A user study on a new Super-Wizard of Oz platform explored in a long-distance survey context

Ritta Baddoura[1, 2]   Gentiane Venture[3]   Guillaume Gibert[1, 2]

[1]INSERM U846, Stem-Cell and Brain Research Institute, Bron, France
[2]Université de Lyon, Université Lyon 1, 69003 Lyon, France
[3]Tokyo University of Agriculture and Technology, Tokyo, Japan

rittabaddoura@yahoo.fr

Abstract

SWoOZ is a new super Wizard of Oz (WoZ) research platform developed to study human-robot interaction (HRI) and mediated human-human interaction. A humanoid robot is used as a proxy between two humans. An experimenter is bound with this proxy and fully controls its head motion with his own movements (live and free of attached sensors). Manipulations can be applied to any motion leaving the rest of the dynamics untouched. This paper presents preliminary results of a user study aiming at evaluating the platform's usability, efficiency and likability. The experimental scenario consists of a realistic long-distance survey conducted by a researcher who interviews Japanese participants on cultural topics (non-deceptive WoZ). The study addresses the possible effects of the remote user's previous experience with robotics (naive vs. non-naive) on the participants' evaluation of the platform.

Keywords

Wizard of Oz; Telerobotics; Social Robotics; Naive vs. non-naive user; Head motion.

Résumé

SWoOZ est une nouvelle plateforme de Super magicien d'Oz (WoZ) visant à l'étude des interactions homme-robot (HRI) et des interactions humain-humain médiatisées. Un robot humanoïde est utilisé comme intermédiaire entre deux humains. L'expérimentateur est lié au robot dont il contrôle entièrement les mouvements de la tête à partir de ses propres mouvements (en direct et sans l'usage de capteurs attachés). Des manipulations peuvent être appliquées à n'importe quel mouvement, laissant intact le reste de la dynamique. Cet article présente les résultats préliminaires d'une étude visant à évaluer la facilité d'utilisation de la plateforme, ainsi que son efficacité et son appréciation du point de vue des utilisateurs. Le scénario expérimental consiste en une enquête menée à distance par un chercheur qui interroge, par le biais de la plateforme, des participants japonais sur des sujets culturels (les participants savent que le robot est téléopéré par un humain). L'étude envisage les effets possibles de l'expérience préalable avec les robots de l'enquêteur (naïf vs. non-naïf) sur les participants.

Mots-Clés

Magicien d'Oz ; Télé-robotique ; Robotique sociale ; Utilisateur naïf vs. non-naïf ; Mouvement de la tête.

1 Introduction & Motivation

During the last few years, there has been a growing attention on exploring teleoperation and telepresence as well as the effects of culture in the social robotics and in the human-robot interaction (HRI) fields. In today's global village where distances are a major component of many personal and professional daily realities, many recent studies focus on showing the interest and the added value of using teleoperated [1, 2] and telepresence robots in many various fields such as remote education [3], health care environments, independent living for the elderly, offices [4], and industrial or military operations that are lead in uncertain and unknown environments [5]. As for the effects of culture, studies address topics such as verbal and non-verbal communication styles [6, 7], user preferences and attitude [8], user beliefs [9] and perception of the robot particularly of its social presence [10, 11], user attribution of personality traits to it [12], and interpretation of facial expressions [13], head motion [14, 15], gaze and gestures [7] and body posture [16] expressed by the robot.

Currently, different kinds of robots with different appearances, capacities and autonomy-levels, are being developed and studied in order to achieve various tasks in different environments. Thus, the importance of building socially-competent robots and mastering the key components of a satisfying and successful interaction with humans has grown wider. The Wizard of Oz (WoZ) technique has been frequently used in this perspective by researchers in the fields of HRI. More precisely, as underlined by [1], WoZ is usually employed to compensate for the robot's insufficient social and/or technical abilities, hence allowing for a smoother interaction and an enhanced vision of future design improvements.

