

INVESTIGATING HUMAN NATURE AND COMMUNICATION THROUGH ROBOTS

EDITED BY : Shuichi Nishio, Hideyuki Nakanishi and Tsutomu Fujinami
PUBLISHED IN: Frontiers in Psychology and Frontiers in ICT





frontiers

Frontiers Copyright Statement

© Copyright 2007-2017 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88945-086-2

DOI 10.3389/978-2-88945-086-2

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

INVESTIGATING HUMAN NATURE AND COMMUNICATION THROUGH ROBOTS

Topic Editors:

Shuichi Nishio, Advanced Telecommunications Research Institute International, Japan

Hideyuki Nakanishi, Osaka University, Japan

Tsutomu Fujinami, Japan Advanced Institute of Science and Technology, Japan



Image by Shuichi Nishio

The development of information technology enabled us to exchange more items of information among us no matter how far we are apart from each other. It also changed our way of communication. Various types of robots recently promoted to be sold to general public hint that these robots may further influence our daily life as they physically interact with us and handle objects in environment. We may even recognize a feel of presence similar to that of human beings when we talk to a robot or when a robot takes part in our conversation. The impact will be strong enough for us to think about the meaning of communication. This e-book consists of various studies that examine our communication influenced by robots. Topics include our attitudes toward robot behaviors, designing robots for better communicating with people, and how people can be affected by communicating through robots.

Citation: Nishio, S., Nakanishi, H., Fujinami, T., eds. (2017). Investigating Human Nature and Communication through Robots. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-086-2

Table of Contents

- 04 Editorial: Investigating Human Nature and Communication through Robots**
Shuichi Nishio, Hideyuki Nakanishi and Tsutomu Fujinami
- 06 Infant discrimination of humanoid robots**
Goh Matsuda, Hiroshi Ishiguro and Kazuo Hiraki
- 13 Iconic Gestures for Robot Avatars, Recognition and Integration with Speech**
Paul Bremner and Ute Leonards
- 27 Physical embodiment can produce robot operator's pseudo presence**
Kazuaki Tanaka, Hideyuki Nakanishi and Hiroshi Ishiguro
- 39 Expression transmission using exaggerated animation for Elfoïd**
Maiya Hori, Yu Tsuruda, Hiroki Yoshimura and Yoshio Iwai
- 49 Attitudinal Change in Elderly Citizens Toward Social Robots: The Role of Personality Traits and Beliefs About Robot Functionality**
Malene F. Damholdt, Marco Nørskov, Ryuji Yamazaki, Raul Hakli, Catharina Vesterager Hansen, Christina Vestergaard and Johanna Seibt
- 62 Can We Talk through a Robot As if Face-to-Face? Long-Term Fieldwork Using Teleoperated Robot for Seniors with Alzheimer's Disease**
Kaiko Kuwamura, Shuichi Nishio and Shinichi Sato
- 75 Intimacy in Phone Conversations: Anxiety Reduction for Danish Seniors with Hugin**
Ryuji Yamazaki, Louise Christensen, Kate Skov, Chi-Chih Chang, Malene F. Damholdt, Hidenobu Sumioka, Shuichi Nishio and Hiroshi Ishiguro
- 84 Impact of Mediated Intimate Interaction on Education: A Huggable Communication Medium that Encourages Listening**
Junya Nakanishi, Hidenobu Sumioka and Hiroshi Ishiguro
- 94 A truly human interface: interacting face-to-face with someone whose words are determined by a computer program**
Kevin Corti and Alex Gillespie



Editorial: Investigating Human Nature and Communication through Robots

Shuichi Nishio^{1*}, Hideyuki Nakanishi² and Tsutomu Fujinami³

¹ Hiroshi Ishiguro Laboratory, Advanced Telecommunications Research Institute International, Kyoto, Japan, ² Department of Adaptive Machine Systems, Osaka University, Osaka, Japan, ³ School of Knowledge Science, Japan Advanced Institute of Science and Technology, Ishikawa, Japan

Keywords: robot, communication, enhancement, human nature, teleoperation, embodiment

Editorial on the Research Topic

Investigating Human Nature and Communication through Robots

The aim of this research topic was to gather findings and hypothesis on how robotic devices have changed, or may change, ways of communication between people. In the last two decades, people acquired new means for communication; cellphones, e-mail, chat, SNS, and so on. With such communication media, along with the progress in information technologies and devices such as World Wide Web (WWW) and smartphones, ones lifestyle has rapidly changed. We can now talk with others anywhere and anytime, can send and receive not just text or voice but also images, movies to express our ideas and feelings in finer detail. Such changes not only increased the bandwidth and relaxed the distance limitation of communication; they also changed how people communicate with each other. Such changes provided researchers with new sources and methods for investigating human nature such as cognitive properties and sociological tendencies.

Now various types of robots that are aimed to work in our daily environment are developed and starting to appear in markets. Some robots can make simple conversation with people autonomously. Some cannot speak but people anthropomorphize them and talk to them. Some work as a mobile video chat system. Robots differ from existing information devices in that they can physically interact with real world objects. They can move round in the world we live, can carry things, can touch people or can be touched by people. You can feel a strong presence of the robot. Having conversation with such robots, or having conversation with other persons through such robots may re-define the meaning of communication.

People are starting to apply this new possibility in various fields. Some are making theater performance and art works with robots. Some are trying to use robots as means to understand and to talk to people with cognitive impairments such as dementia and autism. And some are using robots to refine communication with others. Such trials, as well as efforts to refine robots so that people can easily interact with them, are shedding lights on previously unknown human nature; e.g., how we recognize ourselves and others, what it is to have communication with others.

In this research topic, human communication with robots or through robots were examined from versatile aspects. Two papers examined how robots or their behavior are recognized by people. Matsuda et al. tested if infant can discriminate androids, a robot with very humanlike appearance. It is well known that people tend to feel “uncanniness” toward androids (MacDorman and Ishiguro, 2006). They examined if this uncanny feeling is equipped in people from birth or is developed during growth. Bremner and Leonards examined whether human can process gestures produced by robots in the same way as produced by others humans. When we speak to others, non-verbal elements are generated inevitably due to our body and such elements are processed in combination to speech. The question is, will this multi-modal processing be triggered for robots as well.

OPEN ACCESS

Edited and reviewed by:

Anton Nijholt,
University of Twente, Netherlands

*Correspondence:

Shuichi Nishio
nishio@botransfer.org

Specialty section:

This article was submitted to
Human-Media Interaction,
a section of the journal
Frontiers in Psychology

Received: 03 October 2016

Accepted: 31 October 2016

Published: 18 November 2016

Citation:

Nishio S, Nakanishi H and Fujinami T
(2016) Editorial: Investigating Human
Nature and Communication through
Robots. *Front. Psychol.* 7:1784.
doi: 10.3389/fpsyg.2016.01784

Two papers examined how robots can be made to affect people. Tanaka et al. tested how various factors in a robot, especially its physical embodiment, affect its social telepresence. That is, if robot that have physical bodies are better than telephones or video phones or not, in what way, and how robots can be used to make the effect stronger. Hori et al. tried to express emotion with robots, not by using body motions or facial expressions, but by changing illumination patters.

Four papers tested how communicating through robots would affect people. Damholdt et al. how elderly citizens will respond to a teleoperated robot. That is, when having a conversation through a robot that is controlled by another person, what personal factors of the elderly citizen will affect their attitude toward the robot. Kuwamura et al. focused on elderlies with dementia. They performed a long term testing in a care facility using a teleoperated robot, and checked how the robot is accepted and how people interacted with the robot. Yamazaki et al. examined if having a conversation through a robot, instead of using a telephone, would reduce stress. Nakanishi et al. describes their attempt to use a teleoperated robot as teaching tool in school.

By using a huggable device, they examined if talking to children through the device would help the children to concentrate more on teacher.

And finally, Corti and Gillespie showed a unique setup on “robotic” teleoperation. Instead of having an operator person who controls a robot, they used an artificial chat system (chat bot) to determine what to speak in conversation with others. This “Echoborg” is used to perform several testings and the authors also discuss on the possibility of creating an androids that speak autonomously.

AUTHOR CONTRIBUTIONS

SN wrote the article; HN and TF provided comments on the draft.

FUNDING

Part of this work has been supported by: JST/ERATO; Grants-in-Aid for Scientific Research 25220004; Strategic Platforms for Innovation and Research.

REFERENCES

MacDorman, K. F., and Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interact. Stud.* 7, 297–337. doi: 10.1075/is.7.3.03mac

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Nishio, Nakanishi and Fujinami. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Infant discrimination of humanoid robots

Goh Matsuda^{1,2}, Hiroshi Ishiguro^{3,4} and Kazuo Hiraki^{2,5*}

¹ Department of Medical Education and General Medicine, Kyoto Prefectural University of Medicine, Kyoto, Japan,

² Department of General Systems Studies, Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan,

³ Department of Systems Innovation, Graduate School of Engineering Science, Osaka University, Osaka, Japan, ⁴ Hiroshi Ishiguro Laboratories, Advanced Telecommunications Research Institute International, Kyoto, Japan, ⁵ CREST – Japan Science and Technology Agency, Tokyo, Japan

OPEN ACCESS

Edited by:

Tsutomu Fujinami,
Japan Advanced Institute of Science
and Technology, Japan

Reviewed by:

Alejandro Catala,
University of Castilla-La Mancha,
Spain
Rubén San Segundo Hernández,
Speech Technology Group,
Universidad Politécnica de Madrid,
Spain

*Correspondence:

Kazuo Hiraki,
Department of General Systems
Studies, Graduate School of Arts
and Sciences, The University
of Tokyo, 3-8-1 Komaba, Meguro-ku,
Tokyo 153-8902, Japan
khiraki@idea.c.u-tokyo.ac.jp

Specialty section:

This article was submitted to
Human-Media Interaction,
a section of the journal
Frontiers in Psychology

Received: 15 July 2015

Accepted: 02 September 2015

Published: 22 September 2015

Citation:

Matsuda G, Ishiguro H and Hiraki K
(2015) Infant discrimination
of humanoid robots.
Front. Psychol. 6:1397.
doi: 10.3389/fpsyg.2015.01397

Recently, extremely humanlike robots called “androids” have been developed, some of which are already being used in the field of entertainment. In the context of psychological studies, androids are expected to be used in the future as fully controllable human stimuli to investigate human nature. In this study, we used an android to examine infant discrimination ability between human beings and non-human agents. Participants ($N = 42$ infants) were assigned to three groups based on their age, i.e., 6- to 8-month-olds, 9- to 11-month-olds, and 12- to 14-month-olds, and took part in a preferential looking paradigm. Of three types of agents involved in the paradigm—a human, an android modeled on the human, and a mechanical-looking robot made from the android—two at a time were presented side-by-side as they performed a grasping action. Infants’ looking behavior was measured using an eye tracking system, and the amount of time spent focusing on each of three areas of interest (face, goal, and body) was analyzed. Results showed that all age groups predominantly looked at the robot and at the face area, and that infants aged over 9 months watched the goal area for longer than the body area. There was no difference in looking times and areas focused on between the human and the android. These findings suggest that 6- to 14-month-olds are unable to discriminate between the human and the android, although they can distinguish the mechanical robot from the human.

Keywords: infant, humanoid robot, android, preferential looking paradigm, eye tracking, uncanny valley

Introduction

Over the last decade, various types of humanoid robots have emerged beyond the hypothetical realm of science fiction and into real life. More recently, robots with an extremely humanlike appearance, called “androids,” were developed (Ishiguro, 2006), primarily for interaction with humans. Because the best communicative partner of human beings is undoubtedly other humans, the development of a more humanlike appearance and motion for robots is considered a shortcut to developing robots that will have natural interactions with humans. Thus, investigating how currently available robots are perceived by humans will provide valuable information for this purpose.

The famous “uncanny valley” hypothesis is related to the impression conveyed by robots and their human likeness (Mori, 1970, 2012), and states that extremely humanlike artifacts often elicit negative affect, e.g., a feeling of eeriness, whereas modestly humanlike artifacts evoke familiarity. It was originally a theoretical hypothesis and remains controversial (Burleigh et al., 2013); some

subsequent studies have, however, found empirical evidence supporting the existence of a similar phenomenon in both humans (Seyama and Nagayama, 2007) and other primates (Steckenfinger and Ghazanfar, 2009). In other words, the uncanny valley hypothesis suggests that humans have a sophisticated ability to discriminate between human and non-human beings. In fact, it has been reported that 80% of adult participants recognized that an android with a highly humanlike appearance was not a real human within 1 s (Noma et al., 2006), and that brain activity when viewing a human vs. an android is significantly different, especially in the anterior intraparietal sulcus, which is involved in action perception (Saygin et al., 2011). Currently available androids, therefore, do not seem to have achieved a sufficiently humanlike appearance in the view of human adults.

On the other hand, little is known about infant perception of extremely humanlike artifacts, such as androids. Newborns show primary discrimination abilities in relation to human properties, such as faces, voices, and movements (Goren et al., 1975; DeCasper and Fifer, 1980; Moon et al., 1993; Simion et al., 2008), and gradually gain more expertise during the first year of life. For example, whereas newborns can discriminate their mothers from strangers when the mothers' heads are uncovered (Bushnell et al., 1989), they cannot do so when both women are wearing head scarves (Pascalis et al., 1995), although this only occurs up to 5 weeks of age (Bartrip et al., 2001). Moreover, at around 7 months, infants become able to process detailed facial configurations, such as the distance between eyes and mouth (Cohen and Cashon, 2001), and to identify strangers' faces from a non-frontal view (Fagan, 1976). Discrimination of biological (e.g., a walking hen) from non-biological motion has also been observed in newborns (Simion et al., 2008), but the ability to differentiate human motion (e.g., a walking person) from non-human motion appears around 3 months of age (Bertenthal et al., 1987). By around 12 months of age, infants are able to discriminate possible and impossible human movements, such as fingers or elbows bending in the opposite direction (Christie and Slaughter, 2010; Morita et al., 2012). As mentioned above, although young infants already have primary discrimination abilities in relation to humans, this is not as well-developed as it is in adults. Therefore, it is likely that infant perception of humanoid robots is different from that of adults.

Investigating infant perception of androids inevitably leads to manifesting how infants discriminate human beings from non-human beings. Androids can be regarded as a highly controlled human stimuli for use in investigating human nature in the field of cognitive science (MacDorman and Ishiguro, 2006). Some researchers have already used androids as experimental stimuli (Saygin et al., 2011; Urgen et al., 2013); however, most targeted human adults. To our knowledge, there is only one study in which preschoolers' responses to a real human and an android were compared (Moriguchi et al., 2010), and no studies on younger infants. Therefore, the purpose of this study was to investigate infant discrimination ability in regard to human beings, using humanoid robots and the preferential looking paradigm. When two kinds of stimuli are presented simultaneously in front of infants, a remarkable difference in looking times between both

stimuli indicates that infants can discriminate between each stimulus. This method was devised by Fantz in the 1950s (Fantz, 1958), and is still widely used today in the field of developmental science.

In this study, three agents—a human, an android modeled on the human, and a mechanical-looking robot made from the android—were used as the experimental stimuli. If infants can recognize relatively few differences between the human and the android, significant difference in their looking times to each agent should be observed. Taking the findings of previous studies described above into consideration, it is very likely that younger infants will not realize that the android is not a human, while infants aged over 12 months may be able to discriminate between the two; therefore, this study targeted infants aged between 6 and 14 months. Furthermore, we employed an eye tracking system to measure infant looking times because it allows for more objective measurement and more precise analysis of focused areas than manual coding does. Even if no difference is found in looking times, there may be difference in the regions infants focus on when looking at each agent. Thus, this study will provide new evidence in relation to infants' ability to discriminate human beings from non-human beings, and the pathway by which this ability develops. In addition, from the viewpoint of robotics, this experiment will evaluate the infant's perception of the human likeness of currently available androids. If the uncanny valley hypothesis applies in infancy, particular responses to the android, such as avoiding viewing the android, may be observed.

Materials and Methods

Participants

Infants ($N = 42$; 20 boys, 22 girls; age = 6–14 months) were assigned to three groups based on their age: 6–8 months (six boys, five girls, mean age = 223.73 days, $SD = 20.39$), 9–11 months (eight boys, nine girls, mean age = 291.63 days, $SD = 30.63$), and 12–14 months (six boys, eight girls, mean age = 355.39 days, $SD = 64.43$). A further 22 infants were excluded from analysis following cessation of the experiment due to fussiness, such as crying and inability to stay still ($n = 7$), or a lack of valid gaze data ($n = 15$). Details about the criteria for data exclusion are described in the data analysis subsection below.

This study was approved by the ethics committee of the University of Tokyo. Written informed consent was obtained from the parents of all participants before beginning the experiment.

Stimuli and Apparatus

The visual stimuli were three different black and white video clips (800×800 pixels, 30 fps) that depicted one of three agents (a human, an android, or a mechanical robot) performing a grasping action with their right hand. **Figure 1** shows example frames of each video clip. These clips were made from stimuli used in a previous study (Saygin et al., 2011).

In the human agent clip, a Japanese woman reached her right hand toward a tube of facial wash, grasped it for a moment, and then moved her hand back to the original position. Her facial

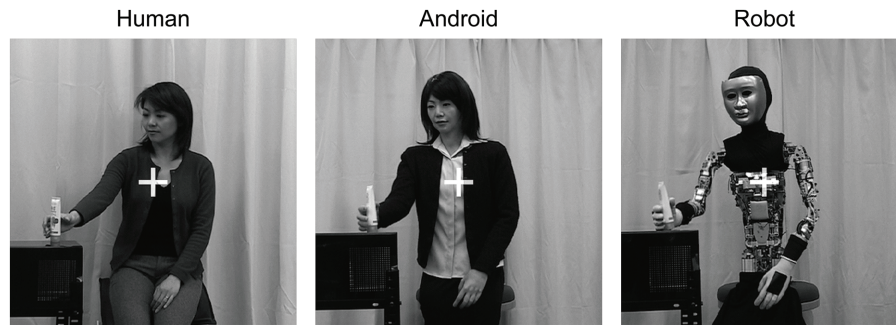


FIGURE 1 | Agents used as experimental stimuli. The android was designed to have the likeness of the human actor, and was identical in internal architecture to the robot. The original face of the robot was covered with a plastic mask to conceal its somewhat bizarre appearance, with naked eyeballs and gums.

expression did not change and her left hand remained on her left thigh. In the android and robot clips, a female android named Repliee Q2 (Osaka University and KOKORO Co. Ltd., Japan) and a mechanical humanoid robot, respectively, performed the same grasping action as the human stimulus. The Repliee Q2 was modeled on the women actor shown in the human stimulus, and its upper body is moved by air actuators. Because the mechanical robot was made by stripping away the clothing and silicone skin from the android, the robots were almost identical in terms of physical size and motion. Although the robots' motions were programmed to resemble the human's action as much as possible, those were actually rather unnatural due to mechanical limitations. In more concrete terms, whereas the human moved her hand straight to the target, the robots moved their hands over the target and then down toward it. All of the video clips were 3.5 s in duration, the second half (1.75 s) of which consisted of the first half (1.75 s) being played backwards. In addition, we used a simple animation with cheerful music that depicts a star changing in color and size as an attention getter.

Gaze data were collected at 300 Hz by the Tobii TX300 (Tobii AB, Sweden) contactless eye tracking system, which was placed at the center of a table. Its back and left and right sides were surrounded with curtains to ensure that the infants' concentration remained on the stimuli. The stimuli were presented on a 23 in liquid crystal display (1920 × 1080 pixels) integrated with the Tobii, and the actual size of each video clip on the display was a 21 cm square. A small video camera (CCD-MC100, Sony Corporation) was additionally attached at the center of the upper frame of the display so that we could observe participants' behavior. During gaze measurement, an experimenter who was located in an area separated by the curtain manipulated the Tobii and the stimuli.

Procedure

Infants viewed the stimuli while sitting on their parent's lap, and the distance between the infants and the display was approximately 60 cm. The tilt angle of the Tobii was adjusted so that it only captured infants' eyes, and then a 5-point calibration was conducted. The parent was instructed not to respond to either the infant or the stimuli. In a single trial, two different video clips were presented at the same time side-by-side on the display,

and were repeated three times without an interval. Thus, a single trial lasted 10.5 s. Each pair of agents (human vs. android: HA, human vs. robot: HR, and android vs. robot: AR) was presented four times, and the distance between two clips was 3.2 cm. The position (left or right) of the stimuli was counterbalanced. We conducted 12 trials if the infant did not become fussy, with the presentation order of each pair randomized. Before every trial, the attention getter was played at the center of the display until the infant looked toward it. Validity of eye tracking was monitored in real time using the "Show Track Status" function of the Tobii. An experimenter determined termination of the attention getter based on this status monitor and live footage from the video camera. In addition, the experimenter asked parents to move infants back to the initial position after a trial in which the Tobii lost infants' eye gaze because they moved vigorously.

Analysis

Trials with invalid (missing) gaze data for more than 50% of the trial duration were excluded from the data analysis. Moreover, participants for whom the data of one or more agent pairs was not obtained at all, were completely excluded. There were 15 infants excluded based on this criterion, primarily due to a hardware failure of the Tobii TX300 eye tracking system. According to the developer of Tobii, when the TX300 is used with a particular firmware (ver. 1.1.0), as we did in this study, it can fail to detect infant gaze during high-frequency measurement because of a problem in its algorithm for gaze detection. This problem does not occur in measurement at lower frequencies, such as at 60 and 120 Hz, and it has been fixed in the latest firmware (ver. 1.1.1). Regrettably, we lost a large amount of data because we were not aware of this important problem and its solution until after the experiment was complete.

We defined three static areas of interest (AOI), corresponding to the face area, a goal area, and the body area (see **Figure 2**). The same three AOI were applied to each agent, and statistical analysis was performed separately for each pair of agents (HA, HR, and AR). To calculate the proportions of looking times toward each AOI of each agent, mean gaze counts were divided by the total gaze count for two agents presented simultaneously. One gaze count corresponds to 3.3 ms viewing at 300 Hz sampling.

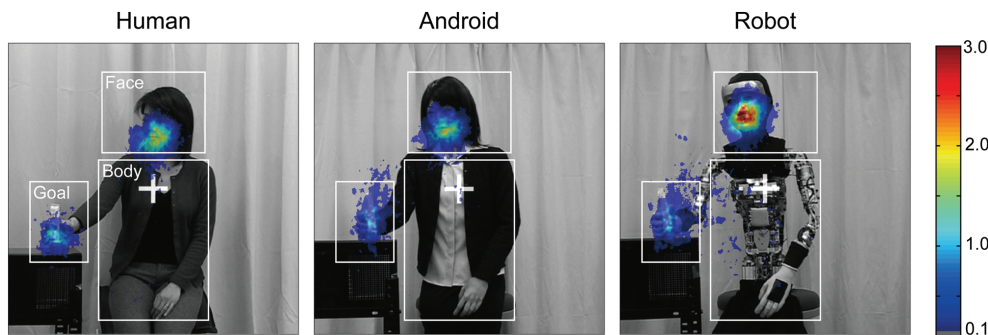


FIGURE 2 | Heat maps of mean gaze count across all trials of all participants, superimposed upon each agent after 7 × 7 pixel Gaussian smoothing was applied. Red represents an area that the greatest number of infants viewed. areas of interest (AOI) are depicted as white rectangles. The reason for the focused areas in the goal area of the android and the robot spreading vertically is probably due to the trajectories of the agents' hands.

A three-way mixed design analysis of variance (ANOVA; age group × agent × AOI) with the arcsine transformation was conducted for the proportions of looking times, and the Huynh-Feldt correction for degrees of freedom was employed as necessary. Multiple comparison with the Bonferroni method was carried out when an interaction was found.

Results

To make it easier to understand the overall trends, heat maps of the mean gaze count across all trials of all participants for each agent are shown in **Figure 2**.

Figure 3 shows the mean proportions of looking time for each AOI in each age group. There were main effects of agent in the HR and AR conditions [HR: $F(1,39) = 22.65$, $p < 0.001$; AR: $F(1,39) = 28.90$, $p < 0.001$], of AOI in all three conditions [HA: $F(1.83, 71.29) = 45.86$, $p < 0.001$; HR: $F(1.70, 66.23) = 50.64$, $p < 0.001$; AR: $F(1.67, 65.29) = 64.46$, $p < 0.001$], and of age group only in the HA condition [$F(2,39) = 5.83$, $p = 0.006$].

Moreover, an interaction between age group and AOI was found in all of the three conditions [HA: $F(3.66, 71.29) = 5.19$,

$p = 0.001$; HR: $F(3.40, 66.23) = 2.77$, $p < 0.05$; AR: $F(3.35, 65.29) = 3.17$, $p < 0.05$]. The details of significant differences between each AOI in each age group and those between each age group at each AOI are described in **Tables 1** and **2**, respectively. **Table 1** shows that infants in all age groups principally watched the face area of each agent, and that infants aged over 9 months watched the goal area for longer than they did the body area. Further, **Table 2** shows the gaze preference for the goal area in infants aged over 9 months, and shows that the 6- to 8-month-old group tended to view the body area for longer than the older groups did.

An interaction of agent and AOI was also found in the HR and AR conditions [HR: $F(2,78) = 3.53$, $p < 0.05$; AR: $F(2,78) = 12.53$, $p < 0.001$]. Multiple comparison revealed that the robot captured the longest looking time among all of the agents in any AOI ($p < 0.05$ for the goal area in the AR condition, $p < 0.01$ for the goal area in the HR condition and for the body area in the AR condition, $p < 0.001$ for the rest), and that infants viewed the face area for significantly longer than they did the other AOI (all $ps < 0.001$).

No second-order interactions were found in any conditions. Further, no effect and interaction involved in the agent factor was

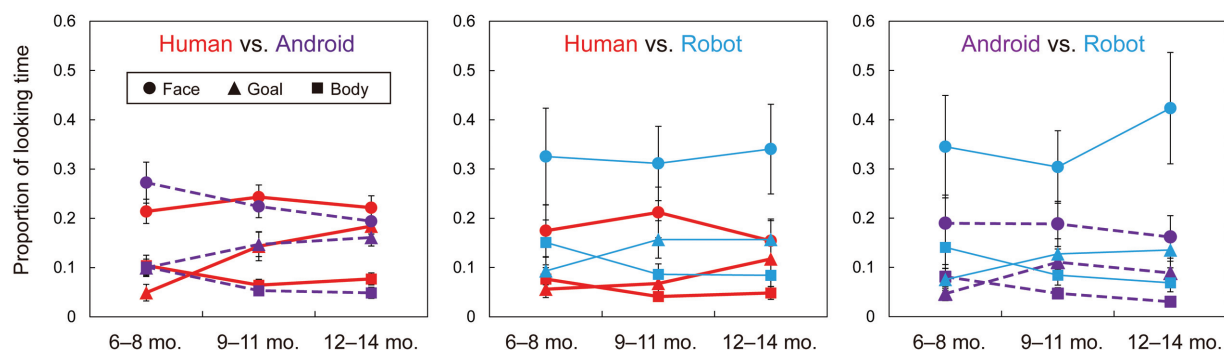


FIGURE 3 | Proportions of total looking times at each AOI of each agent across the three age groups. Red solid lines, purple dotted lines, and blue thin lines represent the human, android, and robot agents, respectively. Circle, triangle, and square markers correspond to AOI of face, goal, and body, respectively. Error bars represent standard errors.

TABLE 1 | The results of multiple comparisons for looking times between each areas of interest (AOI; Face/Goal/Body) in each age group (6–8/9–11/12–14 months).

Age (months)	Human vs. Android	<i>p</i> <	Human vs. Robot	<i>p</i> <	Android vs. Robot	<i>p</i> <
6–8	Face > Goal	0.001	Face > Goal	0.001	Face > Goal	0.001
	Face > Body	0.01	Face > Body	0.01	Face > Body	0.001
			Body > Goal	0.05		
9–11	Face > Goal	0.05	Face > Goal	0.001	Face > Goal	0.01
	Face > Body	0.001	Face > Body	0.001	Face > Body	0.001
	Goal > Body	0.001	Goal > Body	0.05		
12–14	Face > Body	0.001	Face > Goal	0.05	Face > Goal	0.001
	Goal > Body	0.001	Face > Body	0.001	Face > Body	0.001
			Goal > Body	0.01	Goal > Body	0.05

Mean values and standard errors are represented in **Figure 3**, while significant differences are described in this table. An inequality of “A > B” means that the looking time toward the AOI of A was significantly longer than that toward the AOI of B.

TABLE 2 | The results of multiple comparisons for looking times between each age group (6–8/9–11/12–14 months) in each AOI (Face/Goal/Body).

AOI	Human vs. Android	<i>p</i> <	Human vs. Robot	<i>p</i> <	Android vs. Robot	<i>p</i> <
Face	n.s.		n.s.		n.s.	
Goal	9–11 > 6–8	0.01	12–14 > 6–8	0.05		
	12–14 > 6–8	0.001			n.s.	
Body	6–8 > 9–11	0.05	6–8 > 9–11	0.05	6–8 > 9–11	0.05
					6–8 > 12–14	0.01

All mean values and standard errors are represented in **Figure 3**, while significant differences are described in this table. An inequality of “A > B” means that the looking time of group A was significantly longer than that of group B.

detected in the HA condition; that is, there were no significant differences in either looking time or focusing area between the human and the android in any age groups.

Discussion

To examine infant discrimination ability among human and humanlike agents and to test the human likeness of a currently available android, we measured looking times of infants aged between 6 and 14 months in regard to three types of agents of similar body size and motion. The three-way ANOVA revealed that infants of all age groups spent the longest time on viewing the robot, especially its face, compared with the other agents. Further, there was no difference in looking time between the human and android agents. These results suggest that 6- to 14-month-old infants are unable to distinguish the android from the human, although they are able to distinguish the robot from the human.

Infants' gaze preference for the mechanical robot is probably derived from their novelty preference tendency. A considerable number of studies have shown that infants generally prefer unfamiliar to familiar stimuli. The fact that the preference was observed in the AR condition, where the motions of both agents were almost the same, indicates that the visual aspects of the robot, rather than the motion, captured the infants' attention. Although it is likely that the infants who participated in our experiment often saw many women besides their mother in daily life, none had seen the robot before taking part in this study; therefore, the robot must have been the most unfamiliar to them from among the three agents.

Despite the fact that the android is also a rare stimulus for the infants to have observed in reality, there was no gaze preference between the human and android agent. An absence of preference for the looking paradigm does not directly indicate that two stimuli are considered to be identical; hence, it is unclear whether the infants regarded the human and the android as the same person. However, our findings suggest, at least, that the human and the android were regarded as equally humanlike beings.

A similar insensitivity to artificial humanity in infants has been reported by a previous study (Lewkowicz and Ghazanfar, 2012), where it was exhibited that 6- to 12-month-old infants were unable to discriminate a realistic computer graphics (CG) avatar from a real human. Although the authors used the term “realistic” to describe their stimuli, the stimuli actually had a non-photorealistic appearance that any adult could recognize as being a CG avatar at a glance. Our android had a more photorealistic appearance than theirs did; therefore, it should have been difficult for not only 6- to 12-month-old infants but also older infants to discriminate between the human and the android.

The motion of the android used in this study was unnatural due to its mechanical limitations. If infants recognize the unnaturalness of its motion, it is possible that they looked for longer at the android than at the human; however, the results showed that this was not the case. The android's grasping action is somewhat awkward but not impossible for human beings. It is likely that the discrimination ability of infants aged around 1 year for human movement is not yet sophisticated enough to detect this type of awkwardness.

In all the three conditions and for all age groups, infants spent the longest time looking at the face AOI. Infants' preference

for faces has been reported by many previous researchers. Even newborns under 1 week of age prefer face and face-like stimuli to other stimuli (Goren et al., 1975; Macchi et al., 2004; Farroni et al., 2005), and infants gradually focus their attention on faces at between 3 and 9 months of age (Frank et al., 2009). This preference for faces has been observed regardless of the nature of the stimuli, i.e., geometric or photographic images (Farroni et al., 2005), and is, thus, considered to reflect the importance of faces in human communication (Csibra and Gergely, 2009).

Interestingly, looking times in the goal AOI were larger in the older infant groups than in the youngest group. This probably depends on the development of their prediction ability for human action. Falck-Ytter et al. (2011) compared looking behaviors of 6- and 12-month-old infants and adults while watching human goal-directed actions, and revealed that 12-month-olds and adults looked at the goal area significantly faster and for longer than 6-month-olds did. In another similar study (Kanakogi and Itakura, 2011), the authors proposed that this prediction ability for others' actions corresponds to their own motor ability, and demonstrated that infant grasping ability develops gradually after 6 months of age. Our result is highly consistent with these findings. A shorter looking at the goal AOI in the 6- to 8-month-old group may reflect their rudimentary understanding of the goal of the agents' action.

Of course, there are limitations in our study. First, it is possible that the stimuli were too small for infants to detect slight differences in appearance and motion between the human and the android. We used 21 cm square black and white video clips, which were presented 60 cm away from the infants. An agent of this size corresponds to a real agent at about 2.5 m distance. The presentation of a real android may produce different results. In fact, presentation at a realistic size facilitates information processing about the human body in young infants (Heron and Slaughter, 2010). Second, factors that can influence the perceived human likeness of robots are not limited to their appearance and motion. For example, a study using a mechanical humanoid robot reported that infants regarded the robot as a communicative agent only after watching interactions between

a human and the robot (Arita et al., 2005). This finding implies that the interactive functions of robots can influence their human likeness. In addition, infants' characteristics, such as gender, and temperament, influence the perceived human likeness of robots. Because female, compared to male, infants have been reported to show an advantage in processing social stimuli, such as facial expressions (McClure, 2000), and to prefer more human-like stimuli, such as dolls and human faces (Connellan et al., 2000; Lutchmaya and Baron-Cohen, 2002; Alexander et al., 2009), their ability to discriminate between human and non-human beings may mature faster. Finally, gaze measurement is not the only way to investigate infant discrimination ability. Recently, infants' neural response to stimuli has been attracting attention as a new subjective index of their discrimination ability, in association with the development of non-invasive and more simplified technology for measuring brain activity (Csibra et al., 2004; Farroni et al., 2004). Although we did not find differences in infant gaze behaviors between the human and the android agents in this study, infants' neural response to the two types of agent may differ in some brain regions.

To our knowledge, this is the first report concerning infant discrimination of a recently developed android from humans and robots. Our results suggest that discrimination ability in regard to human vs. non-human beings is not as sophisticated in infants younger than 14 months as it is in adults. The uncanny valley effect elicited by the android was not found in infants; in other words, a currently available android may have already reached a humanlike quality for infants, at least with regard to appearance and motion. Androids have great potential as an alternative to human stimuli in future psychological studies.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 25220004. We thank Dr. Yoshihiro Okazaki and Dr. Yasuhiro Kanakogi for their help in conducting the experiment.

References

- Alexander, G. M., Wilcox, T., and Woods, R. (2009). Sex differences in infants' visual interest in toys. *Arch. Sex. Behav.* 38, 427–433.
- Arita, A., Hiraki, K., Kanda, T., and Ishiguro, H. (2005). Can we talk to robots? Ten-month-old infants expected interactive humanoid robots to be talked to by persons. *Cognition* 95, B49–B57. doi: 10.1016/j.cognition.2004.08.001
- Bartrip, J., Morton, J., and Schonen, S. (2001). Responses to mother's face in 3-week to 5-month-old infants. *Br. J. Dev. Psychol.* 19, 219–232. doi: 10.1348/026151001166047
- Bertenthal, B. I., Proffitt, D. R., and Kramer, S. J. (1987). Perception of biomechanical motions by infants: implementation of various processing constraints. *J. Exp. Psychol. Hum. Percept. Perform.* 13, 577–585. doi: 10.1037/0096-1523.13.4.577
- Burleigh, T. J., Schoenherr, J. R., and Lacroix, G. L. (2013). Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Comput. Hum. Behav.* 29, 759–771. doi: 10.1016/j.chb.2012.11.021
- Bushnell, I., Sai, F., and Mullin, J. (1989). Neonatal recognition of the mother's face. *Br. J. Dev. Psychol.* 7, 3–15. doi: 10.1111/j.2044-835X.1989.tb00784.x
- Christie, T., and Slaughter, V. (2010). Movement contributes to infants' recognition of the human form. *Cognition* 114, 329–337. doi: 10.1016/j.cognition.2009.10.004
- Cohen, L. B., and Cashon, C. H. (2001). Do 7-month-old infants process independent features or facial configurations? *Infant Child Dev.* 10, 83–92. doi: 10.1002/icd.250
- Connellan, J., Baron-Cohen, S., Wheelwright, S., Batki, A., and Ahluwalia, J. (2000). Sex differences in human neonatal social perception. *Infant Behav. Dev.* 23, 113–118. doi: 10.1016/S0163-6383(00)00032-1
- Csibra, G., and Gergely, G. (2009). Natural pedagogy. *Trends Cogn. Sci.* 13, 148–153. doi: 10.1016/j.tics.2009.01.005
- Csibra, G., Henty, J., Volein, A., Elwell, C., Tucker, L., Meek, J., et al. (2004). Near infrared spectroscopy reveals neural activation during face perception in infants and adults. *J. Pediatr. Neurol.* 2, 85–90.
- DeCasper, A. J., and Fifer, W. P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science* 208, 1174–1176. doi: 10.1126/science.7375928
- Fagan, J. F. III. (1976). Infants' recognition of invariant features of faces. *Child Dev.* 47, 627–638. doi: 10.1111/j.1467-8624.1976.tb02226.x

- Falck-Ytter, T., Bakker, M., and Von Hofsten, C. (2011). Human infants orient to biological motion rather than audiovisual synchrony. *Neuropsychologia* 49, 2131–2135. doi: 10.1016/j.neuropsychologia.2011.03.040
- Fantz, R. L. (1958). Pattern vision in young infants. *Psychol. Rec.* 8, 43–47.
- Farroni, T., Johnson, M. H., and Csibra, G. (2004). Mechanisms of eye gaze perception during infancy. *J. Cogn. Neurosci.* 16, 1320–1326. doi: 10.1162/089929042304787
- Farroni, T., Johnson, M. H., Menon, E., Zulian, L., Faraguna, D., and Csibra, G. (2005). Newborns' preference for face-relevant stimuli: effects of contrast polarity. *Proc. Natl. Acad. Sci. U.S.A.* 102, 17245–17250. doi: 10.1073/pnas.0502205102
- Frank, M. C., Vul, E., and Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition* 110, 160–170. doi: 10.1016/j.cognition.2008.11.010
- Goren, C. C., Sarty, M., and Wu, P. Y. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics* 56, 544–549.
- Heron, M., and Slaughter, V. (2010). Infants' responses to real humans and representations of humans. *Int. J. Behav. Dev.* 34, 34–45. doi: 10.1177/0165025409345047
- Ishiguro, H. (2006). Android science: conscious and subconscious recognition. *Conn. Sci.* 18, 319–332. doi: 10.1080/09540090600873953
- Kanakogi, Y., and Itakura, S. (2011). Developmental correspondence between action prediction and motor ability in early infancy. *Nat. Commun.* 2, 341. doi: 10.1038/ncomms1342
- Lewkowicz, D. J., and Ghazanfar, A. A. (2012). The development of the uncanny valley in infants. *Dev. Psychobiol.* 54, 124–132. doi: 10.1002/dev.20583
- Lutchmaya, S., and Baron-Cohen, S. (2002). Human sex differences in social and non-social looking preferences, at 12 months of age. *Infant Behav. Dev.* 25, 319–325. doi: 10.1016/S0163-6383(02)00095-4
- Macchi, C. V., Turati, C., and Simion, F. (2004). Can a nonspecific bias toward top-heavy patterns explain newborns' face preference? *Psychol. Sci.* 15, 379–383. doi: 10.1111/j.0956-7976.2004.00688.x
- MacDorman, K. F., and Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interact. Stud.* 7, 297–337. doi: 10.1075/is.7.3.03mac
- McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychol. Bull.* 126, 424–453. doi: 10.1037/0033-2909.126.3.424
- Moon, C., Cooper, R. P., and Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behav. Dev.* 16, 495–500. doi: 10.1016/0163-6383(93)80007-U
- Mori, M. (1970). Bukimi no tani [The uncanny valley]. *Energy* 7, 33–35.
- Mori, M. (2012). The uncanny valley. *IEEE Robot. Autom. Mag.* 19, 98–100. doi: 10.1109/Mra.2012.2192811
- Moriguchi, Y., Minato, T., Ishiguro, H., Shinohara, I., and Itakura, S. (2010). Cues that trigger social transmission of disinhibition in young children. *J. Exp. Child Psychol.* 107, 181–187. doi: 10.1016/j.jecp.2010.04.018
- Morita, T., Slaughter, V., Katayama, N., Kitazaki, M., Kakigi, R., and Itakura, S. (2012). Infant and adult perceptions of possible and impossible body movements: an eye-tracking study. *J. Exp. Child Psychol.* 113, 401–414. doi: 10.1016/j.jecp.2012.07.003
- Noma, M., Saiwaki, N., Itakura, S., and Ishiguro, H. (2006). "Composition and evaluation of the humanlike motions of an android," in *Proceedings of the 2006 6th IEEE-RAS International Conference on Humanoid Robots* (Genova: IEEE), 163–168. doi: 10.1109/ichr.2006.321379
- Pascalis, O., De Schonen, S., Morton, J., Deruelle, C., and Fabre-Grenet, M. (1995). Mother's face recognition by neonates: a replication and an extension. *Infant Behav. Dev.* 18, 79–85. doi: 10.1016/0163-6383(95)90009-8
- Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., and Frith, C. (2011). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Soc. Cogn. Affect. Neurosci.* 7, 413–422. doi: 10.1093/scan/nsr025
- Seyama, J., and Nagayama, R. S. (2007). The Uncanny Valley: effect of realism on the impression of artificial human faces. *Presence* 16, 337–351. doi: 10.1162/pres.16.4.337
- Simion, F., Regolin, L., and Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proc. Natl. Acad. Sci. U.S.A.* 105, 809–813. doi: 10.1073/pnas.0707021105
- Steckenfinger, S. A., and Ghazanfar, A. A. (2009). Monkey visual behavior falls into the uncanny valley. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18362–18366. doi: 10.1073/pnas.0910063106
- Urgen, B. A., Plank, M., Ishiguro, H., Poizner, H., and Saygin, A. P. (2013). EEG theta and Mu oscillations during perception of human and robot actions. *Front. Neurobot.* 7:19. doi: 10.3389/fnbot.2013.00019

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Matsuda, Ishiguro and Hiraki. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Iconic Gestures for Robot Avatars, Recognition and Integration with Speech

Paul Bremner^{1*} and Ute Leonards²

¹ Bristol Robotics Laboratory, University of The West of England, Bristol, UK, ² School of Experimental Psychology, University of Bristol, Bristol, UK

OPEN ACCESS

Edited by:

Shuichi Nishio,
Advanced Telecommunications
Research Institute International, Japan

Reviewed by:

Maria Koutsombogera,
Institute for Language and Speech
Processing, Greece
Kirsten Bergmann,
Bielefeld University, Germany

*Correspondence:

Paul Bremner
paul.bremner@brl.ac.uk

Specialty section:

This article was submitted to
Human-Media Interaction,
a section of the journal
Frontiers in Psychology

Received: 30 October 2015

Accepted: 31 January 2016

Published: 17 February 2016

Citation:

Bremner P and Leonards U (2016)
Iconic Gestures for Robot Avatars,
Recognition and Integration with
Speech. *Front. Psychol.* 7:183.
doi: 10.3389/fpsyg.2016.00183

Co-verbal gestures are an important part of human communication, improving its efficiency and efficacy for information conveyance. One possible means by which such multi-modal communication might be realized remotely is through the use of a tele-operated humanoid robot avatar. Such avatars have been previously shown to enhance social presence and operator salience. We present a motion tracking based tele-operation system for the NAO robot platform that allows direct transmission of speech and gestures produced by the operator. To assess the capabilities of this system for transmitting multi-modal communication, we have conducted a user study that investigated if robot-produced iconic gestures are comprehensible, and are integrated with speech. Robot performed gesture outcomes were compared directly to those for gestures produced by a human actor, using a within participant experimental design. We show that iconic gestures produced by a tele-operated robot are understood by participants when presented alone, almost as well as when produced by a human. More importantly, we show that gestures are integrated with speech when presented as part of a multi-modal communication equally well for human and robot performances.

Keywords: human-robot interaction, gestures, humanoid robotics, tele-operated robot, multi-modal communication

1. INTRODUCTION

Based on the idea that embodiment leads to stronger social engagement than a screen (Adalgeirsson and Breazeal, 2010; Hossen Mamode et al., 2013), we wondered whether a viable alternative for telecommunication is to use a tele-operated humanoid robot as an embodied avatar in a remote location. In previous work with robot avatars they have been shown to improve social presence of a remote operator (Tanaka et al., 2015), and their salience to people in the robot's presence (Hossen Mamode et al., 2013), relative to more traditional telecommunication media (audio and video).

In order for a robot avatar to be a viable communication method it must be capable of transmitting human interactive behavior. In everyday communication people can be observed performing arm gestures alongside their verbal communications (McNeill, 1992; Kendon, 2004). Though there is much debate on whether such gestures have a communicative value for observers, a recent meta-analysis of the literature concluded that they are of communicative value (Hostetter, 2011). Indeed, a number of studies in the human communication literature demonstrate observers of co-verbal gestures comprehend information from them (Cassell et al., 1999; Kelly et al., 1999; Beattie and Shovelton, 2005, 2011; Cocks et al., 2011; Wang and Chu, 2013). Hence, we are motivated to investigate the use of gesturing on a humanoid robot avatar to capitalize

on the reported benefits (salience and social presence), while still maintaining multi-modal communication efficacy.

To transmit the multi-modal communications of a human operator, we have developed a tele-operation interface that uses motion tracking of the operators arms, and audio streaming, to replicate their communication on a NAO robot (Aldebaran Robotics, Gouaillier et al., 2009). By using this implicit control method we aim to allow an operator to communicate as they would face-to-face. Before being able to investigate the benefits of embodiment over video in telecommunication, and interaction benefits of gestures, we first need to demonstrate the capability of the system to reproduce comprehensible gestures on the robot; thus, this is the first aim of the work presented here.

Which kind of gestures are particularly important in human-human communication, and how they can be shown to add communicative value, underpins our approach to evaluating multi-modal communication on a robot avatar. Within the literature on gestures in human interaction a number of schemes have been proposed to classify them according to their form and function (Ekman, 1976; McNeill, 1992; Kendon, 2004).

Iconic gestures are a key class of gestures from the classification scheme proposed by McNeill (1992). Iconic gestures are those that have a distinct meaning, they are of a form that either reiterates or supplements information in the speech they accompany. They typically convey information that is more efficiently and effectively conveyed in gesture than in speech, such as spatial relationships and motion of referents (Beattie and Shovelton, 2005), or the way in which an action is performed (termed manner gestures) (Kelly et al., 1999). Hence, multi-modal communication can be said to be more effective and efficient at conveying information between speaker and listener than uni-modal communication, i.e., taking less time to convey the desired message, and in a clearer way (Beattie and Shovelton, 2005). Given the high communicative value of iconic gestures, here we investigate their use in robot avatar communication.

For human-human communication, a number of approaches have been taken to establish the communicative value of iconic gestures, by examining whether the information understood by observers of multi-modal communication differs from uni-modal communication. One suggested value of gestures is that they improve how memorable the speech they accompany is. Hence, participants' ability to recall details of speech delivered with and without different gestures has been tested (e.g., Cassell et al., 1999; Kelly et al., 1999). Analysis of results for such experiments is non-trivial, and depends strongly on how easy the stimulus material content is to remember.

An alternative approach was suggested by Beattie and Shovelton (2005), whereby participants were asked questions about short multi-modal vignettes, the answers to some of which were only contained in the gestural channel. However, in such an approach it might be difficult to distinguish between speech and gesture integration, and contextual inferences (Beattie and Shovelton, 2011).

To avoid confounds such as the ones potentially inherent in the approaches described above, we decided to base our experiments on a seminal study presented by Cocks et al. (2011). We adapted their design for use with the NAO robot

and our tele-presence control scheme (see Section 2). In their study, participants were presented with a series of actions conveyed either through speech alone, gesture alone, or an iconic (manner) gesture accompanying speech, and asked to select, from a set of images of actions one that best matches what was communicated. The authors were able to clearly distinguish and compare understanding of actions both in uni-modal and multi-modal communication. Hence, their method was able to evaluate integration of information from the two communication channels, a process vital for the utility of co-speech iconic gestures (Cocks et al., 2011).

One of the aims of the work presented here is to investigate whether the integration of speech and gesture occurs for a non-human agent, such as a robot, in the same way that it does for a human. Knowledge in this regard is as yet very limited. Speech and gesture integration for robot-performed pointing (deictic) gestures has been investigated (Ono et al., 2003; Cabibihan et al., 2012b; Saupé and Mutlu, 2014), this showed that relative locations of referents could be better understood by using gestures to supplement speech information. While these studies provide some evidence for speech and deictic gesture integration, iconic gestures have yet to be examined. Moreover, to the best of our knowledge, it has never been investigated whether this integration process is as reliable in robots as it is in people.

A key issue in robot gesturing, is joint coordination and motion timing. Work on how the human brain processes gestures suggests this may be of importance to gesture recognition, and hence in studying speech and gesture integration. In their recent meta-analysis of studies concerning the neural processing of observed arm gestures Yang et al. identified three brain functions associated with gesture processing: mirror neurons, biological motion recognition, and response planning (Yang et al., 2015). Of particular relevance here are mirror neurons, part of the brain associated with performing actions that fire when those actions are recognized. Gazzola et al. showed that mirror neurons still fire when observing some robot motion (Gazzola et al., 2007). However, they suggested that this depends on identification of the goal of the motion. With gesture, the motion goal is often not clear, and so mirror neuron based gesture recognition may instead rely upon identification of motion primitives, component parts of gestural motion based upon muscle synergies in the arm (Bengoetxea et al., 2014).

A potential advantage in our study is we might overcome any scripting-related issues by using our tele-operation control scheme to copy both the shape, timing and joint coordination of human movement. Note, however, that even a tele-operation control system is limited by the design and the degrees of freedom of the robotics system used. Moreover, the non-biological appearance of the robot may interfere with identification of the gestures. Hence, we included testing conditions that allowed us to evaluate the comprehensibility of the gestures produced with our system when presented on their own.

In this paper we aim to address the following research questions: (1) can iconic gestures performed with our tele-operation system be identified?; (2) is performance comparable to when the same gestures are performed by a person?; (3) are iconic gestures performed using our tele-operation system integrated

with speech?; and (4) is integration as efficient for robot performed multi-modal communication as human performed multi-modal communication?

In detail, we pre-recorded a set of communications consisting of verb phrases and appropriate iconic gestures produced by the robot using our tele-operation system, and a matching set by a human actor. The same actor was used for producing the robot stimuli and the human stimuli (recorded on video) to make the conditions as closely matched as possible. The recorded stimuli were then used in an experimental study adapted from the human–human communication literature (Cocks et al., 2011) to investigate whether hand gestures on their own were comprehensible for both robot and human, and whether they could be integrated with speech.

To evaluate integration, we established whether the understanding of the observers' was changed as compared to speech or gesture alone. Understanding was also directly compared for the human (on video) and the robot (embodied replay of recorded communications) within the same observers. We sought to establish the extent of integration benefit achievable with robotic communication, relative to the one observed for a human communicator. We used videos of human gestures in our study to ensure identical stimuli for all participants. We reasoned they would be as efficient as live performances, given high recognition and integration rates (close to ceiling) were observed using video stimuli, in the study on which our work is based (Cocks et al., 2011).

An additional motivation for our comparison of human video communication with a physically present robot is that it allows us to evaluate the differences between these two modes of telecommunication for multi-modal communication. If the performance of gesture understanding and integration for the robot avatar is comparable to video communication, it will enable further work on the salience and utility of these gestures in an interactive context. Beyond the application of the results to the utility of the NAO robot as an avatar, the tele-operated approach allows us to make more general inferences for the design of autonomous communicative robots.

Directly comparing participants' comprehension of iconic gestures and their integration with speech for human and robot performers (in a single experiment) allows us to eliminate a range of confounds that make it difficult to compare findings within the literature. To the best of our knowledge we are the first to make this direct comparison.

This paper is an extended version of our work published in Bremner and Leonards (2015a). We extended our previous work by adding in depth analysis of the gestures used, and the performance of the tele-operation system in reproducing these gestures. Additionally there is far more detailed discussion of our results, including implications of related work in neuroscience on human gesture processing.

2. MATERIALS AND METHODS

We conducted an experimental study with 22 participants (10 female, 12 male), aged 18–55 ($M = 34.80 \pm 10.88SD$), all of

whom were Native English speakers. Participants gave written informed consent to participate in the study, in line with the revised Declarations of Helsinki (2013), and approved by the Ethics Committee of the Faculty of Science, University of Bristol.

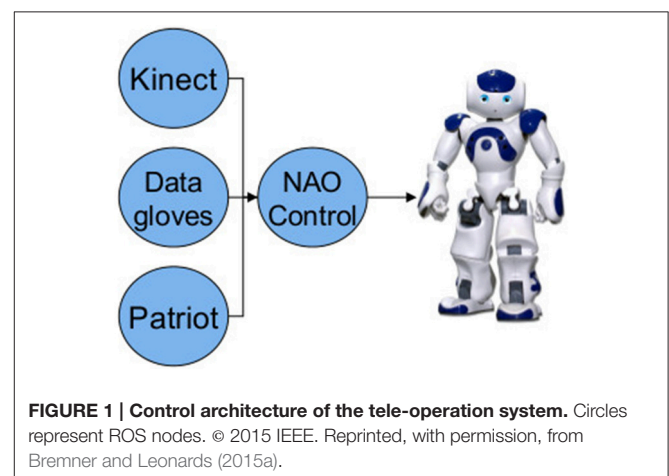
Stimuli consisted of a series of pre-recorded communications, these were either speech alone, gesture alone, or speech and gesture. Each communication was performed by either the human actor (on video) or the NAO robot (physically present). Video was used for the human stimuli to ensure repeatability, and to allow direct comparison of data obtained for speech and gesture integration in dependence of the type of communicator: human or tele-operated robot. Hence, the experiment used a 2 (performer) \times 3 (communication mode) within-subjects design.

2.1. Tele-Operation System

To reproduce gestures performed by a human actor on the NAO humanoid robot platform from Aldebaran Robotics (see **Figure 1**, for specifications see Gouaillier et al., 2009), we designed a motion capture based tele-operation system. The system was built using the ROS framework. Architecturally, ROS can be described as a computation graph made up of software modules (termed nodes), communicating with one another over edges (Quigley et al., 2009). Communication is built on a publisher/subscriber model where a node sends a message by publishing it, and nodes using that message subscribe to it.

ROS offers a number of advantages that make it well suited to our system. Firstly, its communication architecture means that the system is inherently modular, so if one node fails the others can keep running while the failed node is restarted. Secondly, this modularity means nodes can be easily modified independently, only needing to adhere to correct message structure, making the system easily extensible. Thirdly, nodes can be written in different programming languages, here some nodes use C++ and some Python. Finally, ROS is well documented with a large library of existing nodes on which to base our work, speeding development time. Hence its use over viable alternatives such as YARP (Metta et al., 2006) or URBI (Baillie et al., 2008).

In our tele-operation system we have developed separate nodes to gather kinematic information of the human tele-operator from several sensor systems. Each sensor node



then publishes its data as ROS messages, a NAO control node subscribes to these message streams and then calculates the required commands that are then sent to the robot. **Figure 1** shows the system architecture schematic. Audio streaming was handled separately from ROS using the GStreamer media framework to develop a NAO module and corresponding PC application to allow streaming of audio to the robot.

In order to ensure that gestures are reproduced on the robot as closely as possible to the original human motion, hand trajectories, joint coordination and arm link orientations must be maintained. To this end arm link end points (i.e., shoulder, elbow and wrist) are tracked using a Microsoft Kinect sensor; the Nite skeleton tracker API from OpenNI is used to process the Kinect data and produce the needed body points. A Kinect node was written with the Nite API that uses the arm link end points provided by the skeleton tracker to calculate unit vectors for the upper and lower arm in the operator's torso coordinate frame¹, these were then published as ROS messages. Sensor update rate was 30 Hz.

The arm unit vectors are then used by the NAO control node to calculate robot arm joint values that align the arm links of the robot with those same unit vectors in the torso coordinate frame of the robot¹. An example mapping between human and robot arm positions is shown in **Figure 2**. Data from the Kinect were subject to high levels of noise, consequently the joint angles were smoothed using a moving average filter with a 10 frame window.

The filtering process added undesirable delay to the robot commands. Consequently, each filtered value is then modified by adding a trend term, calculated for each joint as a 10 frame moving average of the change in position each frame, then scaled by a factor of 4 (empirically determined) to produce a command similar to, but slightly ahead of, the raw value. To prevent overshoot due to sudden changes in velocity the filtered output was limited to deviate from the un-filtered value by an empirically determined maximum threshold value (0.04 rad). The NAO control module executed these commands to ensure the joints are still in motion when new commands are received, to do this it sent motor demands to execute the motion over a longer period than the update rate would require, so the controller doesn't decelerate more than demanded by the control node. This process utilized the inbuilt NAO position controllers to counteract commands being ahead of the raw value (resulting from the trend term in the filter), and thus allowed smooth handling of the stream of position demands.

