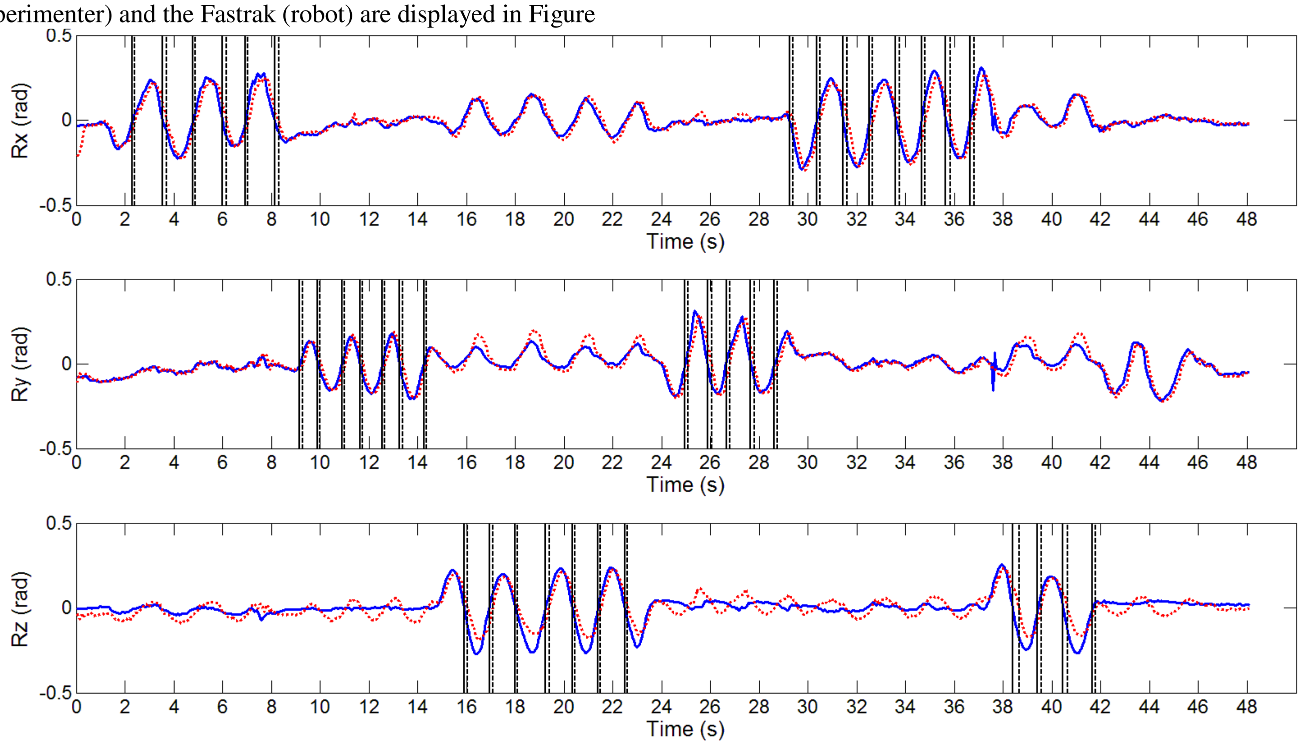

Figure 3: Head rotation signals for the user (blue solid line) and the robot (red dotted line). The robot head rotations were performed after the experimenter head rotation. The delay was estimated using a zero-crossing procedure. The zero-crossing instants are represented by black solid lines for the FaceLab signals and black dotted lines for the Fastrak ones.

Indeed, the robot head is quite heavy compared to the neck joint and there is a module in charge of gravity compensation running continuously. While the head is trying to go to the desired position, the gravity compensation module creates some oscillations that can be seen in Figure 3 on the Rz axis at the beginning of the recording for example. The eyes are not affected by this module and eye rotations are accurately transferred from the confederate to the robot.

V. Damping Head Movements

We have started investigating the role of head movements during human-humanoid robot interaction with the *super WoOZ* platform. Our experiment followed the study proposed by [9]. In this study, confederates' head movements and facial expressions were manipulated in real-time during videoconference conversations by tracking them (using Active Appearance Models) and reconstructing an avatar face [9-11]. Results of this experiment show that increase in head nods and lateral head turns in both naïve participant and confederate during dyadic interaction were noticed if attenuation was applied on the confederate's avatar head movements [9]. Our aim was to see if similar interaction loop effect can be created during human-humanoid robot interaction.

A. *Method*

One person participated in the study. He had self-reported having normal hearing and normal vision. The participant was instructed to freely interact with the iCub robot during 8 minutes. He was informed that the robot face/head and eye movements were controlled by the confederate's own movements. The robot voice was the confederate's voice.