To study human-robot interaction and human-human mediated interaction, and rather than proposing a set of predefined behaviors to be selected by the wizard as in the classical WoZ [1], we developed an enhanced WoZ called SWoOZ (which stands for Super WoZ) setup that has the capacity of mirroring face, eye and head motion on a robot and consequently allowing the generation of a spontaneous movements in order to support a more genuine and realistic interaction [17, 18]. In this

platform, a humanoid robot is used as a proxy between two humans involved in dyadic interactions. One human, called here the remote user [1] (or the interviewer in relation to the interview task performed in our experiment) is bound with the humanoid robot as he controls in real-time and free of attached sensors, by simply performing his own movements, the eye, and face and head motion of this robot. The remote user perceives the scene almost as if he was present, instead of the robot, in the same room as his human interlocutor called here the local user (or the interviewee). The humanoid head motion for instance, as the human interaction partner sees it, is the direct translation of the wizard's motion which is accurately tracked and replicated by the robot with less than 200 ms delay.

SWoOZ can be used to manipulate specific movements without modifying the rest of the dynamics, thus giving an insight on the acceptable limits for the human partner for various parametric manipulations and interactions. Following the study proposed by [19] in a human-human interaction mediated by an avatar, we have first investigated the role of damping head movements during a human-human interaction mediated by a humanoid robot and found as expected that damping head movements affects the interaction [18]. Indeed, naive subjects interacting with a robot controlled in real-time by a confederate's head motion, increased their head movements when the robot's head motion was attenuated.

As reported in [20], many interactions take place simultaneously when humans are communicating through a telerobotic system as the one deployed by SWoOZ. These interactions include HRI between the human users and the remotely controlled communication humanoid proxy, human-human interaction, and human-computer interaction between the remote user and the local user's image on the screen, as it is the case in our setup. In order to further explore these different levels of interaction, as well as to evaluate the SWoOZ platform usability and efficiency when operated by different confederates in the frame of realistic interactive scenarios, we started a study whose preliminary results are presented here.

2 Methods
2.1 Experimental Setup & Equipment

The SWoOZ platform consists of: a) a system able to estimate the remote user's head pose (orientation and location) and rigid/non-rigid motion: The Random Forests Head Tracking system [21] is used in this experiment together with a consumer depth camera (ASUS/Xtion sensor) ; b) a software program to apply online manipulation to specific parameters; c) a humanoid robot: SWoOZ is compatible with the robot NAO (Aldebaran) and iCub (http://www.icub.org/), given that NAO is used in the current study. Once the data are estimated, they are sent to the robot that mimics the estimated remote user's head motion. Further

information about the SWoOZ platform can be found in [17, 18] as well as on the SWoOZ Github page: https://github.com/GuillaumeGibert/swooz.

The remote user's voice captured by a microphone is transmitted to the local user interacting with the teleoperated robot through a small speaker positioned behind it. The head movements of both the remote and the local users are recorded synchronously with an IMU (Inertial Measurement Unit) system. These recorded data will be analyzed later. The IMU sensor is attached around the remote user's and the local user's heads for specific motion analysis such as intensity, jerkiness, velocity and frequency of the human users' movements (IMU sensors are different from the ASUS/Xtion sensor which is only used for the tracking and transmission of the remote user's head motion). To bind the remote user to the robot and enable him as much as possible to sense the scene as if he was seated in its place, auditory and visual feedbacks are transmitted to him using a High Definition (HD) webcam (Creative Live Cam Socialize HD) positioned behind the robot and binaural microphones (MS-TFB-2, The Sound Professionals, Inc.) discreetly placed on the robot's body.