Due to limitations of the resolution of the Kinect when viewing the full body, it is not able to provide all degrees of freedom (DoF) required. Specifically, finger flexion and extension, and hand rotation relative to the forearm (pronation/supination). To overcome these limitations additional sensors were used: a Polhemus Patriot provides pronation/supination, and 5DT data gloves provide finger bend information. ROS nodes were developed for each of the additional sensors, which publish that data as ROS messages at 30 Hz. The NAO node processes this additional data to calculate

¹ calculations are omitted here for brevity as they are relatively trivial.

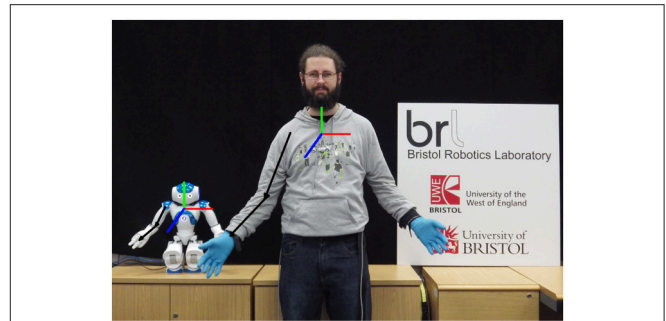


FIGURE 2 | A tele-operator pose reproduced on the NAO robot. Black arrows indicate the directions of the unit vectors along the arm links, the coordinate frame of the torso is shown in RGB (XYZ). © 2015 IEEE. Reprinted, with permission, from Bremner and Leonards (2015a).

the needed joint angles for the robot. It then combines the calculated angles for all arm joints into a single message to send to the robot each command cycle.

2.2. Phrase and Gesture Selection

In order to evaluate whether the tele-operation system could produce comprehensible gestures, and whether the produced gestures were integrated with speech they accompany, we first had to determine a suitable set of phrases and accompanying gestures. We selected 10 verb phrases, depicting common actions (e.g., I played, I opened), chosen from those used by Cocks et al. (2011), see **Table 1** for the full list. An important feature of the phrases selected is that they have more than one manner in which they can be conducted, and these manners can be conveyed with manual gestures.

For each phrase two different iconic (manner) gestures were determined that conveyed manner in which the action was performed. This is an extension of the original design as presented by Cocks et al. (2011), who used only a single gesture for each phrase. We made this modification for two main reasons, firstly to give us a larger range of gestures to evaluate for comprehensibility on the NAO robot; secondly, and more importantly, to better evaluate speech and gesture integration. Indeed, we would argue that showing two different shifts in meaning from a speech only interpretation provides stronger evidence for integration.

To select appropriate gestures there are a number of factors that must be considered. The primary aim for the gestures is that they are sufficiently vague that they might convey multiple possible meanings when viewed without words; at the same time, they must still be interpretable without the need for speech. This requirement also served to increase the ecological validity of the gestures being used, as they were close to those that might be performed in everyday speech. Note that this clearly contrasts with a precise pantomime gesture of a particular action, which is likely to have only one interpretation, and which is rarely used in normal conversation (Cocks et al., 2011).

Another important requirement was that the gestures had to be performable by the NAO robot, such that a fair comparison

TABLE 1 | The 10 verb phrases, their preconceived meanings when accompanied with each of the two manner gestures (integration target), description of the manner gestures.

Phrase	Integration target	Gesture description
I Cleaned	1. Dusting a lamp	One hand open flat, palm down, moves diagonally from center line, at shoulder height, down and outwards toward periphery and then back again twice
	2. Scrubbing a pan	One hand moves in a horizontal circle in center, hand is in a power grip, palm down
I Cut	1. Cutting with a craft knife	One hand moves from center line, horizontally outwards toward periphery, hand in a precision grip
	2. Chopping into a melon	flat vertical palm moves in a downward chopping motion, in periphery
I Fixed	1. Hammering a nail	One hand in a vertical closed power grip moves up and down twice in a curved path, in periphery
	2. Sticking paper with tape	Both hands in precision grip, palm down, hand length apart, move downwards as if pressing something down, in center center
I Lit	1. Pulling a light pull	One hand in a vertical closed power grip moves to shoulder height arm partially extended, then moves vertically downwards, in periphery
	2. Pressing a light switch	One hand, with index finger extended, moves diagonally up and out away from the torso to finish just below shoulder height, in periphery
I Measured	1. Pouring liquid into a measuring jug	One hand adopts an vertical open power grip, the other a vertical precision grip above and to the side of the other hand, the wrist is rotated in a pouring motion, both hands in center center
	2. Using a tape measure	Both hands adopt a precision grip, palm down, and move close together in center center, right hand then moves horizontally away from the stationary left hand, toward periphery
I Opened	1. Pulling open a door	One hand reaches out away from the body, adopts a vertical precision grip then retracts straight backwards, in periphery
	2. Opening a book	one flat hand, horizontal, palm down in center center, hand moves up and out toward periphery with wrist rotation to flip hand over
I Paid	1. Signing a check	one hand in a precision grip tracing a curling path from the center out to the periphery
	2. Handing over cash	One hand open, palm horizontal and face up, hand moves out and up as if presenting an object on the hand, in periphery
I Played	1. Playing chess	One hand adopts a horizontal grip, palm down, in center, near the body then follows an arcing trajectory forwards and releases the grip
	2. Playing a cello	One hand, in a horizontal fist, palm down, moves back and forth across the center-line of the body
I Read	1. Reading a newspaper	Both hands in vertical closed power grip shoulder width apart
	2. Reading a book	Both hands in vertical closed power grip a hand length apart, in centre
I Rubbed	1. Using a pencil eraser	One hand, horizontal closed power grip, palm down, moves left to right rapidly near centreline of body
	2. Rubbing a balloon	One hand partially open power grip moves vertically up and down twice, in periphery

could be made between gestures performed by a person and the robot. While the NAO robot does have degrees of freedom in its arm such that it can cover a wide range of human-like movements (Gouaillier et al., 2009), it does have a number of limitations relevant to the performance of gestures. The most important of these is that the NAO only has three fingered, one degree of freedom hands, where all fingers open and close simultaneously. Hence, NAO is not capable of much in the way of hand-shapes, a key component in many human upper limb gestures. To accommodate for this restriction we selected gestures which mainly comprised arm movements, for which precise hand shape and finger movements were deemed less critical. Note further, the NAO robot also has only one degree of freedom in the wrist (pronation/supination), compared to the 3 degrees of freedom in the wrist of humans, a reduced range of flexion in the elbow, and a safety algorithm to prevent the two

hands from colliding. While we have tried to select gestures that are relatively unaffected by these restrictions, in order to maintain ecological validity, the human performer/tele-operator was not instructed to accommodate any of these factors.

The final selection of gestures are described in **Table 1**. To simplify descriptions, and aid analysis of gesture features, the description of gesture space proposed by McNeill was used (McNeill, 1992). To further aid description we use the terms power grip: gripping with the whole hand, and precision grip: gripping with the finger tips.

2.3. Materials and Procedure

The experiment stimuli consisted of recordings of the 10 verb phrases detailed in **Table 1**. Each verb phrase was performed twice, once for each of the iconic (manner) gestures that portrayed how the action was performed. Two stimulus sets

were recorded, the human performer stimuli was recorded using a digital video camera, the robot stimuli was recorded using the tele-operation system. In order to avoid inter-individual variability in action performance, the same human actor performed both human and robot stimuli.

To avoid possibly distorting participant perceptions due to the presence of the data-gloves necessary for tele-operation, the two stimulus sets were recorded separately. In order to ensure that the stimulus sets were as similar as possible, prior to performing without the data-gloves the actor reviewed the video of each tele-operation performance. The two recordings of each stimulus item were compared, and, where necessary, repeat performances were recorded.

The robot communication stimuli were created by recording the messages transmitted by the sensor nodes using the built in recording capabilities of ROS. Audio was captured using the GStreamer based software module. To allow immediate verification, the robot was controlled and streamed to during recording.

The human video stimuli and the recorded tele-operation stimuli were then edited to produce a set of presentations lasting approximately 5 s each, in three conditions: verbal only condition (V; audio only no performer movement); gesture only condition (G; gesture visible but audio not played); verbal-gesture condition (VG; gesture seen and verbal phrase heard). In both G and VG conditions, there were two different manner gestures so two presentations were created for each verb phrase. Hence, each action phrase came in five different versions per performer (V, G1, G2, VG1, VG2).

To create the human stimuli the audio recorded during the robot performances was added to the videos of the human performance (i.e., replacing the original audio). Hence, identical audio was used for both robot and human performances in the 3 condition with a verbal component. Audio-information was overridden for the human stimuli to make sure that the audio information provided was identical between both human and robot stimuli. To prevent any lip-syncing issues, and eliminate the possibility of facial gesture effects, the human performer's face was obscured in the video. The relative timing of speech and gesture for the robot performances was based on video recorded of the robot captured during stimulus recording with the tele-operation system.

There were 10 experimental conditions in total: five communication modes (V, G1, G2, VG1, VG2) for each of the two performers. Ten action phrases were used in each experimental condition; hence, each participant responded to 100 different trials. The trials were split into 10 blocks, each containing all 10 phrases, and all 10 experimental conditions. To prevent ordering effects, trial presentation order was counterbalanced across and within blocks by means of pseudo-randomization using partial Latin squares.

Following each stimulus presentation, participants were presented with a set of six color photos of people performing actions on the (12.1 inch) screen of a response laptop, and were asked to select one. To do so they clicked with the laptop's mouse cursor on the photo they thought most closely matched what had been communicated; doing so moves on to the next

stimulus presentation. The layout of the images, and hence the location of the target(s) on the response screen, were randomized between conditions and between phrases. Presentation of the response images, and recording of responses was done using the PsychoPy software (Peirce, 2007). Average experiment time was 20 min.

The response image set for each phrase consisted of: a gesture only target for each gesture, that matched the corresponding gesture but not the speech; an integration target for each of the two manner gestures, which matched the corresponding speech and gesture combination; a pair of unrelated foils, not matching either the gesture or the speech, each one linked semantically to one of the gesture-only images (**Figure 3** shows an example set, for "I paid"). For a particular gesture, one gesture only image and one integration target were both semantically congruent with it, so should have been selected with equal likelihood in the G condition. Both of the integration targets were semantically congruent with the speech, so in the V condition each should have been selected with equal likelihood. In each of the VG conditions only a single integration target was congruent for that particular speech and gesture combination, hence it should be the most probable image selection.

Figure 4 shows the experimental set-up. The video screen and the NAO robot were both positioned 57 cm from the participant. A 32 inch wide-screen TV was used to display the video stimuli, thus, the human performer and robot appeared to be of a similar size. The start of each trial was signaled to the participant by playing a tone and displaying either human or robot on the response laptop for 1 s to indicate which presenter was next. This allowed the participant to concentrate on the correct presenter from the outset of each trial. Each trial consisted of playback of the performance of the phrase, followed by automatic display of the response image set. Each trial was initiated by the experimenter after the participant had completed the previous trial; the experimenter was sat out of view of the participant. Prior to the experimental trials, participants



FIGURE 3 | The response images for "I paid": (A,B) match only the gestures; (C,E) are the integration targets, both of which match the speech only condition; (D,F) are the unrelated foils. © 2015 IEEE.

Reprinted, with permission, from Bremner and Leonards (2015a).



FIGURE 4 | Set-up for the experiment. © 2015 IEEE. Reprinted, with permission, from Bremner and Leonards (2015a).

performed two practice trials to ensure they understood the experimental procedure.

3. RESULTS

3.1. Gesture Comprehension

Gesture comprehension was tested by calculating the proportion of correct responses in the conditions with only gestures. To evaluate each gesture, in both performance conditions, a chi-squared test was used to compare the proportion of correct responses for that gesture with chance (of the six images in the response set two were the correct answer, so chance was at 0.33). These results are shown in **Figure 5**. Almost every gesture (with the exception of both the “I lit” gestures in the robot condition) was identified significantly better than chance in both human and robot conditions, with high average proportions of correct responses ($M_{human} = 0.943 \pm 0.065SD$; $M_{robot} = 0.802 \pm 0.17SD$). A Wilcoxon signed rank test (used as the data did not meet assumptions needed for a parametric test) revealed a significant difference between performers ($p < 0.001$) for the same gestures even excluding the “I lit” gestures.

It is apparent from **Figure 5** that sizeable differences in gesture comprehension between performers existed only for some of the gestures examined. Hence, the data were further analyzed, on a per gesture basis, to find for which individual gestures there were significant differences in recognition rate between performers. As the data is binomial and paired (each participant viewed human and robot performances of each gesture), we used an exact McNemar test to evaluate differences. An exact McNemar test for each gesture revealed gestures were identified correctly significantly more frequently in the human performances than in the robot performances for lit1 ($p = 0.00098$), lit2 ($p = 0.00049$), and fixed1 ($p = 0.00781$). Cut2 approached being significantly more frequently correctly identified in human performances than in robot performances ($p = 0.0625$). There were no other significant differences in gesture identification between human and robot performance conditions. Note, however, that these results² need to be treated with caution as performance was almost at ceiling, resulting in small values for the dichotomous variables used in the test calculations.

²For access to results data pertaining to this work please contact the lead author.

In order to investigate possible sources for the difference in gesture comprehension found between human performer and robot, controller performance was further analyzed for two of the gestures; namely those for which significant differences had been reported—lit1 and fixed1. First we compared the physical movement profiles: for this, the recorded robot joint values over the duration of each gesture were plotted along with the joint values for the human performer as recorded by the Kinect (**Figure 6**, Lit1, **Figure 7**, Fixed1). It is clear from the graphs that joint co-ordination and velocity profiles, and hence hand trajectories, are very comparable between human and robot for the two gestures analyzed. However, two common differences can be observed in both plots, firstly the elbow flexion has a limited range of motion on the robot relative to the human, decreasing the amplitude of the peak of the gesture (approximately 15% reduction in vertical travel); further, they have a very brief pause at the top of the stroke.

Secondly, the predictive filter caused the robot joints to accelerate at a slightly different rate to the human joints when the human joint velocity was at certain values; this resulted in those joints finishing their motion approximately 0.1s early. It is hard to quantify the significance of these differences. Although they appear relatively small, critical visual examination of the robot motion on these two gestures may provide further insight. In both cases the hand trajectory is largely as expected and joint coordination appears on visual inspection human-like. However, the slightly shorter vertical travel is noticeably different from what is expected for these two actions, but vertical travel is still clearly perceptible. Further, in the human version of these gestures ulnar/radial deviation in the wrist is used, a degree of freedom lacking in the NAO robot. A pause in the gesture is barely perceptible, and only in the oscillatory motion in fixed1, appearing less smooth than expected.

To provide further insight into differences in gesture performances, the gestures lit2, cut2, played1, and cleaned1 were also analyzed by visual inspection. Though not significantly different in identification between performers, cut2, played1 and cleaned1 all led to differences in identification performance between human and robot performer (5). Similarly to lit1 and fixed1, cut2 and played1 showed reduced vertical travel for the robot performance due to a reliance on elbow flexion. It is also apparent from lit1, lit2, cut2, and cleaned1 that the wrist rotation sensor did not always give accurate readings. As a result, wrist orientation differed visibly from the human version of these gestures. Although we would have thought that hand-shape itself should play only a minor role in these gestures, in lit2, and cut2, a fairly particular hand-shape was adopted by the human which the NAO was unable to approximate well enough.

3.2. Speech and Gesture Integration

To test for speech and gesture integration all stimulus item scores were summed for every participant (the scores for a particular phrase were the combined results for the two gestures that accompanied each), hence we determined the proportion of integration target choices (ITC). **Figure 8** shows the proportion of participant responses where the integration target was selected, in dependence of the presented stimulus mode.

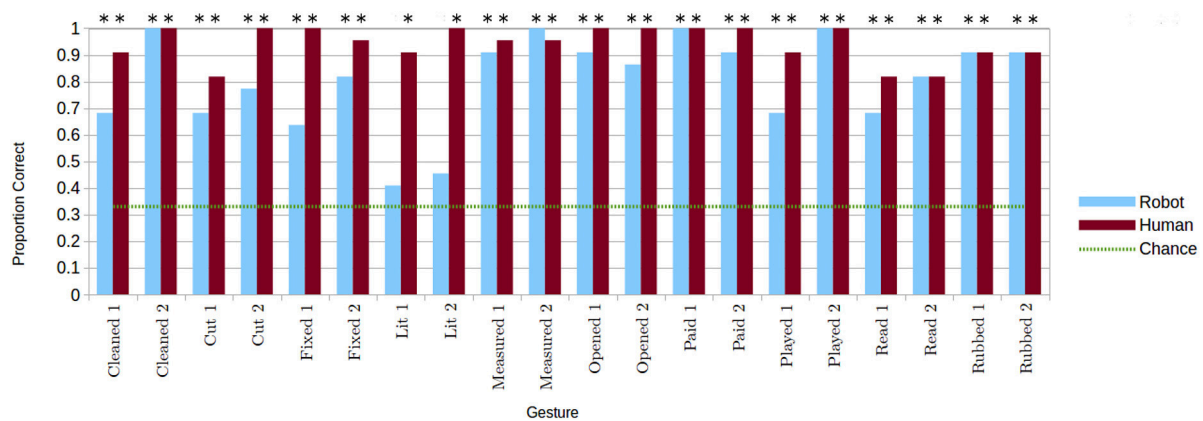


FIGURE 5 | Proportion of correct identifications of each gesture for the two performance conditions, when gestures are presented alone. Correct gesture identifications significantly greater than chance indicated with $*p < 0.05$. © 2015 IEEE. Reprinted, with permission, from Bremner and Leonards (2015a).

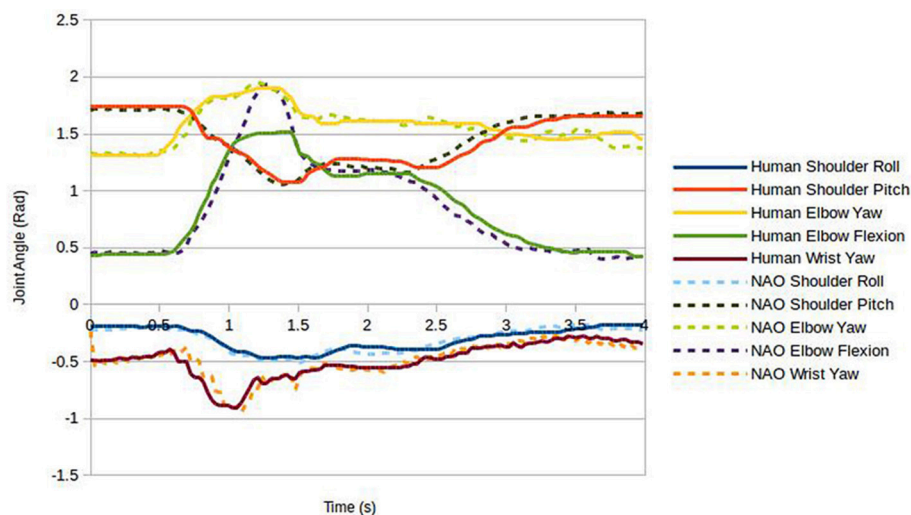


FIGURE 6 | Joint values during the Lit1 gesture for human and robot performers.

Uni-modal presentations had a uni-modal image as a correct answer as well as the integration image. In line with expectations that each was equally likely to be chosen, ITC for the verbal condition were made close to 50% of the time; the gesture conditions favored the non-integration target image, with ITC close to 40%. In the multi-modal presentation condition we observed a distinct increase in the frequency with which the integrated image was selected. Underlying the averaged values for uni-modal image selection, a number of individual stimuli had a particular image of the two viable image choices that was chosen significantly more often than the other. In some cases this was the integration target and in some cases it was not; integration target choice in the multi-modal version of those stimuli did not vary significantly from the value found in less extreme uni-modal cases. Hence, this provides stronger evidence for multi-modal integration in cases where a large change occurred. Moreover, this

shows the robustness of our approach to these variations as the averaged values are close to those expected.

Accordingly, a 2 (presenter) \times 3 (communication modus) repeated measures ANOVA revealed a significant main effect of communication mode [$F_{(2,42)} = 282.57, p < 0.0001$]. *Post-hoc* analysis (Tukey) confirmed that participants chose the integrated images far less often in the gesture only condition ($M = 0.39 \pm 0.11SD$) than in the verbal only condition ($M = 0.49 \pm 0.02SD, p < 0.0005$). More importantly, participants selected the image constituting the integrated information from speech and gesture in the VG condition ($M = 0.82 \pm 0.08SD; p < 0.0005$). Hence, there is clear indication that ambiguity is decreased by means of correct integration of speech and gesture information.

We found no significant main effect for presenter [$F_{(1,21)} = 2.61, p = 0.12$], nor a significant interaction between

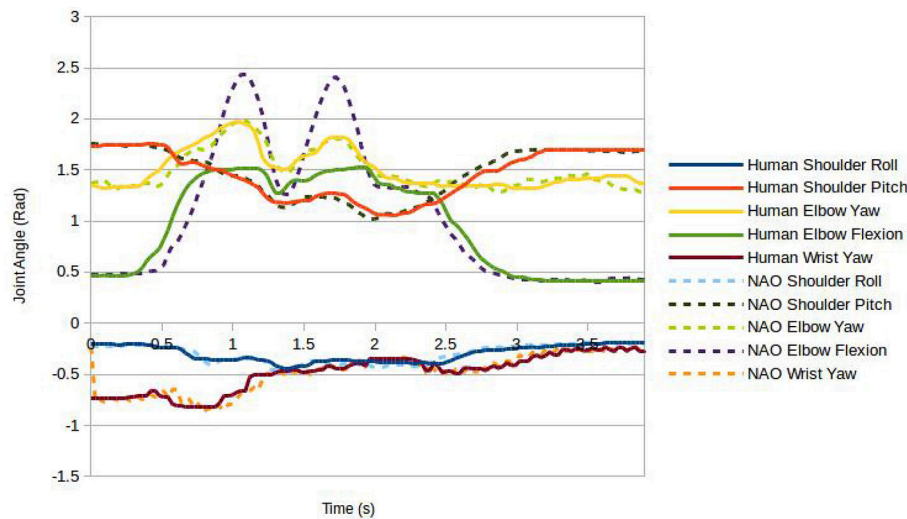


FIGURE 7 | Joint values during the Fixed1 gesture for human and robot performers.

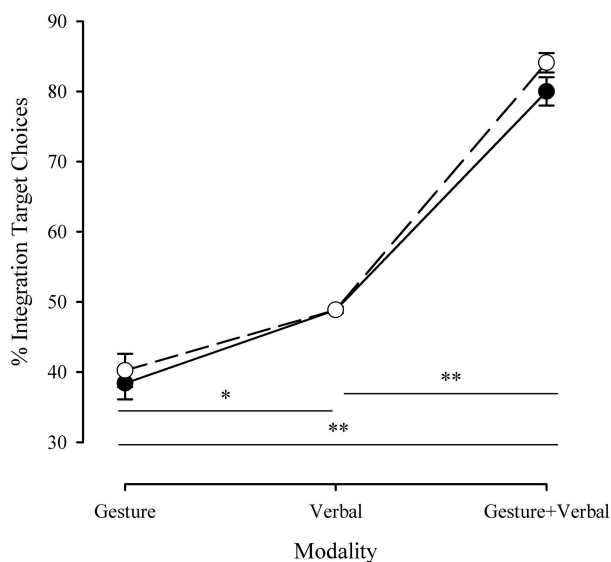


FIGURE 8 | Proportion of integration target image selection for each communication modality, in dependence of the communication performance medium. Shaded symbols: robot communication, empty symbols: human communication. Error bars represent ± 1 SEM. * $p < 0.0005$; ** $p < 0.0001$. © 2015 IEEE. Reprinted, with permission, from Bremner and Leonards (2015a).

communication mode and presenter [$F_{(2,42)} = 1.23, p = 0.30$]. This first analysis seems to indicate that integration of information conveyed in speech and gesture is of similar efficiency for a human communication mediated by video or mediated by a robot avatar.

So that we can gain a clearer picture of the pairwise comparisons of integration target image choices, we propose calculation of an estimate of the effect size of changes in ITC

proportions in dependence of condition. The method we have utilized to do so is based on the method proposed by Cocks et al. (2011) termed multi-modal gain (*MMG*). *MMG* is a means by which we can estimate the change in probability of ITC between uni-modal (speech or gesture alone) and multi-modal conditions (speech and gesture together). To estimate the value of *MMG*, the proportion of ITC in uni-modal communication ($P(Uni)$) is estimated, and then subtracted from the proportion of ITC in the VG conditions ($P(Multi)$), see Equation (1).

$$MMG = P(Multi) - P(Uni) \quad (1)$$

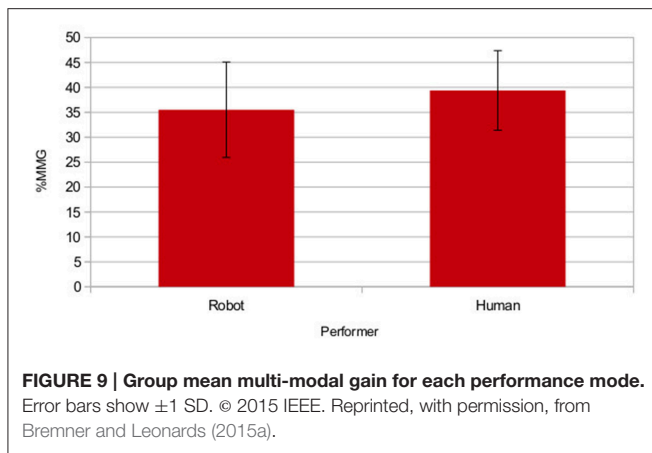
To estimate the proportion of ITC in the uni-modal conditions ($P(Uni)$) the weighted mean of ITC in the verbal (ITC_V) and gesture (ITC_G) conditions are summed, see Equation (2). The basis for this calculation is that the different modalities vary in how likely they are to be utilized by observers, i.e., it is assumed that participants are more likely to be influenced by the modality that they perceive as providing the most useful information. Thus, the two weights, WV and WG , for the verbal and gesture conditions respectively, are calculated as normalized proportions of trials in which integration targets were selected (PCV for V trials and PCG for G trials), see Equations (3) and (4).

$$P(Uni) = WV * ITC_V + WG * ITC_G \quad (2)$$

$$WV = PCV / (PCV + PCG) \quad (3)$$

$$WG = PCG / (PCV + PCG) \quad (4)$$

Hence, *MMG* calculates a single figure for percentage gain, taking into account how often the integration targets were chosen in both uni-modal conditions (the results for both gestures for each phrase were included together). The values for each performer were calculated separately and are shown in Figure 9. By using two gestures per phrase we found that for some phrases in the verbal condition one of the two matching images was selected far



more frequently than the other. Hence, *MMG* for the preferred integration target image was close to zero, i.e., gesture had no effect; conversely, for the other integration target image *MMG* was very high, i.e., gestures had a large effect. This gives us a clear advantage over the original study of Cocks et al. (2011) as we were less vulnerable to the variability of individual meaning preferences, and hence could gain a clearer picture of whether integration effected understanding by incorporating the scores in a single calculation.

We conducted a two tailed *t*-test for each performer against the null hypothesis of $MMG = 0$, the means of both samples ($M_H = 0.393 \pm 0.079SD$; $M_R = 0.355 \pm 0.095SD$) differed significantly from 0 [$t_H(21) = 23.12, p < 0.001, r = 0.98$; $t_R(21) = 17.405, p < 0.001, r = 0.97$]. It is important to be aware that a maximum estimate for *MMG* is given by $1 - P(Uni)$, hence, $MMG_{Rmax} = 0.56$ and $MMG_{Hmax} = 0.56$ (i.e., 56 and 55% for the robot and human respectively). The *MMG* values for both performance modes are approaching ceiling.

The means of the two performers were compared using a paired two tailed *t*-test, and this showed no significant differences [$t_{(21)} = -2.005, Dif = 0.019, p > 0.05, r = 0.21$]. However, for testing the hypothesis that there is no difference between performance modes this analysis was underpowered. In order to allow us to more reliably test this hypothesis, i.e., that the performance mode results are interchangeable, a repeatability measure was used, intraclass correlation coefficient (ICC). The *MMG* scores for each participant were calculated from responses in multiple trials (so can be considered akin to a mean score), hence we used $ICC(2, k)$, as suggested in Shrout and Fleiss (1979). We found significant correlation between the results, indicating fair to substantial reliability [$ICC(2, k) = 0.61, F_{(21,21)} = 2.8, p = 0.011$]. Taking these two analyses together, we thus feel confident that participants' ability to integrate gestures and speech was independent of the performers.

4. DISCUSSION

The findings in this paper address the four research questions proposed in Section 1. We found that (1) human observers

were able to identify upper limb manner gestures the majority of the time when produced by a tele-operated NAO robot. (2) Although identification of robot-performed gestures was worse than that for human-performed gestures, it was still good enough for them to be useful. More importantly, as gesture in human communication is most commonly employed along with speech, we found that (3) when such gestures were performed with speech they were integrated with it; (4) this process was as efficient for the robot as the human performances. Moreover, this integration compensated for any difficulties in identification of robot performed gestures. In the following sections we will discuss these findings in more detail.

4.1. Gesture Comprehension

With the exception of those accompanying "I lit," all gestures used in this experiment were identified clearly above chance for both the human and the robot when they were presented without speech. Though robot gestures were more difficult to identify than human gestures, the general ability to do so is in clear contrast to earlier findings by Cabibihan et al. (2012a) and Zheng and Meng (2012). In both these previous studies they found robot performed gestures were difficult to identify on their own. There are a number of possible causal factors for the differences between our study and previous work. Possible factors are the subtleties in gestures captured by the tele-operation scheme, the different methods of response-gathering (restricted choices as used here, in contrast to free response in related work), the types of gestures used (they used more emblematic gestures, often close to pantomime, in contrast to the iconic manner gestures used here), or some combination of all of these. Whichever the explanatory case, the work presented here provides evidence for the idea that there is communicative value in robot performed gestures.

We suggest that there might be a wider range of gestures than those tested here that will have communicative value for a robot. Therefore, we will look at common features of the gestures used here that were correctly identified. It is also instructive to examine these same features for gestures that were more difficult to identify when performed on the robot than when performed by a human. Differences in the performances likely account for the lower mean recognition rate for robot performed gestures (80.2%, compared to 94.3% for human performances).

The primary common feature is the importance of hand trajectory, including the appropriate hand velocity profile. This is used to convey easily identifiable relative motions that are either part of the action being carried out, or of objects manipulated by the action. This idea is supported by the work of Beattie and Shovelton (2005), who found that gestures portraying relative positions and movements are the most successful at conveying information. Relatedly, when the trajectories could not be correctly perceived gestures were harder to identify. The main reason for this here was due to the reduced range of motion on the NAO elbow flexion, and the lack of the ulnar/radial deviation degree of freedom, resulting in smaller vertical travel for some gestures, and in some cases increased jerk. Moreover, these deviations might also cause difficulties in identifying motion primitives used in gesture recognition (Bengoetxea et al., 2014),

or limit the perception of the movement to being artificial where different mental processes are applied (Yang et al., 2015).

One way in which this issue of gesture recognition has been circumvented, is by having participants evaluate gestures not on their meaning alone, but rather on what action they would do in response, as this activates another area of the brain used in gesture recognition (Yang et al., 2015). This was demonstrated in the findings of Riek et al. where in speeded response trials participants were reported to correctly identify responses to robot performed co-operative gestures; they remained able to do so even when the robot used non-human-like velocity profiles (Riek et al., 2010). This suggests that the context in which the gestures are used may be of importance in the ease with which they are recognized.

A second common feature is hand orientation, as different hand orientations for the same hand trajectory can convey very different actions. Indeed, we found that for gestures where the wrist rotation sensor provided erroneous information, those gestures were less frequently correctly identified. As with deviations in arm trajectory this might mean that movement expected according to muscle synergies observed in human gesture (Bengoetxea et al., 2014) is not observed. A final feature, important for robots that do not possess fully articulated hands such as NAO, is a minimal reliance on hand shapes; i.e., gestures where arm trajectories and the degree to which the hand was open or closed contained sufficient information. We found that for some gestures hand shape was required for the gesture not to be too ambiguous to be correctly identified.

A good illustration of the importance of these features can be found in the gesture lit1, which, while being correctly identified in the human presentation condition, was not identified correctly in the robot presentation condition. The lit1 gesture comprises a vertical hand motion demonstrating pulling a cord to switch on a light (a common action in the UK). In the robot condition the unrelated foil images were selected with close to identical frequency as the target images. Examining the response image set for "I lit," we observed that the main differences between target and foil images was hand orientation, and motion range. Hence, we suggest, if gesture is to be used in uni-modal communication for a robot, as an avatar or autonomously, which gestures are used needs to be carefully examined, and the capabilities of the robot platform taken into account.

While the evidence for the relevance of the aforementioned deviations is limited, it does highlight an important factor both for gestures in HRI and in human communication that merit further investigation. We suggest this key factor is that the differences between human and robot gestures are relatively small, as shown in the performance analysis of the tele-operation control scheme in producing closely matched joint motion. Hence, our data provide further evidence for the notion that people are well conditioned to making subtle gestural discriminations and to identify biological motion and meaning (Kilner et al., 2003; Yang et al., 2015). This is further reinforced by our observations during the development of the range of gestures to be tested.

To test how susceptible observers are to subtle variations in robot performed gesture and how much this depends on

the context (e.g., whether observers are needed to physically or socially interact with the robot) requires more compelling evidence (see also Riek et al., 2010). Further, whether such effects vary between deliberate gesture identification, and the use of such gestures in conversation, also needs to be investigated. Indeed, by testing subtle gesture effects for robot communication we may be able to also learn more about the mechanisms underlying human communication and gesture perception.

4.2. Speech and Gesture Integration

Our findings demonstrate that when performed together speech and gesture are integrated, even when performance is mediated by a tele-operated NAO robot. We observed a larger proportion of integration target choices (ITC) in the multi-modal condition, as compared to either uni-modal condition. Multi-modal communication disambiguates the possible meaning of either gesture or speech on their own. ITC differed between uni-modal conditions, making it difficult to directly evaluate and compare the extent of speech and gesture integration for the two performers. To overcome this difficulty we followed the methodology of Cocks et al. to calculate multi-modal gain (*MMG*) (Cocks et al., 2011). *MMG* incorporates the results from both uni-modal presentation conditions in a calculation to estimate the change in probability of ITC for multi-modal communication as a single value. Highly significant values for *MMG* were found for both performance conditions. More importantly, the extent to which speech and gesture could be integrated was comparable between the two performers, indicating that robot-performed gestures are as efficiently integrated with speech as human-performed multi-modal communication.

As the lit gestures were not identified correctly when presented alone by the robot, it is instructive to examine the image choices when presented alongside speech. For lit1 and lit2 gestures, the correct target image was selected by 82% of participants and 95% of participants, respectively. This shows that participants were able to compensate for the lack of clarity in the gesture performance by using speech information to resolve ambiguity.

These results are somewhat surprising given previous work on speech and gesture integration with mismatched appearance and voice (here there is a clear mismatch of human voice and robot appearance). Kelly et al. showed that when there was a gender mismatch between voice and gesture performer, integration was reduced, and required considered rather than automatic mental processing (Kelly et al., 2010). Hayes et al. replicated these findings with human voice and robot performed gestures (Hayes et al., 2013). Similarly, we found that in speeded trials integration of speech and beat gestures does not occur when using a robot avatar to communicate (Bremner and Leonards, 2015b). The work presented here differs from the aforementioned, in that trials were not speeded.

We suggest that though integration of robot gesture and human speech may not be an automatic process, it occurs nevertheless. Whether there is a difference in mental processing for the gestures examined here, and if there is, whether it effects interaction with robot tele-operators requires further investigation. One way in which this could be tested is to look

not only at information comprehension, but also response times in speeded trials.

As well as being important for tele-communication using humanoid robot avatars, our findings also have implications for design of communicative behavior in autonomous humanoid robots. Perhaps the most important implication is that when a humanoid robot needs to communicate this can be done more accurately and efficiently by splitting semantic information across verbal and gestural communication modalities. In addition, our results demonstrate that multi-modal communications are interpreted similarly whether the gestural component is mediated by video only or by a tele-operated robot. Hence, autonomous robots should, where possible, use gestures to produce more natural seeming human-robot interaction. Thus, our work reinforces findings in the literature that higher subjective ratings are given to robots when they perform gestures (Han et al., 2012; Aly and Tapus, 2013; Salem et al., 2013).

Importantly, the difference in gesture recognition between human video and robot-embodied communication for gesture only communication is compensated for in multi-modal communication. That is to say, a humanoid robot avatar offers comparable performance to video communication when using speech along with gestures. Hence, a robot avatar operator might take advantage of previously observed advantages of robots over 2D communication media, such as enhanced engagement, improved social presence and action awareness (Powers et al., 2007; Adalgeirsson and Breazeal, 2010; Hossen Mamode et al., 2013), while maintaining communicative efficacy.

4.3. Conclusion

We show in this paper, using a fully within subject design, that using our Kinect based tele-operation system iconic manner gestures conveyed on the NAO robot are recognizable. This is despite physical restrictions in the degrees of freedom and movement kinematics of NAO relative to a human. Further, there seem to exist a large range of gestures which might be conveyed successfully. More importantly, we show that such robot-executed gestures can be integrated with simultaneously presented speech as efficiently as human-executed gestures. Whether this is because of, or despite the speech clearly originating from a human operator, remains to be further investigated. Hence, with regard to multi-modal semantic information conveyance, a NAO tele-operated avatar can be close to video mediated human communication in terms of efficacy. These two findings provide strong evidence as to the utility of a tele-operated NAO for conveying multi-modal communication. Although gestures are not recognized quite as well for the robot as they are for the human on video, they are still recognized well enough to make it a viable communication medium. We suggest the slight compromise in uni-modal gesture recognition for a robot performer is compensated for by the potential improvements in social presence and salience to interlocutors.

Our findings also have implications for autonomous communication robots, for which gesturing is an active area of research, and has been shown to offer a number of communicative benefits beyond information conveyance. Huang

and Mutlu found that robot performed deictic gestures improved participants' recall of items in a factual talk; however, gestures other types had minimal effects (Huang and Mutlu, 2014). Bremner et al. showed that although higher certainty in the information recalled was observed for parts of a monolog that were accompanied by (beat and metaphoric) gestures, the amount of information recalled was no better than for parts without gesture (Bremner et al., 2011). However, Van Dijk et al. found there was a positive influence on memory when redundant iconic gestures were performed when describing action performance (Dijk et al., 2013).

Other gesture effects beyond memory have been observed by Chidambaram et al. (2012), who demonstrated a robot was significantly more persuasive when it used gestures and other non-verbal cues. Additionally, hand gestures have been found to improve user ratings of robots on scales such as competence, likeability, and intention for future contact in a number of studies (e.g., Han et al., 2012; Aly and Tapus, 2013; Salem et al., 2013). These findings suggest that performing gestures on a robot avatar may have additional benefits to the robot operator that can be capitalized on, and we are in a position to do so now that we have shown they can be interpreted correctly.

We suggest that, when it is possible, robot communication should be multi-modal to ensure clarity of meaning, and to improve its efficiency and efficacy. This demonstration of the utility of multi-modal communication is not only of importance for our continuing work with tele-operated humanoid robot avatars, but also for socially communicative autonomous humanoid robots. We suggest our results might be generalizable in this way as previous studies showed that participants treat avatars similarly to how they do autonomous systems (von der Pütten et al., 2010). Indeed, one of the applications of humanoid tele-operation is as a tool to test what is important in terms of robot behavior for successful HRI in so-called super Wizard of Oz studies (Gibert et al., 2013).

4.4. Limitations and Future Work

While the work presented here provides initial insight into speech and iconic gesture integration for robotic communicators, it has a number of limitations which we hope to address in future work. Firstly, the range of tested gestures was limited to manner gestures where hand shape was not expected to be critical. In the future we intend to expand on our findings that integration can occur even for gestures that, as a consequence of differences in physical capabilities, can not be realized in a precisely human-like way by a robot. Limited evidence was found for this with the "I lit" gestures which were poorly recognized when performed by the robot.

The degree of similarity between robot performed and the original human gestures was not objectively controlled, other than visual inspection. Given our preliminary findings on the effects of subtle gesture differences, and existing literature on human sensitivity to biological motion, we suggest the examination of the degree of similarity required for comprehension and integration. Doing so would inform robot design and control requirements (extending the ideas in Riek et al., 2010). Additionally, we suggest that by both carefully

controlling gesture motion requirements, and similarity to human motion, one could more easily generalize our results across different robot platforms.

Another limitation of our work was that all gestures used were tested in a laboratory setting, with a limited set of short communications. In future work we aim to improve the ecological validity of our findings by investigating gestures in more interactive settings (extending the ideas in Hossen Mamode et al., 2013). In doing so we aim to look at a larger range of types of gesture, situated within longer sentences, and accompanied by other non-verbal behaviors such as gaze. An important component of this further work will be timing of gestures relative to speech (McNeill, 1992; Kendon, 2004). Though initial testing has shown coordination between speech and gesture to be close to that of the robot operator, whether it is close enough needs to be experimentally verified to fully validate our robot avatar system as a communication medium.

It is also important to note that our results might not be generalizable across cultures. Different nationalities have different gesturing conventions, and semantics (i.e., words that are ambiguous in English are often not in other languages). Further work is required to see if integration varies across different cultures, particularly where gestures are

more (e.g., Italy), or less (e.g., Japan) prevalent in everyday communication.

AUTHOR CONTRIBUTIONS

PB, conception and design of the work; acquisition, analysis, and interpretation of data for the work; drafting of the manuscript. UL, conception and design of the work; analysis, and interpretation of data for the work; revising work critically for important intellectual content. PB and UL, final approval and accountability.

FUNDING

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

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00183>

REFERENCES

- Adalgirsson, S. O., and Breazeal, C. (2010). "MeBot: a robotic platform for socially embodied telepresence," in *Proceedings of International Conference Human Robot Interaction* (Osaka: ACM/IEEE), 15–22. doi: 10.1109/hri.2010.5453272
- Aly, A., and Tapus, A. (2013). "A model for synthesizing a combined verbal and nonverbal behavior based on personality traits in human-robot interaction," in *Proceedings of International Conference Human Robot Interaction* (Tokyo: ACM/IEEE), 325–332. doi: 10.1109/hri.2013.6483606
- Baillie, J.-C., Demaille, A., Hocquet, Q., Nottale, M., and Tardieu, S. (2008). "The urbi universal platform for robotics," in *First International Workshop on Standards and Common Platform for Robotics* (Venice).
- Beattie, G., and Shovelton, H. (2005). Why the spontaneous images created by the hands during talk can help make TV advertisements more effective. *Br. J. Psychol.* 96, 21–37. doi: 10.1348/000712605X103500
- Beattie, G., and Shovelton, H. (2011). An exploration of the other side of semantic communication: how the spontaneous movements of the human hand add crucial meaning to narrative. *Semiotica* 184, 33–51. doi: 10.1515/semi.2011.021
- Bengoetxea, A., Leurs, F., Hoellinger, T., Cebolla, A. M., Dan, B., Cheron, G., et al. (2014). Physiological modules for generating discrete and rhythmic movements: component analysis of EMG signals. *Front. Comput. Neurosci.* 8:169. doi: 10.3389/fncom.2014.00100
- Bremner, P., and Leonards, U. (2015a). "Efficiency of speech and iconic gesture integration for robotic and human communicators—a direct comparison," in *Proceedings of IEEE International Conference on Robotics and Automation* (Seattle, WA: IEEE), 1999–2006. doi: 10.1109/icra.2015.7139460
- Bremner, P., and Leonards, U. (2015b). "Speech and gesture emphasis effects for robotic and human communicators," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (Portland, OR: ACM Press), 255–262. doi: 10.1145/2696454.2696496
- Bremner, P., Pipe, A. G., Melhuish, C., Fraser, M., and Subramanian, S. (2011). "The effects of robot-performed co-verbal gesture on listener behaviour," in *11th IEEE-RAS International Conference on Humanoid Robots*, (IEEE), 458–465. doi: 10.1109/humanoids.2011.6100810
- Cabibihan, J.-J., So, W.-C., and Pramanik, S. (2012a). Human-recognizable robotic gestures. *IEEE Trans. Autom. Mental Dev.* 4, 305–314. doi: 10.1109/TAMD.2012.2208962
- Cabibihan, J.-J., So, W.-C., Saj, S., and Zhang, Z. (2012b). Telerobotic pointing gestures shape human spatial cognition. *Int. J. Soc. Robot.* 4, 263–272. doi: 10.1007/s12369-012-0148-9
- Cassell, J., McNeill, D., and McCullough, K.-E. (1999). Speech-gesture mismatches: evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmat. Cogn.* 7, 1–34. doi: 10.1075/pc.7.1.03cas
- Chidambaram, V., Chiang, Y.-H., and Mutlu, B. (2012). "Designing persuasive robots: how robots might persuade people using vocal and nonverbal cues," in *Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on*, (IEEE), 293–300. doi: 10.1145/2157689.2157798
- Cocks, N., Morgan, G., and Kita, S. (2011). Iconic gesture and speech integration in younger and older adults. *Gesture* 11, 24–39. doi: 10.1075/gest.11.1.02coc
- Dijk, E. T., Torta, E., and Cuijpers, R. H. (2013). Effects of eye contact and iconic gestures on message retention in human-robot interaction. *Int. J. Soc. Robot.* 5, 491–501. doi: 10.1007/s12369-013-0214-y
- Ekman, P. (1976). Movements with precise meanings. *J. Commun.* 26, 14–26. doi: 10.1111/j.1460-2466.1976.tb01898.x
- Gazzola, V., Rizzolatti, G., Wicker, B., and Keysers, C. (2007). The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. *NeuroImage* 35, 1674–1684. doi: 10.1016/j.neuroimage.2007.02.003
- Gibert, G., Petit, M., Lance, F., Pointeau, G., and Dominey, P. F. (2013). "What makes humans so different? Analysis of human-humanoid robot interaction with a super wizard of oz platform," in *Towards Social Humanoid Robots: What makes Interaction Human-Like? Workshop at International Conference on Intelligent Robots and Systems* (Tokyo).
- Gouaillier, D., Hugel, V., Blazejic, P., Kilner, C., Monceaux, J., Lafourcade, P., et al. (2009). "Mechatronic design of NAO humanoid," in *Proceedings of IEEE International Conference on Robotics and Automation* (Kobe: IEEE), 769–774. doi: 10.1109/robot.2009.5152516
- Han, J., Campbell, N., Jokinen, K., and Wilcock, G. (2012). "Investigating the use of non-verbal cues in human-robot interaction with a Nao robot," in *2012 IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom)* (Košice: IEEE), 679–683. doi: 10.1109/CogInfoCom.2012.6421937
- Hayes, C. J., Crowell, C. R., and Riek, L. D. (2013). "Automatic processing of irrelevant co-speech gestures with human but not robot actors," in *Proceedings*

- of the 8th ACM/IEEE International Conference on Human-Robot Interaction (Tokyo: IEEE Press), 333–340. doi: 10.1109/HRI.2013.6483607
- Hossen Mamode, H. Z., Bremner, P., Pipe, A. G., and Carse, B. (2013). “Cooperative tabletop working for humans and humanoid robots: group interaction with an avatar,” in *IEEE International Conference on Robotics and Automation* (Karlsruhe: IEEE), 184–190. doi: 10.1109/icra.2013.6630574
- Hostetter, A. B. (2011). When do gestures communicate? a meta-analysis. *Psychol. Bull.* 137, 297–315. doi: 10.1037/a0022128
- Huang, C.-M., and Mutlu, B. (2014). “Learning-based modeling of multimodal behaviors for humanlike robots,” in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction-HRI’14*, (Bielefeld: ACM Press), 57–64. doi: 10.1145/2559636.2559668
- Kelly, S. D., Barr, D. J., Church, R., and Lynch, K. (1999). “Offering a hand to pragmatic understanding: the role of speech and gesture in comprehension and memory,” *J. Mem. Lang.* 40, 577–592. doi: 10.1006/jmla.1999.2634
- Kelly, S. D., Creigh, P., and Bartolotti, J. (2010). Integrating speech and iconic gestures in a Stroop-like task: evidence for automatic processing. *J. Cogn. Neurosci.* 22, 683–694. doi: 10.1162/jocn.2009.21254
- Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge, UK: Cambridge University Press.
- Kilner, J. M., Paulignan, Y., and Blakemore, S. J. (2003). An interference effect of observed biological movement on action. *Curr. Biol.* 13, 522–525. doi: 10.1016/S0960-9822(03)00165-9
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago, IL: University of Chicago Press.
- Metta, G., Fitzpatrick, P., and Natale, L. (2006). Yarp: yet another robot platform. *Int. J. Adv. Robot. Syst.* 3, 43–48. doi: 10.5772/5761
- Ono, T., Kanda, T., Imai, M., and Ishiguro, H. (2003). “Embodied communications between humans and robots emerging from entrained gestures,” in *Proceedings 2003 IEEE International Symposium on Computational Intelligence in Robotics and Automation* (Kobe: IEEE), 558–563. doi: 10.1109/CIRA.2003.1222241
- Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *J. Neurosci. Methods* 162, 8–13. doi: 10.1016/j.jneumeth.2006.11.017
- Powers, A., Kiesler, S., Fussell, S., and Torrey, C. (2007). “Comparing a computer agent with a humanoid robot,” in *Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference on*, (IEEE), 145–152. doi: 10.1145/1228716.1228736
- Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., et al. (2009). “{ROS}: an open-source Robot Operating System,” in *Open-Source Software Workshop of the International Conference on Robotics and Automation (ICRA)* (Shanghai).
- Riek, L., Rabinowitch, T., Bremner, P., Pipe, A., Fraser, M., and Robinson, P. (2010). “Cooperative gestures: effective signaling for humanoid robots,” in *5th ACM/IEEE International Conference on Human-Robot Interaction* (Osaka). doi: 10.1145/1734454.1734474
- Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., and Joubin, F. (2013). To Err is human(-like): effects of robot gesture on perceived anthropomorphism and likability. *Int. J. Soc. Robot.* 5, 313–323. doi: 10.1007/s12369-013-0196-9
- Sauppé, A., and Mutlu, B. (2014). “Robot deictics,” in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction-HRI’14* (Bielefeld: ACM Press), 342–349. doi: 10.1145/2559636.2559657
- Shrout, P. E., and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* 86, 420–428. doi: 10.1037/0033-2909.86.2.420
- Tanaka, K., Nakanishi, H., and Ishiguro, H. (2015). Physical embodiment can produce robot operator’s pseudo presence. *Front. ICT* 2:8. doi: 10.3389/fict.2015.00008
- von der Pütten, A. M., Krämer, N. C., Gratch, J., and Kang, S.-H. (2010). “It doesn’t matter what you are! Explaining social effects of agents and avatars,” *Comput. Hum. Behav.* 26, 1641–1650. doi: 10.1016/j.chb.2010.06.012
- Wang, L., and Chu, M. (2013). The role of beat gesture and pitch accent in semantic processing: an ERP study. *Neuropsychologia* 51, 2847–2855. doi: 10.1016/j.neuropsychologia.2013.09.027
- Yang, J., Andric, M., and Matthew, M. M. (2015). The neural basis of hand gesture comprehension: a meta-analysis of functional magnetic resonance imaging studies. *Neurosci. Biobehav. Rev.* 57, 88–104. doi: 10.1016/j.neubiorev.2015.08.006
- Zheng, M., and Meng, M. Q.-H. (2012). “Designing gestures with semantic meanings for humanoid robot,” in *2012 IEEE International Conference on Robotics and Biomimetics (ROBIO)* (Guangzhou: IEEE), 287–292. doi: 10.1109/ROBIO

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Bremner and Leonards. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Physical embodiment can produce robot operator's pseudo presence

Kazuaki Tanaka^{1,2}, Hideyuki Nakanishi^{1*} and Hiroshi Ishiguro³

¹ Department of Adaptive Machine Systems, Osaka University, Suita, Japan, ² CREST, Japan Science and Technology Agency, Tokyo, Japan, ³ Department of Systems Innovation, Osaka University, Toyonaka, Japan

OPEN ACCESS

Edited by:

Javier Jaen,
Universitat Politècnica de València,
Spain

Reviewed by:

Andrej Košir,
University of Ljubljana, Slovenia
Khiet Phuong Truong,
University of Twente, Netherlands

*Correspondence:

Hideyuki Nakanishi,
Department of Adaptive Machine
Systems, Osaka University, 2-1
Yamadaoka, Suita, Osaka 565-0871,
Japan
nakanishi@ams.eng.osaka-u.ac.jp

Specialty section:

This article was submitted to
Human-Media Interaction, a section
of the journal *Frontiers in ICT*

Received: 26 February 2015

Accepted: 28 April 2015

Published: 18 May 2015

Citation:

Tanaka K, Nakanishi H and
Ishiguro H (2015) Physical
embodiment can produce robot
operator's pseudo presence.
Front. ICT 2:8.
doi: 10.3389/fict.2015.00008

Recent studies have focused on humanoid robots for improving distant communication. When a user talks with a remote conversation partner through a humanoid robot, the user can see the remote partner's body motions with physical embodiment but not the partner's current appearance. The physical embodiment existing in the same room with the user is the main feature of humanoid robots, but the effects on social telepresence, i.e., the sense of resembling face-to-face interaction, had not yet been well demonstrated. To find the effects, we conducted an experiment in which subjects talked with a partner through robots and various existing communication media (e.g., voice, avatar, and video chats). As a result, we found that the physical embodiment enhances social telepresence. However, in terms of the degree of social telepresence, the humanoid robot remained at the same level as the partner's live video, since presenting partner's appearance also enhances social telepresence. To utilize the anonymity of a humanoid robot, we proposed the way that produces pseudo presence that is the sense of interacting with a remote partner when they are actually interacting with an autonomous robot. Through the second experiment, we discovered that the subjects tended to evaluate the degree of pseudo presence of a remote partner based on their prior experience of watching the partner's body motions reproduced by a robot. When a subject interacted with an autonomous robot after interacting with a teleoperated robot (i.e., a remote operator) that is identical with the autonomous robot, the subjects tended to feel as if they were talking with a remote operator.

Keywords: teleoperated robot, autonomous robot, videoconferencing, avatar, face-to-face, social telepresence, face tracking

Introduction

Currently, we can easily use audio and videoconferencing software. Audio-only conferencing, such as a voice chat, has a problem in that social telepresence decreases. The social telepresence is the sense of resembling face-to-face interaction (Finn et al., 1997). Enhancing social telepresence psychologically makes the physical distance between remote people less and saves time and money on travel. The most common method of enhancing social telepresence is videoconferencing. It had been proposed that live video can transmit the social telepresence of a remote conversation partner (Isaacs and Tang, 1994; de Greef and Ijsselstein, 2001). However, videoconferencing is closer to a situation of talking through a window than face-to-face conferencing due to a display.

To further enhance social telepresence, recent studies have begun on robot conferencing in which people talk with a remote conversation partner through teleoperated humanoid robots. The robots use motion tracking technologies to reflect partner's facial and body motions in real time. The main

features of robot conferencing are to transmit conversation partner's body motions and to present these motions via a physical embodiment. The physical embodiment means the substitution of a partner's body that exists physically in the same place as a user. Thus, it is expected that the user may feel closer to face-to-face interaction. Some studies reported superiorities of robot conferencing to videoconferencing (Morita et al., 2007; Sakamoto et al., 2007). One such study showed that the teleoperated robot, which has a realistic human appearance, enhances social telepresence compared with audio-only conferencing and videoconferencing (Sakamoto et al., 2007). Even so, it is difficult that each user owns a robot with his/her realistic appearance due to the high cost. For this reason, a teleoperated robot that has a human-like face without a specific age or gender is developed (Ogawa et al., 2011). However, there is a question whether such an anonymous robot can produce higher social telepresence compared with videoconferencing in which a user can see the remote partner's motion and appearance.

As the communication medium similar to the robot conferencing, avatar chats are available. Recently, it has become easy and inexpensive to use avatar chats such as avatar Kinect. The avatar chat resembles the robot conferencing in transmitting user's body motions without disclosing the user's appearance, but differs in reflecting these movements onto a computer graphics animation, which does not have a physical embodiment. A lot of studies found positive effects of avatar on distant communication (Garau et al., 2001; Bailenson et al., 2006; Bente et al., 2008; Kang et al., 2008; Tanaka et al., 2013). Several such studies focused on social

telepresence reported that avatar chats are better than audio-only communication (Bente et al., 2008; Kang et al., 2008), but worse than videoconferencing (Kang et al., 2008). Thus, presenting partner's body motion and appearance might contribute to produce social telepresence. If the physical embodiment does not produce social telepresence, the usefulness of humanoid robots would decrease since robots are more expensive than videos and avatars.

