Social Tele-Operation by Confederates: Applying the Actor-Confederate Paradigm to HRI

Paul Bremner[1], Huseyin Cakal[2], Miriam Koschate-Reis[2], Mark Levine[2]

***Abstract*— A standard approach to investigating interactive robot behaviours, prior to autonomous implementation, is the Wizard-of-Oz paradigm, where an operator triggers behaviours in response to actions of a naïve participant. There are however, significant issues pertaining to adaptability and timing of responses, particularly when a lot of interactive communication is required. Here we present an alternative, Social Tele-Operation by Confederates, where trained actors control the robot to allow natural communication, portraying the behaviours to be tested. If similar participant responses can be engendered in face-to-face and robot mediated interaction, it demonstrates their potential utility for autonomous systems. Further, it is possible to *learn from the expert*, in a situation with higher ecological validity than the use of human face-to-face encounters alone; hence, better autonomous systems might be developed. Here we present the methodology, and a first case study to demonstrate its application, where we have begun to investigate its utility.**

I. Introduction

Developing behaviours for socially interactive robots is a significant challenge. A key problem in this process, is that when implemented behaviours fail to elicit expected behaviours in humans, it can be challenging to disambiguate between causes of failure. For an autonomous system causes of failure can be due to sensing issues, algorithm design, or (as is often assumed) that the way people process robot communications is sufficiently different from those of humans for an effect to occur. A common way this is avoided is to use the Wizard of Oz (WoZ) paradigm, where a remote operator (the wizard), unknown to the naïve participant, triggers robot behaviours [1]. However, this can have problems of its own given the difficulty of determining the correct set of behaviours *a priori*, and the challenge for the wizard in executing the correct behaviours in a timely fashion [1]. This issue is exacerbated for experiments that require significant amounts of interactive communication, both in the lab and 'in the wild' where the behaviour of participants is less predictable. In such scenarios communication is likely to seem less natural to participants, and be more prone to failure, or delays, reducing the ecological validity of the testing of behavioural manipulations.

While WoZ is clearly an important and valid methodology, for which the aforementioned issues can be mitigated with careful planning and well trained wizards [1], here we suggest a possible alternative. We term this method social tele-operation by confederates (STOC), whereby a motion capture control scheme allows trained actors to utilise humanoid robots as physical avatars. By means of using such a control scheme, in conjunction with streamed audio, this allows adaptive, natural multi-modal communication. In order to test behavioural manipulations the actor-confederate paradigm from experimental psychology is employed, where an actor-confederate employs the behaviours to be tested in a semi-scripted interaction [2].

While such interactions via a physical proxy are interesting in their own right, they might also be of use in analysis of communicative behaviours for humanoid robots. Work on the use of avatars in virtual reality has demonstrated that the perception of people, and hence behaviour towards them is influenced by the appearance of the avatar which they use to interact [3]. Hence, we suggest that by comparing face-to-face interaction with a robot avatar any differences would be due to the robots appearance, and ability to realise the captured motion. How these differences manifest, and how they relate to the appearance and perceptions of the robot is not yet clear, but is an aim of our ongoing investigations.

In addition, studies comparing interaction with virtual avatars and autonomous agents have shown there to be little difference between the social behaviours exhibited in each case [4]. However, some researchers have observed contradictory results to those in [4], for example in studies on social influence (typically persuasion), though this was found to be context sensitive [5]. If this finding extends to robot avatars, and how it depends on the interaction context requires investigation, but does highlight the potential of our proposed methodology for some interactive contexts.

The other key capability of STOC is that it enables the learning of successful behaviours from the expert operator, in order to facilitate the development of autonomous social robots. A traditional approach to developing social robot behaviours is to analyse human interactions and attempt to model them. We suggest that STOC might provide more useful information on communicative behaviours, by having them performed by the robot, as this takes into account its physical capabilities and limitations; i.e., STOC has higher ecological validity in the context of human-robot interaction than directly observed human-human interactions alone.

Here we describe the proposed methodology and requirements for its use, followed by a more detailed description of the potential capabilities and limitations of it. We then present a case study of an ongoing experiment where we are using this methodology, to demonstrate a scenario in which

This research grant is funded by the EPSRC under its IDEAS Factory Sandpits call on Digital Personhood, grant ref: EP/L00416X/1.