2.2 Participants

Two remote users (one naive and another non-naive) and 20 naive participants (previous exposure to robots was controlled prior to the experiment) volunteered to take part in the study. The 22 candidates are Japanese students from Tokyo University for Agriculture and Technology (TUAT). All of them range in age from 19 to 25 years old. The naive interviewer (never used or interacted with robots or with WoZ setups) interviewed 14 participants (9 males; 5 females); this group will be referred as Participants X in the rest of the text. The non-naive interviewer (previously exposed to manipulating robots and to HRI, he has used NAO previously for research purposes) interviewed 6 participants (5 males; 1 female); this group will be referred as Participants Y in the rest of the text. The remote users/interviewers (both males) have both verbal and non-verbal (head motion) control on the robot, the latter having no autonomy at all. Both were trained to perform the interview (e.g. learning the questions, keeping their behavior consistent with the one a researcher would have, monitoring the participants' answers' duration) and were similarly instructed regarding the technical requirements necessary for the proper functioning of the SWoOZ setup (remain in the field of vision of the depth camera, sit straight etc.).

2.3 Materials, Procedure & Data collection

The scenario design aims at providing a realistic context to the experiment and consists of a cultural user study taking place on a Japanese university campus. The participants volunteered to participate in an anonymous survey lead by a Japanese researcher working in France. The survey investigates, through the participants' answers to an interview and a questionnaire, how the Japanese youth perceives the French and the Japanese cultures. The scenario mainly targets a fluid interaction between the remote user and the local user via the proxy, knowing that the contents of the interviewees'

---

answers (such as e.g. answer's duration, personal opinions or the amount of exact information on the French culture) are not important for the study.

The participants are informed about the following: a) The researcher is unable to be physically present and the interview will be live mediated by a humanoid robot – therefore the scenario involves no deception; b) The interview room is filmed using two cameras, the interviewer's and the interviewee's voices are recorded and IMU sensors are used for head motion capture; c) The same survey will be lead in France for cross-cultural comparison. The remote user is in room A while the robot NAO and the local user are facing each other, seated on either side of a table in a real University meeting room (room B) (see Figure 1).

Figure 1. The remote user (wizard) is in room A while the local user (interviewee) and the robot are in room B.
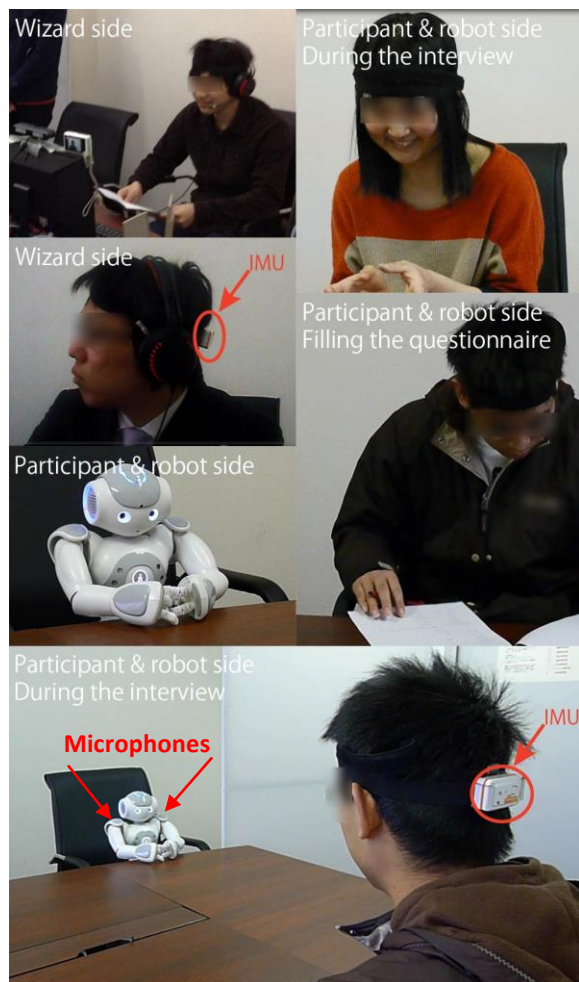
The interview scenario and the related questionnaire were carefully designed for this study, in accordance with Riek's reporting guidelines for WoZ studies in HRI [1], regarding various issues including social deception, rigorous and repeatable design, wizard training, constrained wizard recognition and production abilities, wizard errors, specified user instruction and behavior hypotheses. A pilot study was done on 4 participants in France and 3 participants in Japan to test and improve the questionnaire and the interview questions (including the syntax that had to be in accordance with the researcher's status and role, as well as with the Japanese cultural specificities).