In this study, we conducted two experiments to prove the usefulness of humanoid robot. First, it is necessary to demonstrate that the physical embodiment enhance social telepresence independently from the transmitting information, e.g., audio, motion, and appearance. In the first experiment, we investigated how the physical embodiment and transmitting information factors influence the social telepresence (Tanaka et al., 2014). To analyze the effects of the two factors separately, we prepared six communication methods as shown in **Figure 1**. The voice chat, avatar chat, and videoconferencing that do not have a physical embodiment transmit audio-only, audio + motion, and audio + motion + appearance, respectively. The robot conferencing that has a physical embodiment transmits audio + motion, and so it corresponds to the avatar chat as described above. As the method that corresponds to the voice chat, we set an inactive robot conferencing that transmits audio but no motion. Furthermore, we assumed that the face-to-face interaction corresponds to the videoconferencing.

Another method to prove the usefulness is to demonstrate that humanoid robots produce pseudo presence of remote

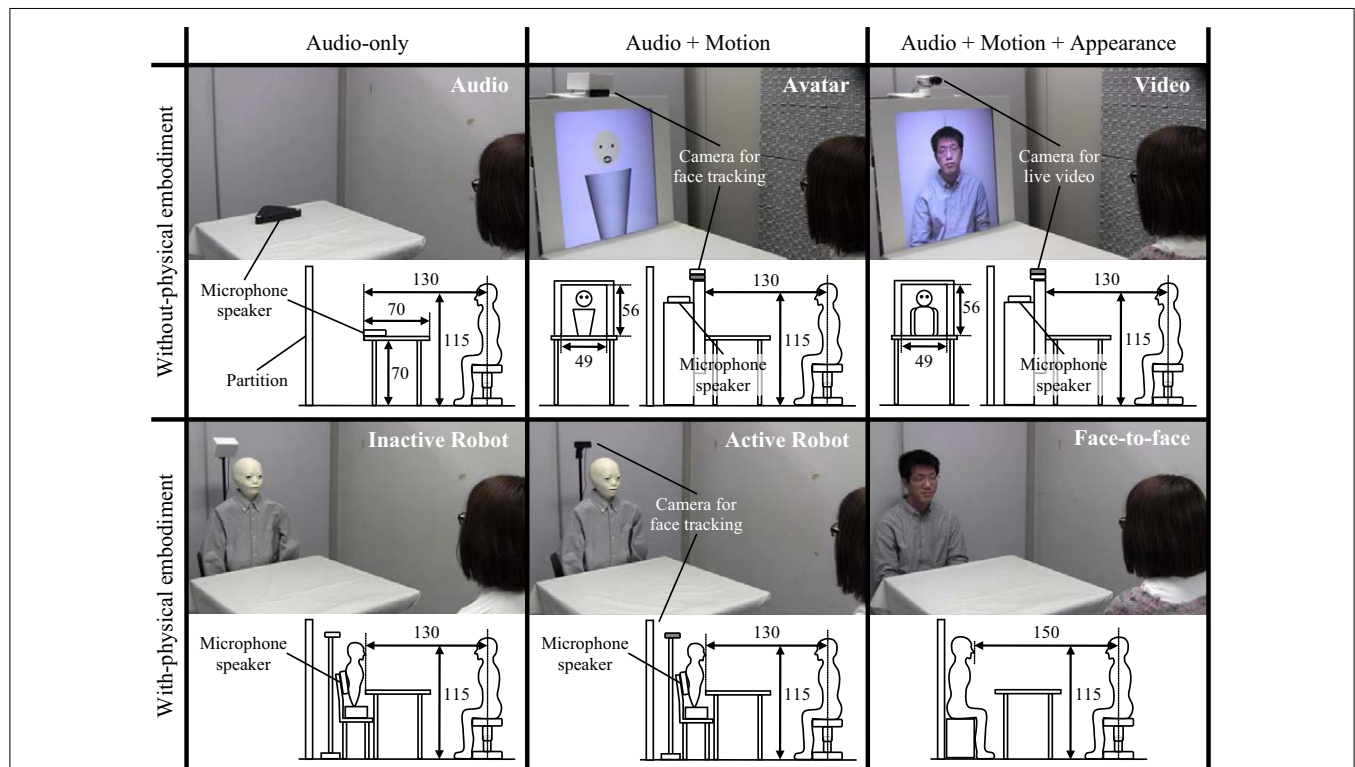


FIGURE 1 | Conditions and the setups of the first experiment (length unit: centimeters): the six communication methods divided into physical embodiment and transmitting information factors.

conversation partner. Pseudo presence means the feeling of interacting with a remote operator when interacting with an autonomous robot. The two main types of humanoid robots are teleoperated and autonomous. Teleoperated robots transmit remote operator's social telepresence by reproducing remote operator's behavior. On the other hand, autonomous robots produce remote operator's pseudo presence by behaving as being controlled by a remote operator. We hence believe that the essential difference between these robots is the presence or absence of a remote operator. If autonomous robots generate human-like behavior comparable to body motions obtained by motion tracking technologies, the user could feel pseudo presence of remote partner even when interacting with the autonomous robot. Thus, to decide presence/absence of conversation partner in interacting with an autonomous robot, prior experience of watching the remote partner's motion reproduced by teleoperated robot whose design is identical as the autonomous robot may be needed.

We expect that the presence of a remote partner in interacting with a teleoperated robot will be recalled in interacting with an autonomous robot by the robot's behaviors and the user might continue to feel the partner's presence. First, a user experiences the teleoperated mode in which the user talks with a remote conversation partner through a robot synchronized to the partner's body motions. After that, the user experiences the autonomous mode in which he/she talks with an autonomous system operating the same robot used in the teleoperated mode. This system generates talking behaviors, e.g., lip motion, from the pre-recorded partner's speech, or nod motions from the user's speech.

There are some applications that a remote operator provides services through a teleoperated robot. An autonomous robot could give the user the same services instead of a remote operator if the robot is able to produce the pseudo presence. In the example of an interaction robot (Ranatunga et al., 2012; Tanaka et al., 2012), if a user who is living alone talked with a remote caregiver through the robot, the user might continue to feel the caregiver's presence even after moving to the autonomous mode. The pseudo presence of remote caregiver may reduce the user's feeling of loneliness more effectively even if the remote caregiver is not talking actually. For a lecture robot (Hashimoto et al., 2011), the students might feel the remote teacher's presence at an autonomous lecture after preliminarily greeting each other in the teleoperated mode. Due to the pseudo presence of remote teacher, the students may pay attention to the lecture even if the lecture is autonomously reproduced. While an autonomous robot interacts with the user, the remote operator does not have to work.

In the second experiment, we compared presence or absence of the prior experience of talking with remote conversation partner through a robot. Our approach that changes the dialog modes of a robot from teleoperated to autonomous utilizes a weakness of anonymous humanoid robot that is the lack of appearance. Therefore, this approach can be applied to media, which do not show partner's current appearance. To confirm the contribution of physically embodied body motion to produce the pseudo presence, we conducted the comparison also in audio-only communication in which the user cannot see the partner.

The paper is organized as follows. The next section presents the related work. Section "Subjects" explains about the subjects

who participated in our experiments. Sections "Experiment 1" and "Experiment 2" explain the methods and the results of the first and second experiment, respectively. The first experiment investigates whether the features of humanoid robot enhance social telepresence. The second experiment investigates the effects of humanoid robot on producing pseudo presence of remote conversation partner. Section "Discussion" discusses the results of these experiments. Finally, Section "Conclusion" concludes the paper.

Related Work

This study is related with the telerobotics and intelligent robotics. In the telerobotics field, many studies have proposed various teleoperated robots that present the operator's facial movements (Kuzuoka et al., 2004; Morita et al., 2007; Sakamoto et al., 2007; Ogawa et al., 2011; Sirkin and Ju, 2012) with a physical embodiment. Several studies reported the superiority of robot conferencing to videoconferencing. One such study showed that the eye-gaze of remote person reproduced by a robot was more recognizable than by a live video (Morita et al., 2007). The study with regard to social telepresence concluded that teleoperated robot transmitted a higher social telepresence of a remote conversation partner than audio-only and videoconferencing (Sakamoto et al., 2007). However, this result seems somewhat obvious, since the teleoperated robot reproduced the whole body of a person, whereas the videoconferencing only showed conversational partner's head. The video image of only a head is harmful to social telepresence (Nguyen and Canny, 2009), so that a superiority of robot conferencing to videoconferencing, which shows the whole body of a person was also not clear. Furthermore, the teleoperated robot that was used in the study had a specific person's appearance, and so it was not clear, which of the factors, the physical embodiment, the appearance, or the ability to present body motions, enhanced social telepresence. To clarify them, we used an anonymous teleoperated robot (Ogawa et al., 2011) that has a human-like face without a specific age or gender, and compared it with partner's life-size video.

In videoconferencing research, it was reported that the remote person's movement that was augmented by a display's physical movement enhanced the social telepresence (Nakanishi et al., 2011). This result implies that the physically embodied body motion enhances social telepresence.

In the intelligent robotics field, there are studies that focused on the effects of the physical embodiment on social presence (Lee et al., 2006; Bainbridge et al., 2011). These studies showed that a humanoid robot produces higher social presence than on-screen agents. These studies evaluated whether people interact with a non-human social agent (i.e., robots and on-screen agent) as if it were an actual human. By contrast, our experiments evaluated whether people feel being with a remote conversation partner in the same room when talking with a humanoid robot. When the teleoperated robot conveyed remote partner's body motions, we estimated the degree of remote partner's social telepresence. On the other hand, when the robot moves automatically, we estimated the degree of the partner's pseudo presence. There is a possibility that the physical embodiment contributes to enhance these presence.

There are some technologies that generate talking behaviors autonomously instead of transmitting remote partner's behaviors. If a robot can generate human-like talking behaviors, a user may believe that it is moving based on the partner's body motions, since the user does not know the partner's current appearance or behavior. Many past studies have proposed algorithms to generate talking behaviors from someone's speech (Cao et al., 2005; Salvi et al., 2009; Lee et al., 2010; Watanabe et al., 2010; Le et al., 2012; Liu et al., 2012). These studies generated human-like talking behaviors that were as natural and various as possible. But no research has investigated whether this approach produces the sense of talking with a remote partner. We predicted that prior experience of watching the remote partner's behavior reproduced by a robot produces the sense when watching talking behavior generated by the same robot. A robot that can be controlled by teleoperated and autonomous modes has been developed (Ranatunga et al., 2012), but the effect of changing these modes on producing remote partner's presence has not been clarified.

Subjects

Thirty-six undergraduates (17 females and 19 males) and 16 undergraduates (9 females and 7 males) participated in our first and second experiments, respectively. We used a recruitment website for part-time workers to collect the subjects who lived near our university campus.

We did not choose students of master's course and upward as subjects to prevent an influence of their expertise on the results. For the same reason, we employed mainly liberal arts undergraduates or science and engineering undergraduates who do not study about robotics. The subjects had never met the experimenter before the experiment.

We recorded the experiments and interviews for the subjects. The subjects were required to sign a consent form that confirmed whether they agree with the recording. The consent form also confirmed whether the recorded movies could be used for presentations, articles, or TV programs. If the subject does not agree with using their movies, he/she could refuse it. We are holding the consent forms and movies under lock and key.

Experiment 1

This section presents the first experiment in which we investigated how the physical embodiment and transmitting information factors influence on social telepresence.

Hypothesis

The main features of robot conferencing are to have a physical embodiment and to transmit conversation partner's body motions. We predicted that these features enhance social telepresence. A previous study showed the superiority of a humanoid robot that has realistic appearance to videoconferencing (Sakamoto et al., 2007). In addition, several previous studies reported that avatars that transmit partner's body motions enhance social telepresence compared with audio-only media (Bente et al., 2008; Kang et al., 2008). Since these findings suggest the contribution of the robot's features on social telepresence, we made the following two hypotheses.

Hypothesis 1: a physical embodiment enhances the social telepresence of the conversation partner.

Hypothesis 2: transmitting body motions enhances the social telepresence of the conversation partner.

Conditions

The hypotheses described in the preceding section consist of these two factors: physical embodiment and transmitting information. The physical embodiment factor had two levels, with/without-physical embodiment, and the transmitting information factor had three levels, audio, audio + motion, and audio + motion + appearance. Thus, to examine the hypotheses, we prepared six conditions of a 2×3 design shown in Figure 1.

As described in Section "Introduction," both robot conferencing and avatar chat transmit remote person's body motions without disclosing the person's appearance. We thus supposed that the avatar chat can become robot conferencing by adding a physical embodiment. Similarly, we assumed that the voice chat becomes an inactive robot conferencing, which does not transmit the body motions of a remote person and the video chat can become face-to-face communication by adding a physical embodiment. In terms of the transmitting information, we assumed that the voice chat and inactive robot transmit only audio, the avatar and robot transmit audio and motion, and the video and face-to-face transmit audio, motion, and appearance. These assumptions allowed us to analyze the effect of adding a physical embodiment to existing communication media. The details of each condition are described below.

Active Robot Condition (Transmitting Audio and Motion with a Physical Embodiment)

The subject talked to the conversation partner while looking at the robot. The robot had a three-degrees-of-freedom neck and a one-degree-of-freedom mouth. The head and lips moved at 30 frames per second according to the sensor data sent from face tracking software (faceAPI), which was running in a remote terminal and capturing the conversation partner's movements. The camera for face tracking was set behind the robot. The microphone speaker was set behind the robot. The robot was dressed with the same gray shirt as the conversation partner.

Avatar Condition (Transmitting Audio and Motion but no Physical Embodiment)

The subject talked to the conversation partner while looking at an anonymous three-dimensional computer graphics avatar that reflected the conversation partner's head and lip motions. The avatar consisted of a skin-colored cylindrical head, black lips, black eyeballs, and a gray conical body, which was the same color as the shirt of the conversation partner. In the preliminary experiment, we used an avatar, which had a spherical head and a realistic shirt, which looked like the robot. However, there were some subjects who felt hard to notice facial movements of the avatar. This problem was solved by changing the design of avatar to a cylindrical head. The recognizable facial movements might improve social telepresence, and so we employed the cylindrical head. In addition, we modified its body to a conical shape to standardize the abstraction level of the looks. The diameter of the

head was equal to the breadth of the robot's head (13.5 cm). The conversation partner's head and lip motions were tracked in the same way as on the active robot condition. The head translated and rotated with three degrees of freedom. The lips were transformed based on the three-dimensional positions of 14 markers. The head and lips moved at 30 frames per second. The avatar was shown on a 40" display. The display was set longitudinally on the other side of the desk. The bezel of the display was covered with a white board, so that the true display area was 49 cm by 56 cm. The microphone speaker was set behind the display. There were two cameras on top of the display. One was for face tracking, and the other was for live video. In this condition, the camera for live video was covered with a white box. The camera was used in the video condition described below.

Face-to-Face Condition (Transmitting Audio, Motion, and Appearance with a Physical Embodiment)

The subject talked to the conversation partner in a normal face-to-face environment. The conversation partner wore a gray shirt. The distance from the subject to the conversation partner was adjusted to 150 cm so that the breadth of the conversation partner's head looked the same as the breadth of the robot's head (13.5 cm).

Video Condition (Transmitting Audio, Motion, and Appearance but no Physical Embodiment)

This condition was identical to a normal video chat. The subject talked to the conversation partner while looking at a live video of the conversation partner. The conversation partner wore a gray shirt. The resolution of the camera for live video was 1280 pixels by 720 pixels, and its frame rate was 30 frames per second. The video was shown on the same display that was used in the avatar condition. Thus, the true display area was 49 cm by 56 cm. The horizontal angle of view was adjusted to 87° so that the breadth of the conversation partner's head was equal to the breadth of the robot's head (13.5 cm) on the display. The camera for face tracking that was used on the avatar condition was covered with a white box.

Inactive Robot Condition (Transmitting Audio with a Physical Embodiment)

The subject talked to the conversation partner while looking at the inactive robot. The camera for face tracking that was used on the active robot condition was covered with a white box. The subject was preliminarily informed that the robot did not move in this condition.

Audio-Only Condition (Transmitting Audio but no Physical Embodiment)

This condition was similar to a normal voice chat. The subject talked to the conversation partner through only a microphone speaker that was set on the desk.

In the preliminary experiment, some subjects doubted that the experimenter would be looking at them from somewhere even if the experimental condition required no camera. We hence informed the subjects that the dialog environments of the subject side and the conversation partner side were the same in all the conditions. To make the subjects believe this bi-directionality of

the dialog environments, the subjects were shown a live video of the subjects' avatar, robot, or video, which were seen by the conversation partner on a 7" display before each experiment. At the same time, the subjects confirmed that their avatar and robot reflected their face and lip movements. The subjects also confirmed that the avatar and robot in front of them reflected the conversation partner's face and lip movements by comparing a live video of the conversation partner that was shown on the 7" display with the avatar and robot. The 7" display for these confirmations was removed before the experiments.

Task

In the experiment, we informed the subjects that they were going to talk with a conversation partner who is in another room through six communication methods described above. An experimenter played the role of the partner. To observe the difference in the social telepresence between the conditions, we asked the subject to answer a questionnaire (which is explained in the next section) after the experiment ended. Since body motions in conversation are mainly speaking and nodding, we set the task in which the subject could see the partner's mouth and neck motions.

The subject was asked by the experimenter to talk about the issue and resolution of a certain gadget and requests for a new function on that gadget at the beginning of each condition. Because all the subjects had to experience the six conditions, we prepared six gadgets as conversational topics, i.e., e-book readers, handheld game consoles, smartphones, robotic vacuum cleaners, portable audio players, and 3D televisions. We did not disclose the next topic beforehand, and the experimenter told the subject which gadget to talk about right when the condition began. While the subject was talking, the experimenter gave back-channel responses with an utterance and a small nod of his head. The nod motions of the robot, avatar, and experimenter are shown in **Figure 2**. As the figure shows, the robot and avatar synchronized with the experimenter.

We did not ask the subject to talk for more than a certain duration, so the subject could stop talking anytime. However, since the six gadgets are attracting considerable attention recently, most subjects knew the issue and resolution of the gadgets to a certain level, and their speech was able to last more than 1 min.

The order of experiencing the conditions and the order of the topics were counterbalanced. The subject trained the task in the face-to-face condition in order to familiarize the subject with the task and the experimenter's motion and appearance, before

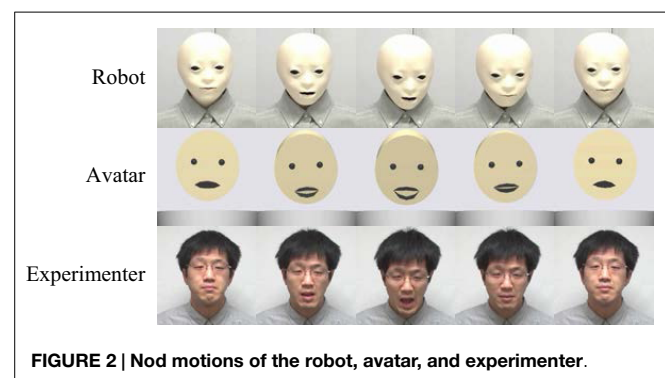


FIGURE 2 | Nod motions of the robot, avatar, and experimenter.

conducting the experiment in the six conditions. The topic of the training was always railway smart cards.

Questionnaire

After experiencing the six conditions, the subjects answered a questionnaire, which asked them to estimate the social telepresence, i.e., the degree of resembling face-to-face interaction (Finn et al., 1997) for each condition. We wanted to obtain the relative comparison of the conditions to avoid a ceiling effect. In the preliminary experiment, we conducted the questionnaire after each condition. However, when a subject marked the highest score for the first condition, the subject was not able to mark higher score for the later conditions even if he/she felt higher social telepresence. In the case of a between-subject design, such a problem will not happen, but another problem here is that there are six conditions, thus a lot of subjects are necessary to enable a between-subjects design.

The questionnaire is shown in **Figure 3**. The questionnaire had six statements that corresponded to the six conditions. The statement was the following: I felt as if I were talking to the conversation partner in the same room. Previous studies showed that the statement which asks a feeling of being in the same room is useful to measure the social telepresence (Nakanishi et al. 2008, 2009, 2011, 2014; Tanaka et al., 2014). The statement was rated on a 9-point Likert scale where 1 = strongly disagree, 3 = disagree, 5 = neutral, 7 = agree, and 9 = strongly agree. The subjects thereby could score the same number on the statements if they felt the same level of social telepresence in the conditions.

The statements were sorted in the order of the conditions and were printed on the questionnaire, with a photo that showed the experimental setup of the corresponding condition. The sort and the photo were good cues to help the subjects remember the feeling of social telepresence in each condition. After conducting

the questionnaire, we interviewed the subjects in order to confirm the reason of scoring. The interview was open-ended. When we received the questionnaire that was against our hypotheses, we asked the subject the reason, e.g., the reason why the avatar condition was higher than the robot condition. Even if the questionnaire followed our hypotheses, we asked the reason, to confirm what point the subject focused on, e.g., physical embodiment, body motion, or appearance.

Result

Thirty-six subjects (17 females and 19 males) participated in our first experiment. The experiment was within-subject design, so each subject experienced all of the six conditions. We did not control the subjects' prior knowledge about the topics of talking with the experimenter. Instead, at the interview following the experiment, we confirmed that their scoring of questionnaire was conducted independently from the difference of topics. There was no subject who mentioned about the topics as the reason of his/her scoring.

Figure 4 shows the result of the questionnaire, in which each point represents the mean value of the scores, and each bar represents the SEM value.

We compared the six conditions to find the effects of the physical embodiment and the transmitting information factors. Since the physical embodiment and the transmitting information factors consisted of two and three levels as shown in **Figure 1** and each subject evaluated all conditions, we conducted 2×3 two-way repeated-measures ANOVA. As a result, we found strong main effects of the physical embodiment factor [$F(1, 35) = 36.955, p < 0.001$] and the transmitting information factor [$F(2, 70) = 279.603, p < 0.001$]. We also found a strong interaction between these factors [$F(2, 70) = 14.794, p < 0.001$]. Regarding this interaction, we calculated the *post hoc* statistical power. First, we calculated the effect size 0.650 from the partial correlation ratio 0.297. Finally, we obtained the sufficiently high-statistical power 0.999 when the significance level was 0.001.

We further analyzed the simple main effects in the interaction with the Bonferroni correction. The physical embodiment significantly improved the social telepresence of the conversation partner, when the transmitting information was audio + motion + appearance [$F(1, 105) = 8.857, p < 0.01$], and audio + motion [$F(1, 105) = 65.470, p < 0.001$]. When the transmitting information was audio only, there was a non-significant tendency for the social telepresence to increase [$F(1, 105) = 3.460, p = 0.086$]. This meant that the subjects felt a higher social telepresence of the conversation partner in the face-to-face condition than in the video condition, and the active robot condition conveyed a higher social telepresence than the avatar. These results support hypothesis 1 that the physical embodiment enhances the social telepresence of the conversation partner. However, the effect of the physical embodiment on the social telepresence was lower in the audio-only communication.

Furthermore, there were significant differences between the three levels of the transmitting information in both cases of without-physical embodiment [$F(2, 140) = 223.095, p < 0.001$] and with-physical embodiment [$F(2, 140) = 107.141, p < 0.001$]. Multiple comparisons showed that the subjects felt a higher social

I felt as if I were talking to the conversation partner in the same room.







	strongly disagree	disagree	neutral	agree	strongly agree				
	1	2	3	4	5	6	7	8	9
	strongly disagree	disagree	neutral	agree	strongly agree				
Inactive	1	2	3	4	5	6	7	8	9
	strongly disagree	disagree	neutral	agree	strongly agree				
	1	2	3	4	5	6	7	8	9
	strongly disagree	disagree	neutral	agree	strongly agree				
Active	1	2	3	4	5	6	7	8	9
	strongly disagree	disagree	neutral	agree	strongly agree				
	1	2	3	4	5	6	7	8	9
	strongly disagree	disagree	neutral	agree	strongly agree				
	1	2	3	4	5	6	7	8	9

FIGURE 3 | Questionnaire to evaluate social telepresence of the six conditions.

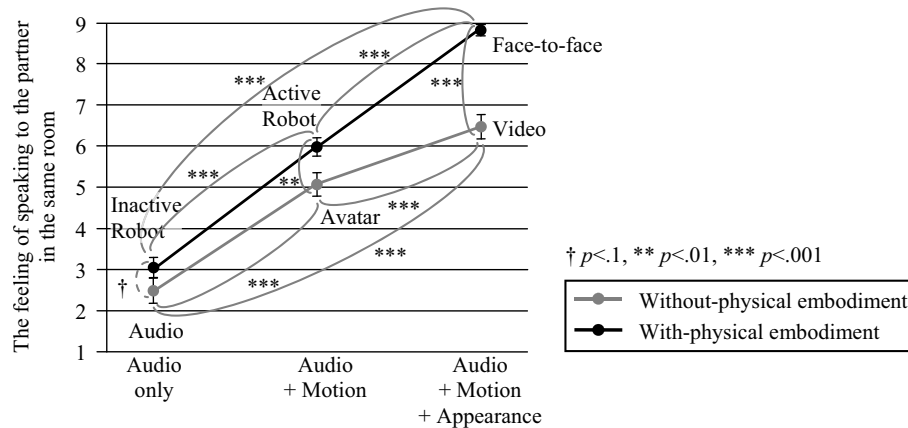


FIGURE 4 | Result of the first experiment: the interaction effect between the physical embodiment and transmitting information factors on social telepresence.

telepresence in the face-to-face condition than in the active robot ($p < 0.001$) and inactive robot ($p < 0.001$) conditions, the active robot condition conveyed a higher social telepresence than the inactive robot condition ($p < 0.001$), the video condition conveyed a higher social telepresence than the avatar ($p < 0.001$) and the audio-only ($p < 0.001$) conditions, and the avatar condition conveyed a higher social telepresence than the audio-only condition ($p < 0.001$). These results prove hypothesis 2 that transmitting body motions enhances the social telepresence of the conversation partner. In addition, transmitting appearance also enhanced the social telepresence.

Experiment 2

The first experiment demonstrated that transmitting the remote partner's current appearance as well as a physical embodiment enhances social telepresence. The result that the robot and video conditions were similar level as shown in **Figure 4** might be because the robot condition could not transmit the appearance. This will be discussed in detail in Section "Discussion." Therefore, the clear usefulness of humanoid robots was not demonstrated. To prove the usefulness of humanoid robots, we had to prove the benefit of both the physical embodiment and the absence of remote partner's appearance.

This section presents the second experiment in which we investigated whether the presence of a remote partner in the teleoperated mode produces a sense of talking with the partner while actually talking with the robot in the autonomous mode. To confirm the contribution of a physical embodiment to produce such a pseudo presence, we compared the robot and audio-only conditions.

Since the first experiment suggested that a physical embodiment is effective when a robot moves, the second experiment dealt with a robot as a single condition. We chose audio-only (without-physical embodiment and body motion) as a condition for the baseline. Audio-only media (e.g., voice chats) can also be automated, since a user cannot see the remote partner's current appearance, like with a robot. If a remote partner's presence in audio-only media can produce the sense of talking with the

partner while actually talking with an autonomous reply system, the physical embodiment, which is a robot, is not needed to produce such a pseudo presence.

Dialog Modes

Almost all the studies that proposed algorithms that generate talking behaviors focused on facial movements, e.g., nodding and lip motions. Since they seem to be the most fundamental facial movements while talking, this study also addressed them. Many of the teleoperated robots proposed in past studies have a face that can present these motions (Sakamoto et al., 2007; Watanabe et al., 2010; Hashimoto et al., 2011; Ogawa et al., 2011). In this experiment, we used telenoid that was used in experiment 1. We controlled it in two modes: teleoperated and autonomous.

Teleoperated Mode

In this mode, the robot's head and mouth were synchronized with the remote operator. The method to control the robot was same as the active robot condition described in Section "Conditions."

Autonomous Mode

The roles of the remote partner in the dialogs are listener and speaker. Their behaviors are mainly nod and lip motions, respectively. We constructed a back-channel system that detects the timing of back-channel feedback from user's speech and a lip-sync system that generates a lip motion synchronized with a remote partner's speech. We simplified these systems for the following reason. If our approach that changes the dialog modes makes subjects feel like they are talking with their remote partner even in simple systems, it would obviously also work on systems that generate more natural and various talking behaviors.

Back-channel system

Many methods detect the timing of back-channel responses. Most used prosodic information, including pause (Noguchi and Den, 1998; Takeuchi et al., 2003; Truong et al., 2010; Watanabe et al., 2010), and fundamental frequency (Noguchi and Den, 1998; Ward and Tsukahara, 2000; Truong et al., 2010). Our method used only the speech pause since it is good cue to identify the break or

end of a sentence, which seems to be the appropriate timing of back-channel responses. One study also used only a speech pause, although their algorithm is more complex than ours in order to enable estimating the timing of back-channel earlier (Watanabe et al., 2010).

The detection rule is shown in **Figure 5**. Each box represents an utterance, and the distance between each box is pause duration t_1 . The utterance and pause parts correspond to higher and lower sound pressure, respectively. The system judged t_1 as a target pause if it exceeded 0.6 s. Speech duration t_2 is the elapsed time from the start of the speech to the time at which the target pause was recognized. If t_2 exceeded 2.0 s, the system judged the target pause as the timing of the back-channel response and reset t_2 to 0. This means that the system reproduces the back-channel response when the pause is continued for 0.6 s after the speech continued for more than 2.0 s.

To adjust the parameters of our back-channel system, we conducted preliminary experiments in which a subject evaluated the timing and frequency of robot's backchannel in the same task to the second experiment. We found that back-channels repeated in short time (<2.0 s) decreases the naturalness of the timing. In addition, the backchannel, which was done more than 0.6 s later from the break or end of sentence tended to be felt late, but the pause that is <0.6 s is not enough to judge the break or end of sentence. We therefore set pause duration t_1 and speech duration t_2 to 0.6 and 2.0 s, respectively.

At the back-channel response, the robot made a nodding motion and a pre-recorded acoustic back-channel. In our preliminary experiment, we used only one nodding motion and an acoustic back-channel, but subjects pointed out that the robot's response seemed constant. We therefore prepared three nodding motions that differ in their degree of pitch and speed and two acoustic back-channels that slightly differ in their tone of voice. This problem that subjects feel constant the robot's response was solved by randomly selecting these nodding motions and acoustic back-channels at the timing.

Lip-sync system

Some lip-sync methods generate lip motions from a human's voice to control a robot (Watanabe et al., 2010; Ishi et al., 2012) and a computer graphic avatar (Cao et al., 2005; Salvi et al., 2009; Watanabe et al., 2010). Our method was simpler because controlling a one-degree-of-freedom mouth does not need highly accurate lip-sync methods.

Our lip-sync system measured the acoustic pressure of the human's voice and related the level to the angle of the robot's chin. In other words, the robot's mouth was synchronized with the

waveform of the human's voice. In our experiments, this system was driven by pre-recorded remote partner's speeches.

Hypothesis

In the first experiment, the active robot condition could convey the same degree of social telepresence as the video condition without transmitting the partner's appearance. We hence thought that physically embodied body motions effectively make the user imagine the remote partner's presence. Due to the experience of imagining the partner's presence, the user might feel the remote partner's pseudo presence when talking with an autonomous robot. On the other hand, when the user talks with an autonomous reply system that uses only speech, it seems harder to feel the partner's presence due to poor information for imagining.

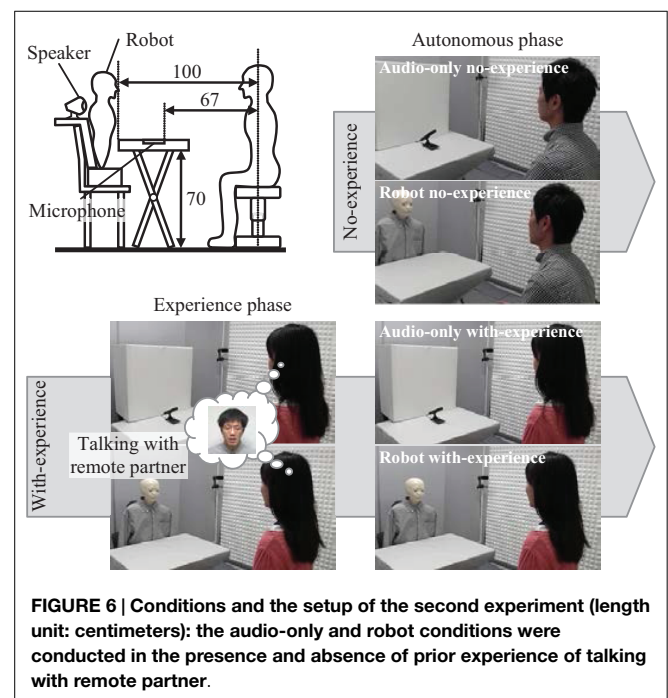
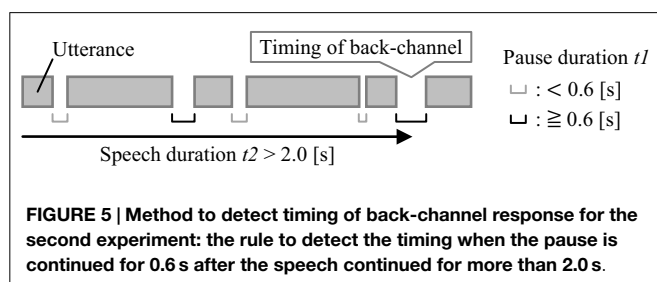
A previous study showed that a teleoperated robot produces higher social telepresence of a remote partner than audio-only communication due to the effect of physical embodiment (Tanaka et al., 2014). We hence predict that physically embodied body motion will improve the sense of talking with a remote partner. The following is the hypothesis of the second experiment.

Hypothesis 3

The user who experienced talking with a remote partner through a teleoperated robot that presents the partner's body motion will feel the sense of talking with the partner even when talking with the same robot that is being autonomously controlled.

Conditions

To examine hypothesis 3, we prepared the four conditions shown in **Figure 6**. The experiment included the experience and autonomous phases. The experience phase was only included in the with-experience conditions. Before experiencing the experience phase, the subjects were told that they would be talking



with a remote partner in the teleoperated mode. However, if an experimenter replied to the subject's speech, the quality of the conversation would differ for each subject. Actually, the experience phase was conducted in the autonomous mode to control the quality of the conversation for each subject. The manipulation check that will be explained in Section "Questionnaire" confirmed that all the subjects believed that the remote partner was listening to their speech through the robot. As described in Section "Autonomous Mode," the back-channel systems proposed by previous works would detect more appropriate timing of backchannel, but our simple algorithm was enough to make the subjects believe the remote partner's presence at a one-turn interaction. Before experiencing the autonomous phase, the subjects were told that they would be talking with an autonomous robot, which autonomously gives back-channel responses. To control the subjects' prior knowledge, we gave them handouts that explained the teleoperated and autonomous modes before the experiment. We also explained our experiments to the subjects.

The figure also shows the experimental setup. In all the experiments, the subject sat in front of a desk. The robot was placed on the other side of it. A directional microphone was embedded in the desk to capture the subject's speech, and the top of it was covered with a cloth to hide the microphone. A speaker was set behind the robot to produce the remote partner's speech.

Task

In the second experiment, the subject was a speaker, and the robot or the system gave a back-channel response to his/her speech as a listener for the following reason. If the subject is a listener, the autonomous system in the audio-only conditions only plays pre-recorded partner's speeches unilaterally from the speaker. In this case, the audio-only conditions seem to have a disadvantage over the robot conditions.

The task was same to the first experiment. The subjects were asked to talk about a gadget at the beginning of each conversation through the robot or the speaker. The lines of asking and the acoustic responses were the pre-recorded voices of a member of our research group. The member greeted the subjects before the experiment to identify the remote partner. The topics in the experience and autonomous phases were portable audio players and robotic vacuum cleaners, and smartphones and 3D TVs, respectively. The order of the topics was counterbalanced.

Questionnaire

After talking about one topic, the subjects were asked to answer manipulation check questions to confirm whether they correctly understand our instructions. For example, after the experience phase, we confirmed that the subjects believed that they were actually talking to a remote partner although it was an autonomous system. The manipulation check consisted of the following two YES/NO statements:

- In the last experiment, your speech was listened to by a remote partner.
- In the last experiment, your speech was recorded instead of being listened to by a remote partner.

After the experiment, the subjects were asked to estimate the pseudo presence that is the remote partner's presence in the autonomous phase. The following was the questionnaire statement:

- I felt as if the conversation partner was listening to me in the same room.

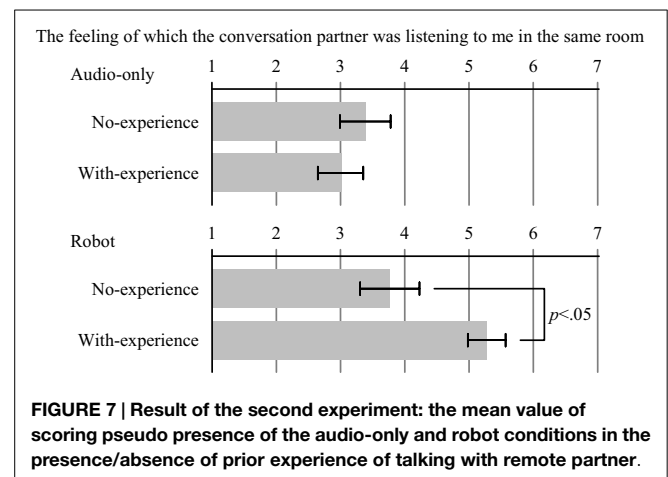
Answers were rated on a 7-point Likert scale: 1 = strongly disagree, 2 = disagree, 3 = slightly disagree, 4 = neutral, 5 = slightly agree, 6 = agree, and 7 = strongly agree. We collected open-ended responses to infer what determined the scores. The statement was accompanied with an entry column where the subjects rationalized their scores.

Result

Sixteen subjects participated in our second experiment. Half (five females and three males) participated in the with-experience conditions and experienced both the experience and autonomous phases. At each phase, they talked in both the audio-only and robot conditions. The order of experiencing the audio-only and robot conditions was counterbalanced. The other half of the subjects (four females and four males) participated in the no-experience conditions and only experienced the autonomous phase. According to the manipulation check, we confirmed that all the subjects believed our instruction.

The result of the second experiment is shown in **Figure 7**, where each box represents the mean value of the responses to the statement, and each bar represents the SEM value. The figure compares the no- and with-experience conditions by a between-subjects *t*-test.

In the audio-only conditions, there was no significant difference between the no- and with-experience conditions [$t(14) = 0.664$, *n.s.*]. On the other hand, in the robot conditions, we found a significant difference between them [$t(14) = 2.575$, $p < 0.05$]. This means that the prior experience in which the subjects talked with the remote partner produced the pseudo presence in the autonomous phase when the subjects could see back-channel responses through the robot. However, the experience did not produce pseudo presence in the audio-only



communication. These results proved hypothesis 3 described in Section “Hypothesis.”

In the with-experience conditions, the subjects had the conversations twice, but in the no-experience condition, they only had them once. Therefore, more conversations might improve the sense of talking with a remote partner. In spite of this, in the audio-only conditions, the difference between the no- and with-experience conditions was not significant. We hence considered that physically embodied motion was the significant factor to produce the pseudo presence regardless of the number of conversations.

Discussion

In the first experiment, the physical embodiment enhanced the social telepresence of the conversation partner. In the interviews, 7 of the 36 subjects said that they felt as if they were facing the conversation partner in the active robot condition compared with the avatar condition because there was a physical object in front of them. However, there was no significant difference between the audio-only condition and the inactive robot condition. In the interviews, 3 of the 36 subjects said that the inactive robot condition was not that different to the audio condition because they could not see the conversation partner's reaction. In fact, in the questionnaire, 8 of the 36 subjects rated the same score for the audio and inactive robot conditions. Moreover, 5 of the 36 subjects said that they felt as if the conversation partner was in front of them when the robot moved. These subjective responses support the experimental result that a physical embodiment enhances social telepresence when transmitting body motions. This result indicates the superiority of robots to avatars, which does not have a physical embodiment. Nevertheless, there are some subjects who rated the same or higher score for the avatar condition than the robot condition. Most such subjects mentioned the uncanny appearance of the robot as the reason for their rating, and they tended to prefer the avatar's design. Thus, if the robot's design was more abstracted, the superiority would appear more significantly.

Presence or absence of motion parallax can be cited as one of the differences between physical embodiment and video. When interacting with the robot, the depth from motion parallax could increase visibility of body motions. The lack of the depth information might be the cause of feeling hard to notice facial movements of the avatar used in the preliminary experiment described in Section “Conditions.” A previous study reported that motion parallax generated by the movement of a camera enhances social telepresence (Nakanishi et al., 2009). The visibility of bodily motion improved by the motion parallax may have contributed to enhance social telepresence.

In terms of the transmitting information, the appearance enhanced social telepresence as well as the body motions. This result shows the disadvantage of robots and avatars that do not transmit the partner's appearance. Although the active robot has this disadvantage, the active robot and video conditions seemed to convey the same degree of social telepresence, as shown in **Figure 4**. In the questionnaire, approximately half of the subjects (16 of the 36) rated the same or higher score for the active robot condition than the video condition. We assumed that the

enhanced social telepresence by the physical embodiment offset the decreased social telepresence by the absence of the partner's appearance. Therefore, the reported superiority of the robot in the social telepresence to the video (Sakamoto et al., 2007) could be caused by the robot's realistic appearance.

We did not investigate the conditions that transmit audio and appearance but not motion. Talking through an inactive robot that has a realistic appearance of a partner, and a partner's photo could correspond to such conditions. Watching the partner's photo while talking is a popular situation since many users of instant messengers put their photos in the buddy list. Although the transmitting appearance enhances social telepresence as mentioned above, it has not been clarified whether the appearance works even if the motion is not transmitted. The effect of presenting appearance on the smoothness of speech had already demonstrated (Tanaka et al., 2013, 2015). The previous study showed that presenting partner's avatar increased the degree of the smoothness of speaking to the partner, but partner's photo did not have such an effect. We hence predict that the appearance also does not enhance social telepresence if the motion is not transmitted as is the case with the physical embodiment. To prove this hypothesis is a future work.

Although the subjects who rated the with-physical embodiment condition higher in all level of the transmitting information factor were less than half of all the subjects (14 of 36), there might possibly be a certain bias toward a preference for physical embodiment. A between-subject design avoids such a bias, but there is a problem that requires a lot of subjects to conduct it as described in Section “Questionnaire.” It is a future work to investigate the effect of physical embodiment without the bias.

The first experiment could not show the superiority of humanoid robots to videos, since humanoid robots cannot present the remote partner's current appearance, which can be transmitted by videos. To prove the usefulness of humanoid robot, we had to demonstrate the benefit of both the physical embodiment and the absence of partner's appearance. There are several studies that partly replaced the partner's video with a robot to obtain the positive effects of both of appearance and physical embodiment (Samani et al., 2012; Nakanishi et al., 2014). This is an opposite approach of ours that utilizes one of the features of humanoid robot that is the absence of the partner's appearance. A humanoid robot can pretend as if it is controlled by a remote operator due to not transmitting the appearance. When interacting with the humanoid robot, the user could feel pseudo presence of the remote operator, and the physical embodiment might be able to enhance the pseudo presence as well as social telepresence. The second experiment investigated these predictions.

The second experiment showed that the interaction with a humanoid robot produces the remote partner's pseudo presence that is the feeling of talking with a remote partner when interacting with an autonomous system compared with audio-only interaction. We found that the subjects tended to deduce the presence/absence of a remote partner according to their prior experience with that same remote partner. However, the experience did not work well at the audio-only interaction. We hence considered that the physically embodied body motions might facilitate recalling the partner's presence based on the experience.

We also found that the subjects' deductions of the presence/absence of the remote partner were influenced by their belief about prior experience. In the experience phase, even though the autonomous system gave back-channel responses under the guise of a remote partner, all the subjects believed that the remote partner was listening to their speech. Such a fake experience produced the sense of talking with the partner when talking with an autonomous robot. Nevertheless, the open-ended responses of all the subjects who participated in the with-experience robot condition did not mention the similarity between the experience and autonomous phases. Almost all the subjects focused on whether the back-channel responses were done in appropriate timing. This result implies that the subjects' deductions were subconsciously influenced by the prior experience.

There is a question whether the real experience that a remote partner is actually replying to the user's speech produce higher pseudo presence. Compared with back-channel system, a real partner can give various responses according to the context of conversation. Such a real experience gives a stronger impression that the remote partner is listening, and the impression would effectively produce the pseudo presence. There is another question whether the prior experience produces the pseudo presence when the robot unilaterally speaks to a subject. The user might feel less presence of a remote partner because the robot is unilaterally reproducing talking behaviors and pre-recorded speech like a video message. In this case, it might be difficult to produce pseudo presence, since the factors in determining the presence/absence of the remote partner (e.g., timing of back-channel response) will be less. Answering these questions is future work. In addition, it is also future work to examine whether a user felt the remote partner's pseudo presence through observation data, e.g., observing whether a user replies to the robot's greeting. If he/she felt that the remote partner had been listening/speaking, they might reply to the greeting; if he/she did not feel that way, they might ignore it.

Conclusion

In this study, to prove the usefulness of humanoid robot, we investigated how the features of humanoid robot contribute to produce remote partner's real/pseudo presence. In the first experiment, we compared robot conferencing with existing communication media divided into physical embodiment and transmitting

information factors. As a result, we found that physically embodied body motions enhance the partner's real presence, i.e., social telepresence, although physical embodiment without presenting body motion does not have such an effect. This result shows the superiority of robots to avatars. However, we also found that the partner's appearance, which robots cannot reproduce, enhances social telepresence. Consequently, humanoid robots were comparable to live videos since the positive effect of the physical embodiment offset the negative effect of lacking appearance.

Previous studies have discussed the superiority of humanoid robots to live videos, but our study noted that humanoid robots in the absence of presenting remote partner's appearance do not always have the superiority in social telepresence. Alternatively, this study proposed the utilization of the anonymity of humanoid robot to produce the partner's pseudo presence that is the feeling of talking with a remote partner when interacting with an autonomous robot. In the second experiment, we evaluate whether an autonomous robot produces a similar presence as a teleoperated robot. From the experiment, we found that the prior experience of talking with the remote partner in teleoperation is effective for producing pseudo presence. If a user watched the remote partner's body motion that reproduced by a robot, the user feels the pseudo presence of the partner even while talking with the same robot in autonomous control.

In terms of conveying social telepresence, live videos are more useful than humanoid robots because operating humanoid robots requires higher cost than using displays. We hence conclude that blurring between teleoperation and autonomous control is desirable for effectively utilizing a humanoid robot. Substituting an autonomous system for the remote operator reduces the operator's task, and at the same time, the user could continue to feel the presence of a remote partner also while interacting with an autonomous system.

Acknowledgments

This study was supported by JST CREST "Studies on Cellphone-type Teleoperated Androids Transmitting Human Presence," JSPS KAKENHI Grant Number 26280076 "Robot-Enhanced Displays for Social Telepresence," SCOPE "Studies and Developments on Remote Bodily Interaction Interfaces," and KDDI Foundation Research Grant Program "Robotic Avatars for Human Cloud."

References

- Bailenson, J. N., Yee, N., Merget, D., and Schroeder, R. (2006). The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *Presence* 15, 359–372. doi:10.1162/pres.15.4.359
- Bainbridge, W. A., Hart, J., Kim, E. S., and Scassellati, B. (2011). The benefits of interactions with physically present robots over video-displayed agents. *Int. J. Soc. Robot.* 1, 41–52. doi:10.1007/s12369-010-0082-7
- Bente, G., Ruggenberg, S., Kramer, N. C., and Eschenburg, F. (2008). Avatar-mediated networking: increasing social presence and interpersonal trust in net-based collaborations. *Hum. Commun. Res.* 34, 287–318. doi:10.1111/j.1468-2958.2008.00322.x
- Cao, Y., Tien, W. C., Faloutsos, P., and Pighin, F. (2005). Expressive speech-driven facial animation. *ACM Trans. Graph.* 24, 1283–1302. doi:10.1109/TNN.2002.1021892
- de Greef, P., and Ijsselstein, W. (2001). Social presence in a home tele-application. *Cyberpsychol. Behav.* 4, 307–315. doi:10.1089/109493101300117974
- Finn, K. E., Sellen, A. J., and Wilbur, S. B. (1997). *Video-Mediated Communication*. New Jersey: Lawrence Erlbaum Associates.
- Garau, M., Slater, M., Bee, S., and Sasse, M. A. (2001). "The impact of eye gaze on communication using humanoid avatars," in *Proc. CHI 2001* (Minneapolis: ACM), 309–316.
- Hashimoto, T., Kato, N., and Kobayashi, H. (2011). Development of educational system with the android robot SAYA and evaluation. *Int. J. Adv. Robot. Syst.* 8, 51–61. doi:10.5772/10667
- Isaacs, E. A., and Tang, J. C. (1994). What video can and can't do for collaboration: a case study. *Multimed. Syst.* 2, 63–73. doi:10.1007/BF01274181
- Ishi, C., Liu, C., Ishiguro, H., and Hagita, N. (2012). "Evaluation of formant-based lip motion generation in tele-operated humanoid robots," in *Proc. IROS 2012*. Vilamoura.

- Kang, S., Watt, J. H., and Ala, S. K. (2008). "Communicators' perceptions of social presence as a function of avatar realism in small display mobile communication devices," in *Proc. HICSS 2008*. Waikoloa.
- Kuzuoka, H., Yamazaki, K., Yamazaki, A., Kosaka, J., Suga, Y., and Heath, C. (2004). "Dual ecologies of robot as communication media: thoughts on coordinating orientations and projectability," in *Proc. CHI 2004* (Vieena: ACM), 183–190.
- Le, B. H., Ma, X., and Deng, Z. (2012). Live speech driven head-and-eye motion generators. *IEEE Trans. Vis. Comput. Graph.* 18, 1902–1914. doi:10.1109/TVCG.2012.74
- Lee, J., Wang, Z., and Marsella, S. (2010). "Evaluating models of speaker head nods for virtual agents," in *Proc. AAMAS 2010* (Toronto: IFAAMAS), 1257–1264.
- Lee, K. M., Jung, Y., Kim, J., and Kim, S. R. (2006). Are physically embodied social agents better than disembodied social agents? The effects of physical embodiment, tactile interaction, and people's loneliness in human-robot interaction. *Int. J. Hum. Comput. Stud.* 64, 962–973. doi:10.1016/j.ijhcs.2006.05.002
- Liu, C., Ishi, C. T., Ishiguro, H., and Hagita, N. (2012). "Generation of nodding, head tilting and eye gazing for human-robot dialogue interaction," in *Proc. HRI 2012* (Boston: ACM), 285–292.
- Morita, T., Mase, K., Hirano, Y., and Kajita, S. (2007). "Reciprocal attentive communication in remote meeting with a humanoid robot," in *Proc. ICMI 2007* (Nagoya: ACM), 228–235.
- Nakanishi, H., Kato, K., and Ishiguro, H. (2011). "Zoom cameras and movable displays enhance social telepresence," in *Proc. CHI 2011* (Vancouver: ACM), 63–72.
- Nakanishi, H., Murakami, Y., and Kato, K. (2009). "Movable cameras enhance social telepresence in media spaces," in *Proc. CHI 2009* (Boston: ACM), 433–442.
- Nakanishi, H., Murakami, Y., Nogami, D., and Ishiguro, H. (2008). "Minimum movement matters: impact of robot-mounted cameras on social telepresence," in *Proc. CSCW 2008* (San Diego: ACM), 303–312.
- Nakanishi, H., Tanaka, K., and Wada, Y. (2014). "Remote handshaking: touch enhances video-mediated social telepresence," in *Proc. CHI 2014* (Toronto: ACM), 2143–2152.
- Nguyen, D. T., and Canny, J. (2009). "More than face-to-face: empathy effects of video framing," in *Proc. CHI 2009* (Boston: ACM), 423–432.
- Noguchi, H., and Den, Y. (1998). "Prosody-based detection of the context of backchannel responses," in *Proc. ICSLP 1998*. Sydney.
- Ogawa, K., Nishio, S., Koda, K., Balistreri, G., Watanabe, T., and Ishiguro, H. (2011). Exploring the natural reaction of young and aged person with telenoid in a real world. *J. Adv. Comput. Intell. Inform.* 15, 592–597.
- Ranatunga, I., Torres, N. A., Patterson, R. M., Bugnariu, N., Stevenson, M., and Popa, D. O. (2012). "RoDiCA: a human-robot interaction system for treatment of childhood autism spectrum disorders," in *Proc. PETRA 2012*. Heraklion.
- Sakamoto, D., Kanda, T., Ono, T., Ishiguro, H., and Hagita, N. (2007). "Android as a telecommunication medium with a human-like presence," in *Proc. HRI 2007* (Washington, DC: ACM), 193–200.
- Salvi, G., Beskow, J., Moubayed, S. A., and Granstrom, B. (2009). SynFace: speech-driven facial animation for virtual speech-reading support. *EURASIP J. Audio Speech Music Process.* 2009:191940. doi:10.1155/2009/191940
- Samani, H. A., Parsani, R., Rodriguez, L. T., Saadatian, E., Dissanayake, K. H., and Cheok, A. D. (2012). "Kissenger: design of a kiss transmission device," in *Proc. DIS 2012* (Newcastle: ACM), 48–57.
- Sirkin, D., and Ju, W. (2012). "Consistency in physical and on-screen action improves perceptions of telepresence robots," in *Proc. HRI 2012* (Boston: ACM), 57–64.
- Takeuchi, M., Kitaoka, N., and Nakagawa, S. (2003). "Generation of natural response timing using decision tree based on prosodic and linguistic information," in *Proc. Interspeech 2003*. Geneva.
- Tanaka, K., Nakanishi, H., and Ishiguro, H. (2014). "Comparing video, avatar, and robot mediated communication: pros and cons of embodiment," in *Proc. CollabTech 2014, CCIS 460* (Santiago: Springer), 96–110.
- Tanaka, K., Nakanishi, H., and Ishiguro, H. (2015). Appearance, motion, and embodiment: unpacking avatars by fine-grained communication analysis. *J. Concurr. Comput.* doi:10.1002/cpe.3442
- Tanaka, K., Onoue, S., Nakanishi, H., and Ishiguro, H. (2013). "Motion is enough: how real-time avatars improve distant communication," in *Proc. CTS 2013* (San Diego: IEEE), 465–472.
- Tanaka, M., Ishii, A., Yamano, E., Ogikubo, H., Okazaki, M., Kamimura, K., et al. (2012). Effect of a human-type communication robot on cognitive function in elderly women living alone. *Med. Sci. Monit.* 18, CR550–CR557. doi:10.12659/MSM.883350
- Truong, K. P., Poppe, R., and Heylen, D. (2010). "A rule-based backchannel prediction model using pitch and pause information," in *Proc. Interspeech 2010* (Makuhari: UTPublications), 26–30.
- Ward, N., and Tsukahara, W. (2000). Prosodic features which cue back-channel responses in English and Japanese. *J. Pragmat.* 32, 1177–1207. doi:10.1016/S0378-2166(99)00109-5
- Watanabe, T., Okubo, M., Nakashige, M., and Danbara, R. (2010). InterActor: speech-driven embodied interactive actor. *Int. J. Hum. Comput. Interact.* 17, 43–60. doi:10.1207/s15327590ijhc1701_4

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Tanaka, Nakanishi and Ishiguro. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Expression transmission using exaggerated animation for Elfoid

Maiya Hori*, Yu Tsuruda, Hiroki Yoshimura and Yoshio Iwai

Electrical Engineering and Computer Science, Graduate School of Engineering, Tottori University, Tottori, Japan

We propose an expression transmission system using a cellular-phone-type teleoperated robot called Elfoid. Elfoid has a soft exterior that provides the look and feel of human skin, and is designed to transmit the speaker's presence to their communication partner using a camera and microphone. To transmit the speaker's presence, Elfoid sends not only the voice of the speaker but also the facial expression captured by the camera. In this research, facial expressions are recognized using a machine learning technique. Elfoid cannot, however, display facial expressions because of its compactness and a lack of sufficiently small actuator motors. To overcome this problem, facial expressions are displayed using Elfoid's head-mounted mobile projector. In an experiment, we built a prototype system and experimentally evaluated its subjective usability.

