The Effect of Anthropomorphism on Social Tele-Embodiment

Christian Becker-Asano[1] and Kai O. Arras[2] and Bernhard Nebel[1] and Hiroshi Ishiguro[3]

Abstract— **This paper outlines our approach to explore the impact of using two different robotic embodiments on an operators ability to convey emotional and conversational nonverbal signals to a distant interlocutor. Although a human's ability to produce and interpret complex, dynamic facial expressions is seen as an important factor for human-human social interaction, it remains controversial in humanoid/android robotics, whether recreating such expressiveness is really worth the technical challenge, or not. In fact, one way to avoid the risk of giving rise to uncanny feelings in human observers is to follow an abstract design for humanoid robots. This question is also relevant in the context of mediated interaction using tele-operation technology, as soon as robotic embodiments are involved. Thus, this paper presents our current project, in which we are comparing the efficiency of transmitting nonverbal signals by means of "Daryl" featuring an abstract, mildly humanized design, against that of "Geminoid F", which features a highly anthropomorphic design. The ability of both of these robots to convey emotions by means of body movements has been successfully evaluated before, but using this ability to transmit nonverbal signals during remote conversation and comparing the resp. efficiencies has not yet been done.**

I. INTRODUCTION AND MOTIVATION

Robotic tele-presence solutions are at the brink of becoming a reality for everyday users including geographically distributed work teams or people with disabilities [1]. The promise is to be remotely present at a distant place and efficiently interact with objects and other people, for instance, in office meetings, factories, or warehouses. However, current systems for such a social tele-embodiment are still not much more than laptops on wheels with limited abilities to display emotional body language and transmit conversational nonverbal cues. Thus, the motivation of this study is to explore how an operator's ability to transmit and then use emotional and conversational nonverbal signals [2] depends on the kind of remote embodiment that is being tele-operated. The following two robotic realizations of such embodiments will be compared against each other:

1) Daryl: A humanoid robot featuring an abstract, partly humanoid and zoomorphic design (cp. Fig. 1, left)
2) Geminoid F: An android-type, very anthropomorphic robot that has proven to be indistinguishable from a real human at first sight (cp. Fig. 1, right)

*This work was partially supported Grant-in Aid for Scientific Research (S), KAKENHI (20220002) and by the German Research Foundation (DFG) under contract number SFB/TR-8

[1]C. Becker-Asano and B. Nebel are with the Department of Computer Science, University of Freiburg, 79110 Freiburg, Germany basano,nebel@informatik.uni-freiburg.de

[2]K. O. Arras is with the Social Robotics Lab, Department of Computer Science, University of Freiburg, 79110 Freiburg, Germany arras@informatik.uni-freiburg.de

[3]H. Ishiguro is with the Department of Adaptive Machine Systems, Osaka University, Osaka, Japan ishiguro@sys.es.osaka-u.ac.jp

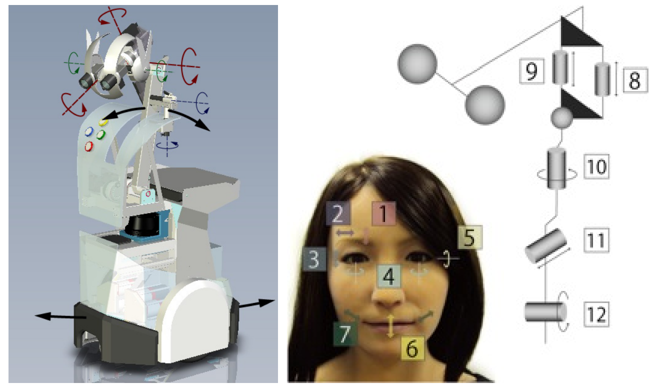

Fig. 1. The robot "Daryl" (left) with ten degrees of freedom and the android robot "Geminoid F" (right) with its twelve degrees of freedom

Fig. 2. The desktop-style tele-operation console currently in use for "Geminoid F" (left) and the head-mounted display (HMD, right), which will be used for tele-operation in this project

In addition, it is planned to compare two different operator modalities (cp. Fig. 2) with respect to their effectiveness in giving the operator a high sense of copresence [3].