[1]with Bristol Robotics Laboratory, University of The West of England, Bristol, BS16 1QY, UK. paul.bremner@brl.ac.uk

[2]with Department of Psychology, College of Life and Environmental Sciences, University of Exeter, Exeter, Devon, EX4 4QG, UK. h.cakal; m.koschate-reis; m.levine; @exeter.ac.uk

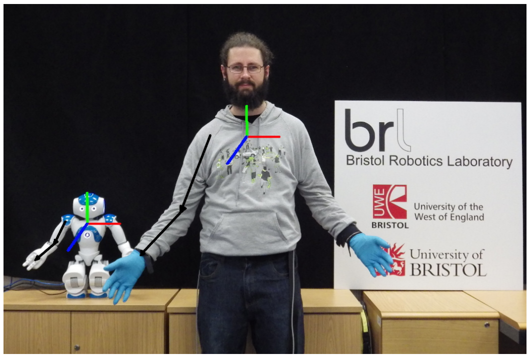

Fig. 1. A matched pose between the tele-operator and the NAO robot. Black arrows mark arm unit vectors, torso coordinate frames in RGB (XYZ).

it might be employed, and present some initial findings.

II. Methodology and Requirements

From a purely methodological point of view, the STOC paradigm has a number of exacting requirements to allow the investigation of social phenomena. These requirements are key for maintaining experimental control, and ecological validity. A systematic study of a phenomenon, requires large scale sampling, repeatability, situational controllability, and experimental manipulation. Hence, STOC utilises the actor-confederate paradigm from psychology, to overcome these methodological limitations, where an actor-confederate portrays the behaviours to be studied to a naïve participant. The actor-confederate paradigm facilitates a high level of experimental control particularly in tasks that involve social interaction, and affords a precise standardization of both the independent and dependent variables [2].

In order to be able to use such a method a tele-operation interface is required that is highly intuitive, and allows an actor to realise their behaviours on the robot system with a minimum of additional cognitive load. Indeed, we suggest the best method to use for this purpose is motion tracking of the actor to transfer physical motion to the operated robot while the operator speaks; so called implicit control [6]. By mapping the operators joint motions on to the robot in this way, joint coordination is maintained such that correct hand trajectories are realised; a limitation here is that some motions of the operator may exceed the joint limits of the robot platform, so these motions cannot be precisely recreated. We have developed such a system based on a Kinect to control a NAO robot (Fig. 1), and have shown that it is able to produce a wide range of comprehensible gestures [7]. It is important to analyse the capabilities of the control scheme at rendering the desired behaviours to aid in the analysis of differences between face-to-face and robot mediated experimental conditions.

Using actors to portray specific behaviours under the instructions of the researcher helps to achieve acceptable levels of spontaneity, and experimental manipulation with a degree of ecological validity. However, actor portrayals have been criticized for stereotypic, infrequent, and decontextualized representations of behaviour [8]. Such limitations, can be remedied for by introducing a semi-controlled scenario, extensive pre-experimental rehearsal and employing a trained actor [9]. Further, as part of the rehearsal and refinement process, actor behaviours should be reviewed by judges external to the experimental team, as a check of convincing behavioural expression, and if the enacted behaviours are perceptible on the robot platform.

A carefully crafted script for the actor and the experimenter is also needed to maintain consistency and to assure that experimental manipulation is not breached across conditions and participants [2]. Such a script should help to establish a discursive context within which the actor-confederate could employ the behaviours to be studied. Furthermore, it must not only contain the required words and behaviour to try and use, but also how the script might be acceptably adapted to a particular setting, while maintaining consistency between participants.

To this end a suitable interaction task is required to ensure the studied behaviours appear to naturally occur from the view of the naïve participant(s). A correctly designed task also aids the actor in producing consistent, natural seeming, behavioural expressions; it should do so by establishing the appropriate social context. Further, it enables the the obfuscation of study goals (study of responses to manipulated social behaviours) under the guise of task performance. Ensuring that the participants are naïve to the aims of the experiment is paramount to avoid participant awareness bias which may prevent them from reacting to stimuli in the way they would in the real world [2]. Hence, a script-based interaction task between a naïve participant (or participants) and a confederate actor enables the researcher to approximate natural social environment reactions to the effects of the experimental manipulations.