OPEN ACCESS

Edited by:

Shuichi Nishio,
Advanced Telecommunications
Research Institute International, Japan

Reviewed by:

Marco Fyfe Pietro Gillies,
Goldsmiths, University of London, UK
Egon L. Van Den Broek,
Utrecht University, Netherlands

*Correspondence:

Maiya Hori,
Electrical Engineering and Computer
Science, Graduate School of
Engineering, Tottori University, 101
Minami 4-chome, Koyama-cho,
Tottori 680-8550, Japan
maiya-h@ieee.org

Specialty section:

This article was submitted to
Human-Media Interaction,
a section of the journal
Frontiers in Psychology

Received: 26 February 2015

Accepted: 03 August 2015

Published: 19 August 2015

Citation:

Hori M, Tsuruda Y, Yoshimura H and
Iwai Y (2015) Expression transmission
using exaggerated animation for
Elfoid. *Front. Psychol.* 6:1219.
doi: 10.3389/fpsyg.2015.01219

Keywords: teleoperated robot, human robot interaction, communication robot, mobile phone, expression transmission

1. Introduction

Robots that have a human appearance have been developed to communicate with people in remote locations. Some studies have used humanoid robots for the transmission of human presence. In particular, teleoperated android robots such as Geminoid F and Geminoid HI-1 (Asano et al., 2010) have appearances similar to an actual person, and were intended to substitute for the presence of actual people. These humanoid robots have high degrees of freedom and can transmit human presence effectively. However, they are expensive and limited to a specific individual target. A robot called Telenoid R1 (Ogawa et al., 2011) was developed to reduce the cost and the number of actuators. Telenoid is not limited to a specific individual target, and is designed to immediately appear as a human. A person can easily recognize Telenoid as human; it can be interpreted as male or female, and old or young. With this design, Telenoid allows people to feel as if a distant acquaintance is next to them. This makes it possible to transmit human presence. Moreover, Telenoid's soft skin and child-like body size make it easy to hold. However, it is difficult to carry around in daily life.

For daily use, a communication medium that is smaller than Telenoid and uses mobile-phone communication technology is now under development. Like a cellular phone, Elfoid is easy to hold in the hand, as shown in **Figure 1**. By combining voice-based conversation with an appearance and touch that is capable of effectively communicating an individual's presence, the user can feel as if they are conversing in a natural fashion with someone directly in front of them (Tanaka et al., 2015). Additionally, when we use such robots for communication, it is important to convey the facial expressions of a speaker to increase the modality of communication (Mehrabian, 1968). If the speaker's facial movements are accurately represented via these robots, human presence can be conveyed. Elfoid has a camera within its chest and the speaker's facial movements are estimated by conventional face-recognition approaches. However, it is difficult to generate the same expression in robots because a large number of actuators are required. Elfoid cannot produce facial expressions



FIGURE 1 | Elfoïd: cellular phone-type teleoperated android.

like a human face can, because it has a compact design that cannot be intricately activated. That is, since Elfoïd's design priority is portability, its modality of communication is less than Telenoid's. For this reason, it is necessary to convey facial expressions some other way.

In the proposed system, facial expressions are displayed using Elfoïd's head-mounted mobile projector. Our hypothesis in this study is that subjective usability, which is composed of satisfaction with the conversation, the impression of the conversational partner, and an impression of the interface, will be improved by adding facial expressions to Elfoïd. In the experiments, we verify whether this hypothesis is correct.

2. Materials and Methods

2.1. Elfoïd: Cellular Phone-type Teleoperated Android

Elfoïd is used as a cellular phone for communication. To convey a human presence, Elfoïd has the following functions. Elfoïd has a body that is easy to hold in a person's hand. The size is about 20 cm. Elfoïd's design is recognizable at first glance to be human-like and can be interpreted equally as male or female, and old or young. Elfoïd has a soft exterior that provides the feel of human skin. Elfoïd is equipped with a camera and microphone in the chest. Additionally, a mobile projector (MicroVision, Inc. SHOWWX+ HDMI) with a mirror is mounted in Elfoïd's head and facial expressions are generated by projecting images from within the head, as shown in **Figure 2**.

2.2. Overview of the Total System

In this research, facial expressions are generated using Elfoïd's head-based mobile projector.



FIGURE 2 | Elfoïd with a mobile projector to convey the facial expressions of a communication partner.

First, individual facial images are captured using a camera mounted within Elfoïd. Next, the facial region is detected in each captured image and feature points on the face are tracked using the Constrained Local Model (CLM) (Saragih et al., 2011). Facial expressions are recognized by a machine-learning technique using the positions of the feature points. Finally, recognized facial expressions are reproduced using Elfoïd's head-based mobile projector.

2.3. Recognition of Facial Expressions

Face tracking techniques that use feature points such as the corners of the eyes and mouth are effective for the recognition of facial expressions because a face is a non-rigid object. Part-based models use local image patches around the landmark points. The part-based model CLM (Saragih et al., 2011) outperforms holistic models in terms of landmark localization accuracy. CLM fitting is the search for point distribution model parameters p that jointly minimize the misalignment error over all feature points. It is formulated as follows:

$$\mathcal{Q}(p) = \mathcal{R}(p) + \sum_{i=1}^n \mathcal{D}_i(x_i; \mathcal{I}), \quad (1)$$

where \mathcal{R} is a regularization term and \mathcal{D}_i denotes the measure of misalignment for the i th landmark at x_i in image \mathcal{I} . In the CLM framework, the objective is to create a shape model from the parameters p . The CLM models the likelihood of alignment at a particular landmark location x . The likelihood of alignment at x is acquired beforehand using the local features of a large number of images. To generate the classifier, Saragih et al. (2011) use logistic regression. Mean-shift vectors from each landmark are computed using the likelihood of alignment, and the parameters p are updated. These processes are iterated until parameters p converge. This method has low computational complexity and is robust to occlusion.

In communication between people, it is important to convey the emotions of the speaker. There have been a considerable number of studies of basic human emotions. Ekman et al. (2002) defined basic facial expressions consisting of anger, disgust, fear, happiness, sadness, and surprise. This shows that these facial expressions are not culturally determined, but are

universal across all human cultures and are thus biological in origin. In this study, six facial expressions that correspond to universal emotions (Ekman et al., 2002)—anger, disgust, fear, happiness, sadness, and surprise—are classified using a hierarchical technique similar to Siddiqi et al. (2013). The facial expressions are hierarchically classified by a Support Vector Machine (SVM). Each classifier is implemented beforehand using the estimated positions of feature points. This study is based on the theory that different expressions can be grouped into three categories (Schmidt and Cohn, 2001; Nusseck et al., 2008) based on the parts of the face that contribute most toward the expression. At the first level, we use 31 feature points around the mouth, eyes, eyebrows, and nose to discriminate the three expression categories: lip-based, lip-eye-based, and lip-eye-eyebrow-based. After the expressions are grouped into the three categories, each category is divided into two emotion classes. In the lip-based category, four feature points around the mouth are used for expressing happiness or sadness. In the lip-eye-based category, 16 feature points around the mouth and eyes are used for expressing surprise or disgust. In the lip-eye-eyebrow-based category, 26 feature points around the mouth and eyes and eyebrows are used for expressing anger or fear.

2.4. Generation of Facial Expressions with Elfoid Using Cartoon Techniques

Recognized facial expressions were reproduced using Elfoid's head-based mobile projector. To represent facial expressions, we generated three projection patterns using the results of emotion estimations. In this study, the projection patterns were stylized using animation techniques (Thomas and Johnston, 1995). It is widely recognized that cartoons are very effective at expressing emotions and feelings. The movements around the mouth and eyebrows were exaggerated. Moreover, color stimuli that convey a particular emotion were added. The three projection patterns are as follows.

2.4.1. With Exaggerated Motion

As an exaggeration of a simple cartoon effect, the movement of the eyes and mouth were exaggerated. The parts required for the movement were determined using the Facial Action Coding System (FACS) (Ekman and Friesen, 1978) that describes the relationships between the emotions and movements of the facial action units. The exaggerated motions added to each facial expression are as follows:

- Anger: brow lowering, upper lid raising, lid tightening, and lip tightening.
- Disgust: nose wrinkling, lip corner depressing, and lower lip depressing.
- Fear: inner brow raising, outer brow raising, brow lowering, upper lid raising, lid tightening, lip stretching, and jaw dropping.
- Happiness: cheek raising and lip corner pulling.
- Sadness: inner brow raising, brow lowering, and lip corner depressing.
- Surprise: inner brow raising, outer brow raising, upper lid raising, and jaw dropping.

2.4.2. With Exaggerated Motion and Color

Color stimuli that convey a particular emotion were added. As in the examples described in Thomas and Johnston (1995), colors were added to the upper part of the face. The color stimuli used in Fujie et al. (2013) are adapted in this study. The colors added to each facial expression were as follows.

- Anger: red.
- Disgust: purple.
- Fear: blue-green.
- Happiness: orange.
- Sadness: blue.
- Surprise: yellow.

Here, these show the colors of Elfoid after projecting images and color calibration had already been performed. Images that use exaggerated motion and color to express the six universal emotions are shown in **Figure 3**.

2.4.3. With Exaggerated Motion and Marks

To investigate the effects of a mark corresponding to each emotion, marks were added to the face. The marks added to each facial expression were as follows.

- Anger: cross-shaped anger sign.
- Disgust: vertical stripes over one side of the face.
- Fear: vertical stripes over the entire face.
- Happiness: blushing cheeks.
- Sadness: tears.
- Surprise: colored highlights in the eyes.

Images that used exaggerated motion and marks to express of the six universal emotions are shown in **Figure 4**. Here, the positions at which the marks were added were determined from candidate positions on the face considering the marks' visibility.

Facial expressions are caused by the movement of muscles in the face. In this study, to create the expression animation, morphing technology was applied. All animation was empirically morphed for 2 s at 30 fps.

2.5. Experiments

In the experiments, we built a prototype system. First, we verified the performance of the prototype system. Next, the subjective evaluations of usability were investigated. It should be confirmed that the study was conducted in accordance with the ethical principles that have their origins in the Declaration of Helsinki.

2.5.1. Recognition Rate of Communication Partner Facial Expressions

We conducted an experiment to verify the recognition rate of facial expressions. In this experiment, Elfoid was used as a cellular phone for communication. Here, assuming that the number of users per Elfoid is limited, we collected the images of one user. We asked the user to display the six basic expressions, assuming a conversation with a person in a remote location. Each expression was captured by the camera (Logicoool HD Pro Webcam C920t) at a resolution of 1024 × 768 pixels in

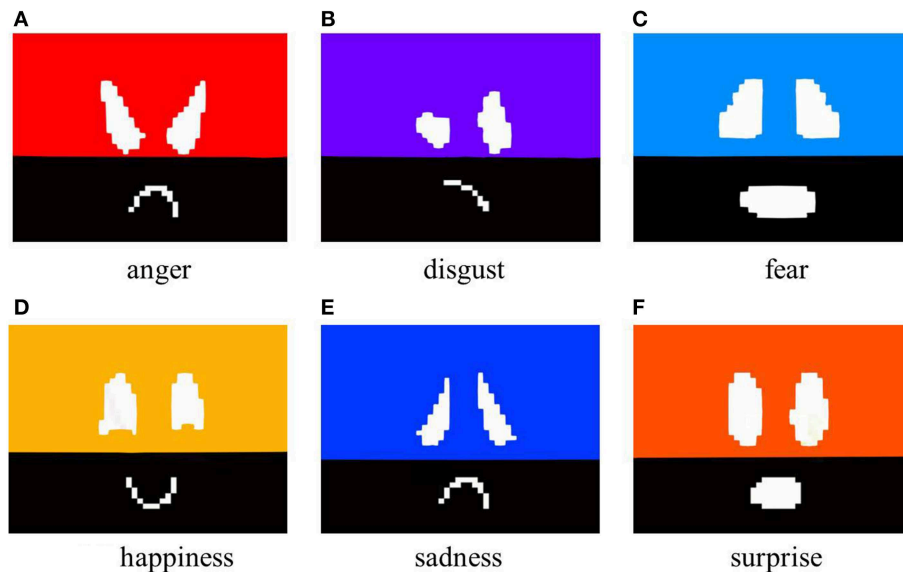


FIGURE 3 | Generated facial images with exaggerated motion and color.

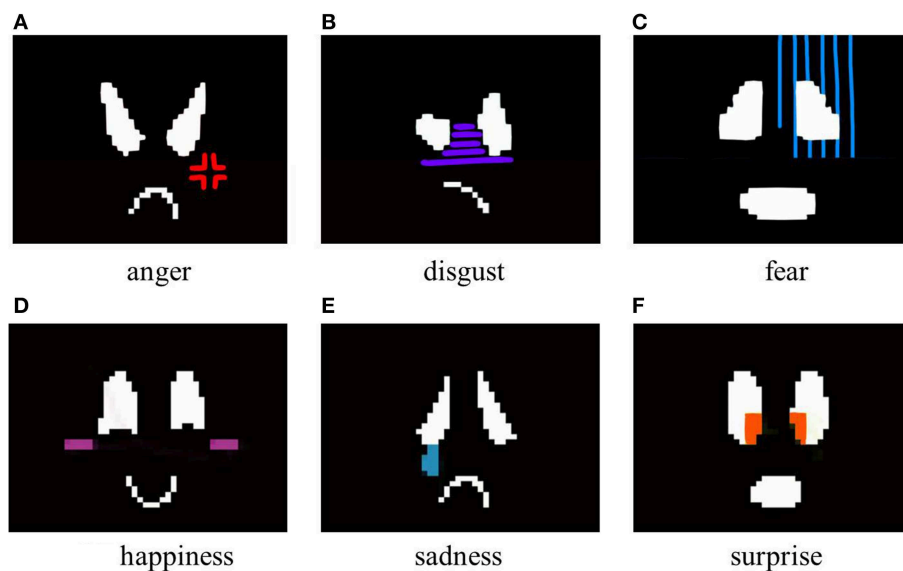


FIGURE 4 | Generated facial images with exaggerated motion and marks.

an indoor environment. The user was asked to sit down in front of the camera and look at it without any constraints on their head movement. Only if the face of the user was out of the field of view of the camera was the user asked to move back into the field of view. The distance between the user and Elfoid was about 30 cm and was not fixed. As training data, we used a total of 8000 images that consisted of 1000 images for each facial expression and 2000 images with no expression. To verify the rate of facial expression recognition, we tested 1000 images for each expression that were different from the training data.

2.5.2. Evaluation of the Accuracy of Emotion Conveyance

In this experiment, to investigate the accuracy of the emotions that are conveyed using Elfoid, various projection patterns were presented to 10 participants (all in their 20s, 8 male and 2 female). We made the participants sit down on a chair and hold the Elfoid in their hands about 30 cm away from their face. Three projection patterns for each emotion, a total of 18 patterns, were displayed to the participants in random order. **Figure 5** shows the facial expressions generated with exaggerated motion and color. **Figure 6** shows the facial expressions generated with exaggerated

motion and marks. To eliminate the influence of environmental disturbances, the experiments were conducted at a particular brightness (measured value: 190 lx). After each presentation, each participant was asked to rate the emotions perceived in the facial expressions of Elfoid. Each emotion was rated from 1 (not felt at all) to 6 (felt extremely strongly). These processes were repeated until all patterns were investigated.

2.5.3. Subjective Usability Evaluation of the Proposed System

Additionally, to verify the validity of this system, we experimentally evaluated its subjective usability. One of

the participants who have a conversation uses Elfoid on communication. Again, we made the participants sit down on a chair and hold Elfoid in their hands about 30 cm away from their face. They then had a conversation with a communication partner at a remote location as they watched Elfoid. To eliminate the influence of disturbances, these experiments were also conducted at a particular brightness (measured value: 190 lx). We used the results of the previous experiments that are specifically described in the Discussion Section to express emotions. As a comparison with the proposed method, we used other two methods. One used an Elfoid whose facial expression was not projected and another used an Elfoid whose

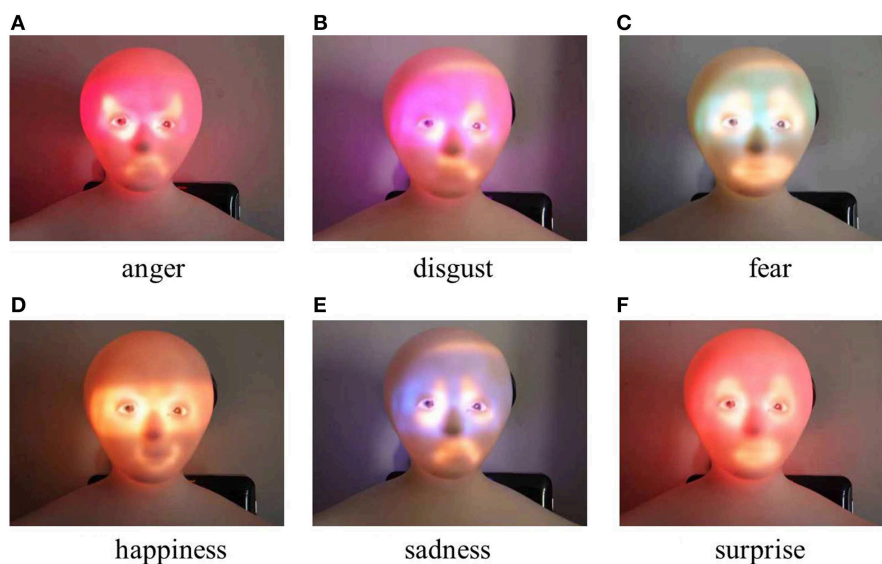


FIGURE 5 | Generated facial expressions with exaggerated motion and color.

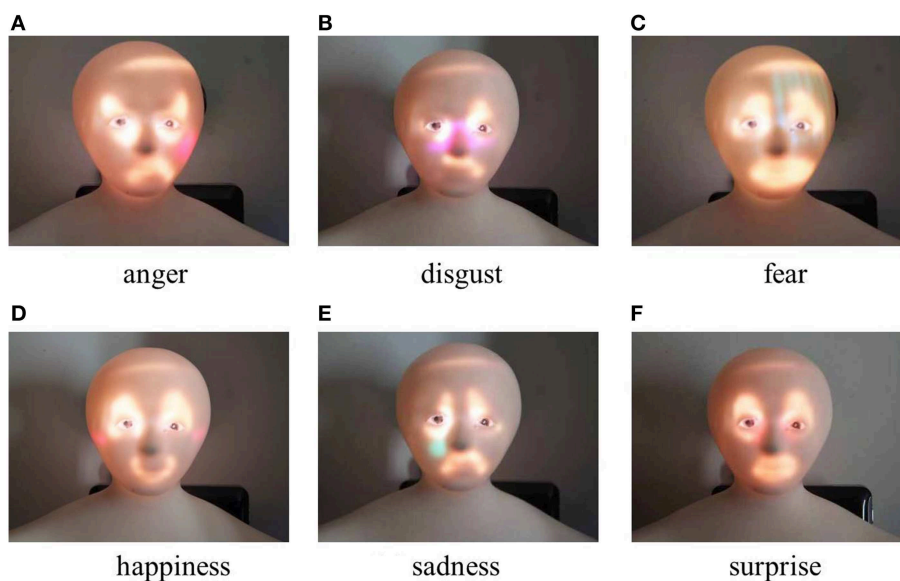


FIGURE 6 | Generated facial expressions with exaggerated motion and marks.

facial expression was generated in a random manner without considering the facial expression of the speaker. We conducted the experiments on 18 participants (all in their 20s, 17 male and 1 female, some of whom participated in both experiments) by changing the methods in random order. We did not tell the participants which emotion was indicated by the presented facial expressions beforehand. In this experiment, to eliminate the influence of false recognition of the speaker's facial expression, the facial expression was recognized manually and the facial expression of Elfoid was generated in real time. We took the delay into consideration as much as possible so that it would be minimized.

Themes of conversation were determined in reference to conventional research (Hara et al., 2014), and those used in this experiment are shown below.

- Who makes more money, men or women?
- Do you think that friendship is possible between men and women?
- Which do you prefer, dogs or cats?
- Should the possession of guns be allowed in Japan?
- Which is more important in the opposite sex, physical attractiveness or personality?
- Do you believe in supernatural powers and hypnosis?

We then gave the participants questionnaires that asked about their level of satisfaction with the conversation, their impression of the conversational partner, and their impression of the interface. After the end of 5 min of conversation, for each condition, participants answered a portion of the questionnaire. In the conversation, participants showed some facial expressions at rates that were not equal. Fifteen items were used to evaluate the proposed system from various viewpoints. The details of the questionnaire items are shown below. Each questionnaire item was used with reference to Sakamoto et al. (2007) and Matsuda et al. (2012).

The impression of the conversation was rated on a scale of 1–8.

1. It was possible to have the conversation cooperatively.
2. It was difficult to make conversation.
3. It was possible to talk while having an interest in each other.

The impression of the communication partner was rated on a scale of 1–7. Each questionnaire item is shown below.

1. Bad impression—good impression.
2. Not serious—serious.
3. Unreliable—reliable.
4. Unhealthy—healthy.
5. Introvert—extrovert.
6. Difficult to talk to—easy to talk to.
7. Their story was poor—their story was good.

The impression of having a conversation with Elfoid was rated on a scale of 1–7, where 7 is the most positive. The items are listed as follows:

1. Presence: presence of the person that a participant feels during a conversation.
2. Humanlike: human likeness of Elfoid's appearance, movements, and behavior.
3. Naturalness: naturalness of Elfoid's appearance, movements, and behavior.
4. Uncanny: uncanniness of Elfoid's appearance, movements, and behavior.
5. Responsiveness: responsiveness of Elfoid to the participant's behavior and conversation.

These questions were asked repeatedly until all methods were investigated.

3. Results

3.1. Recognition Rate of Facial Expressions of the Communication Partner

Table 1 lists the facial expression recognition rates. The accuracy of the facial expression recognition was 83.8% on average.

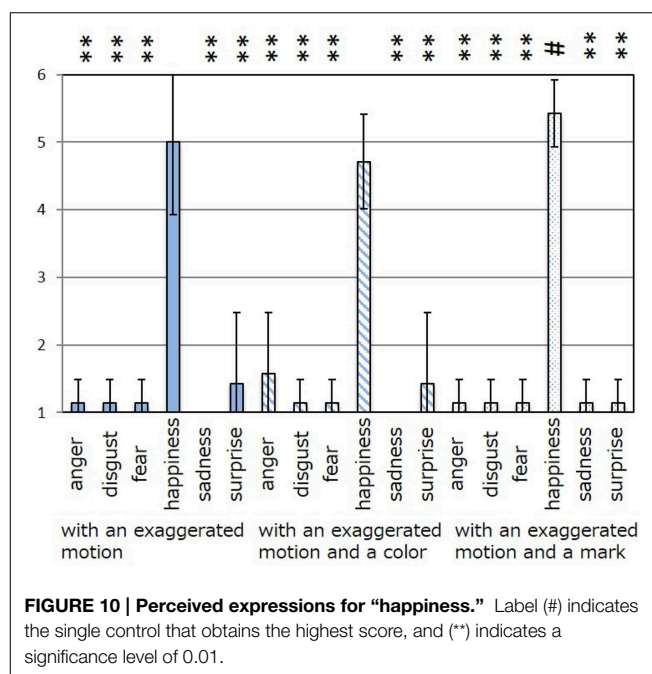
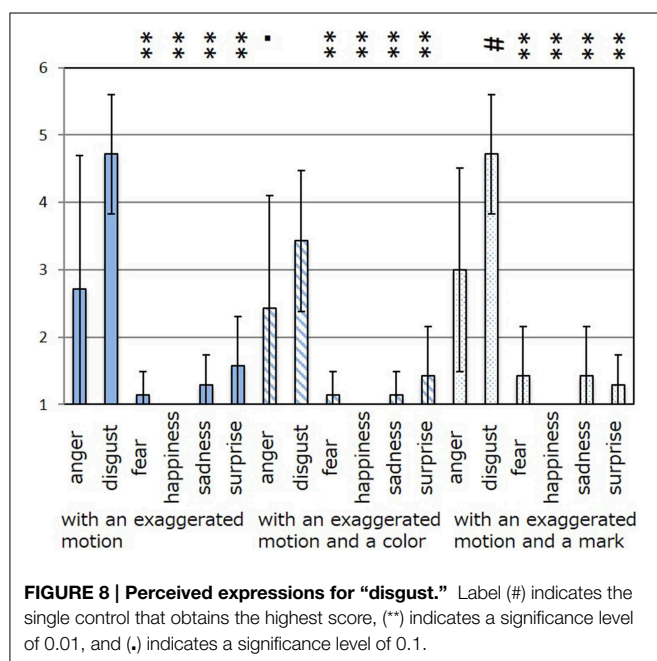
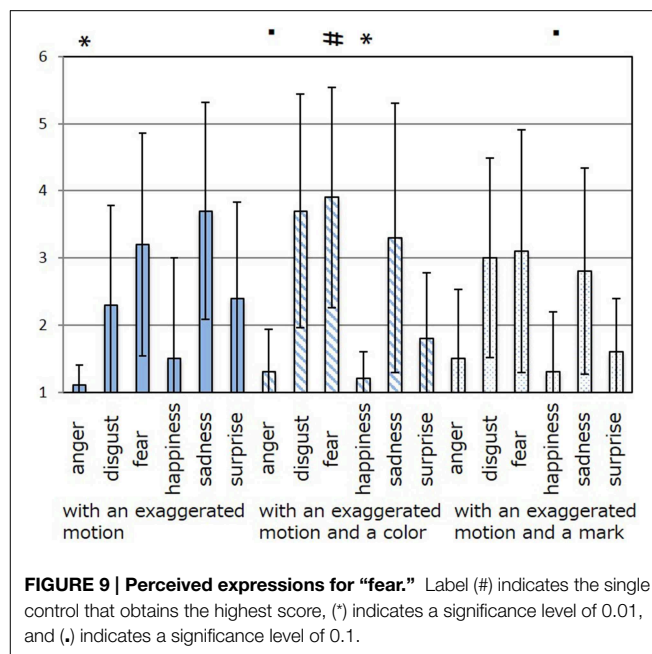
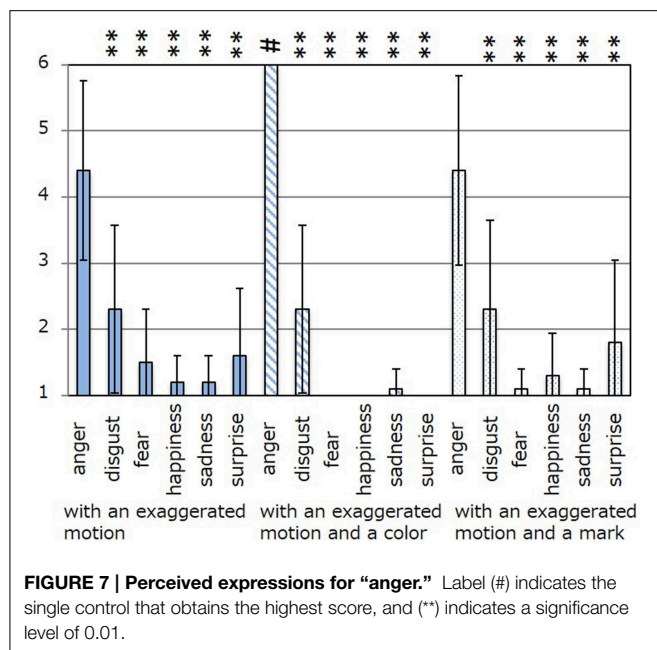
3.2. Evaluation of the Accuracy of Emotion Conveyance

The results of the subjective evaluation process for each facial expression are shown in Figures 7–12. The data shown in these figures are the average score and standard variation of the subjective evaluation. We assume that the population is normally distributed. Bartlett's test was used to check the equality of variances, however, we found that the variances of the results were not the same. Therefore, Dunnett's T3 test was used to compare the average scores. In Figures 7–12, (#) indicates

TABLE 1 | Recognition rate of facial expressions of the communication partner (%).

Truth \ Estimation	Anger	Fear	Disgust	Happiness	Sadness	Surprise	No expression
Anger	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Disgust	17.7	82.3	0.0	0.0	0.0	0.0	0.0
Fear	0.2	0.0	97.9	0.0	0.0	1.9	0.0
Happiness	23.2	0.0	0.0	73.8	3.0	0.0	0.0
Sadness	33.9	0.0	0.0	1.4	60.1	0.0	4.6
Surprise	17.2	0.0	1.9	0.0	0.0	82.8	0.0
No expression	10.4	0.0	0.0	0.0	0.0	0.0	89.6

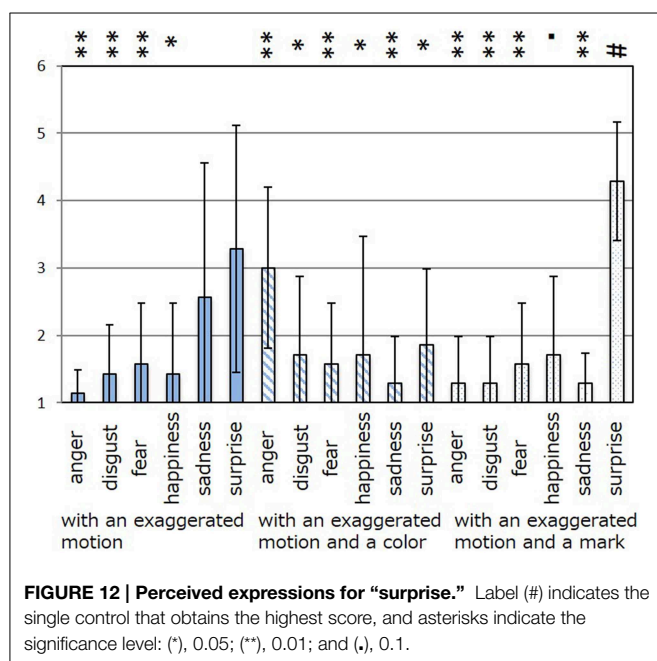
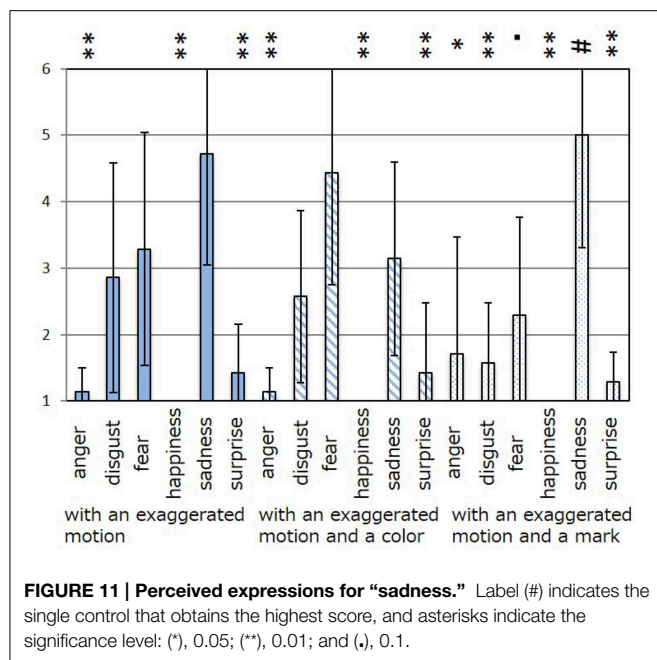
Bold values mean the proportion of correctly recognized facial expressions.



the single control that obtains the highest score, and asterisks indicate the significance level: (*), 0.05; (**), 0.01; and (.), 0.1.

Figure 7 shows the scores obtained when the “anger” expression was generated. The highest score was observed for the emotion “anger” for all patterns of expression. The average score of the participants was 6.00, and Dunnett’s T3 test indicated a significant difference between the score for “anger” and the scores for all other emotions. However, significant differences were not observed between the score for “exaggerated motion and color” and other patterns, as shown in **Figure 7**. With respect to the expression of “disgust,” shown in **Figure 8**, the highest average score given by participants was 4.71, and Dunnett’s T3

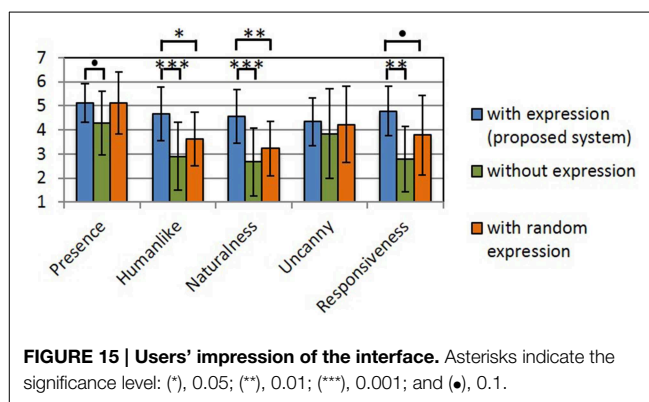
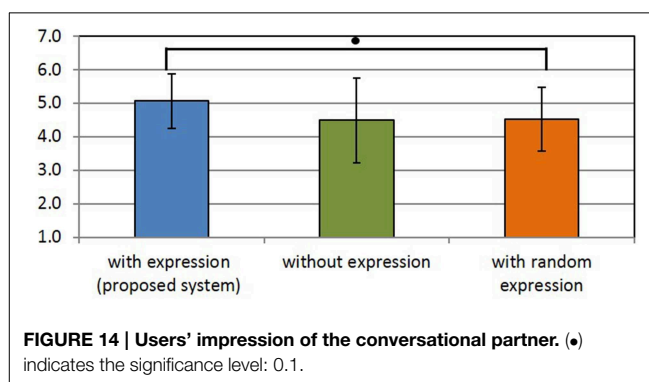
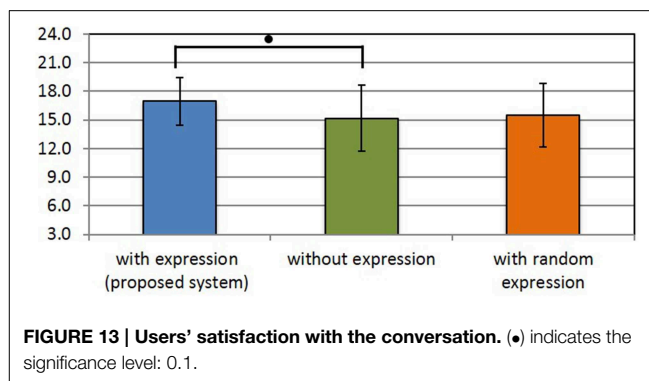
test indicates that there was no significant difference between the scores for “disgust” and “anger.” **Figure 9** shows the least successful case of emotion conveyance, that when “fear” was the intended emotion and the fearful expression generated “with exaggerated motion and color” was displayed. The highest score was observed for the emotion “fear.” However, the average score of the participants was 3.90, and Dunnett’s T3 test indicates that there was no significant difference among the scores for “fear,” “disgust,” and “sadness.” It is also difficult to transmit the expression of fear using the other two expression patterns. This is because the emotional expression of the eyes was close to



that of “sadness,” and a negative emotion was derived from that fact. With respect to “happiness,” “sadness,” and “surprise,” the intended emotion can be conveyed as well as “anger,” as shown in **Figures 10–12**.

3.3. Evaluation of the Subjective Usability of the Proposed System

Figures 13–15 show the experimental results. In each figure, higher scores indicate a better impression. We assume that the population is normally distributed. We found that all scores except for the uncanny score have the same variance. Therefore,



we used Dunnett's test for comparison. In **Figures 13–15**, and asterisks indicate the significance level: (*), 0.05; (**), 0.01; (***), 0.001; and (•), 0.1.

Figure 13 shows the results of the satisfaction level of the conversation. The satisfaction level of the conversation is calculated as the sum of the three items listed in Section 2.5.3, similarly to the previous method in Fujiwara and Daibo (2010). The results show that there is significant difference between the scores for the proposed method and the comparison method that used an Elfoid without any projected facial expression. The proposed system was expected to improve the satisfaction level of the conversation by adding facial expressions.

Figure 14 shows the results for the impression of the communication partner. This impression was calculated as the

average of the scores of seven items, similarly to the previous method in Matsuda et al. (2012). The results show that there was a significant difference between the scores for the proposed method and the comparison method that generated random Elfoid expressions. It was found that the impression of the communication partner could be decreased when Elfoid's facial expression was randomly generated.

Figure 15 shows the results of the impression of the interface. With respect to presence, there was significant difference between the scores for the proposed method and the comparison method that projected no facial expression. With respect to humanlike, naturalness, and responsiveness, there was a significant difference between the scores for the proposed method and the other comparison methods. The proposed system was expected to improve impressions with respect to presence, humanlike attributes, naturalness, and responsiveness. However, the impression of presence was improved even when Elfoid's facial expressions were generated in a random manner.

4. Discussion

The accuracy of the facial expression recognition was 83.8% on average, as described in Section 2.5.1. These results seem to indicate that the accuracy of the proposed method is lower than that of state-of-the-art emotion recognition methods (Janssen et al., 2013). However, in our experiments, the user can move freely, so facial images are not always aligned. To align the facial images, we use one state-of-the-art method for facial alignment, CLM (Saragih et al., 2011). The problem tackled in this study is more difficult than the problem in Janssen et al. (2013), so it is not necessarily true that our method is inferior to this method.

The animation patterns that can efficiently convey an intended emotion are shown as follows.

- Anger: with an exaggerated motion and color.
- Disgust: with an exaggerated motion and a mark (similarly, “with an exaggerated motion”).
- Fear: with an exaggerated motion and color (however, “sadness” and “disgust” are conveyed co-instantaneously with “fear”).
- Happiness: with an exaggerated motion and a mark (similarly, “with an exaggerated motion” and “with an exaggerated motion and color”).
- Sadness: with an exaggerated motion and a mark.
- Surprise: with an exaggerated motion and a mark.

By using the patterns described here there is a high likelihood of transmitting the intended emotion and a low likelihood

of transmitting other emotions. Five facial expressions can be conveyed as the intended emotion. In contrast, a fearful expression cannot be easily conveyed this way. In the case of the fearful expression, negative emotions, such as “sadness” and “disgust” are conveyed co-instantaneously with “fear.” Some studies (Sugano and Ogata, 1996; Ariyoshi et al., 2004; Fujie et al., 2013) have used colors for communication between humans and robots. In comparison with these studies, the face generated by the proposed system is more expressive. Moreover, the proposed method may be able to transmit other emotions (Prinz, 2004) used in conversation. As future work, there is a need to investigate whether it is necessary to transmit other emotions during communication that uses Elfoid.

According to the results of the subjective usability, which is composed of satisfaction with the conversation, an impression of the conversational partner, and an impression of the interface, we found that the subjective usability was improved by adding facial expressions to Elfoid. In particular, by comparing the results of the proposed method with those for randomly generated facial expressions, we determined that the combination of accurate facial recognition of a speaker and the appropriate facial expression of Elfoid is an efficient way to improve its subjective usability.

5. Conclusion

We propose an expression transmission system using a cellular phone-type teleoperated robot with a mobile projector. In this research, facial expressions are recognized using a machine learning technique, and displayed using a mobile projector installed in Elfoid's head to convey emotions. In the experiments, we built a prototype system that generated facial expressions and evaluated the recognition rate of the facial expressions and the subjective evaluations of usability. Given the results, we can conclude that the proposed system is an effective way to improving the subjective usability. For practical use, it will be necessary to realize a stable recognition process that uses relatively little of Elfoid's computing resources. To overcome this problem, we plan to use cloud computing technology.

Acknowledgments

This research was supported by the JST CREST (Core Research for Evolutional Science and Technology) research promotion program “Studies on cellphone-type teleoperated androids transmitting human presence.”

References

- Ariyoshi, T., Nakadai, K., and Tsujino, H. (2004). “Effect of facial colors on humanoids in emotion recognition using speech,” in *International Workshop on Robot and Human Interactive Communication* (Okayama), 59–64.
- Asano, C. B., Ogawa, K., Nishio, S., and Ishiguro, H. (2010). “Exploring the uncanny valley with geminoid HI-1 in a real-world application,” in *Proceedings International Conference of Interfaces and Human Computer Interaction* (Freiburg), 121–128.
- Ekman, P., and Friesen, W. (1978). *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. California, CA: Consulting Psychologists Press.
- Ekman, P., Friesen, W. V., and Hager, J. C. (2002). *Facial Action Coding System (FACS)*. Salt Lake City; London: Research Nexus eBook; Weidenfeld & Nicolson.

- Fujie, Y., Hori, M., Yoshimura, H., and Iwai, Y. (2013). "Emotion transmission by color effects for a teleoperated mobile communication robot," in *Proceedings HRI2013 Workshop on Design of Humanlikeness in HRI from Uncanny Valley to Minimal Design* (Tokyo).
- Fujiwara, K., and Daibo, I. (2010). The function of positive affect in a communication context: satisfaction with conversation and hand movement (in Japanese). *Jpn. J. Res. Emot.* 17, 180–188. doi: 10.4092/jsre.17.180
- Hara, K., Hori, M., Takemura, N., Iwai, Y., and Sato, K. (2014). Construction of an interpersonal interaction system using a real image-based avatar (in Japanese). *IEEE Trans. Electr. Inform. Syst.* 134, 102–111. doi: 10.1541/ieejieiss.134.102
- Janssen, J. H., Tacke, P., de Vries, J. J. G., van den Broek, E. L., Westerink, J. H. D. M., Haselager, P., et al. (2013). Machines outperform lay persons in recognizing emotions elicited by autobiographical recollection. *Hum. Comput. Inter.* 28, 479–517. doi: 10.1080/07370024.2012.755421
- Matsuda, M., Yaegashi, K., Daibo, I., Mikami, D., Kumano, S., Otuka, K., et al. (2012). *An Exploratory Experimental Study on Determinants of Interpersonal Impression Among Video Communication Users: Comparison of the Video and Audio Stimuli (in Japanese)*. Technical Report of IEICE, HCS 111, 49–54.
- Mehrabian, A. (1968). Communication without words. *Psychol. Today* 2, 52–55.
- Nussek, M., Cunningham, D. W., Wallraven, C., and Bülthoff, H. H. (2008). The contribution of different facial regions to the recognition of conversational expressions. *J. Vis.* 8, 1–23. doi: 10.1167/8.8.1
- Ogawa, K., Nishio, S., Koda, K., Balistreri, G., Watanabe, T., and Ishiguro, H. (2011). Exploring the natural reaction of young and aged person with Telenoid in a real world. *J. Adv. Comput. Intell. Intell. Informat.* 15, 592–597.
- Prinz, J. (2004). "Which emotions are basic?" in *Emotion, Evolution, and Rationality*, eds D. Evans and P. Cruse (Oxford, UK: Oxford University Press), 69–87.
- Sakamoto, D., Kanda, T., Ono, T., and Ishiguro, N. H. H. (2007). Android as a telecommunication medium with a human-like presence. *Proceedings of the ACM/IEEE International Conference on Human-robot Interaction* (Washington, DC), 193–200.
- Saragih, J. M., Lucey, S., and Cohn, J. F. (2011). Deformable model fitting by regularized landmark mean-shift. *Int. J. Comput. Vis.* 91, 200–215. doi: 10.1007/s11263-010-0380-4
- Schmidt, K. L., and Cohn, J. F. (2001). Human facial expressions as adaptations: evolutionary questions in facial expression research. *Am. J. Phys. Anthropol.* 116, 3–24. doi: 10.1002/ajpa.20001
- Siddiqi, M. H., Lee, S. L. Y., Khan, A. M., and Truc, P. T. H. (2013). Hierarchical recognition scheme for human facial expression recognition system. *Sensors* 13, 16682–16713. doi: 10.3390/s131216682
- Sugano, S., and Ogata, T. (1996). "Emergence of mind in robots for human interface - research methodology and robot model," in *IEEE International Conference Robotics and Automation* (Minnesota), 1191–1198.
- Tanaka, K., Nakanishi, H., and Ishiguro, H. (2015). Physical embodiment can produce robot operator's pseudo presence. *Front. ICT* 2:8. doi: 10.3389/fict.2015.00008
- Thomas, F., and Johnston, O. (1995). *The Illusion of Life: Disney Animation*. Disney Editions.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Hori, Tsuruda, Yoshimura and Iwai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Attitudinal Change in Elderly Citizens Toward Social Robots: The Role of Personality Traits and Beliefs About Robot Functionality

Malene F. Damholdt^{1,2*}, Marco Nørskov^{1,3}, Ryuji Yamazaki³, Raul Hakli¹, Catharina Vesterager Hansen¹, Christina Vestergaard¹ and Johanna Seibt¹

¹ Department of Philosophy and the History of Ideas, School of Culture and Society, Aarhus University, Aarhus, Denmark,

² Unit for Psychooncology and Health Psychology, Department of Oncology, Aarhus University Hospital and Department of Psychology & Behavioural Science, Aarhus University, Aarhus, Denmark, ³ Hiroshi Ishiguro Laboratories, Advanced Telecommunications Research Institute International, Osaka, Japan

OPEN ACCESS

Edited by:

Tsutomu Fujinami,
Japan Advanced Institute of Science
and Technology, Japan

Reviewed by:

Carryl L. Baldwin,
George Mason University, USA
Hatice Gunes,
Queen Mary University of London, UK

*Correspondence:

Malene F. Damholdt
malenefd@psy.au.dk

Specialty section:

This article was submitted to
Human-Media Interaction,
a section of the journal
Frontiers in Psychology

Received: 23 June 2015

Accepted: 22 October 2015

Published: 20 November 2015

Citation:

Damholdt MF, Nørskov M,
Yamazaki R, Hakli R, Vesterager
Hansen C, Vestergaard C and Seibt J
(2015) Attitudinal Change in Elderly
Citizens Toward Social Robots:
The Role of Personality Traits
and Beliefs About Robot Functionality.
Front. Psychol. 6:1701.
doi: 10.3389/fpsyg.2015.01701

Attitudes toward robots influence the tendency to accept or reject robotic devices. Thus it is important to investigate whether and how attitudes toward robots can change. In this pilot study we investigate attitudinal changes in elderly citizens toward a tele-operated robot in relation to three parameters: (i) the information provided about robot functionality, (ii) the number of encounters, (iii) personality type. Fourteen elderly residents at a rehabilitation center participated. Pre-encounter attitudes toward robots, anthropomorphic thinking, and personality were assessed. Thereafter the participants interacted with a tele-operated robot (Telenoid) during their lunch (c. 30 min.) for up to 3 days. Half of the participants were informed that the robot was tele-operated (IC) whilst the other half were naïve to its functioning (UC). Post-encounter assessments of attitudes toward robots and anthropomorphic thinking were undertaken to assess change. Attitudes toward robots were assessed with a new generic 35-items questionnaire (attitudes toward social robots scale: ASOR-5), offering a differentiated conceptualization of the conditions for social interaction. There was no significant difference between the IC and UC groups in attitude change toward robots though trends were observed. Personality was correlated with some tendencies for attitude changes; Extraversion correlated with positive attitude changes to *intimate-personal relatedness* with the robot ($r = 0.619$) and to *psychological relatedness* ($r = 0.581$) whilst Neuroticism correlated negatively ($r = -0.582$) with *mental relatedness* with the robot. The results tentatively suggest that neither information about functionality nor direct repeated encounters are pivotal in changing attitudes toward robots in elderly citizens. This may reflect a cognitive congruence bias where the robot is experienced in congruence with initial attitudes, or it may support action-based explanations of cognitive dissonance reductions, given that robots, unlike computers, are not yet perceived as action targets. Specific personality traits may be indicators of attitude change relating to specific domains of social interaction. Implications and future directions are discussed.

Keywords: social robots, attitudes toward social robots, personality, anthropomorphism, human-robot interaction

INTRODUCTION

Robotics envisage that by 2020 robotics technology will “influence every aspect of work and home.”¹ According to official projections, by 2025 the market value of robotics will expand to several trillion US\$ per year, mainly due to social robotics, which will be outperforming industrial robotics by a large margin.²

Despite these advances the vast majority of residents in the European Community (87% of 26,751 respondents; Public Attitudes Towards Robots, 2012; Special Eurobarometer 382) has of yet no personal experience with robots (e.g., robotic vacuum cleaners or industrial robots) but report positive attitudes toward robot technologies (70%). However, this positive attitude is relative to the specific context in which the robot is applied, as 60% believe robots should be *banned* from being used as caretakers for children, elderly and disabled people, and 69% would feel uncomfortable having their dog being walked by a robot. In line with this only 3% believe robots *should* be used for education or caretaking of children, elderly or disabled people. This illustrates the challenges that may arise when robots are introduced into the social sphere and assigned assistive functions in direct interaction with humans.

Several studies support that a specific negative attitude—where ‘attitudes’ are defined as “the relatively enduring organization of beliefs, feelings, and behavioral tendencies” (Vaughan and Hogg, 2005, p. 150)—pertains to so-called ‘social’ robots, and their applications (Nomura et al., 2006, 2008). Among the numerous factors that may determine or affect these attitudes are gender (Nomura et al., 2006; Schermerhorn et al., 2008; Kuo et al., 2009), cultural background of the participants (Bartneck et al., 2005, 2007; Nomura et al., 2008), age (Bumby and Dautenhahn, 1999; Kuo et al., 2009; Heerink, 2011; Smarr et al., 2012), initial attitude (Stafford et al., 2014), and previous experience with robots (Nomura et al., 2006; Bartneck et al., 2007). Furthermore, attitudes and assumptions about robots may be determined by their design, as for instance zoomorphic robots give rise to the assumption of pet like functionalities (Nomura et al., 2008) whilst more humanlike features give rise to attribution of human-like capabilities (Nomura et al., 2008; Schermerhorn et al., 2008). Likewise, it appears that the more human features the robot possesses, the greater the expectations (Nomura et al., 2008). This may suggest that the expectation of autonomous function is borne out of a more humanoid robot design. Yamaoka et al. (2007) explored what happens if the expectation of autonomy in a humanoid robot is challenged by explicitly informing participants that a robot is tele-operated, when in fact it is autonomous. Regardless of the information given beforehand, 2/3 of participants felt that they were interacting with an autonomous robot (Yamaoka et al., 2007). As pointed out by the authors this could indicate that the participants became so immersed in the communication that they failed to retain the information about the robot. Several studies have explored how

presumptions about a robot’s autonomy can be influenced by information about the robot’s functionality; this has mainly been investigated by using the so-called ‘Wizard of OZ paradigm’ in which participants are deceived to believe that a robot is autonomous when in fact it is tele-operated to some degree (for a review see, Riek, 2012). However, so far it has not been explored in which way *attitudes toward robots* change if participants are given truthful information about a robot being tele-operated, or are given no information at all about the degree of autonomy.

The aforementioned investigations may be pivotal to determining the mechanisms for attitude change in this particular area of technology. So far it is not well-understood whether, and to what extent, attitudes toward robots can be influenced. Wu et al. (2014) recently reported that attitudes and acceptance toward assistive robots were unchanged despite several encounters with the robots in healthy elderly and elderly with mild cognitive impairment. The lack of change was attributed to social stigma and uneasiness toward technology (Wu et al., 2014). Conversely Stafford et al. (2010) report more positive attitudes toward a healthcare robot amongst elderly residents at a retirement home after interaction with it (Stafford et al., 2010). Although several studies report positive attitudes toward robots after personal encounters (Mirnig et al., 2012; Yamazaki et al., 2012, 2014) most of these studies do not assess pre-encounter baseline attitudes. Hence, it is difficult to infer whether personal encounters *per se* affect attitudes toward robots or whether, for instance, a selection bias may affect the results, i.e., people with more positive attitudes toward robots at baseline volunteer to partake in the studies. Furthermore, due to the lack of baseline assessments, these studies offer little insight into attitude change.

Determining whether attitude change occurs after encounters with robots and identifying variables that impact such changes are important, especially as more positive attitudes might lead to greater acceptance of robot technology (Ezer et al., 2009). One variable that could potentially influence persistence or change of attitudes is personality. Whilst several studies have explored whether the *robot’s* personality has any effect on the human user’s attitudes toward robots, e.g., by matching between robot-user personalities (Goetz et al., 2003; Lee et al., 2006; Syrdal et al., 2007; Brandon, 2012; Aly and Tapus, 2013; Tay et al., 2014), few have studied the extent to which *user’s* personality affects attitudes toward technology (Cassell and Bickmore, 2003; Luczak et al., 2003). In relation to the latter participants with extravert personality traits appear to have an increased likelihood of responding to technology in a social manner (Luczak et al., 2003) and an increased tendency to ascribe personality to robots with a mechanical or basic appearance, as compared to participants with more introvert personalities (Walters et al., 2007). Conversely, people with high trait Neuroticism and low Extraversion scores preferred the robot to have a more mechanical appearance (Walters et al., 2007). Furthermore, personality may impact proximity behaviors toward robots, since a high score on agreeableness was shown to correlate with a tendency to move closer to robots whilst a high score on neuroticism correlates with a tendency to physically distancing oneself from robots (Takayama and Pantofaru, 2009).

¹ Research Agenda 2020 of EuRobotics, a European research conglomerate of 183 robotics firms.

² McKinsey Global Institute (2013). Disruptive technologies: Advances that will transform life, business, and the global economy.

This illustrates how personality traits manifest themselves in explicit behaviors toward robots. The aforementioned studies mainly pertain to studies focused on younger participants and though personality is stable in middle and old age (Roberts and DelVecchio, 2000) the effect of personality on change in attitudes toward robots in elderly populations is as of yet unexplored.

Given that elderly citizens are a particular target user group of social robotics, the current state of the art on attitude research in this area thus calls for more detailed investigation. In particular, so far it is unclear whether attitudinal change in elderly people vis-a-vis other kinds of technology, e.g., computers, translates to the very special case of social robots whose design exploits implicit processes of social cognition. Previous studies on age-related differences in attitude change toward computers showed that “although there were no age differences in overall attitudes, there were age effects for the dimensions of comfort, efficacy, dehumanization, and control” (Czaja and Sharit, 1998). While elderly people can change their attitudes toward computers (Jay and Willis, 1992), both of these studies, as well as others (Igbaria, 1993; Mitra et al., 1999), emphasized that these attitudinal changes depend more on the type of information and training interaction with the computer and less on the temporal duration of the experience. Attitudes toward computer technology in elderly can be changed in the course of 3 days (Czaja and Sharit, 1998) and perhaps also in shorter periods, since attitudinal change in general can occur within minutes (Harmon-Jones and Harmon-Jones, 2002; Harmon-Jones et al., 2009). In short, extant research on attitudinal change on computer technology suggests, first, that elderly users of technology present a sufficiently distinct subgroup, as far as base level attitudes are concerned, to warrant separate investigation; second, attitudinal changes can occur also in elderly people during short temporal periods; and third, changes in attitudes toward computer technology were produced by information and practical interaction. These three insights motivated the basic set-up of our pilot study on change of attitudes toward robot technology in elderly people.

As the term is understood in current research, attitudes have three components: cognitive, affective, and behavioral. Following the set up of previous work on change of attitudes toward computer technology we investigated changes in the first two components, cognitive and affective. An attitude thus can change in two ways—if the emotional involvement changes in degree and kind, or if the conceptual content of the attitude changes. Since attitudes toward social robots involve rather subtle and complex cognitive and affective contents (ascriptions of consciousness, self-consciousness, moral agency, moral patiency, etc.) an assessment of changes in attitudes is best undertaken in an interdisciplinary setting involving quantitative and qualitative methods, as well as conceptual analysis. The pilot study reported here addressed this particular challenge of interdisciplinarity in order to explore (i) how elderly citizen's attitudes toward robots are affected by baseline information about the functionality of robots, (ii) whether they change after repeated direct encounters with a robot and,

(iii), finally whether certain personality traits facilitate attitude changes.

MATERIALS AND METHODS

Subjects

Participants were residing at Vikaergård (VG) Rehabilitation Centre in Jutland, Denmark. VG offers temporary accommodation and secondary rehabilitation after hospitalization for citizens after disease or injury. Patients may stay at VG for up to 6 weeks.

Inclusion criteria: the participants who were invited to partake in the pilot study were deemed “poor eaters” by trained rehabilitation staff. This was an effort to ensure a homogeneous population who could potentially benefit clinically from the study design.

Exclusion criteria were as follows: (a) diagnosis or suspicion of dementia as indicated by a Mini Mental Status Examination (MMSE) score of 23 or less (Folstein et al., 1975), (b) diagnosis of neurological or neurodegenerative disease, (c) macular degeneration or severe hearing loss, (d) inability to self-feed (as indicated by diseases of mouth or throat or severe motor impairment).

Procedure

The pilot study was carried out in accordance with the Declaration of Helsinki and the Regional Committee on Health Research Ethics. Eligible participants were invited to partake in the study by staff at VG who also supplied them with written information about the project. Subjects who agreed to participate and signed written informed consent received a baseline assessment consisting of questionnaires and a structured interview. A trained master-student in psychology undertook the assessments under supervision of a trained psychologist (MFD). In the 3 days following the assessment the participant had lunch (20–40 min) in the company of either a tele-operated robot or a member of staff. Their lunches were video recorded. The participants were randomly assigned to one of three conditions: (a) an informed condition (IC; $n = 7$) where the participants were informed that the social robot would be tele-operated, (b) an uninformed condition (UC; $n = 7$) where the participants were not given any information about the functionality of the robot, (c) a control condition (CT) where the participant had lunch in the company of a member of staff. In all randomization conditions the conversations and conversation topics were non-scripted and mainly focused on the food, weather, health, the stay at VG etc. Hence, the conversation topics did not pertain to attitudes toward robots. The lunch was served in the participants' private rooms at VG. The control condition was canceled due to unforeseen recruitment problems and the participants excluded ($n = 3$).

Finally, the participants received questionnaires and a structured interview 1 week from the baseline assessment. After the encounter the participants were debriefed on the functionality of the robot. The participants were instructed not to discuss the

pilot study with other residents at VG as it could impact the recruitment process and contaminate the data.