The participant was asked to wear headphones to listen to the confederate's voice. An electromagnetic sensor was attached to the headphones to track the participant's head rigid movements. The confederate's head rotations were randomly attenuated by a factor 2 during 4 periods of 1 minute of interaction. The rest of the time the confederate's head rotations were replicated without modification on the robot.

The head movements were recorded synchronously for the confederate and the naïve subject at 60 Hz. For the analysis we focused on the rotations around the 3 axes.

B. Results

Preliminary results show the same tendency as already reported in [9] for the naïve subject (i.e., increase in head nods in the naive participant during attenuation compared to normal interaction) but not for the confederate. We are currently running this experiment on more subjects to assess if damping head movements in human-humanoid robot interaction would create similar interaction loop effect as shown during human-human interaction.

While experimenting with this setup, we realized that two factors brought significant sense of 'liveness' to the robot: blinks and mouth movements while speaking. Even though the robot was able to accurately replicate face/head movements and gaze it did not feel 'alive' until we implemented blinking teleoperation. The voice coming from a speaker placed at the level of the robot waist and facing the naïve subject was not perceived as coming from the robot until the (rudimentary) mouth movements were implemented.

VI. Conclusion & discussion

A new research platform has been developed to study human-robot interaction and communication. The robot is used as a proxy between two humans involved in dyadic interactions. An experimenter is bound with a humanoid robot. He can control in real-time and sensor free the eye and face/head movements performed by a humanoid robot with his own movements. The experimenter can perceive the scene as if he was the robot. The platform was evaluated: the rendering delay was lower than 200 ms and the rotational error was very low. Manipulations can be applied in real-time to any movement leaving the rest of the dynamics untouched.

Several factors will be investigated independently of others: gaze latency, vergence, spatial shifting and head movements synchronization with gaze. Real acceptable limits will be determined by modifying parametrically specific behaviors. These limits may be used to create behavioral models for autonomous robots.

References

[1] L. D. Riek, "Wizard of Oz Studies in HRI: A systematic Review and New Reporting Guidelines," *Journal of Human-Robot Interaction* vol. 1, pp. 119-136, 2012.

[2] H. Ishiguro and S. Nishio, "Building artificial humans to understand humans," *Journal of Artificial Organs,* vol. 10, pp. 133-142, 2007.

[3] I. Straub, S. Nishio, and H. Ishiguro, "Incorporated identity in interaction with a teleoperated android robot: A case study," in *RO-MAN, 2010 IEEE*, 2010, pp. 119-124.

[4] E. Schneider, S. Kohlbecher, K. Bartl, F. Wallhoff, and T. Brandt, "Experimental platform for wizard-of-Oz evaluations of biomimetic active vision in robots," in *2009 IEEE International Conference on Robotics and Biomimetics, ROBIO 2009*, 2009, pp. 1484-1489.

[5] J.-M. Normand, M. V. Sanchez-Vives, C. Waechter, E. Giannopoulos, B. Grosswindhager, B. Spanlang, C. Guger, G. Klinker, M. A. Srinivasan, and M. Slater, "Beaming into the rat world: enabling real-time interaction between rat and human each at their own scale," *PLoS One,* vol. 7, 2012.

[6] C. L. Fernando, M. Furukawa, T. Kurogi, K. Hirota, S. Kamuro, K. Sato, K. Minamizawa, and S. Tachi, "TELESAR V: TELExistence surrogate anthropomorphic robot," presented at the ACM SIGGRAPH 2012 Emerging Technologies, Los Angeles, California, 2012.

[7] G. Metta, G. Sandini, D. Vernon, L. Natale, and F. Nori, "The iCub humanoid robot: an open platform for research in embodied cognition," presented at the Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems, Gaithersburg, Maryland, 2008.

[8] G. Metta, P. Fitzpatrick, and L. Natale, "YARP: Yet Another Robot Platform," *International Journal of Advanced Robotic Systems,* vol. 3, pp. 43-48, 2006.

[9] S. M. Boker, J. F. Cohn, B. J. Theobald, I. Matthews, T. R. Brick, and J. R. Spies, "Effects of damping head movement and facial expression in dyadic conversation using real-time facial expression tracking and synthesized avatars," *Philosophical Transactions of the Royal Society B-Biological Sciences,* vol. 364, pp. 3485-3495, Dec 12 2009.

[10] B.-J. Theobald, I. A. Matthews, J. F. Cohn, and S. M. Boker, "Real-time expression cloning using appearance models," presented at the Proceedings of the 9th international conference on Multimodal interfaces, Nagoya, Aichi, Japan, 2007.

[11] B. J. Theobald, I. Matthews, M. Mangini, J. R. Spies, T. R. Brick, J. F. Cohn, and S. M. Boker, "Mapping and Manipulating Facial Expression," *Language and Speech,* vol. 52, pp. 369-386, 2009.