The following research questions are being addressed:

(a) Does an operator's feeling of presence change significantly depending on the operation modality and which drawbacks and advantages do they have?
(b) Does the operator modality change a human conversation partner's impression of the tele-operated robots independent of the robot type?
(c) Which limitations and advantages result from the two different embodiments with respect to a conversation partner's ability to read an operator's emotional and conversational nonverbal signals?

II. RELATED WORK

Our research question is related to social robotics as well as tele-robotics research. In social robotics a robots ability to simulate and convey emotions [4] is seen as an important factor and with the advent of android robots such as “Geminoid HI-1” [5] or “Geminoid F” [6] it became even more natural to investigate the power of such a robot’s facial expressiveness. In presence research [7], on the other hand, a tele-embodiment’s ability to recreate as many of the non-verbal cues of a distant interactant as possible has also been found to increase the “sense of copresence” [3]. With Geminoid-type android robots resembling their human counterparts to the finest detail, the term “tele-presence” has been coined to refer to the question of whether such tele-operated androids “can represent the authority of the person himself by comparing the person and the android.” [8] The term “copresence” [3], however, has been introduced to specifically label what has also been called social presence, i.e. not only the feeling of being in a distant place, but also the concomitant feeling of sharing this place with others. For both notions of presence the outer appearance of the remote embodiment seems to heavily influence this subjective feeling both on the operator’s as well as the interlocutor’s side.

The effects of an android’s anthropomorphic appearance and its body movements have mainly been investigated in a number of empirical studies within laboratory environments [9]. The high level of control that can be achieved in such a setting is certainly helpful to collect fine-grained data on specific aspects of human-android interaction. To complement the previous research with data from real-life situations, the uncanny valley hypothesis has been investigated with Geminoid HI-1, which was tele-operated in a public location [10]. The feelings reported by the interviewed visitors after they had interacted with Geminoid HI-1 were less negative than those reported during comparable laboratory experiments in Japan. In addition, a number of possible improvements became obvious, some of which are to be addressed in our current research project.

In order to assess a human visitor’s emotional reactions in unscripted interactions, Geminoid HI-1 has been placed secretly in a public café for one month before ARS Electronica 2009. Structured interviews and careful video analyses suggest that in such open scenarios people rather easily mistake an android for a human [11]. In addition, even when Geminoid HI-1 is recognized as being artificial, only a few people report uncanny feelings. These results were complemented by a laboratory study on the emotional effects of Geminoid HI-1, when it starts to laugh [12].

Concerning the expression of emotions, a number of upper body movements of the humanoid robot Daryl have been designed and positively evaluated enabling it to express the emotions happiness, sadness, fear, curiosity, embarrassment, and disappointment [13]. Results of an empirical investigation with Geminoid F provide evidence that some of its facial expressions are reliably interpreted as to depict the emotions anger, happy, sad, and surprised [6].

In all of these scenarios the involved Geminoids are either controlled by manually triggering scripted body movements or by tracking and copying an operator’s head and face motions. In both cases, the operator can only observe his or her Geminoid from a third-person point of view by means of standard two-dimensional display technology similar to the ones described in [14]. The research project outlined here aims to investigate the effects of employing more advanced tele-presence technology, namely a head-mounted display (HMD). Very similar display technology in combination with head-tracking has recently been used for an immersive and interactive presentation of a virtual emergency scenario [15]. The results indicate that similar levels of physiological arousal and fear have been elicited in the players as when passively watching a horror movie clip. Especially the operator’s ability to change the virtual viewing direction simply by moving his or her head seems to increase this effect dramatically. We believe that by employing this kind of sophisticated tele-operation interface we can get significantly closer to achieving emotionally engaging social tele-embodiment [16].

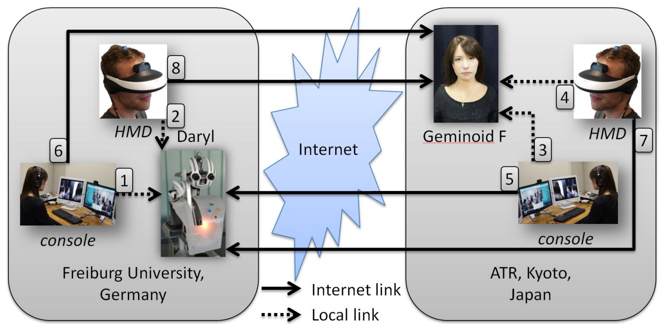

Fig. 3. The humanoid robot “Daryl” (left) and the android robot “Geminoid F” (right), which are used in the project outlined here. The numbers indicate the priority of realization (see also Table I)

III. PROJECT OUTLINE

In this project we are cross-combining the tele-operation modality (HMD versus console, cp. Fig 2) with two different remote embodiments (Geminoid F versus Daryl) as illustrated in Figure 3. In addition, these robots will be tele-operated based on a local as well as a remote, Internet-based video-audio-link. Accordingly, Table I summarizes the resulting eight different conditions.