III. Capabilities and Advantages

The STOC methodology has a number of key capabilities and advantages over more traditional Wizard of Oz methods used for HRI studies where an autonomous system is not yet available. Its primary advantage is that it allows more natural interaction. Human behaviour is difficult to predict *a priori*, so pre-determined responses that are triggered by a wizard may not be well suited to a particular participant or situation. Further, there may be issues related to timing of responses, as determining and executing the correct response requires thought on the part of the wizard; in studies where some adaptation of responses is possible there are often timing problems [1]. Problems related to the issues of appropriateness and timing may result in the confounding of behavioural effects, and reduced ecological validity. While this might be mitigated somewhat by well refined behaviours, and a highly trained wizard, for testing of some behaviours STOC might be more suited. With STOC, responses are produced as in natural conversation, so adaptive responses are utilised in real-time. This makes it particularly suited to experiments 'in the wild', where it is harder to predict how people interacting with the robot will behave. Further, interaction failures and time delays are liable to lead to reduced interaction times, as people may disengage from the interaction at any point.

Another key capability is that it facilitates direct comparison between face-to-face and robot mediated interaction, by using the same actor for both. Hence, any differences observed may be attributed to the robots appearance and ability to realise the controller's behaviours.

Last, but by no means least, is the capability to utilise the behaviours exhibited by the actor during tele-operation in the development of autonomous social robot behaviour, and/or Wizard of Oz protocols. This can be done by observing the degree of success of different operator behaviours in eliciting desired participant responses when performed with the correct timing and interaction context. We suggest that a clearer picture of desirable behaviours can be gained with STOC than through either human observation or WoZ studies alone, i.e., it has higher ecological validity. As mentioned previously we make this assertion on the basis that it allows observation of interactive behaviour taking into account the physical capabilities and limitations of the robot platform, as well as its appearance and the expectations that engenders. Further, the rehearsal period may be used to develop modifications to the behaviours employed for face-to-face interaction such that that they might be more easily recognised when realised by the robot; hence, understanding of the requirements for robot social behaviours might be furthered.

IV. Limitations

It is important to note that there are a number of limitations to the utility of STOC for HRI studies. Most obviously it relies on the actor's ability to consistently produce the desired behaviours. Section II describes methods by which this issue might be minimised.

A more important limitation is that STOC cannot provide definitive evidence for how tested behaviours might operate in autonomous social robots, only suggest behaviours that are worth testing. In particular, while stronger inferences may be made for behaviours relating to physical action (gestures, gaze etc.), those based on the perception of the operators mental state are harder to make.

V. Case Study

In order to better illustrate the described experimental paradigm, we present here an example study where it has been employed. In this study we are primarily examining the effects of emotional expression on eliciting emotion in others. Various studies have shown that observing emotions in others can result in mirroring of emotions (emotion contagion), or elicitation of an empathetic complementary emotion [10].

Emotions are a natural part of human behaviour, and as such have been previously investigated, with varying degrees of success, for social robots [11][12]. The primary reason for this lack of consistency is that studying emotions presents the researcher with a number of challenges that are difficult to overcome. Emotional expressions, are hard to capture in naturalistic settings as they are mostly unpredictable and spontaneous. Further, eliciting emotions using various stimuli, e.g., pictures or video is also problematic as such stimuli lack ecological validity [2]. Hence, finding a basis for a model of how emotional expression might be incorporated in an interactive setting, and whether such emotional expressions effect the emotions of interlocutors is a significant challenge; a challenge that might be addressed by social tele-operation.

A secondary motivation behind our use of STOC is that research suggests that real life manipulations of emotions, i.e., eliciting emotions through direct social experience by the participant, are more likely to be effective [10]. Emotions induced in such experimental designs are also more likely to be experienced as real and natural, or near natural. Social tele-operation by confederates utilises this sort of experimental design.

The study presented here aims to investigate whether emotional expressions portrayed by an actor can be reliably recognised by participants, and elicit emotion in them, whether face-to-face or via the robot avatar. We do so in the context of an interactive instructional task, which allows us to show how valence is added to conveyed instructions, both in verbal and non-verbal communication. Further, the nature of the task means that we can also observe differences in general interaction behaviours in an instructional context, between face-to-face and robot mediated conditions.