The Robot and the Operators

The Telenoid (see **Figure 1**), a tele-operated android robot developed by Hiroshi Ishiguro from Osaka University and the Advanced Telecommunication Research Institute International, was used. This technology enables two persons, A and B (see **Figure 2**), to communicate with each other using the robot as a communication channel. In contrast to a traditional telephone conversation the interaction facilitated by the Telenoid is asymmetric as the interaction interface is not the same for both parties involved. The operator A controls the robot, which is situated at a different location with the interlocutor B. A's head movements and voice are simulated by the robot and via a monitor and headset with sensors. A is supplied with a live audio and video feed of the robot's head and B. The Telenoid's lip movements follow the speech of A and the robot's "arms" can be moved in one direction. Furthermore the Telenoid features an idle movement function for the eyes. The basic idea behind this setup is to empower A with a remote embodiment at B's site via a wireless network connection.

The Telenoid is "designed according to minimum requirements to express humanlike appearance and motion"



FIGURE 1 | The Telenoid robot.

(Geminoid, n.d.). This neutral design approach is supposed to facilitate B's free associations with the cues and information provided by A and attempts to avoid any interference imposed by design features such as gender or age.³

Three female members of staff, all occupational therapists, were trained in operating the robot. The training contained no specific instruction for conversation content but did contain guidelines of how to reply to questions about robot functionality or personal questions. Overall, the operators were instructed to answer truthfully any questions posed about robot functionality. However, it was not necessary to answer any such questions, as they were not posed. Efforts were made that the participants did not have prior encounters with the operator during their stay at VG, thus would not be able to recognize the operator's voice in interactions with the robot.

Whilst the participants were getting lunch in the common room the robot, microphone, and camera were set up in the participants' private room at VG. The camera was mounted on a pole behind the robot overlooking the participant and the lunch table (see **Figure 3**). Thus when the participant returned with their lunch the robot was present on its stand across the table. The robot was controlled from a laptop in an adjacent room with direct video- and audio feed available.

Measures

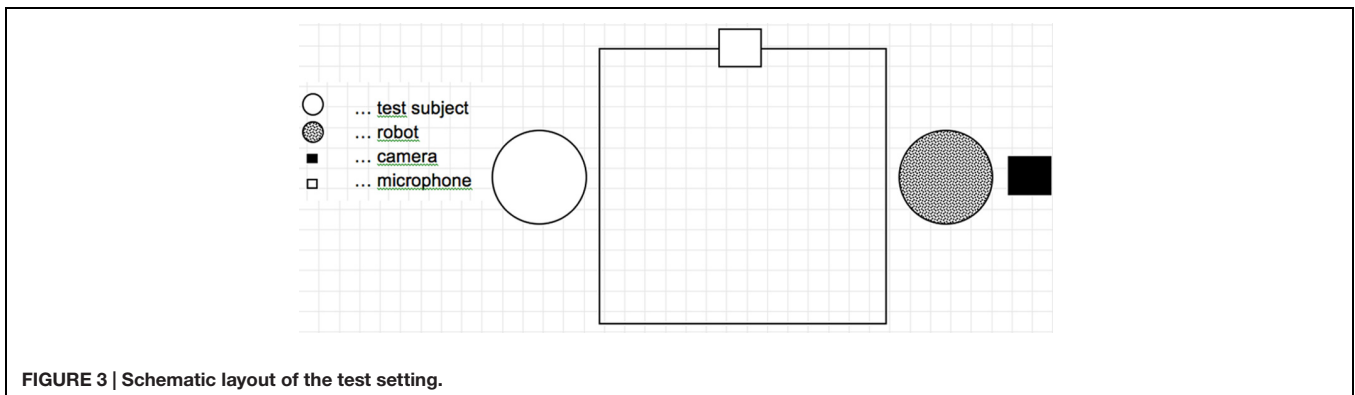
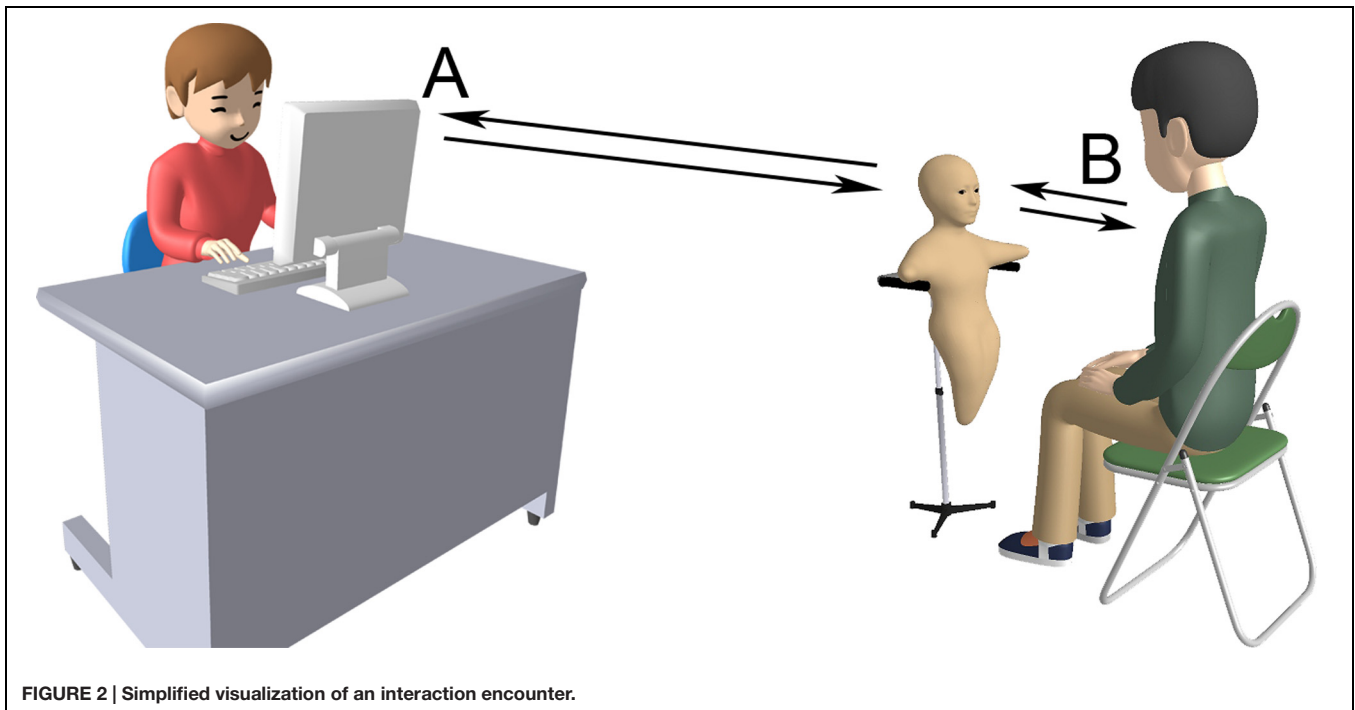
Demographics and Health

Details on age, marital status, general health, eating habits, depression, and perceived stress were obtained from the participants via questionnaires at baseline. Not all questionnaires are included in the current publication.

The NEO-Five Factor Inventory (NEO-FFI)

The NEO-Five Factor Inventory (NEO-FFI; Costa and McCrae, 1992) was used to assess five stable personality dimensions as derived from the five-factor model of personality (NEO-PI-R). The NEO-PI-R is validated cross-culturally (McCrae, 2002) and is available in a validated Danish version. It does not contain items that reflect behavioral, cognitive, or functional well-being of the respondents which would be problematic in the aging study population [for instance the Danish translation of some items in the Tridimensional Personality Questionnaire reads "I have less energy and I am more tired than most people," "I believe in luck for the future" which would not fit the present study given their health status (Cloninger et al., 1991)]. Furthermore, NEO-FFI retains moderate to large correlations with other longer personality questionnaires and has excellent psychometric properties (Larsen, 2007). NEO-FFI was administered at baseline. The questionnaire consists of 60 statements that the respondents rate on a five point Likert scale from "strongly disagree" to "strongly agree." The items were administered verbally whilst the respondent had the five possible answers available in front of them. The five personality dimensions assessed are: Openness (openness to

³Further information on technical aspects of the Telenoid can be found at <http://www.geminoid.jp/projects/kibans/Telenoid-overview.html> (accessed 9 June, 2015).



internal and external stimuli), Conscientiousness (self-discipline and competency), Extraversion (tendency to be sociable and adventurous), Agreeableness (degree of trustfulness, modesty), and Neuroticism (tendency toward experiencing psychological distress or negative affect).

The Attitudes toward Social Robots Scale (ASOR-5)

The ASOR-5 questionnaire is a theoretically based, generic scale of attitudes toward social robotics. The questionnaire was developed in an interdisciplinary taskforce consisting of researchers from psychology, anthropology, and philosophy. ASOR-5 consists of the following subscales: (a) Conceptual relatedness (four items), e.g., “To which degree are you positive about robot technology?” and “please describe in three words your impression of the Telenoid from this picture,” (b) Socio-practical relatedness (eight items), e.g., “Do you think you would take advice from the Telenoid about which medication you should take?,” “Do you think you would be afraid of the

Telenoid?,” (c) Intimate-personal relatedness (five items), e.g., “Can you imagine having a Telenoid in your own home?,” “If you had a Telenoid in your own home would you store it in a broom cupboard?,” (d) Moral relatedness (five items), e.g., “Does it matter how people treat robots?,” “Does the Telenoid have a right to electricity?,” (e) Mental relatedness (five items), e.g., “Do you think the Telenoid can be happy?,” “Do you think the Telenoid can have hobbies and interests?,” and (f) Psychological relatedness (six items), e.g., “I think I would feel sorry for the Telenoid if I saw others be cruel to it,” “I think I would be annoyed if the Telenoid interrupted me in a conversation.” Besides the conceptual relatedness subscale all other items are rated on a five point Likert scale with higher scores indicating more positive attitudes (scores range from 0 to 140). Negative items were reversed before totaling. Furthermore the questionnaire has optional extra items for follow-up assessments (total of 46 items), which are not included in the current publication. The ASOR-5 is integrated in a large validation study alongside the

Godspeed questionnaire (Bartneck et al., 2009), the Negative Attitudes to RobotS questionnaire (Nomura et al., 2005), and the AMPH-10 (see below). For further information please contact the authors.

Anthropomorphism Questionnaire (AMPH-10)

A 10-items questionnaire was developed to assess anthropomorphic thinking. Unlike existing questionnaires of anthropomorphism (e.g., The IDAQ; Waytz et al., 2010) the majority of items (six in total) pertain anthropomorphic thinking toward inanimate objects, e.g., “do you feel grateful toward technology such as a car or computer if you feel it has saved you from a dangerous or difficult situation” or “would you ever give a name to an everyday item, such as a Television”? All items were rated on a four point Likert scale from “very unlikely” to “highly likely” with higher scores indicating more pronounced anthropomorphic thinking (maximum score is 40).

Statistics

The data was analyzed using IBM SPSS Statistics for Macintosh, Version 21.0. (2012; Armonk, NY, USA: IBM Corp). A change score was calculated defined as the difference in ASOR-5 sub-scores from baseline to follow-up. The informed and uninformed conditions were compared on these continuous variables using independent *t*-tests. Paired sample *t*-tests were used to assess changes in the ASOR-5 sub-scores from baseline to follow-up in the informed (IC) and uninformed (UC) condition. The *t*-test is an acceptable statistical approach, even in very small samples (de Winter, 2013). The possible relationship between personality traits and changes or stability in attitudes toward social robotics was explored by Spearman correlations.

Due to the small sample size and exploratory nature of the pilot study Bonferroni corrections for multiple comparisons were not made. Bonferroni adjustments are normally undertaken by dividing the alpha-level by the number of comparisons made in order to reduce the risk of obtaining false positive results as a consequence of multiple analysis of the same data set (Tabachnick and Fidell, 2001). The necessity of Bonferroni corrections are debated and in the present pilot study we opted for reporting the exact alpha-levels and effect sizes (ESs; Rothman, 1990; Feise, 2002). Samples solely relying on the alpha-level can be misleading, as smaller samples will possess less statistical power to detect a difference. To inform on the strength of the effect, ESs are reported (Cohen's *d*) where $d = 0.2$ is considered a small ES, $d = 0.5$ is a medium ES, and $d = 0.8$ or above is deemed a large ES (Cohen, 1988). Due to the modest *n* in the present sample effect-sizes are interpreted conjointly with *p*-values.

A total of 17 elderly participants were enrolled in the study. Three participants were excluded as unforeseen recruitment issues forced us to suspend the control condition.

Repeated *t*-test comparisons showed no significant differences from pre-encounter to post-encounter scores on any of the ASOR-5 domains for the total sample ($n = 14$; see **Table 1**). Hence there was no significant difference in attitude scores on any domains from before they meet the robot till after they had

been in company with it during lunch, for up to 3 days. However, a moderate ES ($d = 0.562$) was observed on the *Intimate-personal relatedness* domain, which indicates a non-overlap between the two groups of 33% (Sullivan and Feinn, 2012).

The participants were assigned to either the IC ($n = 7$) or the UC ($n = 7$) group as they were recruited. The IC and UC groups did not differ significantly in terms of gender distribution (men 71.4% in either group) and there was no significant difference between the IC ($M = 74.83$, $SD = 12.9$) and UC ($M = 75.29$, $SD = 11.7$) groups on age [$t(13) = 0.06$, $p = 0.948$].

Independent two-tailed *t*-tests showed no significant difference between the informed and uninformed condition in attitude change scores on any of the ASOR-5 subscales (see **Table 2**). Hence the change in attitude from pre- to post-encounter did not differ significantly between the two groups who were given different information about robot functionality. However, there was a near significant difference on the socio-practical relatedness subscale where participants who were uninformed about the functionality of the robot, rated it more negatively after meeting it. This is supported by a very large ES ($d = 1.09$), which means that there is a 55% non-overlap between scores in the informed and the uninformed conditions where the latter group was more likely to change their attitude negatively post-encounter.

Spearman correlation analyses were employed to explore possible correlations between attitude change scores on the ASOR-5 questionnaire and personality traits as measured by NEO-FFI. To increase statistical power the IC and UC groups were combined for this analysis. There were significant moderate-high positive correlations between Extraversion ($M = 30.36$, $SD = 4.05$) and the *intimate-personal relatedness* ASOR-5 subscale, and the *psychological relatedness* ASOR-5 subscale (see **Table 3**). There was a significant negative correlation between Neuroticism ($M = 19.14$, $SD = 6.22$) and the ASOR-5 *mental relatedness* subscale. Conscientiousness ($M = 30.93$, $SD = 4.16$), Agreeableness ($M = 31.21$, $SD = 5.92$), and Openness ($M = 23.5$, $SD = 6.19$) did not correlate significantly with any of the ASOR-5 subscales. Furthermore, there was a significant negative correlation between the ASOR-5 *mental relatedness* subscale and anthropomorphic thinking ($M = 8.5$, $SD = 5.32$).

Qualitative Analysis of Video Data Method and Set-up

The pilot study also included video recordings of the lunch sessions; the camera was mounted in the stand of the Telenoid, showing the participant frontal, from the point of view of the Telenoid. The lunch sessions took place in the participant's own room, and the Telenoid was seated at the table when the participant was followed into the room by a staff member of the rehabilitation center carrying the food. The video recordings have been analyzed through content analysis, a method used in both quantitative and qualitative studies to analyze written, verbal, or visual communication messages (Cole, 1988; Elo and Kyngäs, 2007). The material is analyzed through a defined framework, so that it should be possible to reach a result as objective as

TABLE 1 | Repeated two-tailed *t*-test comparisons of the ASOR-5 domains.

	Baseline T1	Post-encounter T2	<i>t</i> -test	<i>p</i> -value	Cohen's <i>d</i>
	Mean (<i>SD</i>)	Mean (<i>SD</i>)	<i>t</i> (13)	<i>p</i> -value	<i>d</i>
ASOR-5 Domains					
SPR	12.43 (3.13)	11.79 (3.75)	0.529	0.606	0.19
IPR	8.21 (3.02)	9.71 (2.27)	−1.9	0.078	0.56
MOR	7.01 (2.04)	6.86 (2.32)	0.224	0.826	0.07
MER	2.79 (3.56)	4.14 (4.07)	−1.24	0.236	0.35
PSR	14.14 (4.57)	13.79 (4.57)	0.340	0.740	0.08
Total scale	43.46 (9.04)	46.23 (4.55)	0.949	0.361	0.39

SPR, socio-practical relatedness; IPR, intimate-personal relatedness; MOR, moral relatedness; MER, mental relatedness; PSR, psychological relatedness.

TABLE 2 | Independent *t*-test comparisons of the IC and UC ASOR-5 change scores.

	Informed condition (IC) (<i>n</i> = 7)	Uninformed condition (UC) (<i>n</i> = 7)	<i>t</i> -test	<i>p</i> -value	Cohen's <i>d</i>
	Mean (<i>SD</i>)	Mean (<i>SD</i>)	<i>t</i> (12)	<i>p</i> -value	<i>d</i>
Demographics					
Age	74.83 (12.9)	75.29 (11.7)	−0.06	0.948	–
Change scores in ASOR-5 scale^a					
SPR	1.57 (4.65)	−2.86 (3.44)	2.03	0.066	1.09
IPR	2.14 (2.04)	0.85 (3.67)	0.81	0.433	0.43
MOR	0.43 (1.72)	−0.71 (2.93)	0.89	0.391	0.48
MER	−0.43 (2.07)	0.29 (0.76)	−1.50	0.160	0.80
PSR	0.14 (2.50)	−0.86 (5.18)	0.46	0.653	0.25
Total scale	3.86 (9.26)	1.50 (12.63)	0.39	0.714	0.21

^aPositive scores reflect positive mean changes in subscale measures after meeting the robot. Subscale change scores = subscale score at time 2 – subscale score at time 1. SPR, socio-practical relatedness; IPR, intimate-personal relatedness; MOR, moral relatedness; MER, mental relatedness; PSR, psychological relatedness.

TABLE 3 | Spearman correlations between the ASOR-5 subscale change scores and personality traits (NEO-FFI) and anthropomorphic thinking.

Variables	ASOR-5IPR	ASOR-5PSR	ASOR-5SPR	ASOR-5MOR	ASOR-5MER
Openness	0.009	−0.257	0.270	−0.258	0.067
Conscientiousness	0.080	0.229	0.055	−0.136	0.330
Extraversion	0.619*	0.581*	0.454	0.511	0.085
Agreeableness	0.134	−0.317	−0.060	−0.165	0.099
Neuroticism	0.479	0.224	0.324	0.281	−0.582*
Anthropomorphic thinking	−0.061	−0.249	−0.243	−0.292	−0.662*

**p* < 0.05. SPR, socio-practical relatedness; IPR, intimate-personal relatedness; MOR, moral relatedness; MER, mental relatedness; PSR, psychological relatedness.

possible, also with different researchers coding and analyzing the material.

The content analysis was framed by two focus points arrived at deductively from the quantitative analysis of the questionnaires. The quantitative analysis showed a lack of change in attitude toward the robot after interaction, which was surprising when seen in relation to studies showing change in attitude after interaction with computers. For the purpose of this paper it was decided to analyze the video data on two specific aspects: attitudes to the Telenoid during the conversation, especially changes in attitudes over the different sessions, and a focus on specific statements about what it was like to talk to the Telenoid. In this way we seek to add a deeper understanding

of some of the interesting findings in the pilot study by triangulating qualitative and quantitative data (Karpatschhof, 2010).

Selected Results of Content Analysis

In all sessions, with both informed and uninformed participants, the participants greeted the Telenoid with hospitable language, answered questions politely, engaged in normal turn-taking. The conversations followed the schema of a normal exchange during lunch as this would be typical at a rehabilitation center, with the general topics being the food being served; whether the participant was able to eat the food; why the participant was at the rehabilitation center; how it was going

with the training sessions; the participant's family situation; the weather. Despite many of the participants volunteering different personal information which the operator could have pursued, the conversations stuck to the frame of a typical conversation between an occupational therapist and a 'patient.'

Participants in general expressed pleasure and curiosity about engaging in the conversation with the Telenoid, and despite there being some technical problems (e. g. bad sound, uncontrolled head movements) the participants consistently retained the social norms of polite conversation and tried to remain in contact with the Telenoid. If the Telenoid suddenly worked again, the participants immediately continued to answer questions. Most participants finished up the last session by expressing positive statements of having enjoyed themselves and being positively surprised about the experience of being in the company of a robot.

The content analysis also revealed that while there were many positive statements about talking to the Telenoid during the sessions, there was no distinctive change in attitude toward the Telenoid in the course of the successive sessions. However, a change did happen, but it happened within the first few minutes of each session, and could be clearly observed by comparing the beginnings of sessions 1 and 2. When participants first entered their room, they had never seen the Telenoid before; they were asked to sit at the table directly in front of it. All participants required some help in taking their place at the table and bringing the food along, and they would often discuss 'it' with the caretaker helping them. Once they were seated the first time they would either greet the Telenoid with some hesitation, or wait until being greeted and then answer. After the first hesitation the conversation would soon follow normal patterns of conversation. The next time the participant came to eat with the Telenoid, there was a significant change in the initial greeting between the participant and the Telenoid. Often the participant would greet the Telenoid already while entering the room, before he/she was in the view of the Telenoid, or they would greet, as if they were greeting someone they knew, as soon as they were sitting at the table, trying to pick up the conversation from yesterday. They showed obvious signs of familiarity and positivity, smiling, waving, looking directly at the Telenoid and seeking eye contact. In the following excerpts from the video recordings it is shown how the initial greetings change between session 1 and 2.

Uninformed male participant #45

First session. The participant is driven in to the table in his wheelchair, he is not really looking at the Telenoid.

T: Hello.

P: Hello.

... (there is a longer pause while the participant is cutting his food.)

T: What are you having for dinner today?

P: I am having filet mignon.

... (there is a little discussion about the food and the participant starts eating).

T: Can you hear what I am saying?

P: ... what, sorry, yes, I can hear you.

... (the participant looks at the Telenoid while answering, but looks away when it is quiet and continues eating. There is a longer pause).

P: But it is a very quiet companion I have.

T: ... (laughs a bit)... It is because she wants to give you time to eat your food.

P: Oh, but that doesn't matter. It is nice and warm, so it won't hurt if it cools down a bit, while I am being interrupted.

Last session. The participant is placed at the table. As soon as the carer/helper leaves he says:

P: Hi Sussi (a name he has given the Telenoid in an earlier session).

T: Hi Ole.⁴

P: Well, here we are again. I can hear you loud and clear again. It wasn't so good yesterday. It is much better. Now you have your own pleasant voice back.

T: That is nice to hear.

Informed female participant #48a

First session. P: Hello, hello... (the participant is coming into the room, but still not visible).

T: (no answer).

P: What is your name?... What is your name?... What is your name?

T: (no answer).

P: Can't you say anything? Yum, It is lovely food I am having. ... Can I take a picture of you? (gets her phone). Is it allowed to take a picture of you? ... I am taking a picture of you. (continues eating).

... (a carer comes in and tells her that there is something wrong with the sound. After a little while the Telenoid makes a sound). ...

P: What are you saying? Are you going to say anything now? I have been excited about talking to you, but you are not answering. ... (continues eating – this happens several times, about 7 min after she has entered the room, the Telenoid is working again).

T: Hello.

P: Hello. Oh so finally you can say something.

T: The sound came on.

P: Yes, what is your name?

T: What do you think is appropriate?

P: Hmm. ... Robert.

T: Robert? That is fine.

P: Okay, let's say that then.

T: I just have to start up. And you are already eating?

D: Yes, thank you. It tastes delicious.

Last session. The participant enters the room and initiates the conversation:

P: Hello Robert.

T: Hi.

⁴Names are changed.

P: Hi. So, here we are again.
 T: Here we are again, yes.
 T: Are things going well?
 P: It is yes. It is going really well, I think.

Uninformed male participant #48b

First session. The Telenoid says hello as the man is being driven into the room. There is no answer. He looks at the Telenoid as he is getting set at the table, but doesn't say anything. He begins eating his meal and the Telenoid says:

T: Hello Martin.
 P: (looks up in surprise and smiles) Hi. It is nice to see you.
 T: What is on the menu?
 P: Asparagus soup. And it actually tastes very good. I am not sure about the other stuff. . . Ham, I think. But I can tell you more about it, when I get to that.
 T: That sounds good. (Pause, the participant continues eating).
 How long have you been here at VG?
 P: 2.5 weeks, I think, and I have to be here for 1.5 weeks more.
 T: And are you happy about being here?
 P: Yes I am. It is actually really nice here. They look after you well, and they are giving me a good training.
 T: It sounds like the purpose for coming here has been fulfilled.
 P: Yes. That is quite right. Actually it is really nice here, and it is also exciting that I got you as a visitor.
(P has some problems with hearing). . .
 P: (leans forward) Sorry, I can't hear what you are saying, I have some problems with hearing.

Second session. As the participant is coming in and the food is being set out on the table the Telenoid says:

T: Hallo Martin.
 P: (in a loud happy voice)...Hallo! It is lovely to see you again.
 . . . (the carer finishes and walks out and says she won't disturb)
 P: (waves dismissively at the carer and looks at the Telenoid with a smile) No, we can easily handle this, right?
 T: Let's hope the food tastes good today.
 P: Yes it is ham I think. It looks good.

Last session. Already as we can hear the participant entering the room, we can hear him shout:

P: Hi!
 . . . (The Telenoid doesn't answer, the participant sits down). . .
 P: Hi. (pause). You are not saying anything today. Haven't you been allowed to. . .
 T: (interrupts) Hi!
 P: Hi! Oh, it is good to see you again (P is clearly happy and smiling).
 T: Yes, same here. Is there still no food for you?
 P: No. But hopefully you have had the electricity you need, so that you are not starving.
 T: I have had what I need . . . (a little laugh in the voice).
 . . . as the session is coming to an end, the participant says:
 P: I can't really eat a lot right now.
 T: It doesn't look like very much. Maybe you can eat a few mouthfuls while we talk.
 P: Aahh noo. . . (The participant hesitates a little, but picks up the fork and takes a little).

P: I can't really eat anymore, but I will try and eat a few mouthfuls when you say so. Oh no, it is not going so well, I am dropping the food. That is not very good.

T: It is ok with me if you drop your food, that doesn't matter.

P: No, I know that. I am not shy in front of you anymore, because I know you are just sitting here as a robot, who is supposed to help me, and you are doing that really well. It is nice to have you here to talk to.

These illustrations are representative for a pattern we could observe across 12 participants, both informed and uninformed. In sum, the content analysis of the initial greetings between participants and the Telenoid in the video recordings showed that during the very first encounter in the first session participants were somewhat hesitant in starting the interaction but quickly accustomed themselves to the new situation by turning to social norms of conversation and consistently retained this pattern of interaction throughout the remaining sessions.

DISCUSSION

To our knowledge this is the first study to assess, for a test population of elderly citizens, change in attitudes toward robots in relation to personality traits as well as taking into account pre- and post-encounter assessments. Overall the pilot study indicates that the elderly participants did not display any statistically significant change in attitude toward robots from pre- to post-encounter. However, a moderate ES ($d = 0.562$) was observed on the *Intimate-personal relatedness* domain, which indicates an effect on this domain. Furthermore, there was no significant difference in attitude change between the participants who were informed about the robot being tele-operated and the participants who were uninformed. The results tentatively suggest that beliefs about robot autonomy and functionality do not significantly impact attitude change toward robots in this population of elderly participants. Participants who were uninformed about the robot functionality at baseline did tend to be more reluctant to rate the robot highly on the socio-practical relatedness scale post-encounter; however, this trend did not reach statistical significance ($p = 0.066$) but the finding is supported by a large ES ($d = 1.06$). Personality was correlated with some changes in attitudes toward robots. There was a moderate correlation between the Extraversion and more positive attitude changes to *intimate-personal relatedness* ($r = 0.619$) and to *psychological relatedness* ($r = 0.581$) whilst Neuroticism and also anthropomorphic thinking correlated negatively ($r = -0.582$) with *mental relatedness*.

The analysis tentatively suggests that the level of information given may impact the way elderly relate to the robot on a socio-practical level as indicated by a large ES on the differences in this domain ($d = 1.06$). Hence, the participants who were uninformed about the robot being tele-operated on average had a negative change in the socio-practical relatedness domain. This domain contains items about whether the Telenoid would be

trusted to give pertinent, coherent, and relevant information. It appears that the elderly participants who were uninformed about its functionality were more reluctant to trust the validity of the advice from the Telenoid compared to the informed group. The interpretation of this finding has to be done with caution though as it did not reach statistical significance ($p = 0.77$).

Overall, the results of this pilot study indicate that the influence of information about functionality of robots is negligible for promoting attitude change toward robots in elderly participants. Several explanations may be offered for this finding. As pointed out by Yamaoka et al. (2007) the participants may become so immersed in communication with the robot that they simply forget the information given to them beforehand. However, this explanation does not accommodate our finding that there is no significant or limited attitude changes from baseline to post-encounter. Arguably, if the participants become so engrossed in conversation with the robot one should have expected that their attitudes would have changed in either positive or negative direction from baseline. Rather, it seems that baseline attitudes are largely retained regardless of the level of information or number of personal encounters with a robot. This is supported by Wu et al. (2014) who also reported stability of attitudes toward robots amongst healthy elderly despite repeated encounters with a robot (encounters of 30 min a week for 4 weeks). These findings can be interpreted as an expression of cognitive conservatism where initial attitudes are retained and new information or experiences are poorly integrated with the existing cognitive schema (Piaget et al., 1952). This effect may have been inadvertently nurtured by the design of the study as one of the main assumptions about attitudes and attitude change is that attitudes can either be mainly founded on cognitions or on affect and that emotionally arousing experiences are best at changing affect-based attitudes (whilst cognitively based attitudes are changeable by both feelings toward and knowledge about the attitude object; Edwards, 1990; Edwards and Von Hippel, 1995; Fabrigar and Petty, 1999). It seems likely that attitudes toward robots as social agents are more reliant upon affect, and that attitude changes borne out of social interaction with a robot may also be driven by emotional arousal. Hence, the rational answers given by the elderly participants on questionnaires or in interviews may be qualitatively different from observable emotional attitudes and their changes over time as displayed by the participant during social interaction with the robot.

The personality trait Extraversion was positively correlated to an increased likelihood of high scores on intimate-personal relatedness post-encounter. This is in line with the relationship between Extraversion, positive emotionality and a preference toward social interactions reported in existing literature (Costa and McCrae, 1980). The correlation between Extraversion and attitude change in the present study is limited to the two domains and seems to reflect a wish to satisfy communicative needs. Neuroticism and anthropomorphic thinking at baseline were negatively correlated to attitude changes in mental relatedness to the robot.

Neuroticism is associated with negative emotionality and an inflexible mind-set (Costa and McCrae, 1980). Hence, higher scores on neuroticism and anthropomorphic thinking appear to “lock” the participants into a certain way of mentally relating to the robot blocking the likelihood for change. Most likely these results are produced by different underlying ‘mechanisms’ for participants with high scores on anthropomorphic thinking and for participants with high scores on neuroticism; where the former may from the very beginning relate to the robot *as-if* it were a person with inherent mental capacities and not change this view, the latter will probably be reluctant to mentally relate to the robot under any circumstance.

The present pilot study offers an interdisciplinary field-based study with one-on-one interaction between could-be end users and a social robot with a repeated measures design. In summary the quantitative results tentatively suggest that (i) explicit attitudes of elderly citizens toward robots are not significantly affected by baseline information about robot functionality, (ii) explicit attitudes to robots do not significantly change after repeated personal encounters with a robot, (iii) higher scores on the personality trait Extraversion are correlated with higher likelihood for positive change on the subscales *intimate-personal relatedness* and *psychological relatedness* whilst higher scores on Neuroticism were associated with a reduced tendency to change on the *mental relatedness* scale.

Several limitations should be mentioned. Despite the technical advances in robot technology malfunctions still occurred possibly because of wireless interference from various appliances in use at the rehabilitation center. This meant that the session with the robot was sometimes canceled, delayed or that the robot did not operate properly (e.g., displayed tremor-like movements of the head or in one case was suddenly unresponsive). The exact effect of such experiences on attitudes and attitude change was not taken into account in this pilot study. Future studies should consider assessing how participants experience technical malfunctions. Secondly, it is possible that all participants knew that the robot was tele-operated simply due to its speech and mannerism. We did not explicitly assess the participants’ beliefs about the functionality of the robot post-encounter. However, the near-significant change on the socio-practical domain of the ASOR-5 questionnaire for the uninformed condition combined with a large ES indicates that the instructions at randomization worked (since this near-significant difference may reflect differing attitudes based on the information given about the robot). Thirdly, the study design did not allow for use of the full functionality of the robot. In particular the participants did not hug (hugging being a key feature of the robot’s functionality) or even touch the robot, which may have affected their level of emotional investment in the interaction. The decision not to include tactile stimulation, specifically hugging, stemmed from the original design of the study where some participants had to eat in the company of a member of staff as a control condition. It would have been unethical to demand the staff to hug the participants.

Fourthly, the moderate N limits the generalizability of the results and the statistical power to detect differences. However, this interdisciplinary pilot study uncovered important trends in the complex relationship between age, attitudes, personality and social robots, which can guide future studies in a larger sample where more complex statistical procedures can be applied.

The interplay between the quantitative and the qualitative results of our study suggest several further implications for future research. Since both Wu et al. (2014) and our pilot study find that older people's attitudes toward robots are largely stable, while studies on the same age group report changes in attitudes toward computers after similar exposure times (Jay and Willis, 1992; Czaja and Sharit, 1998), it is also important to ask whether this difference might have any implications for competing theories of attitudinal change in general. To be sure, if information about functionality and direct interaction changes elderly people's attitudes toward computers but not, *mutatis mutandis*, their attitudes toward robots, this may be attributed to the type in interaction involved in each case. On the other hand, one might also argue that the observed stability of attitudes toward robots fits well with recent explanations of attitudinal change as "action-based discrepancy reduction" (Harmon-Jones and Harmon-Jones, 2002; Harmon-Jones et al., 2009). According to this account, attitudinal change occurs to reduce cognitive-affective discrepancies so as to facilitate future actions. Since computers are already entrenched in our socio-cultural practices, we perceive them as agentively relevant and thus may react to discrepancies between pre-interaction attitudes and cognitive and affective states during experience by adjusting the former to unblock decision and action pathways. In contrast, robots do not yet have agentive relevance—they are not yet perceived as items that figure in test subject's action space and relative to which practical decisions need to be taken, thus the reduction of cognitive-affective discrepancies is practically not yet relevant.

However, in light of the selection of results from our qualitative research as reported in Section "Qualitative Analysis of Video Data" above, another possible explanation of the observed stability of attitudes toward robots is possible. According to the "action-based" explanation of attitudinal change, these processes occur in order to reduce a felt discrepancy among cognitions that carry conflicting action tendencies. More precisely, the reduction of discrepancy occurs to eliminate the negative emotion of dissonance (proximal motivation) and to enable efficient action (distal motivation; Harmon-Jones et al., 2009, p. 128). If no discrepancy in action tendencies is experienced, and if accordingly no emotional dissonance is experienced, on the action-based model there is no reason to change one's attitudes. Based on the qualitative analysis of the video material of our study precisely this appears to be the case. All participants, both informed and uninformed, very quickly (within a few minutes during the first session) settle on the overall interaction pattern of polite social conversation and return to this style of interaction without hesitation, almost eagerly, during subsequent sessions. The fact that several participants choose to give the

Telenoid a first name consolidates the interaction frame of social conversation for the duration of their encounter. Most remarkable perhaps, participants stay with the routines of social conversation even when severe technical problems occur (no sound, uncontrolled head movements of the Telenoid). At no time participants displayed any tendencies to break with the action patterns of social conversation with the Telenoid (e.g., by calling for the caretakers during malfunction, or by ending the session prematurely). In short, the pre-encounter attitudes toward the Telenoid did not have to be corrected since the interaction context did not create any conflicting action tendencies and associated negative emotions.

This explanation would imply that future research on attitudinal change toward social robots cannot use the interaction scenarios that social robots are developed for. Social robots are intentionally designed to engage humans in social interaction patterns, exploiting both explicit and implicit (pre-conscious) "mechanism of social cognition" (Frith and Frith, 2008). Thus mere habituation and increased encounter in everyday social contexts are unlikely to change negative human assessments of social robots. Humans are conditioned to uphold the routines of social interactions precisely because these routines serve the evolutionary function of providing agentive guidance in a large variety of situations where agentive insecurity or conflictedness might otherwise occur. On the assumption that the "action-based" explanation of attitudinal change is on the right track, future experiments on attitudinal change toward social robots thus will need to operate with set ups that involve extraordinary interaction context where genuine conflicts of action-tendencies can arise.

FUNDING SOURCES

This pilot study was realized via the PENSOR project funded by the VELUX FOUNDATION and supported by the Department of Health and Assisted Living Technologies (Municipality of Aarhus), which provided staff and the test venue.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to the Municipality of Aarhus especially Ivan Kjær Lauridsen (Head of Health and Assisted Living Technologies), Birgitte Halle (project leader), and Emilie Vestergaard Kragh Sørensen (intern) as well as to the operators and administrative staff from the rehabilitation center *Sundheds- og Omsorgshotellet Vikærgården*. We are also obliged to the Hiroshi Ishiguro Laboratories (ATR, Japan) for assisting us with guidance, technical expertise and the graphics for **Figure 1**. Furthermore, we are in debt to Stefan K. Larsen (Ph.D. fellow) as well as Thea Puggaard Frederiksen, Rikke Mayland Olsen and Kasper Lund (interns/students) from Aarhus University. The members of the PENSOR project at the School of Culture and Society (Aarhus University) have been creative discussion partners in the crucial project initiation phase.

REFERENCES

- Aly, A., and Tapus, A. (2013). "A model for synthesizing a combined verbal and nonverbal behavior based on personality traits in human-robot interaction," in *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction* (Piscataway, NJ: IEEE Press), 325–332.
- Bartneck, C., Kulic, D., Croft, E., and Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int. J. Soc. Robot.* 1, 71–81. doi: 10.1007/s12369-008-0001-3
- Bartneck, C., Nomura, T., Kanda, T., Suzuki, T., and Kennsuke, K. (2005). "A cross-cultural study on attitudes towards robots," in *Proceeding of the HCI International*, Las Vegas, Nevada.
- Bartneck, C., Suzuki, T., Kanda, T., and Nomura, T. (2007). The influence of people's culture and prior experiences with Aibo on their attitude towards robots. *AI Soc.* 21, 217–230. doi: 10.1007/s00146-006-0052-7
- Brandon, M. (2012). *Effect Personality Matching on Robot Acceptance: Effect of Robot-User Personality Matching on the Acceptance of Domestic Assistant Robots for Elderly*. Master thesis, University of Twente Student, Enschede. Available at: <http://essay.utwente.nl/61971/> [accessed May 28, 2015].
- Bumby, K., and Dautenhahn, K. (1999). "Investigating children's attitudes towards robots: a case study," in *Proceeding of the CT99, The Third International Cognitive Technology Conference* (San Francisco, CA: Citeseer Publisher), 391–410.
- Cassell, J., and Bickmore, T. (2003). Negotiated collusion: modeling social language and its relationship effects in intelligent agents. *User Model. User Adapt. Interact.* 13, 89–132. doi: 10.1023/A:1024026532471
- Cloninger, C. R., Przybeck, T. R., and Svrakic, D. M. (1991). The Tridimensional Personality Questionnaire: U.S. normative data. *Psychol. Rep.* 69(3Pt 1), 1047–1057. doi: 10.2466/pr0.1991.69.3.1047
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edn. (Hillsdale, NJ: Lawrence Erlbaum Associates Inc.), 13.
- Cole, F. L. (1988). Content analysis: process and application. *Clin. Nurse Spec.* 2, 53–57. doi: 10.1097/00002800-198800210-00025
- Costa, P. T. Jr., and McCrae, R. R. (1992). Normal personality assessment in clinical practice: the neo personality inventory. *Psychol. Assess.* 4, 5–13. doi: 10.1037/1040-3590.4.1.5
- Costa, P. T., and McCrae, R. R. (1980). Influence of extraversion and neuroticism on subjective well-being: happy and unhappy people. *J. Pers. Soc. Psychol.* 38, 668. doi: 10.1037/0022-3514.38.4.668
- Czaja, S. J., and Sharit, J. (1998). Age differences in attitudes toward computers. *J. Gerontol. Ser. B Psychol. Sci. Soc. Sci.* 53, 329–340. doi: 10.1093/geronb/53B.5.P329
- de Winter, J. C. (2013). Using the Student's t-test with extremely small sample sizes. *Pract. Assess. Res. Eval.* 18:2.
- Edwards, K. (1990). The interplay of affect and cognition in attitude formation and change. *J. Pers. Soc. Psychol.* 59, 202–216. doi: 10.1037/0022-3514.59.2.202
- Edwards, K., and Von Hippel, W. (1995). Hearts and minds: the priority of affective versus cognitive factors in person perception. *Pers. Soc. Psychol. Bull.* 21, 996–1011. doi: 10.1177/01461672952110001
- Elo, S., and Kyngäs, H. (2007). The qualitative content analysis process. *J. Adv. Nurs.* 62, 107–115. doi: 10.1111/j.1365-2648.2007.04569.x
- Ezer, N., Fisk, A. D., and Rogers, W. A. (2009). "Attitudinal and intentional acceptance of domestic robots by younger and older adults," in *Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments*, ed. C. Stephanidis (Berlin: Springer), 39–48.
- Fabrigar, L. R., and Petty, R. E. (1999). The role of the affective and cognitive bases of attitudes in susceptibility to affectively and cognitively based persuasion. *Pers. Soc. Psychol. Bull.* 25, 363–381. doi: 10.1177/014616729902503008
- Feise, R. J. (2002). Do multiple outcome measures require p-value adjustment? *BMC Med. Res. Methodol.* 2:8. doi: 10.1186/1471-2288-2-8
- Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). 'Mini mental state'. A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-6
- Frith, C., and Frith, U. (2008). Implicit and explicit processes in social cognition. *Neuron* 60, 503–510. doi: 10.1016/j.neuron.2008.10.032
- Geminoid. (n.d.). Available at: <http://www.geminoid.jp/en/robots.html> [accessed June 9, 2015].
- Goetz, J., Kiesler, S., and Powers, A. (2003). "Matching robot appearance and behavior to tasks to improve human-robot cooperation," in *Proceedings of the 12th IEEE International Workshop on Robot and Human Interactive Communication*, 2003 (Millbrae, CA: IEEE), 55–60.
- Harmon-Jones, E., Amodio, D. M., and Harmon-Jones, C. (2009). Action-based model of dissonance: a review, integration, and expansion of conceptions of cognitive conflict. *Adv. Exp. Soc. Psychol.* 41, 119–166. doi: 10.1016/S0065-2601(08)00403-6
- Harmon-Jones, E., and Harmon-Jones, C. (2002). Testing the action-based model of cognitive dissonance: the effect of action orientation on postdecisional attitudes. *Pers. Soc. Psychol. Bull.* 28, 711–723. doi: 10.1177/0146167202289001
- Heerink, M. (2011). "Exploring the influence of age, gender, education and computer experience on robot acceptance by older adults," in *Proceeding of the 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2011 (Lausanne: IEEE), 147–148.
- Igbaria, M. (1993). User acceptance of microcomputer technology: an empirical test. *Omega* 21, 73–90. doi: 10.1016/0305-0483(93)90040-R
- Jay, G. M., and Willis, S. L. (1992). Influence of direct computer experience on older adults' attitudes toward computers. *J. Gerontol.* 47, 250–257. doi: 10.1093/geronj/47.4.P250
- Karpatsch, B. (2010). "Den kvalitative undersøgelsesforms særlige kvaliteter," in *Kvalitative Metoder*, eds S. Brinkmann and L. Tanggaard (Copenhagen: Hans Reitzels Forlag), 409–429.
- Kuo, I. H., Rabindran, J. M., Broadbent, E., Lee, Y. I., Kerse, N., Stafford, R. M. Q., et al. (2009). "Age and gender factors in user acceptance of healthcare robots," in *Proceeding of the 18th IEEE International Symposium on Robot and Human Interactive Communication*, 2009, ROMAN 2009 (IEEE), 214–219.
- Larsen, L. (2007). *Gerontopsykologi: Det Aldrende Menneskes Psykologi*. Aarhus: Aarhus Universitetsforlag.
- Lee, K. M., Peng, W., Jin, S.-A., and Yan, C. (2006). Can robots manifest personality: an empirical test of personality recognition, social responses, and social presence in human-robot interaction. *J. Commun.* 56, 754–772. doi: 10.1111/j.1460-2466.2006.00318.x
- Luczak, H., Roetting, M., and Schmidt, L. (2003). Let's talk: anthropomorphization as means to cope with stress of interacting with technical devices. *Ergonomics* 46, 1361–1374. doi: 10.1080/00140130310001610883
- McCrae, R. R. (2002). "NEO-PI-R data from 36 cultures," in *The Five-Factor Model of Personality Across Cultures*, eds R. R. McCrae and J. Allik (Berlin: Springer), 105–125.
- McKinsey Global Institute (2013). *Disruptive Technologies: Advances that will Transform Life, Business, and the Global Economy*. Available at: http://www.mckinsey.com/insights/business_technology/disruptive_technologies
- Mirnig, N., Strasser, E., Weiss, A., and Tscheligi, M. (2012). "Studies in public places as a means to positively influence people's attitude towards robots," in *Social Robotics*, (Berlin: Springer), 209–218. doi: 10.1007/978-3-642-34103-8_21
- Mitra, A., Steffensmeier, T., Lenzmeier, S., and Massoni, A. (1999). Changes in attitudes toward computers and use of computers by university faculty. *J. Res. Comput. Educ.* 32, 189–202. doi: 10.1080/08886504.1999.10782623
- Nomura, T., Kanda, T., and Suzuki, T. (2005). Experimental investigation into influence of negative attitudes toward robots on human-robot interaction. *AI Soc.* 20, 138–150. doi: 10.1007/s00146-005-0012-7
- Nomura, T., Kanda, T., and Suzuki, T. (2006). Experimental investigation into influence of negative attitudes toward robots on human-robot interaction. *AI Soc.* 20, 138–150. doi: 10.1007/s00146-005-0012-7
- Nomura, T., Suzuki, T., Kanda, T., Han, J., Shin, N., Burke, J., et al. (2008). What people assume about humanoid and animal-type robots: cross-cultural analysis between Japan, Korea, and the United States. *Int. J. Hum. Robot.* 5, 25–46. doi: 10.1142/S0219843608001297
- Piaget, J., Cook, M., and Norton, W. W. (1952). *The Origins of Intelligence in Children*, Vol. 8. New York, NY: International Universities Press.
- Public Attitudes Towards Robots (2012). *Special Eurobarometer 382*. Available at: <http://ec.europa.eu/publicopinion/archives/ebs/ebs382en.pdf>
- Riek, L. D. (2012). Wizard of Oz studies in HRI: a systematic review and new reporting guidelines. *J. Hum. Robot Interact.* 1, 119–136.
- Roberts, B. W., and DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: a quantitative review

- of longitudinal studies. *Psychol. Bull.* 126:3. doi: 10.1037/0033-2909.126.1.3
- Rothman, K. J. (1990). No adjustments are needed for multiple comparisons. *Epidemiology* 1, 43–46. doi: 10.1097/00001648-199001000-00010
- Schermerhorn, P., Scheutz, M., and Crowell, C. R. (2008). “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*, (New York, NY: Association for Computing Machinery), 263–270.
- Smarr, C.-A., Prakash, A., Beer, J. M., Mitzner, T. L., Kemp, C. C., and Rogers, W. A. (2012). “Older adults’ preferences for and acceptance of robot assistance for everyday living tasks,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 56, 153–157.
- Stafford, R. Q., Broadbent, E., Jayawardena, C., Unger, U., Kuo, I. H., Igic, A., et al. (2010). “Improved robot attitudes and emotions at a retirement home after meeting a robot,” in *Proceeding of the RO-MAN, 2010 IEEE*, (Viareggio: IEEE), 82–87.
- Stafford, R. Q., MacDonald, B. A., Li, X., and Broadbent, E. (2014). Older people’s prior robot attitudes influence evaluations of a conversational robot. *Int. J. Soc. Robot.* 6, 281–297. doi: 10.1007/s12369-013-0224-9
- Sullivan, G. M., and Feinn, R. (2012). Using effect size—or why the p value is not enough. *J. Grad. Med. Educ.* 4, 279–282. doi: 10.4300/JGME-D-12-00156.1
- Syrdal, D. S., Dautenhahn, K., Woods, S. N., Walters, M. L., and Koay, K. L. (2007). “Looking good? Appearance preferences and robot personality inferences at zero acquaintance,” in *AAAI Spring Symposium on Multidisciplinary Collaboration for Socially Assistive Robotics*, (Palo Alto, CA), 86–92.
- Tabachnick, B. G., and Fidell, L. S. (2001). *Using Multivariate Statistics*, 4th Edn. New York: Allyn & Bacon.
- Takayama, L., and Pantofaru, C. (2009). “Influences on proxemic behaviors in human-robot interaction,” in *Proceeding of the IEEE/RSJ International Conference on Intelligent Robots and Systems, 2009* (St. Louis: IEEE), 5495–5502.
- Tay, B., Jung, Y., and Park, T. (2014). When stereotypes meet robots: the double-edge sword of robot gender and personality in human–robot interaction. *Comput. Hum. Behav.* 38, 75–84. doi: 10.1016/j.chb.2014.05.014
- Vaughan, G., and Hogg, M. A. (2005). *Introduction to Social Psychology*. Canberra: Pearson Education Australia. Available at: <http://espace.library.uq.edu.au/view/UQ:40925>
- Walters, M. L., Syrdal, D. S., Dautenhahn, K., te Boekhorst, R., and Koay, K. L. (2007). Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. *Auton. Robots* 24, 159–178. doi: 10.1007/s10514-007-9058-3
- Waytz, A., Cacioppo, J., and Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspect. Psychol. Sci.* 5:3. doi: 10.1177/1745691610369336
- Wu, Y.-H., Cristancho-Lacroix, V., Fassert, C., Faucounau, V., de Rotrou, J., and Rigaud, A.-S. (2014). The attitudes and perceptions of older adults with mild cognitive impairment toward an assistive robot. *J. Appl. Gerontol.* doi: 10.1177/0733464813515092 [Epub ahead of print].
- Yamaoka, F., Kanda, T., Ishiguro, H., and Hagita, N. (2007). “Interacting with a human or a humanoid robot?,” in *Proceeding of the IEEE/RSJ International Conference on Intelligent Robots and Systems, 2007* (San Diego, CA: IEEE), 2685–2691.
- Yamazaki, R., Kuwamura, K., Nishio, S., Minato, T., and Ishiguro, H. (2014). Activating embodied communication: a case study of people with dementia using a teleoperated android robot. *Gerontechnology* 13:311. doi: 10.4017/gt.2014.13.02.166.00
- Yamazaki, R., Nishio, S., Ishiguro, H., Nørskov, M., Ishiguro, N., and Balistreri, G. (2012). “Social acceptance of a teleoperated android: field study on elderly’s engagement with an embodied communication medium in Denmark,” in *Proceedings of the 4th International Conference on Social Robotics*, (Berlin: Springer-Verlag), 428–437.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Damholdt, Nørskov, Yamazaki, Hakli, Vesterager Hansen, Vestergaard and Seibt. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Can We Talk through a Robot As if Face-to-Face? Long-Term Fieldwork Using Teleoperated Robot for Seniors with Alzheimer's Disease

Kaiko Kuwamura^{1,2,*}, Shuichi Nishio² and Shinichi Sato³

¹ Graduate School of Engineering Science, Osaka University, Osaka, Japan, ² Hiroshi Ishiguro Laboratory, Advanced Telecommunications Research Institute International, Keihanna Science City, Kyoto, Japan, ³ Graduate School of Human Sciences, Osaka University, Osaka, Japan

OPEN ACCESS

Edited by:

Stefan Kopp,
Bielefeld University, Germany

Reviewed by:

Kirsten Bergmann,
Bielefeld University, Germany
Karola Pitsch,
University of Duisburg-Essen,
Germany

*Correspondence:

Kaiko Kuwamura
kuwamura.kaikou@
irl.sys.es.osaka-u.ac.jp

Specialty section:

This article was submitted to
Human-Media Interaction,
a section of the journal
Frontiers in Psychology

Received: 03 December 2015

Accepted: 29 June 2016

Published: 19 July 2016

Citation:

Kuwamura K, Nishio S and Sato S
(2016) Can We Talk through a Robot
As if Face-to-Face? Long-Term
Fieldwork Using Teleoperated Robot
for Seniors with Alzheimer's Disease.
Front. Psychol. 7:1066.
doi: 10.3389/fpsyg.2016.01066

This work presents a case study on fieldwork in a group home for the elderly with dementia using a teleoperated robot called Telenoid. We compared Telenoid-mediated and face-to-face conditions with three residents with Alzheimer's disease (AD). The result indicates that two of the three residents with moderate AD showed a positive reaction to Telenoid. Both became less nervous while communicating with Telenoid from the time they were first introduced to it. Moreover, they started to use more body gestures in the face-to-face condition and more physical interactions in the Telenoid-mediated condition. In this work, we present all the results and discuss the possibilities of using Telenoid as a tool to provide opportunities for seniors to communicate over the long term.

Keywords: elderly care robot, teleoperated robot, Alzheimer's disease, elderly care facility, gerontology

1. INTRODUCTION

This work presents a case study on fieldwork in a group home for the elderly with dementia using a teleoperated robot. We developed a robot called Telenoid to provide communication support for seniors (**Figure 1**). Telenoid is a teleoperated robot covered with soft vinyl that can transmit a remote operator's physical movements and voice. Telenoid users can physically interact (hug and touch) with the robot while communicating with an operator who can communicate from a remote place through the Internet. From experiments in Japan and Denmark, we found that seniors quickly became fond of interaction with Telenoid, and seniors with dementia also liked it (Yamazaki et al., 2014). However, the effects of using it and how communication differs when talking through Telenoid compared to face-to-face communication are not clear.

In this paper, we describe a long-term fieldwork conducted in a group home (a community-based care facility where mild/moderate demented seniors live together) and compared face-to-face communication with communication mediated through Telenoid. We discuss the possibilities of using Telenoid as a tool to support long-term communication between people and elderly individuals.

1.1. Background

The population of senior citizens is rapidly increasing worldwide. In Japan, more than a quarter of the population is already over 65 (Cabinet Office, Government of Japan, 2014). The number



FIGURE 1 | Telenoid R3b.

of elderly with dementia has reached 4.6 million, and an additional 4 million people probably suffer from mild cognitive impairment (MCI). The Japanese Ministry of Health, Labor and Welfare estimates that the social cost of elderly with dementia was 14.5 trillion yen (approximately 118 billion US dollars) in 2014.

This trend, which is not specific to Japan, can also be seen globally (United Nations, 2013). In the more developed regions, populations aged 60 or over are expected to increase by 45% from 287 million in 2013 to 417 million in 2050. In the less developed regions, populations aged 60 or over are currently increasing even faster, and the numbers are expected to rise from 554 million in 2013 to 1.6 billion in 2050. With an increase of senior citizens, the number of people suffering from dementia is also likely to rise and will impose a severe social cost.

As societies continue to age, the number of seniors living alone will increase. Such changes limit opportunities to communicate with others and weaken their connection to society. Such limited society connections increase the risk of dementia (Fratiglioni et al., 2000). Furthermore, as the degrees of dementia progress, seniors become more withdrawn and experience more difficulty communicating with others including caregivers.

The most common cause of dementia is Alzheimer's disease (AD), which is perhaps responsible for up to 60–70% of all dementia cases (World Health Organization, 2015). AD is a chronic progressive neurodegenerative disorder characterized by the following symptoms: memory loss, language difficulty, executive dysfunction, psychiatric symptoms, such behavioral disturbances as depression, hallucinations, delusions, agitation,

and difficulty performing daily living activities (Burns and Iliffe, 2009). Seniors with AD sometimes reject care and become depressed or belligerent as a result of the behavioral and psychological symptoms of dementia (BPSD). They forget what they have done or said in the short term due to memory impairment. Understanding both the physical and mental conditions of seniors is important for taking care of them. However, accurately determining their mental conditions is difficult since identifying clues that might elucidate their emotional states when they are depressed are complicated. Therefore, it is important for caregivers to motivate seniors with AD to communicate to cope with BPSD and to suppress progress of dementia.

At the same time, the aging of society is exacerbating caregiver shortages. In fact, the lack of caregivers and their job turnover is already severe in both developed and developing countries (Kingma, 2007). According to a survey by a careworker foundation in Japan, 59.3% of caregivers feel overworked due to the actual lack of caregivers whose annual turnover rate has reached 16.5% (Care Work Foundation, 2015). Although the number of seniors who need care is increasing, the number of people who work as caregivers is decreasing, due to low wages (61.3%) and physically/mentally hard work (49.3%). Improving caregivers' working lives and motivating them is crucial (Lu et al., 2012).