TABLE I
OVERVIEW OF THE EIGHT EXPERIMENTAL CONDITIONS; THE NUMBERING OF CONDITIONS ALSO INDICATES THEIR RESPECTIVE PRIORITY OF REALIZATION

	console	HMD
Daryl	(1) local & (5) remote	(2) local & (7) remote
Geminoid	(3) local & (6) remote	(4) local & (8) remote

A. *Local tele-operation of Daryl; conditions (1) and (2)*

We will start with operating the Daryl robot locally by, first, using the classical console-based operator interface (condition 1, cf. Fig. 2, left) and then the HMD-based operator interface (condition 2, cf. Fig. 2, right). For condition one, a webcam in combination with FaceAPI [17] has been used previously to track the operators head (and mouth) movements. These were then instantaneously mapped onto the corresponding actuators of the Geminoid robots and seamlessly combined with a playback of the operator's voice (see also [10] for details). In case of Daryl, we plan to use the 6-DoF (degrees of freedom) inertial motion tracker "colibri" [18] mounted on a headset to track the operator's head movements. This device has proven to provide reliable tracking results in combination with the HMD in a different project [15].

In this setup, the operator can observe the surrounding of her remote embodiment on two desktop displays from two perspectives, first, a third-person overview, and, second, a first-person view provided by a camera directly behind the robot; cp. Fig. 2, left. Notably, the resp. positions and orientations of both views remain static, even when the robot's head orientation is being changed in accordance with the operator's head movements. This rather simple setup has been sufficient to let the operator falsely identify his own location with that of the remote embodiment after some time of continuous tele-operation [10].

For condition two we plan to use the HMD as an interface for tele-operation. Accordingly, the operator will be provided with a stereo, first-person view from the remote embodiment's point of view. These two video streams will be displayed on two 1280x720 pixel OLED displays inside an "HMZ-T1" head-mounted display from Sony (cf. Fig. 2, right). The 6-DoF tracker will be mounted on top of this HMD providing the system with tracking data that are again used to control Daryl's head movements. In contrast to condition one, however, the stereo video stream of Daryl's two inbuilt firewire cameras will be used as visual feedback for the operator. Thus, in this condition the viewpoint's position and orientation will change together with an operator's head movements. We assume that this will lead to a higher degree of copresence [3] in the operator and, in addition, that this kind of interface will be much more intuitive to use.

B. Local tele-operation of Geminoid F; conditions (3) and (4)

Conditions three and four focus on local tele-operation of Geminoid F. The console-based tele-operation has already been realized (condition 3), but it remains an open challenge to evaluate the effectiveness of this interface against the alternative of using an HMD. The latter interface will be realized similar to condition two described above.

At least the following three problems need to be addressed:

1) An operator might need to manually trigger facial expressions in Geminoid F through the interface. Whilst wearing an HMD (condition four) the operator is unable to perceive her local environment and, thus, it may be necessary to implement a graphical user interface as an overlay on top of the 3D visual information presented through the HMD. Then the operator would still be able to press buttons using the mouse in order to trigger pre-defined motion sequences and facial expressions of Geminoid F.
2) An operator's mouth movements might need to be tracked in order to let Geminoid F move its mouth accordingly. The webcam-based approach described above might be impossible, because major parts of an operator's face will be blocked by the HMD. Thus, we might need to rely on the analysis of the operator's voice energy levels.
3) In contrast to Daryl, Geminoid F is not equipped with cameras in its eyes. Thus, we are working on a technical solution that allows us to remotely change the orientation of a stereo camera that will be placed behind Geminoid F. This way an operator's viewpoint can be changed independent of the actual orientation of Geminoid F's head, if needed. Also, if the cameras were located inside Geminoid F's eyes, its blinking would disrupt the operator's vision.