When the participant is seated, the researcher/remote user presents himself and provides a recapitulation of the survey. The participant/local user is reminded that there are no false and right answers: only personal opinions are expected. Then the interview starts.

This oral part of the survey consists of 15 questions revolving around the specificities of the Japanese and French cultures, on their common points and their differences with some focus on communication and interaction questions. Both interviewers/remote users have been trained during three days to master the interview in regards to its contents, general duration, and to the style of questioning, but also in regards to the position in front of the depth camera sensor. More generally, the experimental design aims at defining a precise and repeatable conversational context in which head motion is spontaneously produced by both users; a context that is the same for the 20 interviewees and that validates the comparison between each experimental session.

The interview's overall duration is 10 min while the questionnaire needs 5 to 10 min to be filled (depending on the participant). The whole experiment lasts around 15 to 20 min. When the interview is completed, the remote user asks the participant to fill the questionnaire placed on the table. The questionnaire consists of 40 items divided into 5 sets. 4 sets use a 5-point Likert scale (where 0 = not at all and 4 = to a very high degree). The sets addressed in this paper assess the participant's evaluation of the robot as a proxy, more particularly the participants' evaluation of the robot's efficiency, likability and credibility. An open-ended question ends the questionnaire to allow the local user to express more freely and personally his/her feedback. The Cronbach's alpha of the participants' questionnaire is 0.91 which is above the generally acceptable level 0.7 [22] and shows a very good internal reliability. Additionally, each wizard was asked at the end of the whole round of interviews to fill a questionnaire in 9 items divided into 3 sets, in order to get his feedback on the experiment and on the SWoOZ platform.

3 Hypotheses & Preliminary Results

First of all, we are interested in getting a feedback on the SWoOZ platform efficiency from local users who are not familiar with it and who have no prior experience with robots or with WoZ setups. As for the remote users, we are interested in observing the possible effects of their previous experience with robots and HRI on the local users' feedback. Thus, we explore this experimental scenario with two samples: one is interviewed by a naive remote user and the other by a non-naive remote user. Therefore, we first hypothesize that the remote user's previous exposure to HRI will impact the interviewees' experience of the interaction as well as their ratings (H1) and that X evaluations (interviewed by the naive interviewer) of the proxy will

be significantly different from Y's (interviewed by the non-naive interviewer).

Regarding the local users' evaluation of the humanoid proxy, we focus in this paper on their ratings of its efficiency, usefulness, likability, engagingness, human-likeness and on their satisfaction with it. We are also interested in their evaluation of the humanoid robot's credibility as a proxy/mediator between them and the remote user. Based on the prior study performed with SWoOz [18]; as well as on its ability to enable the proxy to mirror the remote user's motion, we expect the participants' ratings of efficiency, usefulness and satisfaction to be above the average score (H2) (2 being the average on this 5-point Likert scale).

Also, given the fact that the robot has a close to natural head motion as it mirrors the remote user's spontaneous motion, and given the fact that the voice the interviewees hear is the remote user's human voice, we expect the local users to find the proxy engaging and likable thus rating it above the average score (H3). From another perspective, we make the assumption that NAO's credibility as a proxy representing a researcher (in regards to NAO's appearance and given role), as well as its human-likeness (NAO only moves its head but its eyes are rigid, and it has no mouth) to be poorly rated (H4). For instance, [23] showed that a robot's appearance affects its likability and that participants expect the robot appearance to match its task during an interview context.