The lack of caregivers makes caregivers busy and decreases opportunities for caregivers to communicate with residents. If seniors suffer from severe AD, they rarely respond to care. As a result, caregivers have difficulty communicating with their charges and become discouraged. To maintain their motivation, caregivers need skills and adequate time to properly communicate with seniors with dementia. However, this requires experience and training, and it is especially difficult for new/inexperienced caregivers who are often too busy to take time to communicate with their residents.

In Japan, there are volunteers who visit care facilities periodically to have conversation with residents. For smooth communication with the residents with AD, the volunteers need to be trained. Even though they provide opportunities for seniors to have conversation, they cannot attend the facilities every day. The volunteers usually belong to non-profit organizations and can only visit facilities near their houses occasionally. In the facility at which we conducted our experiment, volunteers only visit once or twice a month and talk with just a limited number of residents. Although there are telephones in houses or care facilities, residents with AD rarely use it to have conversation with others. This may be partially due to their weakened hearing ability by aging but also due to their lack of motivation to speak with others. With the progress of AD, one feels difficulty in composing and understanding dialogue properly. By recognizing the decline in their ability, residents with AD quickly lose their motivation to speak with others.

In this paper, we introduce a teleoperated robot Telenoid, which can be teleoperated from remote place. By using Telenoid, seniors living alone or in nursing homes will have more opportunities to communicate with their family or volunteers. The small and soft body of Telenoid allows people to hold

it while having conversation through it, allowing one to have communication with multiple modalities including visual and tactile sensations besides dialogue. Moreover, Telenoid's child-like appearance might attract residents and motivates them to communicate. If Telenoid can motivate residents to communicate, they will become more active or emotional, and caregivers will be able to understand their physical and mental conditions easier.

1.2. Related Works

Recently some attempts have started using information technologies and robots to increase the opportunities for seniors to communicate. One example is the Mobile Robotic Telepresence (MRP) system, which is a video conferencing system mounted on a mobile robotic base. It allows users to telecommunicate with residents from remote locations, and several researches have been carried out with it (Beer and Takayama, 2011; Orha and Oniga, 2012; Kristoffersson et al., 2013). Kuwahara et al. (2006) developed networked reminiscence therapy, which effectively increases the self-esteem of and reduces the behavioral disturbances in seniors with dementia (Kuwahara et al., 2006). Their system combines IP video phones with a photo- and video-sharing facility. In their experimental results, elderly with dementia communicated with therapists by videophone, and networked reminiscence sessions were generally as successful for individuals with dementia as face-to-face reminiscence sessions. We also tried to introduce tablets and video chat to the residents who showed interest in such new devices. However, they soon returned them to us. Although they seemed willing to directly communicate with others, they were discouraged from using such communication tools as phones or video chat. We believe that to increase the opportunities for seniors to communicate, it is important to not just introduce a communication device but also to motivate them to use it.

Perhaps the most famous elderly care companion robot is Paro, a baby seal robot designed for therapy (Wada et al., 2005). It has sensors on its body and reacts with sound and several actuators. Its cute appearance and behavior stimulates the interest of the elderly. Compared to the resident dog, the residents who interacted with Paro significantly felt less loneliness, and they also talked to it and touched it more than the resident dog (Robinson et al., 2013). From seniors with mild/moderate dementia, Paro evokes natural expressions more frequently than stuffed animals and is likely to increase the willingness of the staff members to communicate and work with elderly people with dementia (Takayanagi et al., 2014). However, since it is not designed for verbal communication, seniors talk to Paro, which reacts but cannot have a conversation.

To introduce a robot to elderly care houses, caregivers must constantly use it and residents must be discouraged from losing interest in it. Manuals for use and introduction in care facilities exist for Paro (Wada et al., 2010), and Kanagawa Prefecture in Japan also provides support for introducing robots into care facilities. These allow users to properly employ such robots; otherwise, users will lose interest and stop using them. Tanaka et al. (2007) updated the behavior of a robot called QRIO

during trials to maintain the interest of a classroom of toddlers. Otherwise, children seldom reacted to it. Users might lose interest in robot because of low intelligence, or few variety of reaction in the robot. Sabelli et al. (2011) placed a robot called Robovie2, which was remotely controlled by an operator, in an elderly care center for 3.5 months. Through the ethnographic study, they found that the robot was accepted in the community. However, they provided only ethnographical descriptions and performed no statistical data analysis. As such, although there have been trials to use robots in care facilities for rather long duration, study with objective measurements have been missing and effective methodologies for utilizing robots while keeping people's interest have been unclear.

From experiments of Telenoid in Japan and Denmark, we found that seniors with dementia often showed strong attachment to and liked to communicate with Telenoid (Yamazaki et al., 2014). Although it is difficult to communicate with seniors with dementia, school children were able to communicate with the residents without training by using Telenoid (Yamazaki et al., 2013). We found that Telenoid could motivate seniors with dementia to have conversation with others, while making people talking through Telenoid to be much relaxed compared to face-to-face. However, the quality of the conversation and how third person such as caregivers observing the interaction feels are unrevealed. Also, how people's response to Telenoid changes in longer term is not clear.

In this paper, we described a long-term fieldwork conducted in a group home (a community-based care facility where mild/moderate demented seniors live together) and compared face-to-face communication with communication mediated through Telenoid. We evaluated the quality of the conversation by questionnaire. The questionnaire was answered by the speaker and the observer to reveal the effect of third person. We discussed the possibilities of using Telenoid as a tool to support long-term communication between people and elderly individuals.

2. METHODS

2.1. Participants and Ethics Statement

Three female residents (from 85 to 96 years old) of a senior group home participated in this study. They were all clinically diagnosed as AD. Informed consent was obtained from the group home manager, the doctor in charge, and the participant families. This experiment was approved by the Human Ethics Committee of the Graduate School of Human Sciences, Osaka University (No. 26-60), and the Ethics Committee of the Advanced Telecommunications Research Institute International (No. 14-602-3).

2.2. Procedure

The experiments were conducted once or twice a week for 3 months in a group home for seniors with dementia in Osaka, Japan. The dates and times of the trials were adjusted based on the conditions of the participants and the convenience of the group home. All conversations were exchanged in a public space, either in the dining room or the TV room.

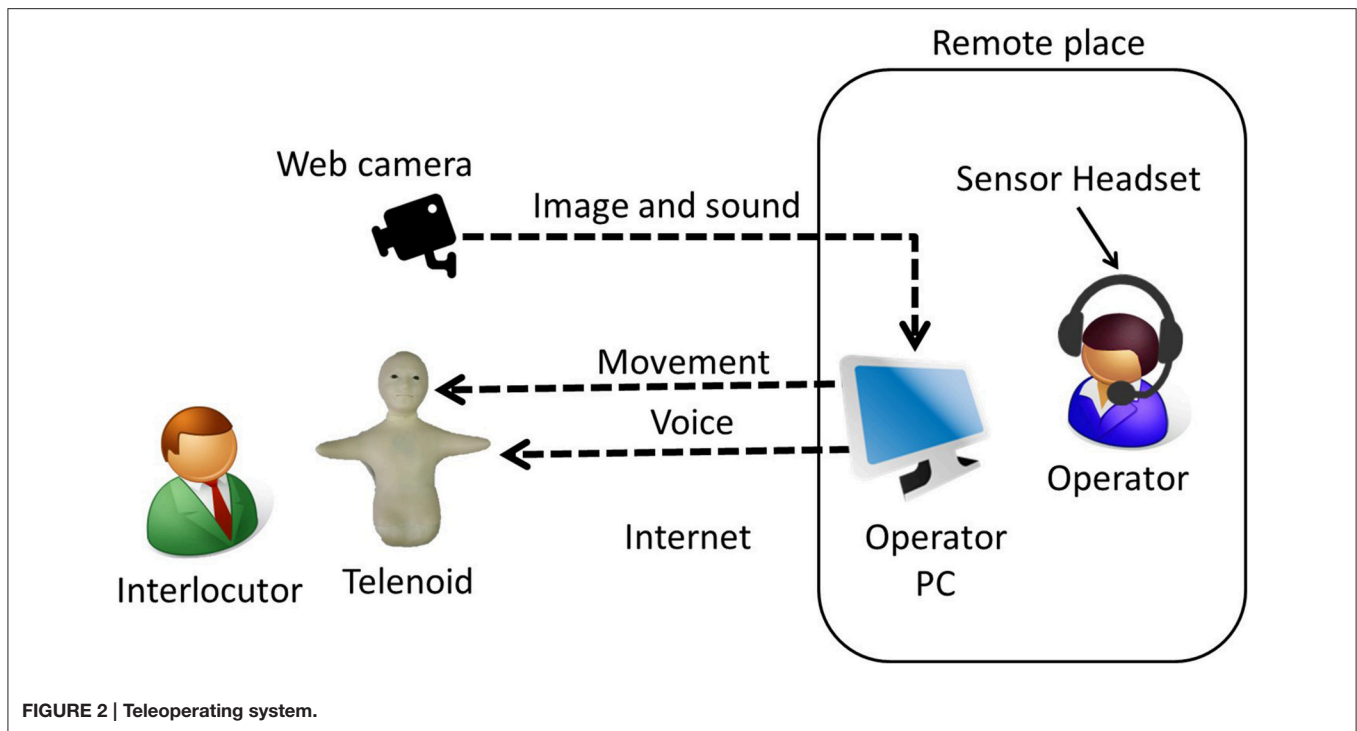


FIGURE 2 | Teleoperating system.

Participants spoke with a person (henceforth *speaker*) in a face-to-face condition (Face condition) and a Telenoid-mediated condition (Telenoid condition). The conditions were randomly ordered and the duration of the conversations was limited to 15 min each. The conversations were suspended when the participant was not feeling well or was unwilling to talk. An *observer* monitored the interaction between the participant and the speaker in both conditions. After both conditions were conducted, the speaker and the observer answered questionnaires. We recruited five university students who major in gerontology as evaluators. None of the evaluators had experience of using robots. They played the speaker and observer roles in turn. We asked them to make evaluation in the quality of conversation and made no further specific instructions.

In the Telenoid condition, the speakers controlled a Telenoid R3b (Figure 1) to communicate with the elderly participants by a teleoperation system from a remote location (Figure 2). Another experimenter first carried Telenoid and sat in front of the participant. During the conversation, the experimenter gave Telenoid to the participant, and if the participant did not refuse it, the participant held it and continued the conversation. When participants held Telenoid, they put it on their laps and sometimes leaned it against a desk. Telenoid has six independent actuators (jaw movement, yaw, pitch, and roll movement for its neck and horizontal movements for each arm) that allow it to synchronize motion with the speaker. The speaker's head motion is captured by sensors (three-axis accelerometer and three-axis magnetometer) embedded in a headset and transmitted to the robot. Speech-driven lip

motion generation, which creates lip motions from the speaker's vocal information, is used to control Telenoid's jaw movement (Ishi et al., 2011).

2.3. Evaluation

2.3.1. Diagnosis of Dementia

The caregivers of the group home answered the following cognitive function tests before and after the experiment. We used these tests to measure the cognitive function of the participants and AD's progress during the experiment.

1. Mini-Mental State Examination (MMSE): 30-point questionnaire that is used extensively in clinical and research settings to measure cognitive impairment (Pangman et al., 2000). Any score greater than or equal to 27 points (out of 30) indicates normal cognition. Scores below indicate severe (≤ 9 points), moderate (10–18 points), or mild (19–24 points) cognitive impairment (Mungas, 1991).
2. Quality of life questionnaire for dementia (QOL-D): 31 items grouped into six response sets to measure six domains of health-related QOL (Terada et al., 2002).
3. Dementia Behavior Disturbance Scale (DBD): 28 items, measured by the frequency of BPSD on a five-point scale (Mizoguchi et al., 1993).
4. Japanese version of the Neuropsychiatric Inventory (NPI-NH): measures 12 symptoms of neuropsychiatric disturbances (Hirono et al., 1997).
5. Barthel Index (BI): measures performances of activities of daily living (ADL) by 10 items (Shah et al., 1989). A total BI score of 0–20 suggests complete dependence, 21–60 indicates

TABLE 1 | Diagnosis of dementia test result of Ms. A.

		Before (11/13/2014)	After (3/26/2015)
MMSE		12/30	13/30
QOL-D	Positive affect	28/28	28/28
	Negative affect and actions	8/24	7/24
	Ability of communication	20/20	20/20
	Restlessness	8/20	7/20
	Attachment with others	10/16	14/16
	Spontaneity and activity	14/16	13/16
DBD		13/112: No major problem in mental and behavioral disorder	8/112: No major problem in mental and behavioral disorder
NPI-NH		Agitation/aggression Frequency 1, severity 1, caregiver distress 1	None
BI		45/100	45/100
VTI		8/10	8/10

severe dependence, 61–90 indicates moderate dependence, and 91–99 indicates slight dependence.

6. Vitality Index (VTI): measures vitality related to ADL in elderly patients with dementia by five subscales (Toba et al., 2002).

2.3.2. Questionnaire

The speaker and the observer filled out the following questionnaire, where each item was rated on a five-point scale:

- Q1 Smoothness of conversation (rough-smooth)
- Q2 Amount of conversation (poor-rich)
- Q3 Quality of conversation (low-high)
- Q4 Impression of participant (gloomy-cheerful)
- Q5 Emotional state of speaker (nervous-relaxed)
- Q6 Emotional expression of speaker (poor-rich)
- Q7 Understanding participant (not understood-understood)

Items in the questionnaire were listed to measure the quality of the conversation. Q1–3 measures the quality of the conversation more quantitative, and Q4–7 measures the impression of the residents and speaker more qualitative. We included these items to measure whether residents were motivated to communicate, and to measure the impression of observer observing the conversation.

Hereafter, we denote a speaker's response to Q_n as Sp_Q_n and an observer's response to Q_n as Ob_Q_n. The questionnaire scores were compared between the Telenoid and Face conditions within subjects by paired *t*-tests to reveal the effect of using Telenoid. We compared the scores of the first and last five trials in each condition (by Student's *t*-test when homoscedasticity was confirmed and Welch's *t*-test when unconfirmed) to determine any long-term effects.

2.3.3. Video Analysis

A surveillance camera in each room (the dining and TV rooms) and one mobile camera were used to record the interactions. From the video recordings, we counted the number of times that the participants used body gestures and made physical contact. Due to limited views, we counted only the number of clear upper body gestures and physical contacts. For control between the Telenoid and Face conditions, hugs in the Telenoid condition were excluded from gestures and physical contacts.

We used a paired *t*-test between the Telenoid and Face conditions within subjects to reveal the behavioral differences using Telenoid. We also compared the frequency of such behaviors of the first and last five trials in each condition (by Student's *t*-test when homoscedasticity was confirmed and Welch's *t*-test when unconfirmed) to determine the long-term effect.

3. RESULTS

We conducted 10 trials (interactions) for each participant. The average duration of an interaction was 709.1 s (*SD* = 316.2) for the Face condition and 798.7 s (*SD* = 383.3) for the Telenoid condition. The Telenoid condition time was longer because residents kept talking to Telenoid even after they were informed of the experiment's end.

3.1. Ms. A: 96 years old

3.1.1. Diagnosis of Dementia

Ms. A was diagnosed as AD in 2006. The test results for the diagnosis of dementia before and after the experiment are shown in Table 1.

Her MMSE score were 12 (before) and 13 (after), indicating that Ms. A had moderate dementia. However, BPSD, which was previously observed when she was staying at home and in another geriatric health service facility, did not appear in

TABLE 2 | Questionnaire results for trials with Ms. A.

				Telenoid		Face	
		Telenoid	Face	First half	Last half	First half	last half
Speaker	Q1	3.6 (0.84)	4.1 (0.57)	3.8 (0.84)	3.4 (0.89)	3.8 (0.45)	4.4 (0.55)*
	Q2	3.6 (0.84)	3.7 (0.95)	3.6 (0.55)	3.6 (1.14)	3.6 (0.89)	3.8 (1.10)
	Q3	3.4 (0.70)	3.4 (0.70)	3.6 (0.55)	3.2 (0.84)	3.2 (0.84)	3.6 (0.55)
	Q4	4.4 (0.70)	3.6 (0.70) **	4.6 (0.55)	4.2 (0.84)	3.6 (0.55)	3.6 (0.89)
	Q5	3.7 (0.82)	3.7 (0.67)	3.6 (0.55)	3.8 (1.10)	3.4 (0.89)	4.0 (0.00)
	Q6	3.3 (1.25)	3.4 (0.70)	3.2 (1.10)	3.4 (1.52)	3.0 (0.71)	3.8 (0.45)*
	Q7	2.8 (0.63)	2.7 (0.95)	2.8 (0.45)	2.8 (0.84)	2.4 (1.14)	3.0 (0.71)
Observer	Q1	4.1 (0.74)	4.0 (0.82)	3.8 (0.84)	4.6 (0.55)	3.6 (0.55)	4.6 (0.89)*
	Q2	4.0 (0.94)	3.9 (0.57)	3.6 (1.14)	4.0 (1.22)	3.6 (0.55)	4.2 (1.30)
	Q3	3.1 (0.88)	3.8 (1.03) ***	2.8 (0.84)	3.6 (0.89)	3.6 (1.14)	4.0 (1.00)
	Q4	4.5 (0.71)	3.5 (0.85) ***	4.8 (0.45)	4.4 (0.55)	3.4 (1.14)	3.6 (1.14)
	Q5	4.0 (0.94)	3.6 (0.70)	4.2 (1.10)	3.4 (0.55)	3.2 (0.45)	3.4 (1.14)
	Q6	3.5 (1.08)	2.9 (0.57)	3.8 (0.84)	3.4 (1.14)	2.8 (0.45)	3.0 (0.71)
	Q7	3.3 (0.82)	3.5 (0.53)	3.4 (0.55)	3.2 (0.84)	3.2 (0.45)	3.4 (0.55)

Left column indicates overall comparison results between the Telenoid condition and the Face condition; Righthand two columns indicate first/last half period summary for each condition. Values in the table indicates: mean score, SD (in parenthesis), t-test result where * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

the current group home. During the experiment, an episodic memory disorder was discovered in Ms. A. No other remarkable cognitive impairments were found.

We conducted several trials, but she did not remember what she had experienced in the previous meetings with Telenoid. In the Face condition she tended to describe the pleasure of her past in a vivid manner. The content of the conversation was only about her past, and not much about the speaker. She did not remember the recent news, and showed a gloomy look on her face when she talked about it. When interacting through Telenoid she seemed to consider that the robot was a child, then she became expressive and started talking aloud with Telenoid. When she talked to Telenoid, she asked about what it wanted to be in the future, displaying conversation fluency. She tended to physically interact with Telenoid by giving hugs and kisses, and touching head to head. Such physical behaviors were not found in the Face condition.

3.1.2. Questionnaire

The questionnaire results are shown in **Table 2**. Comparing the averages from the Telenoid and Face conditions, we found significant differences in Sp_Q4 (Telenoid > Face, $t = -2.75$, $p < 0.05$), Ob_Q3 (Telenoid < Face, $t = 3.28$, $p < 0.01$), and Ob_Q4 (Telenoid > Face, $t = -3.87$, $p < 0.01$). Comparison between the first/last halves showed differences in Face condition's Sp_Q1 ($t = -1.90$, $p < 0.10$), Sp_Q6 ($t = -2.14$, $p < 0.10$), and Ob_Q1 ($t = -2.13$, $p < 0.10$). These results showed improvement in the communication in the later five trials.

3.1.3. Video Analysis

We used a paired t -test between the Telenoid and Face conditions and found significant differences for the frequency of gesture

(Telenoid < Face, $t = 3.75$, $p < 0.01$), and the frequency of physical contact (Telenoid > Face, $t = -5.40$, $p < 0.01$; **Figures 3, 4**). We did not find significant differences for the frequency of gestures or physical contact between the first and last five trials.

3.2. Ms. B: 93 years old

3.2.1. Diagnosis of Dementia

Ms. B was diagnosed as AD in 2010. The test results for the diagnosis of dementia before and after the experiment are shown in **Table 3**.

Ms. B had a gentle personality, but sometimes she rejected care and had problems with other residents and caregivers. She had severe episodic memory disorder and rarely remembered what she experienced in previous meetings with Telenoid and speakers. Her MMSE scores were 17 (before) and 14 (after), which indicates that she had moderate dementia. Mental and physical problems were rarely found by the tests, and she was generally calm during the experiments. She talked about herself in the Face condition, while asking more questions and making physical contact in the Telenoid condition.

3.2.2. Questionnaire

The questionnaire results are shown in **Table 4**. Comparing the averages from the Telenoid and Face conditions, we found significant trends in Sp_Q4 (Telenoid > Face, $t = -1.92$, $p < 0.10$), Sp_Q6 (Telenoid > Face, $t = -1.86$, $p < 0.10$), Ob_Q1 (Telenoid > Face, $t = -2.06$, $p < 0.10$), Ob_Q4 (Telenoid > Face, $t = -2.21$, $p < 0.10$), and Ob_Q5 (Telenoid > Face, $t = -2.23$, $p < 0.10$). We also found significant differences in Ob_Q6 (Telenoid > Face, $t = -2.69$, $p < 0.05$). Comparison between the first/last halves showed significant trends in Telenoid condition's Ob_Q3 ($t = -2.14$, $p < 0.10$) and significant

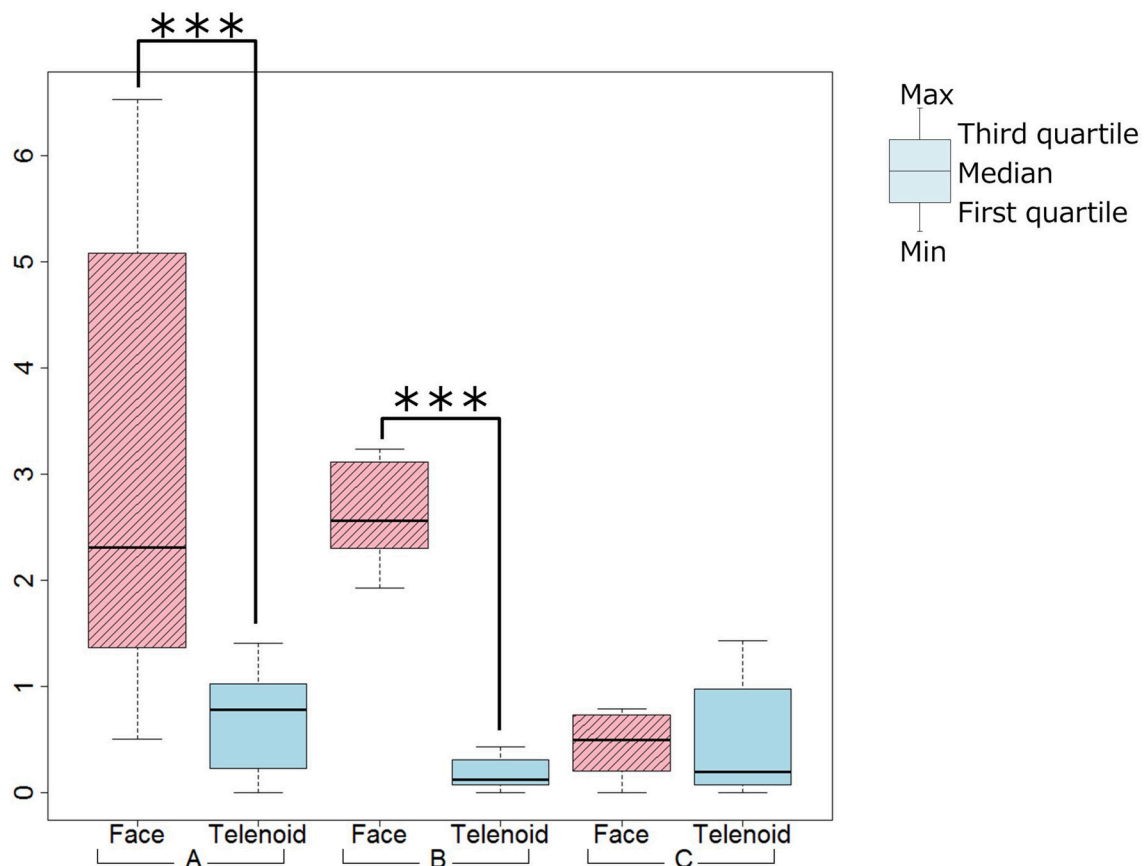


FIGURE 3 | Gesture tendency (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$).

differences in Face condition's Sp_Q7 ($t = -2.36$, $p < 0.05$). These results showed improvement in communication in the last five trials.

3.2.3. Video Analysis

We used a paired t -test between the Telenoid and Face conditions and found significant differences for the frequency of gestures (Telenoid < Face, $t = 11.09$, $p < 0.01$), and the frequency of physical contact (Telenoid > Face, $t = -4.89$, $p < 0.01$; Figures 3, 4). We did not find any significant differences for the frequency of gestures or physical contact between the first and last five trials.

3.3. Ms. C: 85 years old

3.3.1. Diagnosis of Dementia

Ms. C was diagnosed as AD in 2004. Her test results for the diagnosis of dementia before starting the experiments are shown in Table 5. Ms. C was transferred to a special nursing home for the elderly at the end of the experiment and could conduct the test after the experiment. Group home for the elderly with dementia is usually for the seniors with mild dementia, who need a little support to live by themselves. Ms. C was in the home because there was no spare room in the special nursing home at the

beginning of the experiment. She moved to the special nursing home when there was a spare room.

Her MMSE score was 0, indicating severe dementia. She tended to make ambiguous statements and repeat the same phrases. Verbal communication was difficult with her; however, she did not often show a problematic BPSD, and the caregiver distress points were not high. She held eye contact in the Face condition; however, the content of her conversation was difficult to understand. Similar behavior was observed in the Telenoid condition. But she played peekaboo with Telenoid, suggesting that she thought she was interacting with a baby.

3.3.2. Questionnaire

Since Ms. C was transferred to a special nursing home for the elderly at the end of the experiment, we could not measure the diagnosis of dementia after the experiment for Ms. C. However, the questionnaire result and video analysis result during the experiment is measured in a same way as Ms. A and Ms. B.

The questionnaire results are shown in Table 6. Comparing the averages from the Telenoid and Face conditions, we found significant differences in Sp_Q2 (Telenoid < Face, $t = 2.45$, $p < 0.05$). Comparison between the first/last halves showed significant differences in Face condition's Ob_Q1 ($t = -2.75$,

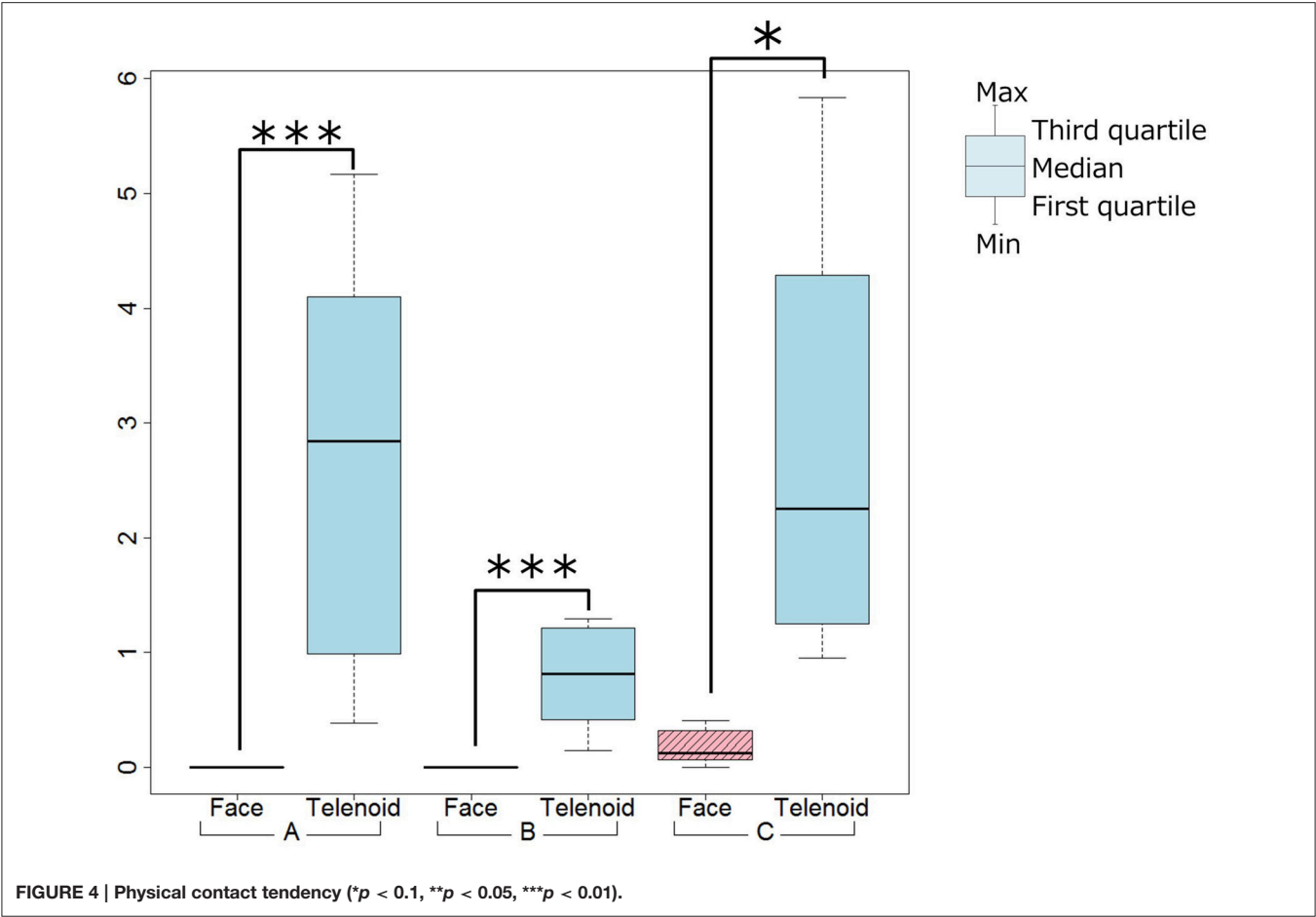


TABLE 3 | Diagnosis of dementia test result of Ms. B.

		Before (9/4/2014)	After (3/26/2015)
MMSE		17/30	14/30
QOL-D	Positive affect	28/28	20/28
	Negative affect and actions	12/24	10/24
	Ability of communication	20/20	20/20
	Restlessness	11/20	8/20
	Attachment with others	16/16	16/16
	Spontaneity and activity	16/16	16/16
DBD		25/112: Defect of memory and fecal incontinence	26/112: Defect of memory and fecal incontinence
NPI-NH		None	Agitation/aggression Frequency 4, severity 1, caregiver distress 1 Anxiety Frequency 4, severity 1, caregiver distress 2
BI		85/100	85/100
VTI		8/10	9/10

TABLE 4 | Questionnaire results for trials with Ms. B.

		Telenoid				Face	
		Telenoid	Face	First half	Last half	First half	Last half
Speaker	Q1	3.4 (1.07)	3.7 (0.82)	3.4 (1.14)	3.4 (1.14)	3.4 (0.89)	4.0 (0.71)
	Q2	3.5 (1.08)	3.5 (0.85)	3.6 (0.89)	3.4 (1.34)	3.2 (0.84)	3.8 (0.84)
	Q3	3.3 (0.67)	3.1 (0.74)	3.0 (0.71)	3.6 (0.55)	3.2 (0.84)	3.0 (0.71)
	Q4	4.1 (0.88)	3.3 (0.67)*	4.0 (0.71)	4.2 (1.10)	3.2 (0.45)	3.4 (0.89)
	Q5	2.9 (0.88)	3.1 (0.99)	3.0 (0.71)	2.8 (1.10)	3.0 (1.00)	3.2 (1.10)
	Q6	3.6 (0.84)	3.1 (0.88)*	3.4 (0.89)	3.8 (0.84)	3.2 (0.84)	3.0 (1.00)
	Q7	3.0 (0.82)	3.3 (0.82)	3.0 (1.00)	3.0 (0.71)	2.8 (0.84)	3.8 (0.45)**
Observer	Q1	4.0 (0.94)	3.2 (1.03)*	4.2 (1.10)	3.8 (0.84)	3.2 (0.84)	3.2 (1.30)
	Q2	3.7 (0.95)	3.4 (0.97)	3.8 (1.30)	3.6 (0.55)	3.2 (0.84)	3.6 (1.14)
	Q3	3.4 (0.70)	3.3 (0.67)	3.0 (0.71)	3.8 (0.45) *	3.0 (0.00)	3.6 (0.89)
	Q4	4.1 (0.74)	3.2 (1.03)*	3.8 (0.84)	4.4 (0.55)	2.8 (0.84)	3.6 (1.14)
	Q5	3.7 (0.82)	2.9 (0.74)*	4.0 (0.71)	3.4 (0.89)	2.6 (0.55)	3.2 (0.84)
	Q6	3.5 (0.71)	2.8 (0.63)**	3.6 (0.89)	3.4 (0.55)	2.6 (0.55)	3.0 (0.71)
	Q7	3.1 (0.74)	3.3 (0.48)	2.8 (0.84)	3.4 (0.55)	3.2 (0.45)	3.4 (0.55)

Left column indicates overall comparison results between the Telenoid condition and the Face condition; Righthand two columns indicate first/latter half period summary for each condition. Values in the table indicates: mean score, SD (in parenthesis), t-test result where * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

$p < 0.05$), Ob_Q2 ($t = -5.77$, $p < 0.01$), and Ob_Q3 ($t = -2.89$, $p < 0.05$). These results showed improvement in the communication in the last five trials.

3.3.3. Video Analysis

We used a paired t -test between the Telenoid and Face conditions and did not find significant differences for the frequency of gestures. However, we did find a significant trend in the frequency of physical contact (Telenoid > Face, $t = -2.06$, $p < 0.10$; **Figures 3, 4**). We also did not find significant differences in the frequency of gestures or physical contact between the first and last five trials.

4. DISCUSSION

4.1. Ms. A

When we compared the scores between the Telenoid and Face conditions, both Q4s from the speakers and observers were significantly positive for the Telenoid condition (**Table 2**). This means that Ms. A showed a more positive reaction when talking to Telenoid than talking face-to-face. In the Telenoid condition, she changed her voice tone as if talking to a child. She seemed to treat Telenoid like a child, which allowed her to communicate in a more relaxed manner, leading to a positive Q4 score for the Telenoid condition. In fact, there were comments on the questionnaire. Immediately after she met Telenoid, she said, "You are so cute. I love you." Whereas in the face-to-face condition, even though she seemed nervous at the beginning of the interaction, she gradually managed to have a smooth conversation. We compared the questionnaire scores for the first and last five trials. In the Face condition, Sp_Q1, Sp_Q6, and Ob_Q1 had significant differences; they increased in the latter trials. The participant talked cheerfully with Telenoid from the

beginning and did not have any significant differences between the first and latter five trials.

For Ob_Q3 (Quality of conversation), the Telenoid condition's score was negative compared with that in the Face condition. This might be because the participant recognized Telenoid as a child and the conversations content was playful. From the video analysis results, the participant tended to make physical contact in the Telenoid condition and used gestures in the Face condition. This indicates that she used physical interactions with Telenoid instead of verbal communication, as if taking care of a child. In fact, she tended to physically interact with Telenoid by hugs and kisses and touching its head. Such physical behaviors were not found in the Face condition.

4.2. Ms. B

When we compared the Telenoid and Face conditions, both Q4s and Q6s from the speakers and observers were significantly or marginally positive for the Telenoid condition (**Table 4**). The speakers also often adapted to the participants by changing their voice using a voice changer to sound more like a child.

The video analysis showed that in the Telenoid condition the participant made more physical contact, which was rarely observed in the Face condition. This was expected since physical interactions are usually only held among close relations. The speaker observed such interactions through the monitor, which might cause her to have better conversations with more emotional expressions. During the conversation, Ms. B seemed to interact with Telenoid as if it were a child, as in the case of Ms. A. Ms. B became calm when talking with Telenoid, which might explain the positive result in Q4 in the Telenoid condition. In fact, several questionnaire comments said that the participant seemed to become nervous at the beginning of the interaction in the Face condition with less eye contact, while conversely other

TABLE 5 | Diagnosis of dementia test result of Ms. C .

Before (9/4/2014)		
MMSE		0/30
QOL-D	Positive affect	23/28
	Negative affect and actions	15/24
	Ability of communication	6/20
	Restlessness	5/20
	Attachment with others	12/16
	Spontaneity and activity	6/16
DBD		22/112: Apathy, refusal, and incontinence were found
NPI-NH	Hallucinations	Frequency 4, severity 1, caregiver distress 0
	Agitation/aggression	Frequency 4, severity 1, caregiver distress 1
	Anxiety	Frequency 4, severity 1, caregiver distress 1
	Apathy	Frequency 4, severity 1, caregiver distress 0
	Disinhibition	Frequency 4, severity 1, caregiver distress 1
	Irritability	Frequency 4, severity 3, caregiver distress 1
	Aberrant motor behavior	Frequency 4, severity 1, caregiver distress 1
BI		45/100
VTI		8/10

comments said that the participant was relaxed and smiled more often to Telenoid.

We found that the emotional state of the speaker (Q6) became positive because the speaker experienced a more positive reaction from Ms. B through Telenoid. Telenoid affected the participant positively, resulting in a different quality of interaction, which the speaker enjoyed. Thus, Telenoid improved the conversation of both the participant and the speaker. There were also positive face-to-face conversations between Ms. B and the speaker; however, in the Telenoid condition the speaker observed the interactions from a third-person point of view, which allowed the speaker to participate in conversations objectively and have more positive feedback than in the Face condition.

4.3. Ms. C

When comparing the questionnaire scores for the first and last five trials, Face condition's Ob_Q1, Ob_Q2, and Ob_Q3 had significantly positive points for the latter half (Table 6). This suggests that the speaker adapted to the participant in the latter half, although it was difficult at the beginning.

Compared with the Face condition, Sp_Q2 (Amount of conversation) was significantly negative in the Telenoid condition. For the participant who had difficulty in the conversations, non-verbal information becomes more important. In the Telenoid condition, the speaker operating Telenoid only received limited information through the camera. The limited information may cause difficulty for the speaker during conversation, lowering scores. One of the speaker's comments on the questionnaire said, "Non-verbal information, like holding

hands and eye contact, is important, but communicating this through Telenoid was difficult." There were no such comments by the observers.

In the video analysis, we found no significant differences between the Telenoid and Face conditions for the frequency of gesture tendency, while the frequency of physical contact was significantly higher for the Telenoid condition. This indicates that the participant was also attempting to have non-verbal communication with Telenoid, the same as in the Face condition. Therefore, the speaker's questionnaire scores might rise by improving the Telenoid operating system to support more non-verbal communication. The results also indicate that Telenoid might be a viable platform for communicating with seniors with severe dementia.

4.4. Overall Discussion

All three participants tended to have more physical contact in the Telenoid condition. This result also implies that participants interacting with Telenoid were less nervous from the beginning of the conversation. They treated Telenoid as a child, which is huggable and easier to touch. Since it is huggable, they felt free to interact with it from the beginning.

The results suggest that because Telenoid has a physical presence, the elderly can hold it and they also like its child-like appearance. We believe such results cannot be seen by existing robots, including telepresence robots or Paro. To support communication by robots, especially for seniors with dementia, the robots appearance has to be in a form that the elderly can recognize and talk to at a relatively close distance that simplifies physical interaction. The close distance allows elderly to recognize a robot easily and enable to touch, which is important to establish a good relationship (Caris-Verhallen et al., 1999).

The Q4 scores (participant's impression) from both Ms. A and Ms. B supported the Telenoid condition. Ms. A and Ms. B tended to make more gestures in the Face condition. The reason might be because the participants had difficulty moving their upper body to make gestures while holding Telenoid.

Compared with Ms. A and Ms. B, since Ms. C has severe AD, verbal communication is more difficult with her. Ms. C tended to use more gestures in conversation and showed no significant differences in gesture tendency between the Telenoid and Face conditions.

5. CONCLUSION

We discussed the possibility of introducing a teleoperated robot into an elderly care house for long-term interaction. We compared two conversation conditions: face-to-face and using a teleoperated robot, Telenoid. Our experiment results showed that two participants with moderate AD had positive reactions from talking with Telenoid. The result supports the previous research about positive reaction of elderly using Telenoid (Yamazaki et al., 2014), and moreover, we found the result compared to face-to-face communication for long term.

The third participant had severe AD, and it was difficult to verbally communicate with her. However, she interacted

TABLE 6 | Questionnaire results for trials with Ms. C.

		Telenoid				Face	
		Telenoid	Face	First half	Last half	First half	Last half
Speaker	Q1	2.2 (1.03)	2.8 (0.92)	2.4 (1.14)	2.0 (1.00)	2.8 (1.10)	2.8 (0.84)
	Q2	2.2 (1.14)	3.0 (1.05)**	2.4 (1.52)	2.0 (0.71)	2.8 (1.30)	3.2 (0.84)
	Q3	2.0 (0.94)	2.2 (0.92)	2.2 (1.30)	1.8 (0.45)	2.4 (1.14)	2.0 (0.71)
	Q4	3.8 (1.23)	3.6 (0.52)	3.4 (1.34)	4.2 (1.10)	3.6 (0.55)	3.6 (0.55)
	Q5	3.1 (0.88)	3.0 (0.94)	3.2 (0.84)	3.0 (1.00)	2.6 (0.89)	3.4 (0.89)
	Q6	2.7 (0.95)	2.2 (0.63)	2.6 (1.14)	2.8 (0.84)	2.0 (0.71)	2.4 (0.55)
	Q7	2.3 (0.95)	2.1 (0.99)	2.2 (1.30)	2.4 (0.55)	1.8 (0.84)	2.4 (1.14)
Observer	Q1	2.3 (0.95)	2.7 (1.06)	2.2 (0.84)	2.4 (1.14)	2.0 (1.00)	3.4 (0.55)**
	Q2	2.1 (0.88)	2.6 (1.17)	2.0 (1.00)	2.2 (0.84)	1.6 (0.55)	3.6 (0.55)***
	Q3	2.3 (1.06)	2.1 (0.74)	2.0 (1.00)	2.6 (1.14)	1.6 (0.55)	2.6 (0.55)**
	Q4	3.5 (0.85)	3.3 (0.95)	3.8 (0.84)	3.2 (0.84)	3.4 (1.14)	3.2 (0.84)
	Q5	2.9 (0.99)	2.7 (0.67)	3.0 (1.00)	2.8 (1.10)	2.6 (0.55)	2.8 (0.84)
	Q6	2.7 (0.82)	2.3 (0.67)	2.8 (0.84)	2.6 (0.89)	2.4 (0.55)	2.2 (0.84)
	Q7	2.4 (0.70)	2.6 (0.70)	2.6 (0.55)	2.2 (0.84)	2.6 (0.89)	2.6 (0.55)

Left column indicates overall comparison results between the Telenoid condition and the Face condition; Righthand two columns indicate first/latter half period summary for each condition. Values in the table indicates: mean score, SD (in parenthesis), t-test result where * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

with Telenoid using non-verbal communication in a way that resembled the face-to-face condition. Thus, we conclude that Telenoid may trigger positive emotions in residents with moderate AD and suggest the possibilities of non-verbal communication with residents with severe AD as well.

To introduce a robot to elderly care houses, caregivers must constantly use it and residents must be discouraged from losing interest in it. We compared the questionnaire results of the first and latter five trials and found significant differences or tendencies for five items in the Face condition and one item in the Telenoid condition, indicating that in the Face condition, people had better conversations as the experiment went on, and in the Telenoid condition, the quality of the conversation remained high. As for the face-to-face conversation, we believe this is because both the seniors and the speakers felt nervous at the beginning and took time to have effective conversations. On the other hand, people communicated smoothly through Telenoid from the beginning. The Telenoid condition had fewer items to improve in the latter five trials; however, no item worsened. This indicates that Telenoid did not lose the interest of the residents, not even at the end of the experiment. The robot we used in this study, Telenoid, is teleoperated and the operator can behave and speak in a variety of ways. Such nature of Telenoid may make it more alive, and interacting with Telenoid will likely to appear to be closer to human-human interaction than other robots such as Paro, QRIO, or Robovie as mentioned in the previous section. Since the state-of-art of artificial intelligence technology is quite limited, especially for having conversation with people, the teleoperation system used in Telenoid seems to be a very effective and practical solution.

The Telenoid users monitored the positive reactions of participants through a camera. The speaker may become motivated to better care for the patients by watching such

interactions that cannot be seen in face-to-face communication. If caregivers were to use Telenoid, they might become emotionally expressive and enjoy conversations with seniors, boosting their motivation to care for those living with dementia. Observing the residents from a third-person point of view and communicating in a manner that is not possible face-to-face might improve caregiver attitudes, resulting in better relationships and an improved atmosphere in the facility. This could help caregivers and facility residents get to know each other better and eventually lower the turnover rate for the former.

If seniors suffer from severe AD, they rarely respond to care. As a result, caregivers have difficulty communicating with their charges and become discouraged. Observer's questionnaire result shows that the impression of residents with mild AD will become more cheerful when talking to Telenoid. This indicates that caregivers observing the interaction between Telenoid and the residents can notice the cheerful behavior of the residents, which might motivates caregivers. Also, if the caregiver met the resident for first time, the caregiver might have difficulty talking to the resident. By using Telenoid, the caregiver can easily have a conversation and understand the characteristics of the resident, which can be useful for the next meeting.

Caregivers sometimes have difficulty telling residents to do something. Residents sometimes refuse to wake up in the morning or eat lunch. Such refusal, which is caused by BPSD, can sometimes be solved by interacting with others. In such cases, Telenoid might be used as other people and interact with residents.

However, the experiment did not prove the effect of Telenoid itself, since the speakers had conversations both with Telenoid and face-to-face. Having conversations through Telenoid might reduce the nervousness of a speaker who is talking face-to-face, or the opposite effect might have happened. Although the

speakers and residents experienced conversations in both forms, the Telenoid results showed significantly higher evaluations. Therefore, the Telenoid conversations outperformed the face-to-face conversations, but no cross effect are clear from the results here. We have to add a speaker-only condition using the Telenoid condition and only the face-to-face condition to reveal such an effect.

Another limitation of the current study lies in that its results do not show the effect of using Telenoid in comparison with other robots. We found positive results in the Telenoid conditions, perhaps not because of Telenoid, but since seniors with AD forgot the previous meetings. Future work has to include other robots and compare them to reveal long-term effects. So far we have only acquired a partial result with Telenoid because experimenters and volunteers were necessary for supporting the experiment. Caregivers had difficulty setting up Telenoid and using it properly since they were too busy with other tasks. If

the volunteers at the facility can operate Telenoid from their homes, the load of using it will decrease. We will consider a plan that introduces Telenoid and its appropriate usage in future work.

AUTHOR CONTRIBUTIONS

SN and SS designed the experiment. KK, SN, and SS carried out the experiment at the care facility. KK and SS analyzed the results. KK and SN mainly prepared the manuscript.

ACKNOWLEDGMENTS

This work was partially supported by the Mitsubishi Foundation (25319), JSPS KAKENHI Grant Number 14J04848, and Strategic Platforms for Innovation and Research, Innovation fund Denmark.

REFERENCES

- Beer, J. M., and Takayama, L. (2011). "Mobile remote presence systems for older adults: acceptance, benefits, and concerns," in *Proceedings of the 6th International Conference on Human-Robot Interaction, HRI '11* (New York, NY: ACM), 19–26.
- Burns, A., and Iliffe, S. (2009). Alzheimer's disease. *BMJ* 338:b158. doi: 10.1136/bmj.b158
- Cabinet Office, Government of Japan (2014). *Annual Report on the Aging Society*. Available online at: <http://www8.cao.go.jp/kourei/english/annualreport/2014/pdf/c1-1.pdf> (accessed: November 30, 2015).
- Care Work Foundation (2015). *Fact-Finding Survey on Working Conditions of Care Workers* (in Japanese). Available online at: <http://www.kaigo-center.or.jp/> (accessed: December 3, 2015).
- Caris-Verhallen, W. M., Kerkstra, A., and Bensing, J. M. (1999). Non-verbal behaviour in nurse-elderly patient communication. *J. Adv. Nurs.* 29, 808–818. doi: 10.1046/j.1365-2648.1999.00965.x
- Fratiglioni, L., Wang, H.-X., Ericsson, K., Maytan, M., and Winblad, B. (2000). Influence of social network on occurrence of dementia: a community-based longitudinal study. *Lancet* 355, 1315–1319. doi: 10.1016/S0140-6736(00)02113-9
- Hirono, N., Mori, E., Ikejiri, Y., Imamura, T., Shimomura, T., Hashimoto, M., et al. (1997). [Japanese version of the neuropsychiatric inventory—a scoring system for neuropsychiatric disturbance in dementia patients]. *No To Shinkei* 49, 266–271 (in Japanese).
- Ishi, C. T., Liu, C., Ishiguro, H., and Hagita, N. (2011). "Speech-driven lip motion generation for tele-operated humanoid robots," in *International Conference on Auditory-Visual Speech Processing* (Volterra) 131–135.
- Kingma, M. (2007). Nurses on the move: a global overview. *Health Serv. Res.* 42(3 Pt 2), 1281–1298. doi: 10.1111/j.1475-6773.2007.00711.x
- Kristoffersson, A., Coradeschi, S., and Loutfi, A. (2013). A review of mobile robotic telepresence. *Adv. Hum. Comput. Interact.* 2013, 17. doi: 10.1155/2013/902316
- Kuwahara, N., Abe, S., Yasuda, K., and Kuwabara, K. (2006). "Networked reminiscence therapy for individuals with dementia by using photo and video sharing," in *Proceedings of the 8th international ACM SIGACCESS Conference on Computers and accessibility, Assets '06* (New York, NY: ACM), 125–132.
- Lu, H., Barriball, K. L., Zhang, X., and While, A. E. (2012). Job satisfaction among hospital nurses revisited: a systematic review. *Int. J. Nurs. Stud.* 49, 1017–1038. doi: 10.1016/j.ijnurstu.2011.11.009
- Mizoguchi, T., Iijima, S., Eto, F., Ishizuka, A., and Orimo, H. (1993). Reliability and validity of a Japanese version of the dementia behavior disturbance scale. *Nihon Ronen Igakkai Zasshi.* 30, 835–840 (in Japanese). doi: 10.3143/geriatrics.30.835
- Mungas, D. (1991). In-office mental status testing: a practical guide. *Geriatrics* 46, 54–58.
- Orha, I., and Oniga, S. (2012). Assistance and telepresence robots: a solution for elderly people. *Carpathian J. Electron. Comput. Eng.* 5, 87–90
- Pangman, V. C., Sloan, J., and Guse, L. (2000). An examination of psychometric properties of the mini-mental state examination and the standardized mini-mental state examination: implications for clinical practice. *Appl. Nurs. Res.* 13, 209–213. doi: 10.1053/apnr.2000.9231
- Robinson, H., MacDonald, B., Kerse, N., and Broadbent, E. (2013). The psychosocial effects of a companion robot: a randomized controlled trial. *J. Am. Med. Dir. Assoc.* 14, 661–667. doi: 10.1016/j.jamda.2013.02.007
- Sabelli, A. M., Kanda, T., and Hagita, N. (2011). "A conversational robot in an elderly care center: an ethnographic study," in *Proceedings of the 6th International Conference on Human-Robot Interaction, HRI '11* (New York, NY: ACM), 37–44.
- Shah, S., Vanclay, F., and Cooper, B. (1989). Improving the sensitivity of the barthel index for stroke rehabilitation. *J. Clin. Epidemiol.* 42, 703–709. doi: 10.1016/0895-4356(89)90065-6
- Takayanagi, K., Kirita, T., and Shibata, T. (2014). Comparison of verbal and emotional responses of elderly people with mild/moderate dementia and those with severe dementia in responses to seal robot, PARO. *Front. Aging Neurosci.* 6:257. doi: 10.3389/fnagi.2014.00257
- Tanaka, F., Cicourel, A., and Movellan, J. R. (2007). Socialization between toddlers and robots at an early childhood education center. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17954–17958. doi: 10.1073/pnas.0707769104
- Terada, S., Ishizu, H., Fujisawa, Y., Fujita, D., Yokota, O., Nakashima, H., et al. (2002). Development and evaluation of a health-related quality of life questionnaire for the elderly with dementia in Japan. *Int. J. Geriatr. Psychiatry* 17, 851–858. doi: 10.1002/gps.711
- Toba, K., Nakai, R., Akishita, M., Iijima, S., Nishinaga, M., Mizoguchi, T., et al. (2002). Vitality index as a useful tool to assess elderly with dementia. *Geriatr. Gerontol. Int.* 2, 23–29. doi: 10.1046/j.1444-1586.2002.00016.x
- United Nations (2013). *World Population Prospects: The 2012 Revision. Population Division, Department of Economic and Social Affairs*. New York, NY.
- Wada, K., Ikeda, Y., Inoue, K., and Uehara, R. (2010). "Development and preliminary evaluation of a caregiver's manual for robot therapy using the therapeutic seal robot Paro," in *Proceedings of RO-MAN, 2010 IEEE* (Viareggio), 533–538.
- Wada, K., Shibata, T., Saito, T., Sakamoto, K., and Tanie, K. (2005). "Psychological and social effects of one year robot assisted activity on elderly people at

- a health service facility for the aged,” in *Proceedings of the 2005 IEEE International Conference on Robotics and Automation, 2005 (ICRA 2005)*, 2785–2790.
- World Health Organization (2015). *Dementia* [Fact Sheet]. Fact Sheet n°362. Available online at: <http://www.who.int/mediacentre/factsheets/fs362/en/> (accessed: November 30, 2015).
- Yamazaki, R., Nishio, S., Ishiguro, H., Nørskov, M., Ishiguro, N., and Balistreri, G. (2014). Acceptability of a teleoperated android by senior citizens in Danish society: a case study on the application of an embodied communication medium to home care. *Int. J. Soc. Rob.* 6, 429–442. doi: 10.1007/s12369-014-0247-x
- Yamazaki, R., Nishio, S., Ogawa, K., Matsumura, K., Minato, T., Ishiguro, H., et al. (2013). Promoting socialization of schoolchildren using a teleoperated android: an interaction study. *Int. J. Human. Robot.* 10:13500072. doi: 10.1142/S0219843613500072

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer KB and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2016 Kuwamura, Nishio and Sato. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Intimacy in Phone Conversations: Anxiety Reduction for Danish Seniors with Hugvie

Ryuji Yamazaki^{1*}, Louise Christensen², Kate Skov³, Chi-Chih Chang⁴,
Malene F. Damholdt^{5,6}, Hidenobu Sumioka¹, Shuichi Nishio¹ and Hiroshi Ishiguro^{1,7}

¹ Hiroshi Ishiguro Laboratories, Advanced Telecommunications Research Institute International, Kyoto, Japan, ² Section for Aesthetics and Culture, Department of Aesthetics and Communication, Aarhus University, Aarhus, Denmark, ³ Section for Global Studies, Department of Culture and Society, Aarhus University, Aarhus, Denmark, ⁴ Interdisciplinary Nano Science Center, Aarhus University, Aarhus, Denmark, ⁵ Department of Philosophy and the History of Ideas, Institute for Culture and Society, Aarhus University, Aarhus, Denmark, ⁶ Unit for Psychooncology and Health Psychology, Department of Psychology, Aarhus University, Aarhus, Denmark, ⁷ Department of Systems Innovation, Graduate School of Engineering Science, Osaka University, Toyonaka, Japan

OPEN ACCESS

Edited by:

Gerrit C. Van Der Veer,
University of Twente, Netherlands

Reviewed by:

Andrej Košir,
University of Ljubljana, Slovenia
Shin'ichi Konomi,
University of Tokyo, Japan

*Correspondence:

Ryuji Yamazaki
ryuji-y@atr.jp

Specialty section:

This article was submitted to
Human-Media Interaction,
a section of the journal
Frontiers in Psychology

Received: 03 December 2015

Accepted: 31 March 2016

Published: 19 April 2016

Citation:

Yamazaki R, Christensen L, Skov K,
Chang C-C, Damholdt MF,
Sumioka H, Nishio S and Ishiguro H
(2016) Intimacy in Phone
Conversations: Anxiety Reduction
for Danish Seniors with Hugvie.
Front. Psychol. 7:537.
doi: 10.3389/fpsyg.2016.00537

There is a lack of physical contact in current telecommunications such as text messaging and Internet access. To challenge the limitation and re-embody telecommunication, researchers have attempted to introduce tactile stimulation to media and developed huggable devices. Previous experiments in Japan showed that a huggable communication technology, i.e., Hugvie decreased stress level of its female users. In the present experiment in Denmark, we aim to investigate (i) whether Hugvie can decrease stress cross-culturally, i.e., Japanese vs. Danish participants (ii), investigate whether gender plays a role in this psychological effect (stress reduction) and (iii) if there is a preference of this type of communication technology (Hugvie vs. a regular telephone). Twenty-nine healthy elderly participated (15 female and 14 male, $M = 64.52$ years, $SD = 5.67$) in Jutland, Denmark. The participants filled out questionnaires including State-Trait Anxiety Inventory, NEO Five Factor Inventory (NEO-FFI), and Becks Depression Inventory, had a 15 min conversation via phone or Hugvie and were interviewed afterward. They spoke with an unknown person of opposite gender during the conversation; the same two conversation partners were used during the experiment and the Phone and Hugvie groups were equally balanced. There was no baseline difference between the Hugvie and Phone groups on age or anxiety or depression scores. In the Hugvie group, there was a statistically significant reduction on state anxiety after meeting Hugvie ($p = 0.013$). The change in state anxiety for the Hugvie group was positively correlated with openness ($r = 0.532$, $p = 0.041$) as measured by the NEO-FFI. This indicates that openness to experiences may increase the chances of having an anxiety reduction from being with Hugvie. Based on the results, we see that personality may affect the participants' engagement and benefits from Hugvie. We discuss the implications of the results and further elaborations.