C. Internet-based tele-operation of Daryl and Geminoid F; conditions (5) to (8)

In a final step, we also want to test the performance of our technical solution when the remote embodiments are tele-operated through an Internet connection; cf. Fig. 3. Although many technical details remain the same, establishing a reliable, low-latency network link between Germany and Japan for both the movement commands as well as the real-time video and audio streams will be challenging. An operator's head movements (Fig. 4, left) will be mapped on the robot's axis inside a control client, before they will be transmitted together with the audio stream via a TCP socket. The "Geminoid Sever" developed and maintained by the Ishiguro laboratory at ATR, Japan, is currently being extended to interface with Daryl. It will receive the continuous stream of movement commands and translate them into appropriate movement commands for Geminoid F or Daryl, either of which will be used to recreate the operator's head movements in synchrony with the audio transmission (Fig. 4, right).

The back-channel for the operator is realized either by two independent cameras (console) or a stereo camera (HMD) and a stereo microphone at the tele-embodiment's distant location (Fig. 4, upper right). Both video streams together with the audio stream are transmitted by the robot control server using UDP-based Real-time Transport Protocol (RTP) back to the Tele-Op Client, where they are played back to the operator.

IV. EMOTIONAL AND CONVERSATIONAL NONVERBAL SIGNALS

According to [2], five types of emotional and conversational nonverbal signals can be distinguished: emblems, illustrators, manipulators, regulators, and emotional expressions.

Emblems as well as manipulators are of minor interest to our project, because they involve hand and arm movements and our robotic embodiments don't allow for the recreation of those movements. Similarly, we plan to concentrate our

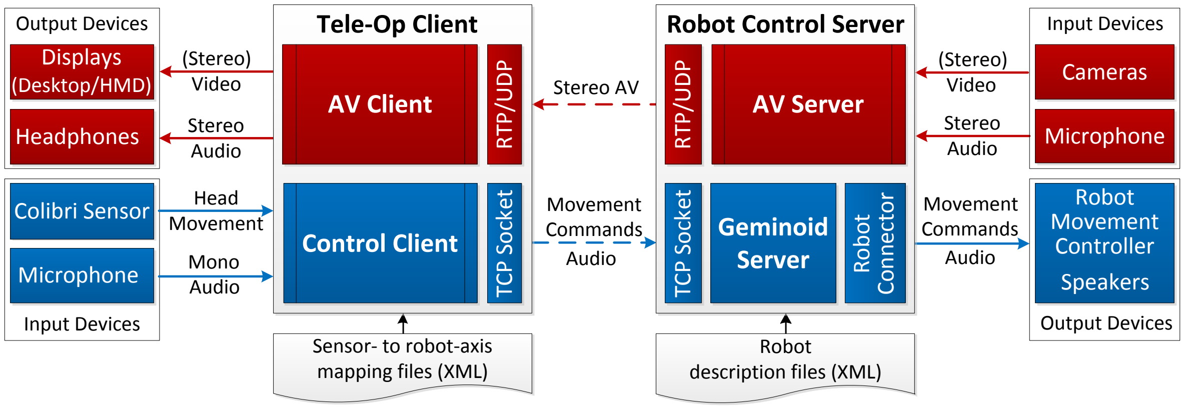

Fig. 4. Technical realization overview of the project with the operator on the left and the robotic tele-embodiment on the right

study on only one of the seven types of illustrators distinguished by [2], namely, the "batons". They involve head and also facial movements and serve to emphasize a phrase or a word. Interestingly, most commonly "facial batons involve either brow raising or brow lowering" [2, p. 41], which are both challenging to detect when the operator uses an HMD.

Both regulators and emotional expressions will be equally important for our project as the former refers to all those head and facial movements that regulate the flow of conversation and the latter can serve as both a direct nonverbal feedback to and a modulator of conversational content. Accordingly, an operator's head nods and at best also her facial expressions would need to be recognized and replayed in synchrony with the audio signal to achieve a transmission of such nonverbal signals. This is clearly a challenging task in itself, but it becomes even more difficult in combination with an HMD, by which parts of an operator's face are occluded (cp. Fig. 2, right).