We calculated the descriptive statistics (95% CI) based on the interviewees' ratings of the proxy's performance. Participants X and participants Y gave generally medium-to-low ratings (see Figure 2). Y (interviewed by the non-naive remote user) seem to have given more generous scores than X.

X found the human-like aspect of the robot to be very poor (X: M= 1.00, SD= 1.13) and considered that the robot failed in being credible (X: M= 1.28, SD= 1.03). Nevertheless, they found the proxy rather satisfying (X: M= 2.14, SD= 1.30), likable (X: M= 2.07, SD= 1.10) and useful (X: M= 2.5, SD= 1.11). Y gave also low scores to the robot's human-likeness (Y: M= 1.5, SD= 1.11) and efficiency (Y: M= 1.66, SD= 0.94) and found it unsatisfactory (Y: M= 2.00, SD= 0.81). Despite these ratings, they considered it to be likable (Y: M= 2.83, SD= 0.68), engaging (Y: M= 2.5, SD= 0.95) and useful (Y: M= 2.5, SD= 1.11).

We did a Mann-Whitney test to ascertain if the differences between X and Y scores are statistically significant, thus implying an effect of the remote user's previous experience with robots and HRI. The observed U-values failed to be significant at p≤ 0.05 (see

Table 1), thus invalidating a possible effect of the remote user's previous experience on the participants' ratings.

As some recent studies have underlined the human partner gender effects in HRI [10, 24, 25], we also did a Mann-Whitney test to rule out possible gender effects on the participants' ratings. Results showed the gender factor to be statistically not significant, which can be possibly attributed to the low representation of women in both samples as well as to the small size of the samples.

Figure 2. X's and Y's evaluation of the SWoOZ proxy on a 5-point Likert scale (0 = not at all; 4 = to a very high degree)
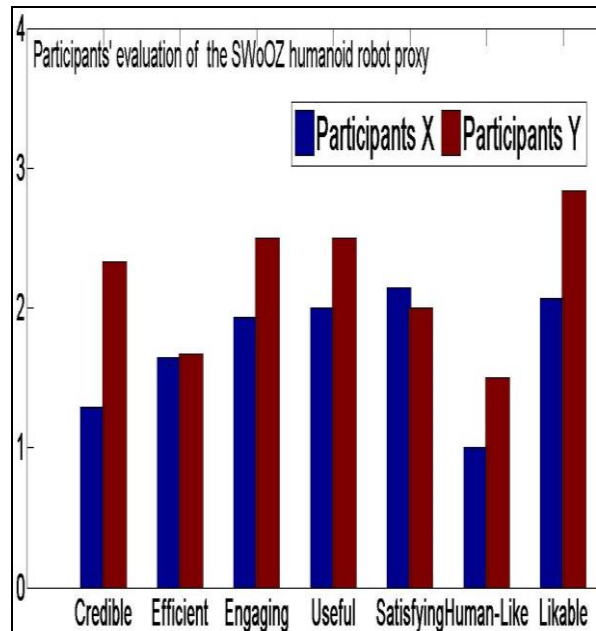
Table 1. Mann-Whitney test results based on X's and Y's evaluations of the proxy (U critical = 17)

The Proxy is ╱ U observed	Credible	Efficient	Engaging
	22	41.5	32
Useful	Human-Like	Likable	Satisfying
32	32.5	31	38

4 Discussion and Conclusion

Generally, the participants showed a moderate to low appraisal of the proxy's performance. H3 was validated as the participants found the proxy's rather likable and engaging (among the highest scores for X and Y). H2 was partly validated since the participants were moderately satisfied with the proxy (the highest score for X). They found it rather useful as well. But despite that, they judged it as poorly efficient. H4 was partly validated: as expected, the proxy's human-likeness was poorly rated, but X and Y did not seem to have similar feedback on the robot's credibility. X average rating of this dimension was very low, whereas Y gave moderate ratings to it.