Keywords: telecommunication, Hugvie, anxiety, stress, personality

INTRODUCTION

How can interpersonal communication media shape our social connections? People are engaged in communication through computer screens, tablets, and cell phones to the extent that every day human contact is turning digital. People express concern that our communication has become increasingly shallow and we forget to spend time in the natural human way, together with other humans. Body-contact such as hugging and face-to-face interactions is outsourced to global networks like Skype, Facebook, and Twitter. For some, the digitalization of human contact can be a threat to personal relationships, because of the lacking face-to-face contact. Yet at the same time, such media offer the promise of more opportunity for connection with more people, and others find benefits with new connections and stronger bonds through new communication media (Baym, 2010). In both perspectives, communication media are regarded to change the nature of our interpersonal connections.

The development of telecommunication technologies has given us many beneficial opportunities to communicate with people worldwide. With the start of phones, we were able to talk with people who were not in our presence. To encounter the spatial and temporal flexibility, new technologies such as internet and smartphones have been developed and widely used as preferred communication tools, because it improved people's ability to stay in touch at anywhere and anytime, but information we can send and receive is typically limited to text based messages and video.

This means that tactile stimulation is still absent, so recent researches are attempting to introduce intimacy in form of physical contact to remote communication media by introducing wearable devices like "HugMe" (Cha et al., 2009) and robots to assist people in everyday life to facilitate social interactions. The HugMe is an interpersonal haptic teleconferencing system. By using it, the passive users like children with the haptic jackets are to feel the "touch" from the active users like their parents in remote. An example of robotic communication medium is a human-like robot "Telenoid" for telecommunication (Ogawa et al., 2011). It has a huggable design and can be used as welfare technology to enhance the social interactions of seniors, especially those who are cognitively impaired in order to remotely communicate with the appeal of intimacy, i.e., close relationship in embodiment. For those who have been affected by the digital divide, embodied communication technology like the Telenoid robot can provide an easy and attractive way to remotely communicate with others, and promote social interactions in both verbal and non-verbal ways. Recent attempts to re-embodiment the internet, with help from robotics, has left a question of determining in what way aspects of physical contact could be optimal conditions for communication (Seibt and Nørskov, 2012). In the line of these attempts, we explore the efficacy of telecommunication media that provide physical contacts and that acceptance might differ in different environments and culture.

Touch is one of our first senses and is our most fundamental means of contact with the world (Barnett, 1972). Studies demonstrate that interpersonal touch plays a crucial role in the development and well-being of humans (Field, 2001). For

example, the simple act of touching a patient by a nurse can result in a decrease in the patient's level of stress (Whitcher and Fisher, 1979), and those infants whose mothers use more stimulating touch are reported to have better visual-motor skills (Weiss et al., 2004). Tactile sensations can have powerful effects on people's behaviors and emotions, and facilitate bonding between pairs in a couple or groups in both animals and human (Boccia, 1986; Light et al., 2005). However, tactile aspects of communication are lacking in long-distance interactions as in phone conversation. Intimacy here can be defined as having a feeling of close relationship with others, for example, talking to your loved one through phone can be heartfelt and we ask how we can realize intimacy in phone conversation as or more than in face-to-face interactions. Due to the limitations of interpersonal touch in communication devices, new communication tools like wearable devices (e.g., google glass) have been designed to provide the opportunity to add physical contact to internet and phone users so that distance no longer would be a limitation (Bonanni et al., 2006). Also, researchers have been attempting to introduce assisting robots as communication tools to achieve psychological effects by physical contact (Kanda et al., 2002; DiSalvo et al., 2003; Stiehl and Breazeal, 2005).

Pet-like social robotic companions are introduced in elderly care, such as the small therapeutic seal-typed robot called "Paro." The usage of Paro in nursing homes demonstrated a sense of companionship and decreased loneliness (Wada et al., 2005; Wada and Shibata, 2006). Media technologies have the potential for promoting communication and having positive psychological effects on elderly, even though many elderly are part of the group, who has trouble keeping up with new technologies. Paro promotes a feeling of comfort when touched, almost like a living animal, which are reported to have a therapeutic effect on stress both mentally and physically. In relation to this, studies have shown that touches, hugs and massages from people, even animal-assisted therapy, has a physiological effect on stress reduction (Beetz et al., 2012; Morhenn et al., 2012).

Many studies have reported endocrine responses to psychological stress and stressful tasks, such as public speaking and mental arithmetic, can increase cortisol levels (e.g., Kirschbaum et al., 1993). Cortisol, known as the stress hormone, is produced in the adrenal glands and regulates many processes that occur in the body in response to stress within an effort to maintain homeostasis. Reviews have highlighted the effects of psychological stressors on this physiological system are variable and inconsistent (e.g., Biondi and Picardi, 1999). However, assessment of cortisol in blood and saliva to see psychological stress levels is a widely accepted and commonly used method in psychoneuroendocrinology (Kirschbaum and Hellhammer, 1994; Hellhammer et al., 2009).

In a previous experiment in Japan, a significant reduction in cortisol levels was shown for those who had conversations through the huggable communication medium Hugvie (Sumioka et al., 2013). However, the study had several limitations. Gender, age, and cultural background could all influence the endocrine changes and modulate people's interpretation of, and hence their response to, interpersonal touch, as reported (Gallace and Spence, 2010). The Japanese culture does not belong to the cultures

in which people often touch each other, e.g., handshakes and hugs (Finnegan, 2005). The interpersonal touch is a cultural phenomenon, therefore in Japan the custom of bowing and limiting touches could have affected the result of the previous study, as the participants might, due to the lack of body contact, have reacted excessively resulting in overstimulation while hugging the communication media.

Culturally Europeans such as Danes have a longer history with body contact, hugs and handshakes, in everyday life. The societal openness to new technologies makes Denmark ideal for testing Hugvie on a cross-cultural and gender basis. In the experiment in Japan, all the participants were female, which therefore did not allow for exploring whether decrease in cortisol level was due to gender differences. Therefore, the results of the Japanese experiment needs to be replicated and tested in other social settings. In the Danish experiment, we aim to investigate if (1) the decrease of stress when using Hugvie is different or the same between Japanese and Danish participants as well as (2), investigate whether reduction of the cortisol level and preference for this type of communication technology is influenced by gender (by including males in the study).

Hence the purpose of this experiment is to investigate (i) whether Hugvie can decrease stress cross-culturally, i.e., Japanese vs. Danes, (ii) investigate whether gender plays a role in this psychological effect (stress reduction) and (iii) if there is a preference of this type of communication technology (Hugvie vs. a regular telephone).

MATERIALS AND METHODS

In this experiment, we replicate a previous experiment in Japan by using a similar method, so in this present study the human-shaped Hugvie pillow-phone was compared with a regular phone. We setup almost identical experimental conditions as described in the following article: “Huggable communication medium decreases cortisol levels” (Sumioka et al., 2013). The study by Sumioka et al. (2013) found strong correlations between saliva and blood cortisol levels. Therefore, in our experiment, we only took saliva samples, as blood samples would be redundant.

Both male and female participants were equally and randomly divided into two groups who all had to talk with the same stranger of opposite gender while either hugging Hugvie (Hug Group) or talking in a regular mobile phone on speaker (Phone Group). The latter was the control group. To evaluate participants' psychological and physiological responses to the social interaction through the communication medium, we measured cortisol levels through saliva samples and had the participants answer questionnaires at baseline and after the conversation session. The participants were video-recorded during the session in order to follow and evaluate their reactions and behavior in comparison with cortisol level. In addition, we decided to interview the participants after the sessions to see how they perceived the communication media.

As the results were positive for the Japanese women in the previous study, we predicted that we would come to a similar result, but with gender differences. For the purpose of the

psychological effect, we evaluated the questionnaires outcome and expected to find changes in answers after the sessions in the Hugvie group. We assumed personality traits, as well as cultural background, might be related to the result.

An ethical committee in Jutland, De Videnskabssetiske Komiteer, For Region Midtjylland decided that an approval was not needed for this experiment. It was also checked and approved by the committee at Cognition and Behavior Lab, Aarhus University.

Communication Device

The Hugvie® (Figure 1) was developed by Osaka University and ATR Hiroshi Ishiguro Laboratory. It is a pillow-phone in a minimalistic human form for talking whilst hugging. Its height is 75 cm and its weight is 600 g. It is designed to enable users to feel the presence of any remote partners strongly while communicating with them. The human-like robot with minimalistic characteristic traits called the Telenoid (Ogawa et al., 2011), was the inspiration for creating the Hugvie pillow-phone.

In studies with the Telenoid it was reported that physical contact, e.g., hugging, has an impact on the psychological state of the user (Ogawa et al., 2011). In line with these findings, the Hugvie was designed with focus on the hugging experience. Its pillow like feeling stems from the spandex fiber cover with polystyrene microbead filling. Like the Telenoid, there are no actuators inside it, Hugvie appears like a person with open arms ready for a hug. With a pocket design, it is possible to place a mobile phone inside its head. This is intended to give the user a feeling of hugging their conversation partner while talking through the pillow. Because of its design, it is possible to investigate the effect of touches.

Subjects

The experiment included totally 29 healthy participants (15 female and 14 male). They were elderly healthy subjects (fine elder citizens; $M = 64.52$ years, $SD = 5.67$), who were invited to evaluate our communication media. We used flyers and posters, which were distributed throughout the city in places like activity centers and libraries where elderly people gather. We also asked staff at elderly care centers and officials with broad networks within the senior community to help gather



participants. We targeted elderly because they have a higher hormonal stability than younger people do, especially in the case of women. Exclusion of participants would happen in case of acute or chronic hormonal dysregulation or if they were on any kind of hormonal medication. The participants received oral and written information about the study and gave their written informed consent. Furthermore they were informed that the experiment included several prohibitions such as alcohol intake or smoking 1 day ahead of the study and to refrain from drinking, eating, or exercising 1 h before the session. The Hug and Phone groups were randomly selected, yet evenly spread in morning and afternoon sessions. They were not informed prior to the session of which group they were assigned to.

Conversation Partner

We selected two capable conversation partners among students at Aarhus University. They proved to possess good conversational skills and could with ease fill out 15 min of conversation in the experiment. As in the previous experiment in Japan, the conversation partners were university students in their 20 s a, but not only a male (27 years old) but also female (28 years old). They received basic information about the experiment and gave informed consent in the same way as the participants.

Experimental Environment

The experiment took place at COBE Lab at Aarhus University. The participants filled out the questionnaires and had the conversation in separate rooms. In the questionnaire room, we prepared and accepted the informed consent, the pre-conversation questionnaires and one of our staff assisted with the first saliva sample in this room. In the conversation room, we prepared a big cozy chair, a camera, post-conversation questionnaires and the second saliva sampling tube. We brought either the phone or Hugvie depending on the participant's group right before the conversation session started (**Figure 2**). The participants never met their conversation partners, who made the phone calls from another room. During the conversations and the questionnaires, participants were left alone.

Experimental Procedure

The experiments were conducted evenly in the morning and the afternoon (8:00–12:00 and 13:00–17:00) in according to the Japanese experiment. After filling out pre-conversation questionnaires about their feelings of anxiety, stress level, and personality by using questionnaires the State-Trait Anxiety Inventory (STAI), Becks depression inventory (BDI-II), Geriatric Depression Scale (GDS), Perceived Stress Scale (PSS), and NEO Five Factor Inventory (NEO-FFI), saliva samples were given by all participants.

After this, the participants came to the conversation room and could relax for about 5 min before they were given either the Hugvie or phone on speaker. All participants conducted a 15-min conversation with the conversation partner of opposite sex with the given communication media. The conversation partner has prior been informed to introduce himself/herself and ask for the participants' name, where after they ask the participant about their best memories of the past year. It was allowed to have free

conversation during the session with the exception of questions about their educational background, parents' jobs, and political views for ethical reasons. After the conversation, the second saliva sample was retrieved and the participant was asked to fill out the post-conversation questionnaire about their feelings of anxiety throughout the conversation and finally we conducted a brief interview to hear the participants' opinion about the conversation and media usage.

Questionnaires

We asked the participants about their feelings in the pre- and post-conversation questionnaires. For the pre-conversation questionnaires, we used the STAI, the PSS, the BDI-II, the GDS, and the NEO-FFI. For the post-conversation questionnaires, we used the STAI and PSS. The STAI is a commonly used measure of trait and state anxiety (Spielberger, 1983). It consists of 40 questions and differentiates between the temporary condition of state anxiety and the more general and long-standing quality of trait anxiety. We used all the questions of the STAI before the conversation-session and repeated only the state part after the session. To obtain subjective stress measure, we also used the 10-item PSS that is a measure of the degree to which situations in one's life are appraised as stressful (Cohen et al., 1983).

We used the GDS and BDI-II only at baseline to ensure clinical depression did not interfere with results. To assess the level of depression in elderly, we used the BDI-II composed of 21 items (Beck et al., 1996) and a short version of the GDS containing 15 questions with simple yes/no response set (Sheikh and Yesavage, 1986). To investigate the relations between the participant's personality and the changes in stress level, we also used the 60-item NEO-FFI that provides a concise measure of the five basic personality factors (Costa and McCrae, 1989).

The NEO-FFI was used to assess five stable personality dimensions as derived from the five-factor model of personality (NEO-PI-R). The NEO-PI-R is validated cross-culturally (McCrae, 2002) and is available in a validated Danish version. The NEO-FFI items were administered verbally and the respondents rated them on a five point Likert scale from "strongly disagree" to "strongly agree." The five personality dimensions are Openness (openness to internal and external stimuli), Conscientiousness (self-discipline and competency), Extraversion (tendency to be sociable and adventurous), Agreeableness (degree of trustfulness, modesty), and Neuroticism (tendency toward experiencing psychological distress or negative affect).

Cortisol Collection and Analysis

The saliva samples were assayed for cortisol determination by a cortisol enzyme immunoassay (Cortisol EIA; Arbor Assays, USA) using a standard curve method with reported detection limits of 45.4 pg/ml. The assay was performed as instructed by the manufacturer. The cross-reactivity of the assay is 18.8% with dexamethasone, 1.2% with cortisone, 7.8% with prednisolone, 1.2% with corticosterone and <0.1% with progesterone. Saliva was obtained at least 2 h after eating. The participant's mouth was rinsed prior to saliva collection to avoid food borne antigens or other materials that may affect cortisol analysis.



FIGURE 2 | Experimental settings.

Once the saliva was collected, protease inhibitors (Complete protease inhibitor cocktail tablets, Roche) were added according to the manufacturer's protocols to prevent protein degradation and it was stored at -80 until the assay. For the saliva assay, thawed samples were centrifuged at 2500 g for 20 min and the supernatant was collected for the assay. All samples from the participants were included in the same assay batch to eliminate within subject inter-assay variance. All the samples were assayed in duplicates and averaged. The effect of physical touch was measured as the decrease of cortisol that was calculated by subtracting the cortisol levels before the conversion from those after the conversion.

RESULTS

Data was analyzed using IBM SPSS Statistics for Macintosh, Version 21.0. (2012; Armonk, NY: IBM Corp). Paired and unpaired two-tailed T -tests were used for group comparisons on continuous variables as this statistical procedure have been found to be robust even in very small samples (de Winter, 2013). Bonferroni adjustments were not applied. The relationship between personality traits and changes in anxiety level was explored by Spearman correlations.

There was no significant age difference between the Hugvie ($M = 64.9$, $SD = 6.4$) and Phone groups ($M = 64.1$, $SD = 5.2$), $t(27) = 0.396$, $p = 0.695$. There was no significant baseline differences in reported depression as assessed by Becks Depression Inventory between the Hugvie ($M = 4.2$, $SD = 4.69$) and phone group ($M = 2.42$, $SD = 2.56$), $t(27) = 0.223$, $p = 0.176$.

Paired samples t -test showed no statistically significant differences between baseline ($M = 8.50$, $SD = 4.23$) and post-encounter scores ($M = 9.21$, $SD = 4.37$) on the Perceived Stress Scale for the Hugvie group, $t(13) = -1.046$, $p = 0.315$ or for the

TABLE 1 | Description of personality traits for Phone and Hugvie groups.

	Phone group (<i>M</i> , <i>SD</i>)	Hugvie group (<i>M</i> , <i>SD</i>)
Openness	31.54 (8.11)	32.20 (4.16)
Conscientiousness	36.57 (3.90)	31.93 (5.04)
Extraversion	32.50 (5.24)	31.73 (5.85)
Agreeableness	36.31 (4.05)	34.20 (4.60)
Neuroticism	13.50 (5.92)	15.73 (6.53)

Phone group (Baseline: $M = 9.21$, $SD = 3.47$; Post-encounter: $M = 9.14$, $SD = 3.99$), $t(13) = -0.126$, $p = 0.902$.

Paired samples t -test showed a statistically significant difference between baseline ($M = 32.55$, $SD = 6.4$) and post-encounter scores ($M = 27.86$, $SD = 4.53$) on STAI for the Hugvie group, $t(14) = -3.362$, $p = 0.005$ but not for the Phone group (Baseline: $M = 28.78$, $SD = 3.42$; Post-encounter: $M = 28.07$, $SD = 3.35$), $t(15) = -0.924$, $p = 0.372$.

A STAI change score was calculated (Baseline STAI score minus post-encounter STAI scores) and correlated using a Spearman correlation to personality traits as assessed with the NEO-P-IR (for descriptives of personality traits for the Phone and Hugvie groups see **Table 1**). In the Hugvie group ($n = 15$) there was no significant correlation between STAI change score and extraversion ($r = 0.151$, $p = 0.591$), agreeableness ($r = 0.213$, $p = 0.446$) or neuroticism ($r = 0.432$, $p = 0.108$). However, there was a positive correlation with openness ($r = 0.532$, $p = 0.041$) and a near-significant correlation with conscientiousness ($r = -0.509$, $p = 0.053$). For the phone group ($n = 14$) there was no correlations between anxiety state changes and any of the personality traits: extraversion ($r = 0.156$, $p = 0.595$), conscientiousness ($r = -0.050$, $p = 0.864$), agreeableness ($r = -0.064$, $p = 0.836$) neuroticism ($r = 0.98$, $p = 0.739$) or openness ($r = 0.257$, $p = 0.397$).

A Chi-square test for independence with Yates Continuity Correction indicated that there was no significant association between whether the participant was assessed in the morning or afternoon and which group he/she was ascribed to (Hugvie or phone group), $X^2(1, N = 29) = 0.03, p = 0.87, \phi = -0.100$. There was no significant difference in cortisol level from baseline to post-encounter for the Hugvie (Baseline: $M = 1.86$ ng/ml., $SD = 0.95$; Post-encounter: $M = 1.70$ ng/ml., $SD = 0.96$) $t(13) = -1.01, p = 0.330$ or the phone group (Baseline: $M = 1.79$ ng/ml., $SD = 1.18$; Post-encounter: $M = 1.63$ ng/ml., $SD = 0.89$) $t(12) = 1.96, p = 0.457$ using paired samples t -test.

A cortisol change score was calculated (post-encounter cortisol level minus baseline cortisol level) and using a Spearman correlation it was correlated to the STAI change score. There was no significant correlation with STAI change score in the Hugvie group ($r = 0.043, p = 0.884$) or the phone group ($r = -0.077, p = 0.803$) nor was cortisol change correlated significantly to changes in perceived stress score for the Hugvie group ($r = -0.436, p = 0.136$) or phone group ($r = 0.443, p = 0.130$).

DISCUSSION

The purpose of this paper is to describe how the huggable communication medium Hugvie could be perceived and effective among Danish people. In the Hug group there was a significant difference between baseline and post-encounter scores and we found that Hugvie is effective in reducing anxiety for Danes as well cross-culturally. No significant differences were found in the phone group. Another finding was that the difference was related to personal traits, namely openness. There was also a near-significant correlation between STAI change score and conscientiousness. There was no gender difference to see in the questionnaires.

The cortisol level showed null-results which could be a lack of sensitivity of the cortisol measures or even reactivity in light of cultural differences. From a previous study, the cortisol level was increased with a lesser extent compared with that of amylase after participants were exposed to a stressful video (Takai et al., 2004). It took longer for the cortisol to show changes than amylase, so the sessions for our Hugvie experiment might not have been long enough to make use of cortisol and it might have been better to include changes in amylase levels too. Therefore, we will focus our discussion on the questionnaire results and interviews.

Personality Matters

The relation between anxiety reduction and openness as seen in the questionnaire results could indicate that users with sensitivities, such as openness to new experiences, would be the main group among Danes to benefit by using the huggable medium. If the user, for example, would have an active imagination and would be sensitive to aesthetics, Hugvie could be helpful in reducing anxiety and hereby stress. The greater likelihood of experiencing an anxiety reduction when using Hugvie would therefore be, in this case, elderly who score higher on openness.

We did not see any positive outcome with other personalities in this experiment, but on the other side, there was also a near-significant negative correlation between anxiety state changes and conscientiousness in the Hug group. This indicates that people who are high on conscientiousness have a greater likelihood of becoming anxious with Hugvie. It has been found that conscientiousness is negatively related to creativity; whereas, openness to experience relates to it positively (George and Zhou, 2001). We see these personality traits affect on the effects of Hugvie in both ways, positively and negatively, thus, could be an important factor for media usage.

Openness to experience is a personality trait in the five factor model of personality and is fundamental for aesthetic appreciation and creativity. Openness consists of a set of specific tendencies that cluster together, involving six facets like imagination. A model of openness divides the trait into the two groups, i.e., openness and intellect (DeYoung et al., 2007). The openness aspect is about the heart that includes aesthetic sensitivity, creativity, and imaginativeness while the intellect aspect is a brain division that includes fluid intelligence, vocabulary knowledge, and an intellectual life approach.

Openness to experience has consistently predicted aesthetic appreciation and engagement in the arts where people immerse in aesthetic activities such as reading, painting, visiting art galleries, and valuing the arts (McManus and Furnham, 2006). People with high openness draw more enjoyment from, have more positive attitudes toward, and are gladly more exposed to the arts (Fayn and Silvia, 2015). Exposure to new media can for some people be joyful and beneficial whilst others cannot gain the same experience. This can be seen in parallel with aesthetics in artifacts as a singular experience meaning the experience that differs according to time, place, mood, and so on.

We consider the idea of aesthetic experience as a singular experience. Every meeting or experience with an artifact, for instance a piece of art, like a painting is singular, and it differs from individual to individual how the artifact is experienced; the individual's state of mind on that specific day, as well as the setting. If you see a painting more than once you will never have the same experience thereby making it singular. Maybe you were in a different mood, maybe you were with different people, or you see the painting a different place than the first time.

This theory is used in situations when meeting aesthetic artifacts like artwork. Furthermore, there is the idea that you have to be open to the aesthetic potential or value of an artifact to experience it as an aesthetic artifact. If you are not open to the artifact's aesthetic potential, you cannot have an aesthetic experience. Therefore, it can also be that having an open mind to Hugvie and the very experience is necessary. You have to be open to accept the way holding Hugvie makes you feel. For instance, one female participant did not like holding Hugvie because she felt she had outgrown soft toys, while a male participant started feeling it being natural to hold Hugvie quite fast. It was a positive experience for some participants, but also an utterly negative experience for a few, while some did not really think about and were indifferent to it. It could therefore mean that some of the participants were not even very open to the idea in the first place or it could also mean that they just had a negative experience

though still understanding that Hugvie could be helpful for others.

There were instances where people were skeptical and did not like the idea of Hugvie prior to the session based on the information about Hugvie provided in the flyers and posters. After the session, there was a significant anxiety reduction as a result and some even changed their opinions. However, only short experiments were conducted, so there is a possibility that the novelty of the medium attracted participants, and if they use it in a longer term the effect of reducing anxiety might disappear. To reduce Hugvie's novelty effects, a longitudinal study is necessary for establishing its efficacy thoroughly. It will also be required to test it in settings that are more naturalistic like home for people to act freely.

Social Norms and Media Usage

If I limit to statistically significant results reported in this paper, I can find no evidence for cross-cultural effect as indicated by this paper title and conclusion.

There seems to be a small difference regarding gender in the interviews where overall the male participants seemed more positive toward Hugvie. Both culturally and in terms of gender it might seem socially inappropriate in Denmark for men to use Hugvie, as some of the participants also expressed, to the point that soft toys may be seen as feminine or childish things. Therefore, it is interesting that it did make a positive difference for most of the male participants – using Hugvie did make them feel more comfortable according to their comments after the sessions. In the previous Japanese experiment participants, who were all female, have not mentioned social acceptance regarding usage of Hugvie, but most seemed positive toward Hugvie.

Although there may be some cultural differences, it is difficult to say whether the reason is that Japanese people tend to be more polite and wrap things up, or if they are just more positive toward the Hugvie experience in general. The Danish participants seemed to give their honest opinions toward Hugvie. One man in particular said that it would be taboo for men to use it because of its toy-like appearance for him, and the women from the group interview raised the question of whether it could be acceptable for men to use Hugvie. We did not have the same comparative conditions such as gender, both male and female, in Japan. The Japanese experiment did not use the STAI questionnaire, so we need to conduct the experiment in Japan again. In such same conditions, we have to carry forward cross-culturally comparative experiments in our future work.

The medium could not be effective for everybody, but we see a development in attitude toward Hugvie, for instance, in a group interview. We had three nurses, two from phone-group and one from Hugvie-group, share their experiences with and without Hugvie and their thoughts prior to the experiment. Hereafter they could all see, touch and hold Hugvie while discussing their viewpoints on it. The nurses started giving Hugvie a gender instead of just calling it a “thing” or “it,” Hugvie became him. This could indicate a form of closeness.

Prior to the experiments, they were all skeptical of Hugvie and the intentions for usage, thinking that this artificial thing would not be able to replace the warm hands of a human when

considering their field. But after sitting with it they had Hugvie associated with a child or an animal (penguin) because of its size and shape. One even mentioned it felt like sitting with her grandchild because of the warmth Hugvie provided although the feeling was too artificial in the start and therefore it could use more softness perhaps a fur cover.

Because of the warmth that one felt touched her heart, they suggested it would be good to use Hugvie for residents at nursing homes. Especially the comfort really could be used in such establishments and for demented people. The color of this Hugvie was bright orange and this was also welcomed as a happy color. Though the nurses approved of Hugvie they still felt it might be limited to certain groups of people and that men might refrain from using it and therefore mainly would be for women.

With this interview, we see the negative or rather skeptical attitude toward Hugvie prior to the experiment turning into an open talk about the usage of it. The three ladies became very open and accepting of the medium, but only one had the actual experience talking through Hugvie. While sitting with Hugvie, though the two others did not share the same experience as the hug-group participant, they developed their opinions in a positive way where they seemed very open minded when listening to her experience.

Many participants described a comfortable feeling and the sense of being with someone, like a pet or a child, when talking through Hugvie, but when we asked whether they would like to use it at home, there was surprisingly few who would even consider it.

A male participant said “If Hugvie should be used outside of the home, it has to become a trend so everyone would use it or else people would think the user would be weird,” “It could be smaller maybe,” “But, I think that there is some meaning to the big size in terms of how calm it makes you, because it actually feels almost like two people sitting and talking. You don't feel alone.”

Overall when asking about the usage at home, we got comments similar to this – “It is hard to imagine using it in practice, if you had one at home.”

The opinions about who could or could not use Hugvie should be considered according to social norms since these questions and comments are made. This should not be limited to one country, but should be studied cross-culturally because there would possibly be a different opinion and attitude depending on the way people are raised, and which environment they have lived in. Age could also be a factor for the assumption that men would not be able to use Hugvie as standard viewpoints could change depending on the generation and media exposure.

CONCLUSION

We found that Hugvie was effective in reducing anxiety for Danes, with significant difference in the questionnaire between baseline and post-encounter scores in the Hugvie group, but no significant differences in the phone group. Essentially, we found that the difference was related to the personal trait openness. Statistically, no significant gender differences were found, but it might be due to the small number of participants in each group.

Still there seemed a slight difference in the interviews according to who can use Hugvie or not, but this is more a question of norm than whether it can have a stress reducing effect. We suggest this for further research.

The indication of openness in the results suggest users of the huggable medium to have a certain sensitivity, like active imagination and aesthetic sensitivity, to reach a stress reducing effect from using Hugvie. Participants who score higher on openness and spend time with Hugvie had a greater likelihood of experiencing anxiety reduction.

Although Hugvie resembled an animal, a child or just felt comfortable to sit with for some, others did not like it at all. The participants had a common opinion when it came to whether they could imagine using the communication medium at home; they did not want to use it by themselves, but many suggested care facilities or lonely elderly for primary users. Mainly because they thought it was impractical to use when moving around and having to find it to insert the phone before usage, which could be the reason why immobile people came to mind.

The participants mentioned improvements such as it becoming softer, more mobile/handy and the pocket/head part more stable to make Hugvie more preferable. If they were lonely or demented, it might be more acceptable to use a toy like phone or else it could just be that the Hugvie should be redesigned if it should be introduced to people who are used to a more flexible way of communicating.

There are such limitations as lack of comparison of the Danish study with a Japanese one that investigates the effect of Hugvie in all the same conditions and the cortisol test did not give a clear result, so we should have used other methods like amylase to supplement and other types of hormones. We could have used a larger group to investigate the details of gender difference and the interviews could have touched different mentalities and norms cross-culturally. Social norms related to usage of Hugvie and its cultural relation to personalities needs to be investigated in terms of the acceptability of new media in societies.

In the results, we found that personality matters for the usage of communication media. We suggest that when we

apply communication media to people, we should investigate personality traits that could affect the effects of the media, perhaps with a possible matrix of personality and various types of communication media. There are other factors such as gender and cultural differences, which might affect the effects as well and we could look into the components of these factors. This could be tested through experiments with different types of classified communication media such as regular phones, video-conferencing systems, huggable communication media, and various types of social robots – mechanoid, zoomorphic, and humanoid.

AUTHOR CONTRIBUTIONS

RY: project leader; LC, and KS: experiment staff; C-CC, MD, and HS: analysis of data; SN: adviser; HI: media inventor.

FUNDING

This study was partially supported by a Strategic Platform for Innovation and Research (SPIR), the Danish Council for Strategic Research and The Danish Council for Technology and Innovation, and the PENSOR project funded by the VELUX FOUNDATION.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to the Municipality of Aarhus especially Ivan Kjær Lauridsen (Head of Health and Assisted Living Technologies) and Birgitte Halle (project leader) as well as to COBE Lab (Aarhus University). Furthermore, we are in debt to Marco Nørskov, Raul Hakli, Stefan K. Larsen, Christina Vestergård, Johanna Seibt, Glenda Hannibal, Rikke Mayland Olsen, Thea Puggaard Frederiksen from Aarhus University and all the members of the PENSOR project at the Department of Culture and Society (Aarhus University).

REFERENCES

- Barnett, K. (1972). A theoretical construct of the concepts of touch as they relate to nursing. *Nurs. Res.* 21, 102–110.
- Baym, N. K. (2010). *Personal Connections in the Digital Age: Digital Media and Society Series*. Cambridge: Polity Press.
- Beck, A. T., Brown, G., and Steer, R. A. (1996). *Beck Depression Inventory II Manual*. San Antonio, TX: The Psychological Corporation.
- Beetz, A., Uvnäs-Moberg, K., Julius, H., and Kotrschal, K. (2012). Psychosocial and psychophysiological effects of human-animal interactions: the possible role of oxytocin. *Front. Psychol.* 3:234. doi: 10.3389/fpsyg.2012.00234
- Biondi, M., and Picardi, A. (1999). Psychological stress and neuroendocrine function in humans: the last two decades of research. *Psychother. Psychosom.* 68, 114–150. doi: 10.1159/000012323
- Boccia, M. L. (1986). "Grooming site preferences as a form of tactile communication and their role in the social relations of rhesusmonkeys," in *Current Perspectives in Primate Social Dynamics*, eds D. M. Taub and F. A. King (New York, NY: Van Nostrand Reinhold), 505–518.
- Bonanni, L., Vaumobilee, C., Lieberman, J., and Zuckerman, O. (2006). "TapTap: a haptic wearable for asynchronous distributed touch therapy," in *Proceedings of the Extended Abstracts on Human Factors in Computing Systems (Montréal, Québec, Canada, April 22 – 27, 2006)*. CHI '06 (New York, NY: ACM Digital Library), 580–585.
- Cha, J., Eid, M., Barghout, A., Rahman, A. M., and El Saddik, A. (2009). "HugMe: synchronous haptic teleconferencing," in *Proceedings of the 17th ACM International Conference on Multimedia, MM' 09* (New York, NY: ACM Digital Library), 1135–1136.
- Cohen, S., Kamarck, T., and Mermelstein, R. (1983). A global measure of perceived stress. *J. Health Soc. Behav.* 24, 385–396. doi: 10.2307/2136404
- Costa, Jr., and McCrae, R. R. (1989). *The NEO-PI/NEO-FFI Manual Supplement*. Odessa, FL: Psychological Assessment Resources.
- de Winter, J. C. (2013). Using the Student's t-test with extremely small sample sizes. *Pract. Assess. Res. Eval.* 18, 1–12.
- DeYoung, C. G., Quilty, L. C., and Peterson, J. B. (2007). Between facets and domains: 10 aspects of the big five. *J. Pers. Soc. Psychol.* 93, 880–896. doi: 10.1037/0022-3514.93.5.880

- DiSalvo, C., Gemperle, F., Forlizzi, J., and Montgomery, E. (2003). "The Hug: an exploration of robotic form for intimate communication," in *Proceedings of the 12th IEEE International Workshop on Robot and Human Interactive Communication* (New York, NY: IEEE Press), 403–408.
- Fayn, K., and Silvia, P. J. (2015). "States, people, and contexts: three psychological challenges for the neuroscience of aesthetics," in *Aesthetic Art, Aesthetics, and the Brain*, eds J. P. Huston, M. Nadal, F. Mora, L. F. Agnati, and C. J. C. Conde (Oxford: Oxford University Press), 40–56.
- Field, T. (2001). *Touch*. Cambridge, MA: MIT Press.
- Finnegan, R. (2005). *Communicating: The Multiple Modes of Human Interconnection*. New York, NY: Routledge.
- Gallace, A., and Spence, C. (2010). The science of interpersonal touch: an overview. *Neurosci. Biobehav. Rev.* 34, 246–259. doi: 10.1016/j.neubiorev.2008.10.004
- George, J. M., and Zhou, J. (2001). When openness to experience and conscientiousness are related to creative behavior: an interactional approach. *J. Appl. Psychol.* 86, 513–524. doi: 10.1037//0021-9010.86.3.513
- Hellhammer, D. H., Wüst, S., and Kudielka, B. M. (2009). Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinology* 34, 163–171. doi: 10.1016/j.psyneuen.2008.10.026
- Kanda, T., Ishiguro, H., Ono, T., Imai, M., and Nakatsu, R. (2002). "Development and evaluation of an interactive humanoid robot," in *Proceedings of the ICRA'02: IEEE International Conference on Robotics and Automation* (Roman: IEEE), 1848–1855.
- Kirschbaum, C., and Hellhammer, D. H. (1994). Salivary cortisol in psychoneuroendocrine research: recent developments and applications. *Psychoneuroendocrinology* 19, 313–333. doi: 10.1016/0306-4530(94)90013-2
- Kirschbaum, C., Pirke, K. M., and Hellhammer, D. H. (1993). The 'Trier Social Stress Test' — A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology* 28, 76–81. doi: 10.1159/000119004
- Light, K. C., Grewen, K. M., and Amico, J. A. (2005). More frequent partner hugs and higher oxytocin levels are linked to lower blood pressure and heart rate in premenopausal women. *Biol. Psychol.* 69, 5–21. doi: 10.1016/j.biopsycho.2004.11.002
- McCrae, R. R. (2002). "NEO-PI-R data from 36 cultures: further intercultural comparisons," in *The Five-Factor Model of Personality Across Cultures*, eds R. R. McCrae and J. Allik (New York, NY: Springer), 105–126.
- McManus, I., and Furnham, A. (2006). Aesthetic activities and aesthetic attitudes: influences of education, background and personality on interest and involvement in the arts. *Br. J. Psychol.* 97, 555–587. doi: 10.1348/000712606X101088
- Morhenn, V. B., Beavin, L. E., and Zak, P. J. (2012). Massage increases oxytocin and reduces adrenocorticotropin hormone in humans. *Altern. Ther. Health Med.* 18, 11–18.
- Ogawa, K., Nishio, S., Koda, K., Balistreri, G., Watanabe, T., and Ishiguro, H. (2011). Exploring the natural reaction of young and aged person with Telenoid in a real world. *J. Advan. Comput. Intell. Inform.* 15, 592–597.
- Seibt, J., and Nørskov, M. (2012). "Embodying" the internet: towards the moral self via communication robots? *Philos. Technol.* 25, 285–307. doi: 10.1007/s13347-012-0064-9
- Sheikh, J. I., and Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. *Clin. Gerontol.* 5, 165–173. doi: 10.3109/09638288.2010.503835
- Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory STAI (form Y) ("Self-Evaluation Questionnaire")*. California, CA: Consulting Psychology Press.
- Stiehl, W., and Breazeal, C. (2005). "Design of a therapeutic robotic companion for relational, affective touch," in *Proceedings of the 14th IEEE International Workshop on Robot and Human Interactive Communication* (Nashville, TN: IEEE), 408–415.
- Sumioka, H., Nakae, A., Kanai, R., and Ishiguro, H. (2013). Huggable communication medium decreases cortisol levels. *Sci. Rep.* 3, 3034. doi: 10.1038/srep03034
- Takai, N., Yamaguchi, M., Aragaki, T., Eto, K., Uchihashi, K., and Nishikawa, Y. (2004). Effect of psychological stress on the salivary cortisol and amylase levels in healthy young adults. *Arch. Oral. Biol.* 49, 963–968.
- Wada, K., and Shibata, T. (2006). "Robot therapy in a care house: results of case studies," in *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication* (Roman: IEEE), 581–586.
- Wada, K., Shibata, T., Saito, T., Sakamoto, K., and Tanie, K. (2005). "Psychological and social effects of one year robot assisted activity on elderly people at a health service facility for the aged," in *Proceedings of the IEEE International Conference on Robotics and Automation* (Roman: IEEE), 2785–2790.
- Weiss, W. J., Wilson, P. W., and Morrison, D. (2004). Maternal tactile stimulation and the neurodevelopment of low birth weight infants. *Infancy* 5, 85–107. doi: 10.1207/s15327078in0501_4
- Whitcher, S. J., and Fisher, J. D. (1979). Multidimensional reaction to therapeutic touch in a hospital setting. *J. Pers. Soc. Psychol.* 37, 87–96. doi: 10.1037/0022-3514.37.1.87

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Yamazaki, Christensen, Skov, Chang, Damholdt, Sumioka, Nishio and Ishiguro. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Impact of Mediated Intimate Interaction on Education: A Huggable Communication Medium that Encourages Listening

Junya Nakanishi^{1,2*}, Hidenobu Sumioka² and Hiroshi Ishiguro^{1,2}

¹ Intelligent Robotics Laboratory, Graduated School of Engineering Science, Osaka University, Toyonaka, Japan, ² Hiroshi Ishiguro Laboratory, Advanced Telecommunication Research Institute International, Kyoto, Japan

OPEN ACCESS

Edited by:

Tsutomu Fujinami,
Japan Advanced Institute of Science
and Technology, Japan

Reviewed by:

Marco Fyfe Pietro Gillies,
Goldsmiths, University of London, UK
Ian Oakley,
Ulsan National Institute of Science and
Technology, South Korea

*Correspondence:

Junya Nakanishi
nakanishi.junya@
irl.sys.es.osaka-u.ac.jp

Specialty section:

This article was submitted to
Human-Media Interaction,
a section of the journal
Frontiers in Psychology

Received: 06 November 2015

Accepted: 24 March 2016

Published: 19 April 2016

Citation:

Nakanishi J, Sumioka H and
Ishiguro H (2016) Impact of Mediated
Intimate Interaction on Education: A
Huggable Communication Medium
that Encourages Listening.
Front. Psychol. 7:510.
doi: 10.3389/fpsyg.2016.00510

In this paper, we propose the introduction of human-like communication media as a proxy for teachers to support the listening of children in school education. Three case studies are presented on storytime fieldwork for children using our huggable communication medium called Hugvie, through which children are encouraged to concentrate on listening by intimate interaction between children and storytellers. We investigate the effect of Hugvie on children's listening and how they and their teachers react to it through observations and interviews. Our results suggest that Hugvie increased the number of children who concentrated on listening to a story and was welcomed by almost all the children and educators. We also discuss improvement and research issues to introduce huggable communication media into classrooms, potential applications, and their contributions to other education situations through improved listening.

Keywords: listening, child education, huggable communication medium, mediated intimate interaction, mental states, classroom communication

1. INTRODUCTION

Communication with others is an important process for acquiring generic knowledge in society, such as language, communication skills, and social manners. After learners receive and interpret the information presented by caregivers or teachers, they sometimes acquire new knowledge and skills based on feedback. Obviously, a learner's ability for information comprehension is fundamental in the initial learning phase to acquire generic knowledge.

Listening is one such crucial skill, especially in school education since the information that must be learned is generally provided verbally. For example, 68% of the class time in German primary school classes and 53% in U.S. college students is spent listening (Bohlken, 1999; Imhof and Weinhard, 2004). However, investigations have reported that many first graders in several countries start school unprepared for learning, including an inability to listen during class lessons (McClelland et al., 2000; Rimm-Kaufman et al., 2000; Sakakihara, 2010). Two other studies reported that at most only half of kindergarteners have mastered the basic skills that are involved in regulating behavior, including paying attention, following instructions, and controlling inappropriate actions (McClelland et al., 2000; Rimm-Kaufman et al., 2000). In Japan, this is called the first-grader problem (Sakakihara, 2010), which denotes that teachers assigned to first grade face teaching obstacles, because an increasing number of children suffer from such behavioral problems as being noisy, leaving their seats, and disrupting class activities. The Tokyo metropolitan board of

education surveyed 1313 Tokyo public primary schools in 2009 and discovered such problems in about one-quarter of the schools. Not surprisingly, many studies have reported that such classroom behavior problems negatively influence student performance in reading, writing, and math (Klein, 2002; Lutz and Intermediate Unit, 2003; Spira and Fischel, 2005; Miles and Stipek, 2006; Bub et al., 2007; McClelland et al., 2007; Raver et al., 2008; Bulotsky-Shearer and Fantuzzo, 2011). This problem must be solved to avoid impeding children's development.

Teachers and researchers have addressed the development of a curriculum for school readiness that includes listening training (Brigman and Webb, 2003; Webster-Stratton and Reid, 2004; Denham, 2006). For example, the *Incredible Years Child Training Program* guides children in learning how to make friends and follow school rules, how to listen, wait, avoid interruptions, and quietly raise their hands to ask questions through practical training and small group discussions (Webster-Stratton and Reid, 2004). Unlike specific training, using supportive systems that improve the classroom's listening environment allows teachers to bypass the training time for school readiness because these systems can support children's listening in parallel with lessons, allowing teachers to devote more class time to regular lessons. For example, the *sound field amplification system* (Millett, 2008; Dockrell and Shield, 2012), which offers the possibility of immediately minimizing the impact of poor classroom acoustics on student learning, projects the teacher's voice so that children will have a better opportunity to clearly hear his/her instructions. This system does not reduce exposure to external sound sources. But importantly, raising the volume of the teacher's voice increases the speech signal levels relative to the levels of other sound sources. The impact of these systems was expanded to support children with hearing loss and to meet the recommended acoustical standards for noise levels and reverberation times. They also facilitate children's ability to discriminate words and spoken language more accurately and achieve better standardized test scores in early literacy and statistically and significantly improve attention, communication, and classroom behavior ratings (see Millett, 2008 for a review).

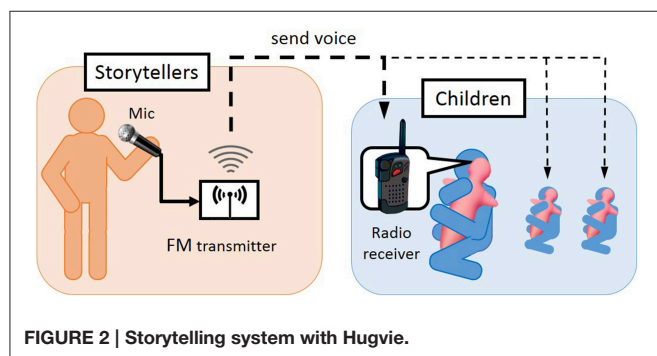
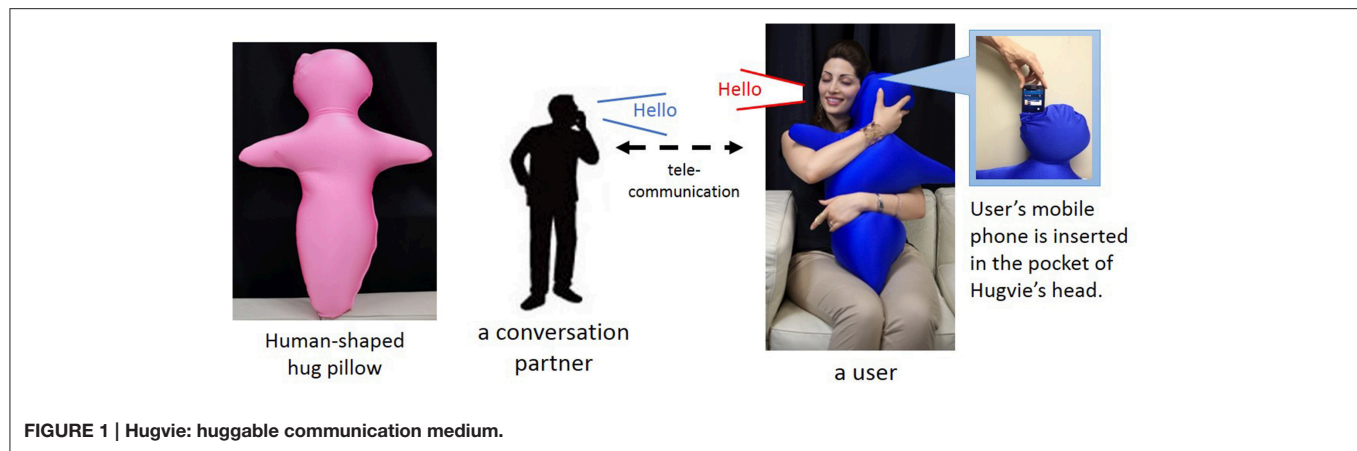
Although they successfully provided opportunities to acquire listening skills by improving the external conditions of classrooms, they do not help students prepare their own internal states for listening. Human mental states are important in the educational curriculum for readiness to learn, including listening (Raver and Knitzer, 2002; Brigman and Webb, 2003; Webster-Stratton and Reid, 2004; Denham, 2006; Thompson and Raikes, 2007) because they influence our ability for self-control (Baumeister and Heatherton, 1996). Stress and anxiety make it difficult for people to control themselves and concentrate on speakers (Vogely, 1998). This is a serious problem for children due to their limited ability to exercise self-control. Actually children, especially first graders, often feel stress in their school environment, relationships with classmates, and lessons (Fabian and Dunlop, 2007; Wong, 2015). Systems that support both the internal and external conditions of listeners must be developed.

In this context, we focus on social interactions where people touch each other, such as a caregiver holding a child and reading a story with a picture book to her/him. Such interactions have

two advantages for encouraging children to concentrate on listening. First is the impact of the tactile channel on stress reduction, which is one known effect of interpersonal touch (Gallace and Spence, 2010). Unlike other methods for decreasing stress by visual or auditory stimulation (Katcher et al., 1984; Pelletier, 2004; Labbé et al., 2007), tactile stimulation reduces stress without disturbing the audiovisual information provided by speakers in typical lectures. We can listen to and look at a lecture while touching something; however, that is difficult while listening to or looking at others. Second is the intimate distance shared by a speaker and listener. Such distance easily draws the listener's attention to the speaker's voice because it might be the strongest stimuli among others, as in sound field amplification systems (Millett, 2008; Dockrell and Shield, 2012). Another problem is that teachers cannot simultaneously establish close interactions with every student. Even when just a few children crave physical contact from their teachers, physical contact limits the teacher's behaviors, such as writing on the blackboard. Therefore, that solution cannot be achieved in the present educational environment.

We introduce a human-like communication medium as a proxy for teachers to achieve intimate social interaction in classrooms and support both forming external information and preparing mental states for listening. In this study, we use a huggable communication medium called Hugvie with which users can strongly experience the presence of remote partners while hugging it (Minato et al., 2013) (**Figure 1**). Hugvie, whose body is mainly a cushion in a human-like shape, allows users to feel as if they are hugging conversation partners by squeezing something human-like and hearing a voice near their ears. Since a previous study has already shown that conversation with Hugvie reduces stress (Sumioka et al., 2013), we expect that it will also help children prepare themselves for listening to others by improving both their external conditions and mental states.

However, since this is the first study that introduces a huggable communication medium into classroom activities, it remains unclear how children and educators will react to it and whether they will accept it. In this paper, we present three case studies where we introduced Hugvie in storytime settings and observed how children react to investigate whether it improves children's listening. We also investigated its acceptability by children and storytellers because acceptability to new information systems indicates their successful introduction into our lives (Nickerson, 1981; Gould et al., 1991; Davis, 1993). In particular, human-like devices might be rejected, as implied by the "uncanny valley" effect, which suggests that people have uncomfortable feelings to human-like robots as their appearances become more human-like (Mori et al., 2012). This effect is usually discussed in interaction between adults and very human-like robots. But one study implied that children do not exhibit positive responses to a robot with a more abstract human representation (Yamamoto et al., 2009). Furthermore, children may hesitate to hug such devices because they can feel their teacher's presence from Hugvie. Therefore, in this paper, we qualitatively and quantitatively investigate these two possibilities, the improvement of children's listening with Hugvie and social acceptance to Hugvie, through field observations and



discuss supporting children's listening by a communication medium.

2. MATERIALS AND METHODS

2.1. Hugvie

Hugvie, a huggable communication medium, is a human-shaped cushion (75-cm high and 600 g) that was designed as a communication device to give users a hugging experience. It is a soft cushion filled with polystyrene microbeads and covered with spandex fiber. Putting a hands-free mobile phone inside a pocket of its head enables users to talk while hugging it (Figure 1), increasing the feeling they are actually hugging a distant conversation partner.

2.2. Storytelling System with Hugvie

We focused on storytelling as a typical activity since teachers spend more than half of their class time on verbal instruction from elementary school to college school in different countries (Janusik and Wolvin, 2009) and it is often used as a teaching tool for organizational learning and received wisdom (Haigh and Hardy, 2011). Storytelling in elementary schools is usually done in one-to-many communication; a storyteller reads a picture book to many children, while Hugvie is used in one-to-one interactive communication (e.g., Minato et al., 2013; Sumioka et al., 2013). Therefore, we applied radio broadcasting

for one-to-many storytelling by putting a radio receiver inside Hugvie instead of a mobile phone.

Figure 2 shows our radio broadcasting system for storytelling. Storytellers tell the child listeners a story by showing a picture book through a microphone connected to a FM radio transmitter. All of the children listen to the storyteller's voice near their ears through radio receivers while hugging their Hugvies. Note that children can also directly listen to the storyteller's voice since both are in the same room. However, they will probably feel that the storyteller is whispering to them since they simultaneously hear the storyteller both directly and through the radio receivers.

2.3. Case Study 1: Introducing Huggable Communication Media into General Storytime for Children

2.3.1. Aim

For investigating the impact of a huggable communication medium on children's listening and its acceptability by children and teachers, we introduced Hugvie into a storytime activity and observed the responses of children and storytellers. Storytime includes just storytelling and one with using tools such as pictures, books, and toys. We observed storytime to allow us to get much information about children's listening because they are mainly listening during storytime. We conducted a field experiment to observe the natural responses of children and teachers to Hugvie.

2.3.2. Subjects and Procedure

Thirty-three preschool children who are 5 or 6 years old participated in a storytime event. This study was approved by the ethics committee of the Advanced Telecommunications Research Institute International (Kyoto, Japan). Since the subjects were young children, we explained this study to all the parents and received informed consent from them. We received permission from the parents and the school to include the image records of the children for research purposes. The child participants were given Hugvies at the school's library and shown the correct posture for using them by a male experimenter: sitting straight and hugging their Hugvie to enable a device at its head to contact the children's own ear (Figure 1). We confirmed that all

of the children could hear the male experimenter's voice from Hugvie at a comfortable volume after adjusting the volume on the radio receiver inside each child's device. Female volunteers with much story-time experience did storytime for children. At the beginning, a volunteer did a few tricks and sang rhyming songs with the children for about 4 min to make sure that the children realized how Hugvie works. Then three other volunteers told them a story illustrated with picture cards for about 7 min (**Figure 4**). After another 3-min trick show, another story was told for about 11 min. We call these trials where the children used Hugvie the *Hugvie condition*. After that, we collected the Hugvies from the children and two paper-cutting activities were performed for about 26 min (**Figure 5**), where two different volunteers told two stories while cutting colored paper and combined them into the characters and the scenery from the stories (*typical condition*). Finally, all the children sang while a volunteer played the piano.

2.3.3. Measurement

Two coders who did not know the purpose of the experiment analyzed the recorded movies to identify the behavioral differences between the typical and Hugvie conditions. Since the time between the two conditions was different, we used the first 25 min of the movies in each condition. Children sometimes moved beyond the video camera or overlapped with another child since they were more active than we expected in the typical condition. We eliminated their data from further analysis when at least one of the coders had difficulty judging their face directions. After this preprocessing, the data collected from some children became much smaller than in the storytime because they disappeared many times from the video camera. Therefore, we used the data collected from 29 children who were observed more than the 75 percent of the whole movie in each condition for our analysis.

We defined not listening to the storytellers as children who did not direct their faces toward the storytellers as captured from the movie data. The coders coded whether each child listened to the storytellers on a second-by-second basis through the movies. The inter-coder agreement score through the used data was $\kappa = 0.62$, indicating substantial inter-observer reliability (Viera and Garrett, 2005). We calculated the not-listening rate (NLR) for each child in each condition to evaluate the behavior of the children with the data where both coders agreed on child (not) listening: $NLR = NL/(NL + L)$, where NL indicates the total not-listening time and L is the total listening time.

2.3.4. Results

Hugvie produced big changes in the children's behaviors. **Figure 3** shows the listening scores in the typical and Hugvie conditions. We found significant differences between them with a paired t -test ($t = -6.83$, $p < 0.001$, ES: $d = 1.27$). **Figures 4**, **5** show the typical behaviors of children whose attention was drawn to something else in the two conditions. In the typical condition, some children walked around the room and talked or played with others after losing interest in the storytellers. The children who were far from the storytellers tended to engage in such behavior. On the other hand, such behaviors did not occur

when children used Hugvie, although a few children looked away from the storyteller. Interestingly, the children at the back of the room seemed to listen to the volunteers' voices from Hugvies

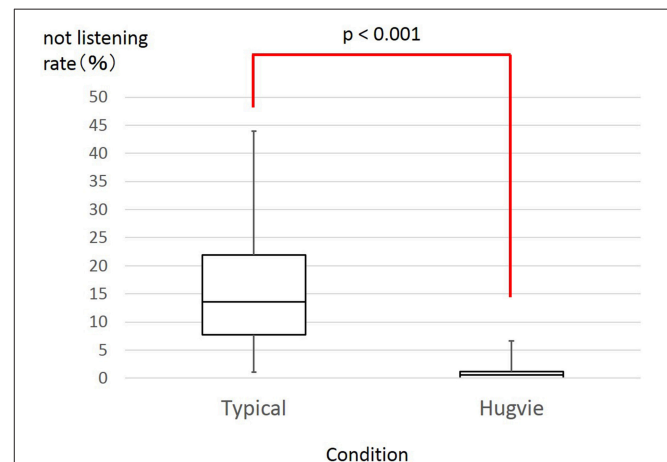


FIGURE 3 | Rate of children who directed their faces at something other than volunteers.



FIGURE 4 | Storytime with Hugvie (12 min. later): two of 30 children (total number countable from this figure) became distracted (white dotted circles).



FIGURE 5 | Typical storytime (57 min. later): 10 of 30 children became distracted (total number countable from this figure) (white dotted circles).

without any complaints even though they had difficulty seeing the picture cards.

No children rejected or showed dislike of Hugvie, though some children did not use it. Two children who did not understand how to use it were helped by volunteers, and some near the storytellers listened directly to the story instead of through their Hugvies. They seemed to feel comfort and fun from Hugvie. For example, some children said *“It really feels good!”* when they hugged their Hugvies. One girl in the back also appeared to be having fun during storytime and expressed her feeling to an experimenter.