A. *Study Outline*

We conducted an experimental study with 84 participants (34 male, 50 female, age $M = 20.01 \pm 6.17SD$). Participants gave written informed consent to participate in the study which was in line with the revised Declarations of Helsinki (2013), and approved by the Ethics Committee of the College of Life and Environmental Sciences, University of Exeter.

Participants took part in a series of three interactive instructional tasks, adapted from Serholt, Basedow, Barendregt, & Obaid [13], they consisted of three simple Lego constructions of equal complexity, for which a confederate actor conveyed a series of 6 instructions to be carried out by the participant. Participants were allowed to ask for clarifications, with the aim being successful assembly of each Lego structure. Participants were instructed by the confederate either face-to-face (F2F) or via the tele-operated robot (TR), and either happy or sad emotion was added to the second task (task order was counter-balanced between participants). The same actor was used for all conditions to ensure behavioural consistency. Hence, the study used a 2 (presentation mode) x 2 (emotion) fully between subjects design.

For the TR conditions we used a NAO humanoid robot platform from Aldebaran Robotics (see Fig. 1, for specifications see [14]). The robots arms were controlled using our Kinect based tele-operation scheme (as described in [7]), with the addition of an Oculus Rift to track head movements; images from cameras mounted on the robots head were displayed in the Rift allowing the actor to see the

Fig. 2. Experimental set-up. C1 and C2 cameras, P participant, A-C actor-confederate. Pieces for the Lego tasks are in cups at P end of table.

participant. Speech was streamed using an external speaker and microphone hidden behind the robot, and connected to the operator via speech over IP software.

In the F2F conditions the actor sat at one end of a table where they were given the instructions to convey, and in the TR conditions they were shown into an adjoining room where the robot controls were; in the TR conditions the robot was standing on the end of the table. Fig. 2 shows the experimental set-up for the two conditions.

With regards to emotions, we hypothesize that (H1) participants will be able to recognize expressions of two basic emotions, sad and happy, equally well both in F2F and TR conditions: tone of voice and gestures will likely be sufficient to convey emotion despite the NAO lacking facial expression. Further, (H2) emotion contagion will be less effective in the TR condition: we suggest a robot is not expected to have emotions, and is able to display fewer emotional cues and this will reduce the effects on participant emotions.

With regards to the interactive context, we hypothesize (H3) that the participants in all conditions will evaluate the task difficulty and instructions similarly: all the information was contained in speech and was the same between conditions. We also hypothesize (H4) that interactive behaviour with the actor in the two conditions will be the same: previous work with tele-operated human-like robots has shown there to be similar interaction behaviour to human interaction [15], and this might also apply here.

B. Protocol

We used a professional actor as our confederate. It is suggested that professional actors are able to provide more natural expressions of emotions without exaggeration and are generally more comfortable in recorded sessions [8]. We followed the guidelines suggested by Busso & Narayanan (2008) and the actor went through an extensive pre-experimental rehearsal process under the experimenters guidance [9]. This helped the actor to get used to the material and feel more confident leading to a more naturalistic portrayal of emotions.

During the pre-experimental phase the actor-confederate was given a set of discrete emotions from GEMEP (happy, sad, angry etc.) [8] and asked to express each emotion in the set. These expressions were then recorded and each expression was evaluated for its naturalness, its appropriateness and adaptability to the interactional context, i.e., giving instructions. External judges then watched the recordings and were asked to recognize each emotion. This allowed us to evaluate the capabilities of the actor for various emotions beyond those required, but also to ensure that happiness and sadness could be differentiated from within a full range of emotions. We therefore felt confident in our selection of these two emotions and the actors ability to portray them naturally.

In order to familiarise the actor with the capabilities of the robot and the tele-operation system, the pre-experimental phase also included practice sessions of robot operation. Happy and sad performances were recorded of robot operation and verified by external judges. It is important to note that as a result of these sessions it was found that more exaggerated arm motions were required to convey emotion than in the F2F condition.

We then developed a script for three sets of instructions which were emotionally neutral in their semantic content, and hence any set can be performed with either of the selected emotions. Also scripted were hand gestures to go with each instruction that contained redundant information (i.e., information that was already in speech). This was to allow emotional body motion within the context of the task. In order to allow the actor to be able to adapt to behaviours of the participants while maintaining experimental consistency, we also specified a set of acceptable deviations from the pre-written script. These included alternate ways of explaining each instruction, and how to format corrections.