Nevertheless, the Mann-Whitney test results failed to validate the remote user's previous experience effect on the participants, thus infirming H1 and showing that the observed differences between X and Y ratings are

statistically insignificant. This might be probably due to the small size of the samples, especially of the one interviewed by the non-naive remote user. Therefore, it would be of real interest to proceed to further experiment with more participants, at least with the non-naive remote user, in order to reexamine the previous experience effect. Another explanation would be that the preparation/training phase of both remote users' has reduced the gap between them, but this does not seem to be a sufficient reason. Also, the mediation/teleoperation characteristics of the SWoOZ platform might mitigate the remote user's effect. Indeed, the proxy's mediation interferes in the human dyad and is in favor of rendering a rather homogeneous/similar behavior of the robot, especially that the remote user's head motion is only mirrored here. Thus, using a humanoid robot with richer face expressions, such as the iCub for example (it is able to move its eyes and mouth) might better render some differences between the remote users and enable us to more precisely assess the impact of their previous experience with robots on the participants.

These results are only preliminary and have to be completed with further analysis of the remaining experimental data. However, some important aspects are underlined such as the failure to prove significant effect of the remote user's previous experience with robots. Nevertheless, this result is interesting in regards to the platform's usability as it suggests that any naive person could, with some preparation, successfully use the SWoOZ platform and be as equivalently effective as a more experienced person. The question of the proxy's choice, in relation to its appearance and to its assigned task needs to be given more attention for credibility reasons. Last but not least, the participants' satisfaction and their moderate appreciation of the proxy's (and therefore of the SWoOZ device) usefulness and of its engaging and likable behavior, are encouraging feedbacks to work towards improving the features of the SWoOZ platform for more efficiency and smoother ability to mediate human-human interaction.

5   Perspectives and Future works

The analysis of the remaining parts of the questionnaire (that were not addressed here) and the processing of the IMU data to obtain various head motion characteristics, will be addressed in future papers to shed more light on this study's results. Furthermore, running other experiments using human-human non-mediated interaction, or videoconference mediated interaction, or another robot, would open up to constructive comparisons with our collected data. Investigating technical aspects such as the localization of the camera in the setup: behind the robot vs. on the head of the robot, could be also of interest to understand the impact of having egocentric vs. non-egocentric visual input on the remote user's sense of "immersion".

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]   S. Nishio, H. Ishiguro, M. Anderson, and N. Hagita, "Representing Personal Presence with a Teleoperated Android: A Case Study with Family," in *AAAI Spring Symposium: Emotion, Personality, and Social Behavior*, 2008, pp. 96-103.

[3]   F. Tanaka, T. Takahashi, S. Matsuzoe, N. Tazawa, and M. Morita, "Telepresence robot helps children in communicating with teachers who speak a different language," presented at the Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction, Bielefeld, Germany, 2014.

[4]   A. Kristoffersson, S. Coradeschi, and A. Loutfi, "A review of mobile robotic telepresence," *Advances in Human-Computer Interaction,* vol. 2013, p. 3, 2013.

[5]   V. Harutyunyan, V. Manohar, I. Gezehei, and J. W. Crandall, "Cognitive Telepresence in Human-Robot Interactions," *Journal of Human-Robot Interaction,* vol. 1, pp. 158-182, 2012.

[6]   P. Rau, Y. Li, and D. Li, "Effects of communication style and culture on ability to accept recommendations from robots," *Computers in Human Behavior,* vol. 25, pp. 587-595, 2009.

[7]   M. Fukushima, R. Fujita, M. Kurihara, T. Suzuki, K. Yamazaki, A. Yamazaki, K. Ikeda, Y. Kuno, Y. Kobayashi, and T. Ohyama, "Question strategy and interculturality in human-robot interaction," in *Human-Robot Interaction (HRI), 2013 8th ACM/IEEE International Conference on*, 2013, pp. 125-126.