V. Conclusion

We are confident that the two robotic embodiments presented above will provide sophisticated means to explore the research questions raised in Section I. It remains open, however, how exactly the multivariate effects of combining two embodiments with two tele-operation modalities will be assessed and analyzed. Perhaps questionnaire based post-hoc interviews together with video based online analyses will be the most suitable combination, because the high level cognitive evaluations resulting from the former could be combined with an interpretation of the more spontaneous reactions shown by the latter.

ACKNOWLEDGMENT

We would like to thank Severin Gustorff for his contribution to this project.

References

[1] E. Guizzo, "When my avatar went to work," *IEEE Spectrum Magazine*, vol. 47, no. 9, pp. 26–50, 2010.

[2] P. Ekman, "Emotional and conversational nonverbal signals," in *Language, knowledge, and representation*, M. Larrazabal and L. Miranda, Eds. Kluwer Academic Publishers, 2004, pp. 39–50.

[3] S. Zhao, "Toward a taxonomy of copresence," *Presence: Teleoperators and Virtual Environments*, vol. 12, no. 5, pp. 445–455, 2003.

[4] C. Breazeal, "Function meets style: Insights from Emotion Theory applied to HRI," *IEEE SMC Transactions, Part C*, vol. 32, pp. 187–194, 2004.

[5] S. Nishio, H. Ishiguro, and N. Hagita, "Geminoid: Teleoperated android of an existing person," in *Humanoid Robots, New Developments*, A. C. de Pina Filho, Ed. Vienna, Austria: I-Tech, June 2007, pp. 343–352.

[6] C. Becker-Asano and H. Ishiguro, "Evaluating facial displays of emotion for the android robot Geminoid F," in *IEEE SSCI Workshop on Affective Computational Intelligence*, ser. IEEE 2011 Symposium Series on Computational Intelligence, 2011, pp. 22–29.

[7] M. Slater and S. Wilbur, "A framework for immersive virtual environments (five) - speculations on the role of presence in virtual environments," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 6, pp. 603–616, 1997.

[8] S. Nishio, H. Ishiguro, and N. Hagita, "Can a teleoperated android represent personal presence? — a case study with children," *PSYCHOLOGIA*, vol. 50, no. 4, pp. 330–342, 2007.

[9] C. Becker-Asano, "Affective computing combined with android science," *KI - Künstliche Intelligenz*, vol. 25, pp. 245–250, 2011.

[10] C. Becker-Asano, K. Ogawa, S. Nishio, and H. Ishiguro, "Exploring the uncanny valley with Geminoid HI-1 in a real-world application," in *IADIS Intl. Conf. Interfaces and Human Computer Interaction*, 2010, pp. 121–128.

[11] A. M. von der Pütten, N. C. Krämer, C. Becker-Asano, and H. Ishiguro, "An android in the field," in *ACM/IEEE Intl. Conf. on Human-Robot Interaction*, Lausanne, Switzerland, 2011, pp. 283–284.

[12] C. Becker-Asano and H. Ishiguro, "Laughter in social robotics – no laughing matter," in *8th intl. workshop on Social Intelligence Design (SID2009)*, 2009, pp. 287–300.

[13] S. Embgen, M. Luber, C. Becker-Asano, M. Ragni, V. Evers, and K. O. Arras, "Robot-specific social cues in emotional body language," in *IEEE Intl. Symp. on Robot and Human Interactive Communication (Ro-Man 2012)*, 2012, accepted.

[14] T. Fong and C. Thorpe, "Vehicle teleoperation interfaces," *Autonomous Robots*, vol. 11, pp. 9–18, 2001.

[15] C. Becker-Asano, D. Sun, B. Kleim, C. N. Scheel, B. Tuschen-Caffier, and B. Nebel, "Outline of an empirical study on the effects of emotions on strategic behavior in virtual emergencies," in *Affective Computing and Intelligent Interaction*, ser. Lecture Notes in Computer Science, S. D'Mello, A. Graesser, B. Schuller, and J.-C. Martin, Eds., vol. 6975. Springer, 2011, pp. 508–517.

[16] E. Paulos and J. Canny, "Social tele-embodiment: Understanding presence," *Autonomous Robots*, vol. 11, pp. 87–95, 2001.

[17] Seeing Machines, "Faceapi product website," CTAN http://www.seeingmachines.com/product/faceapi/, July 2012.

[18] Trivisio Prototyping GmbH, "Colibri - inertial motion tracker," CTAN http://www.trivisio.com/products/motiontracking/colibri, July 2012.