In order for the emotion manipulations to seem natural the participants had to be convinced that the actor was another participant, an experimental protocol was developed to this end. Upon their arrival participant and actor were greeted as participants by the experimenter, and both invited to read and sign the consent form, and given initial instructions. Though the actor was always assigned to the role of instructor, it was presented to the participant as a random selection.

We contextualized the emotional expressions using a mild deception technique. In each condition, the actor delivered the instructions for the first task in a neutral manner. Upon completion, the participant was instructed to complete the questionnaire to evaluate the instructions, their own emotional condition and how they perceive the emotional state of the *other participant*. Then shortly after the start of the second task, the actor reached to his mobile ostensibly checking a text message that he just received. Depending on the condition, the confederate acted as if the message is a happy or a sad one and delivered the instructions for the second task coloured by the emotion. The final set of instructions were again delivered in a neutral manner. To verify the efficacy of our protocol, an initial pilot group of participants were probed after their experimental sessions for their belief in the confederate being a participant and gave no indication they were suspicious.

C. Measures

To assess self-experienced emotions and the emotions experienced by the confederate actor we adapted a scale from Geneva Appraisal Questionnaire developed by Geneva Emotion Research Group. [8]. After each task the participants were asked to rate how they felt, i.e.. angry, irritated, happy, pleased, amused, compassionate, tender, and sad, and how they perceive the confederates emotions using a 9 point Likert scale for emotional intensity. In order to disguise the focus on emotions, the emotion questions were embedded within a larger set probing the difficulty of each task and the ability of the instructor to provide good instructions.

Both participant and confederate were recorded on video during the sessions. The recorded material was coded for verbal and non-verbal behaviour. Specifically we developed a coding scheme which included focus of attention, i.e. frequency of looking at the confederate or robot, requests for clarification, and task performance, measured as average amount of time it takes to complete the task.

D. Results and Analysis

Here we present preliminary results analysis based on our first parsing of the data. The study is still ongoing with the need for more in depth data analysis, particularly of video data, and an additional condition (voice only, see section VI).

1) Recognition of Basic Emotions (H1): A 2x2 factorial ANOVA on the perception of emotional intensity results showed that there was a significant main effect of Emotion condition for both basic emotions Happy $F(1, 76) = 6.747, p = .011$; Sad $F(1, 76) = 9.949, p = .002$. This shows that participants successfully recognized the two basic emotions across human and robot conditions. There was no significant difference across robot and human conditions.

Examining the video data of both conditions and the rehearsal sessions, we observed consistent features of behaviour exhibited by the actor. Gestures that accompanied instructions conveyed in the happiness condition were larger, faster, and performed with more energy than in the neutral or sad conditions. Whereas, gestures in the sadness condition used opposite characteristics. More detailed analysis of the video and Kinect data is required to better characterise happy and sad versions of gestures.

2) Emotion Contagion (H2): A 2x2 factorial ANOVA on experienced emotional intensity showed that there was a near significant main effect of emotion condition on happiness intensity $F(1, 75) = 3.393, p = .069$ and a significant effect on sadness intensity $F(1, 75) = 3.984, p = .050$. We also detected a significant interaction effect $F(1, 75) = 4.099, p = .046$. Simple effects analysis revealed that there is a significant difference between participants in the happy and sad conditions only in the human condition $F(1, 76) = 7.112, p = .009$. Specifically, participants in the human-happy condition felt significantly happier in the human-condition ($M = 4.89, SD = 1.85$) compared to the participants in the human-sad condition ($M = 3.15, SD = 1.88$) whereas the level of happiness that the participants reported to have experienced did not differ across happy and sad conditions in the robot condition. The data indicates that in the robot condition participants felt more amused than in the F2F conditions, which in turn made participants experience happiness irrespective of the emotion condition; further work is needed to determine if this is a mediating effect.

Taken together these results provide evidence that whereas contagion of positive affect, as measured by happiness, is limited to human-human encounters, contagion of negative affect, measured here as sadness, appears to be robust both in face-to-face and robot mediated communication.

3) Instructions (H3): In line with our expectations we found no statistically significant difference in terms of how our participants evaluated the instructions and the task difficulty. This suggests that participants in either condition did not find the task and instructions more or less difficult.