[8]   C. Bartneck, T. Nomura, T. Kanda, T. Suzuki, and K. Kennsuke, "A cross-cultural study on attitudes towards robots," in *HCI International*, 2005.

[9]   F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita, "Interacting with a human or a humanoid robot?," in *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, 2007, pp. 2685-2691.

[10]   P. Schermerhorn, M. Scheutz, and C. R. Crowell, "Robot social presence and gender: Do females view robots differently than males?," in *Proceedings of the 3rd ACM/IEEE*

international conference on Human robot interaction, 2008, pp. 263-270.

[11] A. Kristoffersson, K. S. Eklundh, and A. Loutfi, "Towards measurement of interaction quality in social robotic telepresence," in *Proceedings of the Ro-Man Workshop on Social Robotic Telepresence*, 2012, pp. 24-31.

[12] A. Weiss, B. van Dijk, and V. Evers, "Knowing me knowing you: Exploring effects of culture and context on perception of robot personality," in *Proceedings of the 4th international conference on Intercultural Collaboration*, 2012, pp. 133-136.

[13] C. Becker-Asano and H. Ishiguro, "Intercultural differences in decoding facial expressions of the android robot Geminoid F," *Journal of Artificial Intelligence and Soft Computing Research,* p. 215, 2011.

[14] G. Trovato, T. Kishi, N. Endo, M. Zecca, K. Hashimoto, and A. Takanishi, "Cross-Cultural Perspectives on Emotion Expressive Humanoid Robotic Head: Recognition of Facial Expressions and Symbols," *International Journal of Social Robotics,* vol. 5, pp. 515-527, 2013 2013.

[15] C. L. Sidner, C. Lee, L.-P. Morency, and C. Forlines, "The effect of head-nod recognition in human-robot conversation," in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, 2006, pp. 290-296.

[16] A. Kleinsmith, P. R. De Silva, and N. Bianchi-Berthouze, "Cross-cultural differences in recognizing affect from body posture," *Interacting with Computers,* vol. 18, pp. 1371-1389, 2006.

[17] G. Gibert, M. Petit, F. Lance, G. Pointeau, and P. F. Dominey, "What makes human so different? Analysis of human-humanoid robot interaction with a super Wizard of Oz platform," in *International Conference on Intelligent Robots and Systems*, Tokyo, Japan, 2013.

[18] G. Gibert, F. Lance, M. Petit, G. Pointeau, and P. F. Dominey, "Damping robot's head movements affects human-robot interaction," presented at the Human-Robot Interaction (HRI), Bielefeld, Germany, 2014.

[19] 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.

[20] A. Kiselev and A. Loutfi, "Using a mental workload index as a measure of usability of a user interface for social robotic telepresence," in *Workshop in Social Robotics Telepresence*, 2012.

[21] G. Fanelli, J. Gall, and L. Van Gool, "Real time head pose estimation with random regression forests," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 2011, pp. 617-624.

[22] J. C. Nunnally, *Psychometric theory*: McGraw-Hill, 1978.

[23] D. Li, P. P. Rau, and Y. Li, "A cross-cultural study: effect of robot appearance and task," *International Journal of Social Robotics,* vol. 2, pp. 175-186, 2010.

[24] F. Eyssel, D. Kuchenbrandt, S. Bobinger, L. d. Ruiter, and F. Hegel, "'If you sound like me, you must be more human': on the interplay of robot and user features on human-robot acceptance and anthropomorphism," presented at the Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction, Boston, Massachusetts, USA, 2012.

[25] C. R. Crowell, M. Scheutz, P. Schermerhorn, and M. Villano, "Gendered voice and robot entities: perceptions and reactions of male and female subjects," presented at the Proceedings of the 2009 IEEE/RSJ international conference on Intelligent robots and systems, St. Louis, MO, USA, 2009.