In line with our script, the actor adapted the instructions to suit the task performance and understanding of each participant. Adaptations included correcting mistaken understanding, and responding to questions raised by participants. The actor also adapted the pace of the instructions to align with the performance rate of the participant, particularly as the tasks progressed and participants learned how the actor phrased instructions, and sometimes guessed instructions before they were complete.

4) Verbal and Non-verbal Behaviour (H4): We also analysed the video data for any possible differences for focus of attention, frequency of gaze at Robot vs Actor, average number of requests for clarification, and task performance measured as average amount of time it takes to complete the tasks. T-test results revealed that participants in the Robot condition are more focused on the robot which was measured as Gaze Frequency at the Robot ($M = 29, SD = 4.48$) compared to the participants in the Human condition ($M = 11.20, SD = 10.03, t(17) = 5.061, p < .001$; requested more clarifications ($M = 21, SD = 11.03$) compared to the participants in the Human condition ($M = 9.22, SD = 5.76, t(17) = 3.051, p = .007$; and needed more time to complete the tasks ($M = 169.25, SD = 22.19$, in seconds) compared to the participants in the Human condition ($M = 135.30, SD = 11.39, t(17) = 4.263, p = .001$.

E. Discussion

Our preliminary data analysis has shown that emotions are recognised via the robot, despite the lack of facial expressivity on the NAO robot. However, this may just be due to the ease of recognition with tone of voice. Hence, we aim to have external judges evaluate videos of the emotional task performances without sound to investigate if they successfully convey emotion. If the gestures are able to successfully convey emotion we will analyse in more detail the video and Kinect data of the gestures performed, given that they have consistent characteristics across experimental sessions, and they appear to largely agree with the suggestions for emotional motion in the literature [12]. Further, more detailed Kinect analysis will allow more general rules for gesture performance modification to be established and see if they match other findings in the literature.

Interestingly we also found that there was a difference in emotion contagion between the F2F and TR conditions. Several possible explanations are apparent. One explanation is that participants were in general amused by the appearance and motion of the robot so this affected the influence of the actors emotions upon them. In further analysis we aim look for other common factors in robot perception between conditions. An alternative explanation we suggest is that the reduced range of cues, while not affecting conscious recognition, may have reduced the efficacy of the portrayed emotions. Another explanation is that expectations engendered by a robot with no facial expressions, i.e., an unemotional machine, may changed the way the emotions are processed. Which of these is the most reasonable explanations requires significant further investigation.

Another important finding was that there are clear differences in the social interaction behaviours between the F2F and TR conditions. This was despite the fact that the instructions and a set of adaptations were performed consistently between conditions. Differences included frequent looking at the actor in the TR condition, as well as more requests for clarifications. A possible explanation is that social cues for success and attention were harder to perceive, requiring conscious rather than subconscious attention. Indeed, there is some evidence in the videos of the actor unconsciously giving feedback on the success or failure of construction in facial gestures. More systematic analysis of the videos for these behaviours are required to verify this. However, the fact they looked to the robot for clarifications shows natural interaction behaviour using the robot avatar. One means we are pursuing for determining reasons for this is a voice only instruction condition, to evaluate whether the behaviour differences can be attributed to the robot, or remote interaction.

VI. Conclusions and Future Work

In this paper we have detailed social tele-operation with confederates, an alternative to standard Wizard of Oz approaches, in which the actor-confederate paradigm from psychology is adapted for use with implicitly controlled tele-operated robot avatars. Doing so allows us not only to investigate robot mediated interactions, but also to gain understanding of social interaction behaviours for developing autonomous systems, i.e., to learn from the expert operator, in a setting with high ecological validity. It is particularly suited to investigations where a high degree of adaptive social interaction is required, and where user behaviour is harder to predict *a priori*, such as experiments 'in the wild'. Further, as differences between face-to-face and robot mediated interaction are largely due to robot appearance and physical capabilities, some inferences can be made as to behaviours that may or may not be useful for a socially interactive robot, directing future investigations.

The example study presented gives a more concrete idea of the methods application, and is our first investigation as to the utility of the methodology. Our findings show that while the method shows promise, a great deal more work is needed to understand the data it provides. The study is ongoing work, and as such in the future we will be adding an additional voice only communication condition, and doing much more in depth data analysis. Another avenue of future work is to use voice changing software with the tele-operation system to conduct super Wizard of Oz experiments, but to do so it will be important to ascertain whether naïve participants can be convinced the robot is autonomous.

In future work we also aim to modify the control scheme to have some degree of semi-autonomy, allowing us to test components of an autonomous social robot system through comparison with a fully tele-operated system. To do so we will use the concept of transformed social interaction proposed by Bailenson et al. [16], whereby elements of an avatars interactive behaviours are modified from the control inputs of the operator.

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, no. 1, Jan. 2012.

[2] R. E. Heyman, M. F. Lorber, J. M. Eddy, and T. V. West, "Behavioral observation and coding," in *Handbook of Research Methods in Social and Personality Psychology*, H. T. Reis and C. M. Judd, Eds. Cambridge University Press, 2014, pp. 345–372, cambridge Books Online.

[3] M. Slater, "Real People Meeting Virtually Real People a Review of Some Experiments in Shared Virtual Environments," *BT Technology Journal*, vol. 17, no. 1, pp. 120–127, 1999.

[4] A. M. von der Pütten, N. C. Krämer, J. Gratch, and S.-H. Kang, "It doesnt matter what you are! Explaining social effects of agents and avatars," *Computers in Human Behavior*, vol. 26, no. 6, pp. 1641–1650, Nov. 2010.

[5] J. Fox, S. J. Ahn, J. H. Janssen, L. Yeykelis, K. Y. Segovia, and J. N. Bailenson, "Avatars Versus Agents: A Meta-Analysis Quantifying the Effect of Agency on Social Influence," *HumanComputer Interaction*, vol. 30, no. 5, pp. 401–432, May 2014.

[6] S. Coradeschi and A. Loutfi, "Social Robotic Telepresence," *Technology*, pp. 5–6, 2010.

[7] P. Bremner and U. Leonards, "Efficiency of speech and iconic gesture integration for robotic and human communicators - a direct comparison," in *Proceedings of 2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015.

[8] T. Bänziger and K. Scherer, "Using Actor Portrayals to Systematically Study Multimodal Emotion Expression: The GEMEP Corpus," in *Affective Computing and Intelligent Interaction: Second International Conference, ACII 2007 Lisbon, Portugal, September 12-14, 2007 Proceedings*, A. Paiva, R. Prada, and R. W. Picard, Eds., 2007, pp. 476–487.

[9] C. Busso and S. Narayanan, "Scripted dialogs versus improvisation: lessons learned about emotional elicitation techniques from the IEMOCAP database." *INTERSPEECH*, 2008.

[10] B. Parkinson, "Interpersonal Emotion Transfer: Contagion and Social Appraisal," *Social and Personality Psychology Compass*, vol. 5, no. 7, pp. 428–439, July 2011.

[11] H. Miwa, K. Itoh, M. Matsumoto, M. Zecca, H. Takariobu, S. Roccella, M. Carrozza, P. Dario, and A. Takanishi, "Effective emotional expressions with emotion expression humanoid robot WE-4RII," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566)*, vol. 3. IEEE, 2004, pp. 2203–2208.

[12] A. Lim, T. Ogata, and H. G. Okuno, "Converting emotional voice to motion for robot telepresence," in *2011 11th IEEE-RAS International Conference on Humanoid Robots*. IEEE, Oct. 2011, pp. 472–479.

[13] S. Serholt, C. A. Basedow, W. Barendregt, and M. Obaid, "Comparing a humanoid tutor to a human tutor delivering an instructional task to children," *2014 IEEE-RAS International Conference on Humanoid Robots*, pp. 1134–1141, Nov. 2014.

[14] D. Gouaillier, V. Hugel, P. Blazevic, C. Kilner, J. Monceaux, P. Lafourcade, B. Marnier, J. Serre, and B. Maisonnier, "Mechatronic design of NAO humanoid," in *2009 IEEE International Conference on Robotics and Automation*. IEEE, May 2009, pp. 769–774.

[15] S. Nishio, H. Ishiguro, and N. Hagita, "Geminoid: Teleoperated android of an existing person," in *Humanoid Robots: New Developments*, 2007, pp. 343–352.

[16] J. N. Bailenson, A. C. Beall, J. Loomis, J. Blascovich, and M. Turk, "Transformed Social Interaction: Decoupling Representation from Behavior and Form in Collaborative Virtual Environments," *Presence: Teleoperators and Virtual Environments*, vol. 13, no. 4, pp. 428–441, Aug. 2004.