The teachers and volunteers who observed the event were surprised at the result. A female volunteer said, *“I’m really surprised that Hugvie easily calmed the children because we usually spend lots of effort relaxing the children and keeping them calm so that they can pay attention to the story. I want to introduce Hugvie into other activities like storytime to toddlers or elderly people.”* Two other volunteers made similar comments.

2.3.5. Discussion

We found that listening through Hugvie decreased the number of children who didn’t seem to listen. While children were often distracted during typical storytime, children with Hugvie paid more attention to the storytellers. This effect appears stronger for children in the back of the room since they tend to lose focus without Hugvie due to their distance from the storyteller. On the other hand, the closer the children are to the storytellers, the weaker this effect might be since some children near the storytellers listened without their Hugvies.

Children showed no negative impressions toward Hugvie. Rather, they often expressed positive impressions such as comfort and fun. Note that the children accepted Hugvie not only in the storytime sessions by the female volunteers but also in instruction about it by a male experimenter. Perhaps, Hugvie is basically accepted by children in storytime regardless of the gender of adult storytellers.

The educators and the volunteers were also surprised at the changes in the children. This implies that the introduction of Hugvie is useful in school education. One teacher suggested that Hugvie was cast as a proxy of the storyteller: *“Basically, the students are listening to their teacher in a one-to-one conversation even though some have difficulty focusing on their teacher in class. Listening through Hugvie might enhance their feeling of a storyteller who’s talking directly to them.”* A volunteer pointed out a change in her storytime: *“I concentrated on reading the book since I didn’t need to read loudly so that the children in the back could hear me.”* Storytime with Hugvie might facilitate children’s listening by allowing storytellers to devote more concentration on telling a story.

Hugvie showed the potential of a huggable communication medium to facilitate children’s listening. However, we need more trials to test its effects because this case study is not a perfect comparison; the two environmental conditions are different. For example, storytime with Hugvie was done before the typical condition. Children might be nervous because they have few experiences of being in school, so that they might not talk and play in the former condition. Another difference is that the

length of the concentration required in typical storytime is longer than with Hugvie because it is hard for children to maintain concentration for a long time. Since the rest time is also less in the typical condition than in the Hugvie condition, children might be so tired that they became easily distracted in the latter. The storytime contents were also different. For storytime without Hugvie, the volunteers often said nothing while cutting paper. Such a boring time might cause children to lose interest in the storytime. However, given the fact that volunteers with much storytime experience felt surprised by the children’s behavior, Hugvie might still positively impact listening. Such surprises reflected the children’s changes from more than just a few of them who didn’t listen.

Practically, storytime styles vary in certain situations and such differences might change the listening support effect. For example, various persons can be storytellers. In this case study, the storytellers were mainly women with much storytime experience. Their expertise might induce Hugvie’s effect. Can amateur storytellers promote the Hugvie effect? Another example is a group activity that is often performed as a class activity. While only one story was told to children at the same time in this experiment, members of different groups tell different stories to other group members in parallel. In such a situation, children have to listen to their storyteller in a noisier situation than in this experiment. Can Hugvie still support children? To address these questions and investigate how different storytime situations affect Hugvie’s supportive effect, we introduced it into storytime by child storytellers as a group activity in case study 2.

2.4. Case Study 2: Introducing Huggable Communication Media into Simultaneous Storytime in Children Groups

2.4.1. Aim

To investigate whether Hugvie encourages children to concentrate on listening in such noisier situations as group activities, we introduced it into simultaneous storytime in children groups as a different storytime style from case study 1. Since most children’s speaking skills are less advanced than those of adults, casting a child as the storyteller can investigate whether, regardless of a storyteller’s speaking skills, Hugvie prompts listening. Additionally, we set at most four storytime groups at the same time to observe Hugvie’s effect in a noisier situation. Such an investigation is valuable not only because it is the first such trial of a huggable communication medium but also because it is more difficult to pay attention to a story without Hugvie in those situations; lesser speaking skills disturb precise listening comprehension, and in simultaneous storytime groups, storyteller voices offset each other.

2.4.2. Subjects and Procedure

We introduced Hugvie into storytime sessions in the elementary school event to 139 preschool children who are 5 or 6 years old. They were divided the children into 34 groups of three to five kids with two or three 5th graders as guides of the school. Each group could freely join several sessions (including storytime) in the event. This study was approved by the ethics committee of the

Advanced Telecommunications Research Institute International (Kyoto, Japan). We explained our study to all of the children's parents and received informed consent from them. In the storytime events at the school's library, they were given Hugvies and instructed how to use them by showing the correct posture as described in case study 1. We confirmed that all children could comfortably hear the experimenter's voice from their Hugvies. After that the 5th graders told stories with picture books to the preschool children for 10 min (**Figure 6**). At most four groups of storytime were held at the same time in the same room. The other groups waited in the room until some of the four groups had finished and moved on to other events.

2.4.3. Measurement

We video-taped the storytime sessions and observed the children through the recorded movies. We received permission to include image records of the children from their parents and the school for research purposes. We categorized the children who did not direct their faces to the storytellers as children who did not listen to the story.

2.4.4. Results

As the event continued, the room got louder owing to the children who were waiting to join the storytime with Hugvie. Some children chased each other around the room, and others played and/or talked with their friends or their fifth-grade guides. A few waited in silence. However, most children concentrated on the listening to the story in silence once they joined the Hugvie storytime session. Only 6% sometimes lost their attention, but they soon resumed listening without walking around or talking with others. No children rejected Hugvie. They showed such positive impressions as looking comfortable, as we observed in case study 1 when they held Hugvies.

2.4.5. Discussion

Our results showed new potential applicable occasions for Hugvie. Regardless of the low speaking skills of the 5th grade storytellers, the preschool children listened with Hugvie. This means that anyone can be a storyteller regardless of speaking skills.



FIGURE 6 | Simultaneous storytime in children groups.

This result also suggests that Hugvie reduces not only impediments in the listening process but also the requirement needed for speaking. Although the experiment room was quite noisy due to simultaneous storytime and children who were waiting to join storytime sessions, they concentrated on listening with Hugvie. Hugvie enabled us to hold storytime in noisy environments because it produced the speaker's voice near the user's ears and relaxed the children. This achievement is completely different from what was reached by a listening support device, such as the sound field amplification system, because such devices drown out other sounds in the entire room by amplifying the speaker's voice.

We did find one negative aspect of storytime with Hugvie with respect to body posture from recorded movies. As the storytime continued, a few children showed incorrect postures, although they were correctly holding Hugvie at the beginning of the storytime: leaning on or lying astride it. While 83% held Hugvie as instructed by the experimenter, 10% leaned on Hugvie and 7% lay astride it (**Figure 7**). This might be a problem for its introduction into school education because posture is important for health management related to physical development and visual loss (Kratěnová et al., 2007). Therefore, we need to improve Hugvie to prompt children to maintain good posture. Our observation suggests that its softness caused bad posture. Since Hugvie is easy to bend and fold, children sitting on the ground tended to bend their backs and lie on Hugvie.

As with case study 1, none of the 139 children rejected Hugvie. However, this does not mean that all of the children were pleased with it. Some might have used it because the adults asked them to do so. There is room to investigate acceptability; performance may fall if children are unwilling to use a device. Since previous introductions of support devices into schools (Tanaka et al., 2013; Komatsubara et al., 2014) showed the importance of willingness to use, in case study 3 we asked the children whether they are willing to use Hugvie after storytime with it.

2.5. Case Study 3: Willingness to Use Huggable Communication Media

2.5.1. Aim

For investigating how willing children are to use Hugvie, we gave them the option of using it or not in storytime after they and their parents experienced storytime with Hugvie once. Observing whether children used it in that situation shows their willingness



leaning on Hugvie



lying astride Hugvie

FIGURE 7 | Two listening behaviors.

to use it. We also asked the children about their impressions of Hugvie.

2.5.2. Subjects and Procedure

We introduced Hugvie into storytime for children at a science museum in Tokyo called Miraikan. Our participants, 29 children and their parents, were gathered in an open space of the museum by its staff members who explained the event. This study was approved by the ethics committee of the Advanced Telecommunications Research Institute International (Kyoto, Japan). We explained this study to all the parents of the subjects and received informed consent from them.

The experiments consisted of two sessions: forced and free. In the forced session, the participants were divided into two groups by families, and one group was given Hugvies and instructed how to use them. After we confirmed that all children could comfortably hear Hugvie's voice, a female volunteer with much storytime experience with children read a story with a picture book. Then we collected the Hugvies and gave them to the other group, and the volunteer told a story with an another picture book. Each storytime session lasted about 5 min.

After the forced session (storytime with parents) finished, a free session was conducted. A female staff member of the museum gathered only the children and asked them whether they wanted to use Hugvie for another storytime session (Figure 8). She also asked them to express their thoughts about Hugvie. Then she read another book with/without Hugvie according to their own willingness to use. During the free session, an experimenter explained the purpose of the experiments and studies with Hugvie to their parents in the back of the area and asked them by a questionnaire for their impressions about their children using Hugvie. The experiments were recorded. The event was held twice: 15 children participated in the first event and 14 in the second. The child participants ranged in age from 3 to 10.

2.5.3. Measurement

We counted the number of children who used Hugvie in the free session of each event and also checked their impressions of it in the free sessions. We collected comments from 21 parents

about their impressions of their children using Hugvie. Two coders who did not know the purpose of our experiment read all the comments and categorized the parent impressions of children using Hugvie as positive, negative, or neutral. The inter-coder agreement score was $\kappa = 0.83$, indicating almost perfect inter-observer reliability (Viera and Garrett, 2005). In addition, we extracted the behavioral differences of the children between storytime with and without Hugvie in the forced session through the recorded movies. We received permission to include the image records of the children from their parents and the museum for research purposes.

2.5.4. Results

When we asked children whether they wanted to use Hugvie, six of 15 and seven of 14 used Hugvie in the first event. In interviewing the children in the second event, the children who were pleased with it made such comments as, "Using it allowed me to listen more clearly" and "It's so cute." On the other hand, the children who were unwilling to use it commented that "It's difficult for me to hug it and listen" and "Hugvie's voice was so loud that it gave me a headache" (Table 1).

The results of the parents' impressions showed that more parents had positive impressions than negative. Twelve felt Hugvie had a positive effect on their children. Eight of 12 recognized that their children concentrated more on listening to the story with Hugvie. One mother reported that her child seemed to come back to her to be comforted during the storytime session without Hugvie. Three others hoped to use Hugvie in kindergartens or while their children were alone at home. Seven parents showed negative impressions of Hugvie. One father said he did not notice any Hugvie effect on his child. One mother found that her child looked sleepy. Two parents were worried that their children would get bored with Hugvie, and three parents wanted the interface to be improved, such as the sound quality and ease of use. The rest of the parents reported partial positive impressions of Hugvie for its usefulness for children who are far away from the storytellers although one of two coders categorized their impressions as negative or neutral.

We found some interesting behaviors of the children in the forced sessions. During storytime, nine ran up to and grabbed their parents when they were not using Hugvie, although they did not do that while using it. Two children with slightly smaller bodies than Hugvie repeatedly quit paying attention to a storyteller regardless of the conditions, and the other children almost always concentrated on listening to the storytime in both conditions.



FIGURE 8 | Asking children whether they want to use Hugvie after storytime.

TABLE 1 | Number of children who willingly used Hugvie and the reasons of their decisions.

	Willing to use	Unwilling to use
Number of children	13	16
Reason	Able to listen clearly Hugvie is cute	Difficult to hug and listen Too noisy Hugvie is not cute

2.5.5. Discussion

Approximately half of the children were willing to use Hugvie, which means that a fair number of them were attracted to it after using it just once. It remains unclear whether the device's rate is high enough to introduce it into schools because no studies exist on the educational applications of similar communication devices. However, the rate is important as a baseline to improve Hugvie in respect to willingness to use it.

Our interviews and questionnaires showed that many children and their parents felt that Hugvie prompted users to concentrate on listening. In other words, Hugvie had such a strong effect that users noticed the difference caused by it. On the other hand, a few users did not feel any effect. We infer that this was mainly caused by the interface problems, including unsuitable size, sound quality, and/or ease of understanding how to use it. For example, Hugvie requires users to place their ears near the speaker because it is not very loud. Thus, misunderstanding the speaker location prevents adequate listening to the story through Hugvie, reducing its effect. These findings are important for improving Hugvie and the design policy of such support devices for telecommunication and physical interaction.

Children often run to and grab their parents, suggesting a desire to reduce their feeling of loneliness by making physical contact (Gallace and Spence, 2010). However, after using Hugvie, the children did not rush to greet their parents. We infer that this shows that using Hugvie reduced their feeling of loneliness the children felt during the storytime. If their parents were not near them when they were not using Hugvie, they would be distracted away from the storytime. Perhaps Hugvie encourages listening by improving not only the external condition but also the internal condition. On the other hand, two young children did not listen calmly in either storytime condition. Their bodies were too literally small to use Hugvie. In this case, its unsuitable size disrupted its use and reduced its effect.

2.6. General Discussion

Out of the 201 children in all the case studies, none rejected our huggable communication medium, which suggests Hugvie might be accepted by most preschool children. Yamamoto et al. reported that 2- to 3-year-old children did not exhibit positive responses to a small robot with a non-human-like appearance that showed human-like contingent actions. They argued that perhaps the children experienced the uncanny valley effect due to the conflict between the robot's appearance and its human-like actions (Yamamoto et al., 2009). Although Hugvie has such an intrinsic conflict between its abstract human form and a human voice from its inner communication device, our results suggest that it does not produce negative feelings in children. Schools might be receptive to introducing huggable communication media into their curriculums.

However, not all of the children were satisfied with Hugvie. Some were unwilling to use it due to the difficulty of hugging and listening. Observations and user opinions suggested that the difficulty was caused by Hugvie's usability, including its size and stiffness, sound quality, and/or user-friendliness. This feedback

provides insights into ways to improve Hugvie and highlights future research issues to be addressed before we introduce it into school education.

Case studies 2 and 3 suggested that such physical features as stiffness and size must be suitable for users. In case study 2, we found a potential problem when children use improper listening postures with Hugvie. Adult users never showed such postures since Hugvie was designed to be suitable for them. Children lean on Hugvie for support due to the immaturity of their musculoskeletal systems while adults can maintain their posture by themselves. We will verify our inferences in the future using another version of Hugvie that is stiff enough to support a child's body.

As reported in case study 3, Hugvie distracted children from listening if its size is inappropriate since small children with smaller bodies who had difficulty holding Hugvie often became distracted away from the storyteller. Another possible reason is that such distraction is caused not by a size mismatch but age. Younger children lacked the ability to sustain attention for a long time. Therefore, interesting future work might investigate the influence of size mismatch between users and Hugvie for listening with a smaller type of Hugvie.

The interviews and questionnaires of case study 3 also suggest that some users could not listen well with it because they did not understand how to use it. We need to improve Hugvie's interface to reduce such future misunderstandings. For example, marking where users should place their ears is a possible improvement. An automatic volume control system while holding Hugvie would allow each user to adjust Hugvie's volume.

As reported in case study 1, the children near the storytellers attentively listened without holding Hugvie because the storyteller's voice was louder than the sound from Hugvie. Perhaps Hugvie's voice should be the strongest stimuli among the surrounding sounds, including the storyteller's direct voice, to encourage children to use Hugvie. Although children do not need to use it when they are near a storyteller, they might benefit from using it in other aspects, as suggested in studies on interpersonal touch. For example, a brief touch from teachers motivates children to participate in lessons (Guéguen, 2004). We expect similar effects on children when they are holding Hugvie. Tactile stimulation from it would encourage the voluntary behavior of children when they listen to a teacher's request through Hugvie. Further investigation is needed.

We also found evidence that Hugvie might benefit both teachers and children. In case study 1, as pointed out by a volunteer, the storytellers concentrated more on the story's content with Hugvie since they did not need to speak so loudly. Previous studies report that teachers often suffer from such voice problems as phonation difficulties, hoarseness, and throat pain because they have to speak loudly to control their classrooms (Yiu, 2002). Sound field amplification systems provide a possible solution to this problem. However, increasing the sound volume in a classroom might disturb the class in the next classroom if rooms are not properly soundproofed. On the other hand, Hugvie reduces the noise level in class and improves the external conditions because it enables teachers to talk in a lower voice and children to concentrate on listening in class.

Therefore, Hugvie helps reduce the voice problems experienced by teachers and enables them to concentrate on improving their teaching.

Our results also show the possibility of Hugvie's future applications. We found that our proposed storytelling system enables children to become immersed in a story even with an inexperienced storyteller in noisy environments. We also expect that Hugvie can be introduced into other activities, such as interaction with senior citizens and group work. Interaction between children and seniors is difficult because the listening skills of the latter are often poor and children's speaking skills are immature. The results of case study 2 tell us that Hugvie can deal with both problems.

Group work requires concentration on conversation among the group members. However, usually some voices are drowned out by other group conversations. Hugvie's vocal sounds can overcome the surrounding conversations without offsetting them. It can also evoke interest in a speaker (Nakanishi et al., 2013), indicating that it encourages the involvement of each member in group discussions.

Although all of our case studies suggest that a huggable communication medium has a possibility to support children's listening skill, further investigation is needed. First, we have to evaluate Hugvie's effect on children in more controlled conditions. Another important issue to be addressed is the investigation of how deeply Hugvie affects cognition. In this paper, we focused on the changes in the children's behaviors and social acceptability to Hugvie since this is the first study that introduced a huggable communication medium into educational situations. However, perhaps listening through Hugvie enhances information comprehension and memory more than usual listening. Actually, such enhancements are needed in education. Many graduate school students of college have high listening skills (McDevitt et al., 1991), and most college students who fail examinations lack listening skills (Conaway, 1982). As a next step, we have to verify a story's comprehension with some sort of listening comprehension quiz.

REFERENCES

- Baumeister, R. F., and Heatherton, T. F. (1996). Self-regulation failure: an overview. *Psychol. Inq.* 7, 1–15.
- Bohlken, B. (1999). Substantiating the fact that listening is proportionately most used language skill. *The Listening Post*, 70, 5.
- Brigman, G. A., and Webb, L. D. (2003). Ready to learn: teaching kindergarten students school success skills. *J. Educ. Res.* 96, 286–292. doi: 10.1080/00220670309597641
- Bub, K. L., McCartney, K., and Willett, J. B. (2007). Behavior problem trajectories and first-grade cognitive ability and achievement skills: a latent growth curve analysis. *J. Educ. Psychol.* 99:653. doi: 10.1037/0022-0663.99.3.653
- Bulotsky-Shearer, R. J., and Fantuzzo, J. W. (2011). Preschool behavior problems in classroom learning situations and literacy outcomes in kindergarten and first grade. *Early Childhood Res. Q.* 26, 61–73. doi: 10.1016/j.ecresq.2010.04.004
- Conaway, M. S. (1982). "Listening: learning tool and retention agent," in *Improving Reading and Study Skills*, eds A. S. Algier and K. W. Algier (San Francisco, CA: Jossey-Bass), 51–63.

3. CONCLUSION

Through three case studies, we demonstrated that huggable communication media show possibilities to encourage children to listen to others. Our huggable communication medium, Hugvie, virtually enables intimate interactions with conversation partners to improve external and internal conditions for listening. Our results showed that Hugvie, which addressed the classroom problem where children did not listen to a speaker, is accepted by children, their caregivers, and their educators. Our results also suggest that Hugvie can support communication between people who sometimes suffer from low speaking skills and low listening skills, such as children and seniors. We hope the intimate interactions mediated by huggable communication media can reduce problems of school education and other situations where listening skills are crucial and encourage people to learn from others.

AUTHOR CONTRIBUTIONS

JN and HS wrote the main manuscript text. All authors designed research.

ACKNOWLEDGMENTS

This work was mainly supported by the Japan Science and Technology Agency (JST), the Core Research of Evolutional Science and Technology (CREST) research promotion program. Part of this work was supported by JST, the Exploratory Research for Advanced Technology (ERATO), and the ISHIGURO symbiotic human-robot interaction project. The authors would like to thank the staff at the Higashi-Hikari elementary school and Miraikan and all participants (children and their parents) for their cooperation. The authors also would like to thank the reviewers for their valuable comments and suggestions to improve the quality of the manuscript.

- Davis, F. D. (1993). User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *Int. J. Man-Machine Stud.* 38, 475–487.
- Denham, S. A. (2006). Social-emotional competence as support for school readiness: what is it and how do we assess it? *Early Educ. Dev.* 17, 57–89. doi: 10.1207/s15566935eed1701_4
- Dockrell, J. E., and Shield, B. (2012). The impact of sound-field systems on learning and attention in elementary school classrooms. *J. Speech Lang. Hear. Res.* 55, 1163–1176. doi: 10.1044/1092-4388(2011/11-0026)
- Fabian, H., and Dunlop, A. W. (2007). "Outcomes of good practice in transition processes for children entering primary school," in *Working Papers in Early Childhood Development*, No. 42 (The Hague: Bernard van Leer Foundation).
- Gallace, A., and Spence, C. (2010). The science of interpersonal touch: an overview. *Neurosci. Biobehav. Rev.* 34, 246–259. doi: 10.1016/j.neubiorev.2008.10.004
- Gould, J. D., Boies, S. J., and Lewis, C. (1991). Making usable, useful, productivity-enhancing computer applications. *Commun. ACM* 34, 74–85.

- Guéguen, N. (2004). Nonverbal encouragement of participation in a course: the effect of touching. *Soc. Psychol. Educ.* 7, 89–98. doi: 10.1023/B:SPOE.0000010691.30834.14
- Haigh, C., and Hardy, P. (2011). Tell me a story—a conceptual exploration of storytelling in healthcare education. *Nurse Educ. Today* 31, 408–411. doi: 10.1016/j.nedt.2010.08.001
- Imhof, M., and Weinhard, T. (2004). “What did you listen to in school today,” in *25th Annual Convention of the International Listening Association* (Ft. Myers, FL).
- Janusik, L. A., and Wolvin, A. D. (2009). 24 hours in a day a listening update to the time studies. *Int. J. Listening* 23, 104–120. doi: 10.1080/10904010903014442
- Katcher, A., Segal, H., and Beck, A. (1984). Comparison of contemplation and hypnosis for the reduction of anxiety and discomfort during dental surgery. *Am. J. Clin. Hypn.* 27, 14–21.
- Klein, L. G. (2002). Set for success: building a strong foundation for school readiness based on the social-emotional development of young children. *Kauffman Early Educ. Exch.* 1, 1–5. doi: 10.2139/ssrn.2355477
- Komatsubara, T., Shiomi, M., Kanda, T., Ishiguro, H., and Hagita, N. (2014). “Can a social robot help children’s understanding of science in classrooms?” in *Proceedings of the Second International Conference on Human-Agent Interaction* (Tsukuba: ACM), 83–90.
- Kraténová, J., Žejglicová, K., Malý, M., and Filipová, V. (2007). Prevalence and risk factors of poor posture in school children in the Czech republic. *J. School Health* 77, 131–137. doi: 10.1111/j.1746-1561.2007.00182.x
- Labbé, E., Schmidt, N., Babin, J., and Pharr, M. (2007). Coping with stress: the effectiveness of different types of music. *Appl. Psychophysiol. Biofeedback* 32, 163–168. doi: 10.1007/s10484-007-9043-9
- Lutz, M. N., and Intermediate Unit, C. C. (2003). A multivariate analysis of emotional and behavioral adjustment and preschool educational outcomes. *School Psychol. Rev.* 32, 185–203.
- McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., and Morrison, F. J. (2007). Links between behavioral regulation and preschoolers’ literacy, vocabulary, and math skills. *Dev. Psychol.* 43:947. doi: 10.1037/0012-1649.43.4.947
- McClelland, M. M., Morrison, F. J., and Holmes, D. L. (2000). Children at risk for early academic problems: the role of learning-related social skills. *Early Childhood Res. Q.* 15, 307–329. doi: 10.1016/S0885-2006(00)00069-7
- McDevitt, T. M., Sheehan, E. P., and McMenamin, N. (1991). Self-reports of academic listening activities by traditional and nontraditional college students. *College Stud. J.* 25, 478–486.
- Miles, S. B., and Stipek, D. (2006). Contemporaneous and longitudinal associations between social behavior and literacy achievement in a sample of low-income elementary school children. *Child Dev.* 77, 103–117. doi: 10.1111/j.1467-8624.2006.00859.x
- Millett, P. (2008). *Sound Field Amplification Research Summary*. Deaf and Hard of Hearing Program, Faculty of Education York. Available online at: <http://research.epicoustics.com> on 10.30.2015
- Minato, T., Nishio, S., and Ishiguro, H. (2013). “Evoking an affection for communication partner by a robotic communication medium,” in *Demonstration Session Proceeding of the 8th ACM/IEEE International Conference on Human-Robot Interaction D*, Vol. 7 (Tokyo).
- Mori, M., MacDorman, K. F., and Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robot. Automat. Mag.* 19, 98–100. doi: 10.1109/MRA.2012.2192811
- Nakanishi, J., Kuwamura, K., Minato, T., Nishio, S., and Ishiguro, H. (2013). “Evoking affection for a communication partner by a robotic communication medium,” in *The First International Conference on Human-Agent Interaction (IHAI 2013)* (Hokkaido).
- Nickerson, R. S. (1981). Why interactive computer systems are sometimes not used by people who might benefit from them. *Int. J. Man-Machine Stud.* 15, 469–483.
- Pelletier, C. L. (2004). The effect of music on decreasing arousal due to stress: a meta-analysis. *J. Music Ther.* 41, 192–214. doi: 10.1093/jmt/41.3.192
- Raver, C. C., Jones, S. M., Li-Grining, C. P., Metzger, M., Champion, K. M., and Sardin, L. (2008). Improving preschool classroom processes: preliminary findings from a randomized trial implemented in head start settings. *Early Childhood Res. Q.* 23, 10–26. doi: 10.1016/j.ecresq.2007.09.001
- Raver, C. C., and Knitzer, J. (2002). *Ready to Enter: What Research Tells Policymakers about Strategies to Promote Social and Emotional School Readiness Among Three- and Four-Year-Old Children*. New York, NY: National Center for Children in Poverty.
- Rimm-Kaufman, S. E., Pianta, R. C., and Cox, M. J. (2000). Teachers’ judgments of problems in the transition to kindergarten. *Early Childhood Res. Q.* 15, 147–166. doi: 10.1016/S0885-2006(00)00049-1
- Sakakihara, Y. (2010). *East Asia Child Science Exchange Program 6: First-Grader Problems at Elementary School and Developmental Disorders*. Available online at: <http://www.childresearch.net/events/exchange/on10.30.2015>
- Spira, E. G., and Fischel, J. E. (2005). The impact of preschool inattention, hyperactivity, and impulsivity on social and academic development: a review. *J. Child Psychol. Psychiatry* 46, 755–773. doi: 10.1111/j.1469-7610.2005.01466.x
- Sumioka, H., Nakae, A., Kanai, R., and Ishiguro, H. (2013). Huggable communication medium decreases cortisol levels. *Sci. Rep.* 3:3034. doi: 10.1038/srep03034
- Tanaka, F., Takahashi, T., and Morita, M. (2013). Tricycle-style operation interface for children to control a telepresence robot. *Adv. Robot.* 27, 1375–1384. doi: 10.1080/01691864.2013.826782
- Thompson, R. A., and Raikes, H. A. (2007). “The social and emotional foundations of school readiness,” in *Social and Emotional Health in Early Childhood: Building Bridges between Services and Systems*, eds D. F. Perry, R. K. Kaufmann, and J. Knitzerpages (Baltimore, MD: Paul H. Brookes Publishing Co.), 13–36.
- Viera, A. J., and Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam. Med.* 37, 360–363.
- Vogely, A. J. (1998). Listening comprehension anxiety: students’ reported sources and solutions. *Foreign Lang. Ann.* 31, 67–80.
- Webster-Stratton, C., and Reid, M. J. (2004). Strengthening social and emotional competence in young children? the foundation for early school readiness and success: incredible years classroom social skills and problem-solving curriculum. *Infants Young Child.* 17, 96–113. doi: 10.1097/00001163-200404000-00002
- Wong, M. (2015). Voices of children, parents and teachers: how children cope with stress during school transition. *Early Child Dev. Care* 185, 658–678. doi: 10.1080/03004430.2014.948872
- Yamamoto, K., Tanaka, S., Kobayashi, H., Kozima, H., and Hashiya, K. (2009). A non-humanoid robot in the uncanny valley: experimental analysis of the reaction to behavioral contingency in 2–3 year old children. *PLoS ONE* 4:e6974. doi: 10.1371/journal.pone.0006974
- Yiu, E. M. (2002). Impact and prevention of voice problems in the teaching profession: embracing the consumers’ view. *J. Voice* 16, 215–229. doi: 10.1016/S0892-1997(02)00091-7

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

JN declare no potential conflict of interest. HS and HI are employed by ATR. ATR has patents on Huggy. HI has consulted for Vstone Co., Ltd., which sells Huggy, and received compensation. He also owns stock in the company.

Copyright © 2016 Nakanishi, Sumioka and Ishiguro. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A truly human interface: interacting face-to-face with someone whose words are determined by a computer program

Kevin Corti* and Alex Gillespie

Department of Social Psychology, London School of Economics and Political Science, London, UK

OPEN ACCESS

Edited by:

Shuichi Nishio,
Advanced Telecommunications
Research Institute International, Japan

Reviewed by:

Takashi Minato,
Advanced Telecommunications
Research Institute International, Japan
Marco Nørskov,
Aarhus University, Denmark

*Correspondence:

Kevin Corti,
Department of Social Psychology,
London School of Economics and
Political Science, Houghton Street,
London WC2A 2AE, UK
kevin@kevincorti.org

Specialty section:

This article was submitted to
Human-Media Interaction,
a section of the journal
Frontiers in Psychology

Received: 25 March 2015

Accepted: 30 April 2015

Published: 18 May 2015

Citation:

Corti K and Gillespie A (2015)
A truly human interface: interacting
face-to-face with someone whose
words are determined
by a computer program.
Front. Psychol. 6:634.
doi: 10.3389/fpsyg.2015.00634

We use speech shadowing to create situations wherein people converse in person with a human whose words are determined by a conversational agent computer program. Speech shadowing involves a person (the shadower) repeating vocal stimuli originating from a separate communication source in real-time. Humans shadowing for conversational agent sources (e.g., chat bots) become hybrid agents (“echoborgs”) capable of face-to-face interlocution. We report three studies that investigated people’s experiences interacting with echoborgs and the extent to which echoborgs pass as autonomous humans. First, participants in a Turing Test spoke with a chat bot via either a text interface or an echoborg. Human shadowing did not improve the chat bot’s chance of passing but did increase interrogators’ ratings of how human-like the chat bot seemed. In our second study, participants had to decide whether their interlocutor produced words generated by a chat bot or simply pretended to be one. Compared to those who engaged a text interface, participants who engaged an echoborg were more likely to perceive their interlocutor as pretending to be a chat bot. In our third study, participants were naïve to the fact that their interlocutor produced words generated by a chat bot. Unlike those who engaged a text interface, the vast majority of participants who engaged an echoborg did not sense a robotic interaction. These findings have implications for android science, the Turing Test paradigm, and human–computer interaction. The human body, as the delivery mechanism of communication, fundamentally alters the social psychological dynamics of interactions with machine intelligence.

Keywords: android science, cyranoid, dialog systems, embodiment, human–computer interaction, speech shadowing, Turing Test, uncanny valley

Introduction

“Meaning is the face of the Other, and all recourse to words takes place already within the primordial face to face of language”

(Levinas, 1991, p. 206).

In comparison to other forms of interaction, face-to-face communication between humans is characterized by more social emotion, higher demands for comprehensibility, and increased social obligation; the face of the other commands an ethical relation that is absent in people’s interaction with “things” (Levinas, 1991). Face-to-face, close-proximity interaction between tangible bodies is

the primordial human inter-*face* and is the format of exchange most conducive for shared understanding (Linell, 2009). Computer technologies specifically designed to simulate human social functioning (e.g., conversational agents) have to date communicated with people via technical interfaces such as screens, buttons, robotic devices, avatars, interactive voice response systems, and so on. This leaves a need to explore human perception of and interaction with these technologies under conditions that replicate the full complexity of face-to-face human–human communication. The present article introduces a means of doing so. We demonstrate a methodology that allows a person to interact “in the flesh” with a conversational agent whose interface is an actual human body.

Contemporary Android Science

Android science aims to develop artificial systems identical to humans in both appearance and behavior (verbal and non-verbal) for the purposes of exploring human nature and investigating the ways in which these systems might integrate into human society (MacDorman and Ishiguro, 2006a; Ishiguro and Nishio, 2007). The field is as interested in better understanding people through their interacting with anthropomorphic technology as it is in further developing the technology itself. Considerable progress has been made in these endeavors, with perhaps the most notable work being that undertaken and inspired by Hiroshi Ishiguro of Osaka University’s Intelligent Robotics Laboratory, whose research and engineering teams have developed highly lifelike autonomous and semi-autonomous androids. MacDorman and Ishiguro (2006b) argue that in being controllable, programmable, and replicable, androids are in certain respects superior to human actors as social and cognitive experimental stimuli. They further contend that androids can evoke in humans expectations and emotions that attenuate the psychological barrier between people and machines.

The motor behaviors of autonomous androids are controlled by technologies that perceive and orient to the physical environment while their speech is controlled by a conversational agent. As autonomous technologies are still quite limited in terms of functionality, the social capacities of these types of androids are severely constrained. Tele-operated androids, meanwhile, overcome the limitations of fully autonomous models by way of a human operator controlling the android’s speech and movement (Nishio et al., 2007b). On account of their enhanced social capabilities, tele-operated androids have stimulated ample research in psychology and other domains of social and cognitive science. For instance, researchers have investigated the extent to which a person’s presence with remote others is amplified or weakened when tele-operating an android compared to when communicating in person or via more distal technological mediators such as video conferencing (Nishio et al., 2007a; Sakamoto et al., 2007). Researchers have also explored the extent to which tele-operators perceive their android to be extensions of themselves, sensing physical stimuli administered to the android as if the stimuli had been administered to their own body (Ogawa et al., 2012). Perhaps the most discussed phenomenon in the field of android science is the “uncanny valley,” posited by Mori (1970). This idea suggests

that the affinity a person has for an artificial agent will increase as the appearance and motor behavior of the agent becomes more human-like; however, at a certain point along the human-likeness continuum (where the agent begins to look more or less human but for slight, yet telling, signs of artificiality) feelings of affinity will sharply decline, before rapidly rising again as the agent becomes indistinguishable from an actual human (MacDorman and Ishiguro, 2006b; Seyama and Nagayama, 2007).

We propose inverting the composition of tele-operated android systems in order to create hybrid entities consisting of a human whose words (and potentially motor actions) are entirely or partially determined by a computer program. We refer to such hybrids as “echoborgs,” which can be classified as a type of “cyranoid”—Milgram’s (2010) term for a hybrid composed of a person who speaks the words of a separate person in real-time. Echoborgs can be used to examine the role of the human body, as the delivery mechanism of communication, in mediating social emotions, attributions, and other interpersonal phenomena emergent in face-to-face interaction. Furthermore, echoborgs can be used to evaluate the performance and perception of artificial conversational agents under conditions wherein people assume they are interacting with an autonomously communicating human being. To ground these claims, however, we shall first discuss the tools and constraints of contemporary android science in order to identify where echoborg methodology can contribute.

The Challenge of Creating Androids that Speak Autonomously

Examples of autonomous androids include Repliee Q1 and Repliee Q2, which were developed jointly by Osaka University and the Kokoro Corporation (see Ishiguro, 2005; Ranky and Ranky, 2005). Because androids of this nature attempt to replicate humans at both an outer/physical level as well as an inner/dispositional level, they can be evaluated against what Harnad (1991) defined as the *Total Turing Test* (also referred to as the *Robotic Turing Test*; Harnad, 2000), which establishes the entire repertoire of human linguistic and sensorimotor abilities as the appropriate criteria for judging machine imitations of human intelligence. The development of an autonomous android capable of passing such a test, however, remains a distant holy grail.

One source of current constraints concerns how artificial agents in general interpret and participate in dialog. Various terminologies describe technology that interacts with humans via natural language. “Dialog system,” “conversational agent,” and “conversational AI,” for instance, are terms used to denote the linguistic subsystems of artificial agents, though no clear consensus exists with regard to how non-overlapping these and other terms are. “Conversational agent,” the term we have employed thus far, is perhaps the most convenient term for conceptualizing the echoborg because it has been adopted by a parallel project—the development of embodied conversational agents (software that interfaces through onscreen anthropomorphic avatars). Much of the literature that distinguishes the functionality of various linguistic subsystems, however, couches these technologies as dialog systems. Types of dialog systems include high-level systems of integrated artificial intelligence that employ advanced learning and reasoning algorithms enabling a user and a machine to jointly

accomplish specific tasks within a formal dialog structure (e.g., logistics and navigation planning agents), low-level systems that use basic algorithms to simply mimic, rather than understand, casual human conversation (e.g., web-based “chat bots”), and mid-level systems that strike a balance between high-level and low-level functionality (e.g., agents designed to field queries from and respond to pedestrians in transit centers; for a discussion of dialog system hierarchy, see Schumaker et al., 2007). Dialog systems can also be differentiated in terms of the level of initiative they take when interacting with users (Zue and Glass, 2000). System-initiative agents are those that control the parameters of dialog and elicit information from the user that must be compatible with certain response formats (e.g., interactive voice response telephone systems). User-initiative agents, on the other hand, are those in which the user presents queries to a passive agent (e.g., Apple’s Siri application). Mixed-initiative agents (by far the least developed variety; Mavridis, 2015) involve both the user and agent taking active roles in a joint task with the nature of dialog being qualitatively more conversational relative to other types of dialog systems.

If we treat, as Turing (1950) did, discourse capacity as a basic proxy for an interlocutor’s “mind,” then even today’s most advanced dialog system technologies render available to artificial agents such as androids minds that are at best starkly non-human (though potentially very powerful), and at worst extremely impoverished relative to that of humans. Though contemporary high-level and mid-level dialog systems are indeed impressive and their functionality continues to expand rapidly, they are not, in principle, attempts to mimic a human interlocutor capable of casual conversation. On the contrary, they are presently intended to interact with humans in specific domains and generally do not operate outside of these contexts (e.g., such a system cannot spontaneously switch from being a logistics planning agent to having a conversation about an ongoing basketball game). No human would be expected to communicate in a manner similar to these types of artificial intelligence, nor are humans necessarily constrained in terms of only being capable of communicating from within a fixed and narrow language-game. System-initiative and user-initiative agents also deviate from the norms of human–human interaction as they grant to one interlocutor total and unbreakable communicative control.

Though we can perhaps imagine high-level and mid-level dialog systems capable of engaging humans in casual conversation someday being ubiquitous throughout social robotics, at present only certain low-level and primarily text-based systems are engineered specifically for this purpose. An early but well known example of such a system is ELIZA, a chat bot with the persona of a Rogerian psychotherapist (Weizenbaum, 1966). Modern examples include A.L.I.C.E. (Artificial Linguistic Internet Chat Entity; Wallace, 2015), Cleverbot (Carpenter, 2015), Mitsuku (Worswick, 2015), and Rose (Wilcox, 2015). Many chat bots make use of the highly customizable AIML (Artificial Intelligence Markup Language) XML dialect developed by Wallace (2008) and operate by recognizing word patterns delivered by a user and matching them to response templates defined by the bot’s programmer. Increasingly sophisticated mechanisms for generating response corpora have been developed for chat bots in recent years. For instance,

some developers have turned to real-time crowdsourcing of online communication repositories, such as Twitter and Facebook, as a means of producing responses appropriate for a given user input (see Mavridis et al., 2010; Bessho et al., 2012).

Chat bots are widely available on the internet and feature regularly in events such as the annual Loebner Prize competition (Loebner, 2008), a contest held to determine which chat bot performs most successfully on a Turing Test. This test involves a human interrogator simultaneously communicating via text with two hidden interlocutors while attempting to uncover which of the two is a bot and which is a real person. To date, no chat bot has reliably passed as a human being, and we are unlikely to see this feat accomplished in the near future (Dennett, 2004; French, 2012).

Generally, human interactions with chat bots fail to arrive at what conversation analysts refer to as “anchor points”: mutually attended to topics of shared focus that establish an implicit “center of gravity” during moments of conversation following routine canonical openings (Schegloff, 1986; Friesen, 2009). As chat bots tend to be user-initiative agents, they cannot engage in the type of fluid mixed-initiative conversation that is natural to mundane human–human interaction (Mavridis, 2015). Chat bots demonstrate a poor capacity to reason about conversation, cannot consistently identify and repair misunderstandings, and generally talk at an entirely superficial level (Perlis et al., 1998; Shahri and Perlis, 2008). According to Raine (2009), many chat bots work “based on an assumption that the basic components of a communication are on a phrase-by-phrase basis and that the most immediate input will be the most relevant stimulus for the upcoming output” (p. 399), an operative model that can lead conversation to irreparably fall apart when the perspectives of parties to a conversation diverge in terms of the meaning or intention each party assigns to an utterance. Human communication is fundamentally temporal and sequential, with many past and possible future utterances feeding into the meaning of a given utterance (Linell, 2009).

Developing acoustic technology that can accurately perceive spoken discourse remains a related challenge. The error rate of speech recognition technology is dramatically compounded by, among other things, variation in a speaker’s accent, the lengthiness and spontaneity of their speech, their use of contextually specific vocabulary, the presence of multiple and overlapping speakers, speech speed, and so on (Pieraccini, 2012). Thus, speech recognition systems within artificial agents perform best not when discerning casual conversational dialog, but when discerning brief and predictable utterances. Microphone array technologies and software capable of identifying and isolating multiple speakers continue to improve (e.g., the “HARK” robot audition system; Nakadai et al., 2010; Mizumoto et al., 2011), but demonstrations of these systems have essentially involved stationary apparatuses confined to laboratory environments.

Tele-Operated Androids: Mechanical Bodies, Human Operators

Tele-operated androids were developed in part to overcome a social research bottleneck within android science born of the various limitations of conversational agents and perception technologies (Nishio et al., 2007b; Watanabe et al., 2014). They

thus constitute a methodological trade-off: rather than being both physically artificial *and* having computer-controlled behavior (a combination that currently results in poor social functioning), the tele-operated paradigm cedes behavioral control to a human and in doing so augments the speech and motor capabilities of the android.

Perhaps the most well-known tele-operated android is Geminoid HI-1, a robot modeled in the likeness of its creator, Hiroshi Ishiguro. From a remote console, the tele-operator is able to transmit their voice through the geminoid (derived from the Latin word “geminus,” meaning “double”) while software analyzing video footage of the tele-operator’s body and lip movements replicate this motor behavior in the geminoid. The tele-operator can also manually control specified behaviors such as nodding and gaze-direction. Video monitors and microphones capture the audio-visual perspective of the geminoid and transmit to the tele-operation console, allowing the tele-operator to observe the geminoid’s social environment (Nishio et al., 2007b; Becker-Asano et al., 2010).

Relative to their fully-autonomous counterparts, the enhanced conversational capacities of tele-operated androids allow researchers to study communicatively rich human–android interactions as well as offer a means of operationally separating the behavioral control unit of an agent (the tele-operator) from the body, or interface, of the agent (the android). As Nishio et al. (2007b) contend:

“The strength of connection, or what kind of information is transmitted between the body and mind, can be easily reconfigured. This is especially important when taking a top-down approach that adds/deletes elements from a person to discover the “critical elements” that comprise human characteristics” (p. 347).

These methodological assets have inspired an abundance of exploratory laboratory and field work in recent years. Abildgaard and Scharfe (2012), for instance, used Geminoid-DK to conduct university lectures and reported on how perceptions of the android differed between male and female students. Research involving android-mediated conversations between parents and children has explored to what extent children sense the personal presence of a tele-operator (Nishio et al., 2008). Straub et al. (2010) studied how tele-operators and those they communicate with jointly construct the social identity of an android. Dougherty and Scharfe (2011), meanwhile, explored whether touch influences a person’s trust in a tele-operated android.

Despite the progress and promise of tele-operated androids, this line of research faces particular constraints. The non-verbal behaviors of autonomous and semi-autonomous androids are more mechanical and less fluid relative to humans. In their neuroimaging analysis of how people perceive geminoid movement, Saygin et al. (2012) show how incongruity between appearance (human-like) and motion (non-human-like) implicitly violates people’s expectations. Developing tools for matching an android’s bodily movements to those of its tele-operator is a major research priority (Nishio et al., 2007b), and improving techniques for achieving facial synchrony is particularly necessary given the intricate facial musculature of humans and the role of facial

expression in conveying emotion and facilitating social interaction (Ekman, 1992; Bänziger et al., 2009; for a discussion of robot emotion conveyance, see Nitsch and Popp, 2014). Current anthropomorphic androids are relatively limited in terms of their capacity for human-like facial expressivity (Becker-Asano, 2011). For instance, Geminoid F’s face can successfully express the emotions *sad*, *happy*, and *neutral*, but the model struggles to convincingly convey *angry*, *surprised*, and *fearful* (Becker-Asano and Ishiguro, 2011). Also, the inexactness of an android’s lip movements in relation to the words spoken by its tele-operator has been discussed as possibly degrading the quality of social interactions (Abildgaard and Scharfe, 2012). Moreover, geminoids and other android models cannot walk on account of their having large air compressors facilitating numerous pneumatic actuators (Ishiguro and Nishio, 2007).

The imperfect appearance of tele-operated androids remains a barrier to replicating the social psychological conditions of face-to-face human–human interaction. Despite painstaking efforts to create realistic silicone android models (Ishiguro and Nishio, 2007), people are minutely attuned to subtle deviations from true humanness (e.g., eyes that lack glossy wetness). In a field study conducted to test whether people would notice an inactive or relatively passive geminoid in a social space, a majority of people reported having seen a robot in their surroundings (von der Pütten et al., 2011), a finding which suggests that most people are not easily fooled into believing an android is an actual person even in social situations where they do not engage the android directly. Moreover, though geminoids and other highly anthropomorphic androids are seen as the most human-like and least unfamiliar of robot types, people nonetheless perceive these androids as more threatening than less anthropomorphic models (Rosenthal-von der Pütten and Krämer, 2014).

There is also an important practical constraint characterizing the tele-operated and autonomous android paradigms. As Ziemke and Lindblom (2006) point out, it is quite time consuming and costly to produce android experimental apparatuses. This raises issues as to the scalability of the current android science research model and the extent to which experiments making use of a particular device in one laboratory can be replicated elsewhere.

The Echoborg

An echoborg is composed of a human whose words (and potentially motor actions) are entirely or partially determined by a computer program. Echoborgs constitute a methodological trade-off inverse to that of the tele-operated paradigm discussed above, as they allow the possibility of studying social interactions with artificial agents that have truly human interfaces. The unique affordances of echoborgs can complement those of tele-operated and fully-autonomous androids and contribute to our understanding of the social psychological dynamics of human–agent interaction.

Speech Shadowing and the Cyranoid Method

The echoborg concept stems from work conducted by Corti and Gillespie (2015), whose application of Milgram’s (2010) “cyranoid method” of social interaction demonstrates a means of creating

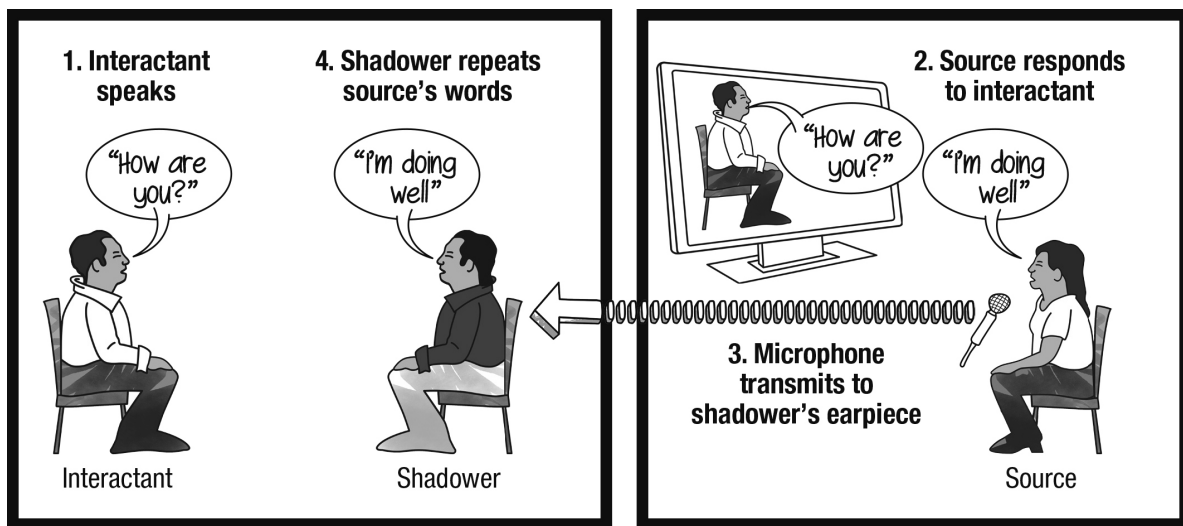


FIGURE 1 | Illustration of a basic cyranoid interaction. The shadower voices words provided by the source while engaging with the interactant in person.

hybrid human entities via an audio-vocal technique known as “speech shadowing.” Speech shadowing involves a person (the shadower) voicing the words of an external source simultaneously as those words are heard (Schwitzgebel and Taylor, 1980). This can be facilitated by way of an inner-ear monitor worn by the shadower that receives audio from the source. Research has shown that native-language shadowers can repeat the words of a source at latencies as low as a few hundred milliseconds (Marslen-Wilson, 1973, 1985; Bailly, 2003) and can perform the technique while simultaneously attending to other tasks (Spence and Read, 2003). Shadowers tend to reflexively imitate certain gestural elements of their source (e.g., stress, accent, and so on)—a phenomenon known as “phonetic convergence” (Goldinger, 1998; Shockley et al., 2004; Pardo et al., 2013).

One finds the use of speech shadowing as a research tool primarily in psycholinguistics and the study of second-language acquisition. In the late 1970s, however, Milgram—famous for his controversial studies on obedience to authority (Milgram, 1974)—began using speech shadowing to investigate social scenarios involving people communicating through shadowers. He saw the technique as a means of pairing sources and shadowers whose identities differed in terms of race, age, gender, and so on, thus allowing sources to directly experience an interaction in which their outer appearance was markedly transformed (see **Figure 1**). From the point of view of the shadower, the method enabled exploration into the sensation of contributing to an unscripted conversation not one’s self-authored thoughts, but entirely those of a remote source. Inspired by the play *Cyrano de Bergerac*, the story of a poet (Cyrano) who assists a handsome but inarticulate nobleman (Christian) in wooing a woman by telling him what to say to her, Milgram referred to these source-shadower pairs as “cyranoids.”

As speech shadowing proved to be a relatively simple task that research participants were quick to grasp, Milgram quickly began exploring a variety of cyranic interactions. For instance, in several pilot studies he examined whether “interactants” (Milgram’s

term for those who encountered a cyranoid) would notice if the source was changed mid-conversation (Milgram, 1977). Milgram (2010) also sourced for 11- and 12-year-old children during interviews with teachers naïve to the manipulation. Following these interactions, all of the teachers seemed to take the interviews at face value—they neither picked up on the true nature of the interactions nor sensed that the child they interviewed had behaved non-autonomously. The teachers had succumbed to the “cyranic illusion,” that is, the tendency to perceive interlocutors as autonomous communicators and thus fail to notice an interlocutor that is a cyranoid.

Corti and Gillespie (2015) argue that one of the cyranoid method’s primary strengths is that it allows the researcher to manipulate one component of the cyranoid, either the shadower or the source, while keeping the other component fixed. Thus, one can study how the same source is perceived when interacting through a variety of shadower-types. Conversely, a researcher can opt to keep the shadower constant and vary the identity of the source across experimental conditions. This capacity mirrors the functionality of tele-operated androids as well as similar methods for studying transformed social interactions (e.g., using 3D immersive virtual environment technology to alter people’s identities; see Blascovich et al., 2002; Bailenson et al., 2005; Yee and Bailenson, 2007). A unique benefit of the cyranoid method is that it allows for in person, face-to-face interactions between an interactant and a hybrid. When interacting with a cyranoid, one is not interacting with an onscreen person, or a human-like machine, or a virtual representation of a human, but with an actual human body.

While Corti and Gillespie’s (2015) recent work was conducted in the laboratory, it follows recent field explorations of cyranoids in experiential art installations (Mitchell, 2009) and as classroom learning tools (Raudaskoski and Mitchell, 2013). Taken together, these studies outline a number of basic protocols for constructing cyranic interactions and discuss the devices necessary for creating a basic cyranoid apparatus, which involves both a means of

discreetly transmitting audio from the source to the shadower as well as a means for the source to hear (and, if possible, see) the interaction between the shadower and the interactant. The amalgam of devices one uses toward these requirements depends upon the type of interaction the researcher wishes to create. For instance, if a researcher wants to keep hidden from interactants the fact that a cyranoid is present in an interaction, then the cyranoid apparatus should be discreet and non-visible/audible to interactants. If the researcher wants the shadower to be mobile, then the devices that compose the cyranoid apparatus must transmit wirelessly. Minimizing the audio latency in the communication loop is crucial to any cyranoid apparatus; interactant→source and source→shadower audio transfer must be accomplished in a realistic amount of time.

A cyranic interaction involving a covert cyranoid is typically accomplished using an apparatus similar to the following. A wireless “bug” microphone placed near where the shadower and interactant engage each other transmits to a radio receiver listened to by the source in an adjacent soundproof room. The source speaks into a microphone connected to a short-range radio transmitter which relays to a receiver worn in the pocket of the shadower. Connected to the shadower’s receiver is a neck-loop induction coil worn underneath their clothing. The shadower wears a wireless, flesh-colored inner-ear monitor that sits in their ear canal and receives the signal emanating from the induction coil, allowing the shadower to hear and thus voice the source’s speech. This amalgam of devices is neither visible nor audible to interactants.

Ceding Verbal Agency to a Machine

Echoborg methodology takes the original cyranoid model and replaces the human source with an artificial conversational agent. The words produced by the conversational agent are thus voiced and embodied by a human shadower. Echoborgs have at least four main research affordances:

Interchangeability of Shadowers and Conversational Agents

Both the shadower and the conversational agent that comprise an echoborg are easily customizable and interchangeable. The researcher need only train a confederate with the desired physical attributes to speech shadow sufficiently and then couple them with a conversational agent. This gives the researcher the freedom to construct many echoborgs, each differentiated from one another in terms their particular conversational agent, gender, age, and so on. Thus, one can observe how the same conversational agent is perceived depending on the identity of the shadower by holding the conversational agent constant across experimental conditions and varying the shadower (e.g., female shadower vs. male shadower). Alternatively, the researcher can hold the shadower constant and vary the conversational agent (e.g., ELIZA vs. A.L.I.C.E.).

Visual Realism

Echoborgs offer a means of studying interactions under conditions where the interactant’s cognitive sense of the interaction is

undistorted by any esthetic, acoustic, non-verbal, or motor non-humanness of the physical agent they encounter (e.g., lips that do not exactly align with the words they utter or eyes that do not perfectly make contact with the interactant’s). Speech shadowing is not a cognitively demanding task; it is rather simple for a well-rehearsed speech shadower to attend to other behaviors while replicating the speech of their source, including matching their body language to the words they find themselves repeating (e.g., shaking their head from side-to-side upon articulating the word “no”).

Mobility

Echoborgs can take advantage of the shadower’s physical mobility and need not be confined to stationary interactions—they can walk or otherwise move about while communicating with interactants. Human communication did not evolve for having conversations *per se*; it evolved for coordinating joint activity (Tomasello, 2008). Research on everyday language use shows that communication is a means of doing (Clark, 1996). Accordingly, mobile echoborgs open up the possibility of testing conversational agents in the context of performing a joint non-stationary activity.

Covert Capacity

Taking advantage of the cyranic illusion, echoborgs can interact with people covertly (i.e., under conditions wherein interactants assume they are encountering an autonomously communicating person). This affordance can be juxtaposed with the fact that at present, those who interact with tele-operated or autonomous androids are under no illusion that they are interacting with a fully-autonomous human being. The covert capacity of echoborgs thus presents a new means of researching interactions with conversational agents. It is one thing to evaluate interactions with conversational agents in contexts where people are cognitively aware, or at least primed to believe, that they are speaking to something artificial, but it is entirely different to study these systems under conditions where the interface one encounters (an actual human body) creates the visceral impression that one is dealing with an autonomous person.

Overview of Studies

We conducted three experiments in which participants interacted with echoborgs. These studies explored the ways in which echoborgs, as human interfaces, mediate the experience of conversing with a chat bot in various contexts, as well as the extent to which echoborgs improve a chat bot’s ability to pass as human (i.e., be taken for a human rather than a robot). Each study was approved by an ethics review board at the London School of Economics and Political Science and conducted at the university’s Behavioral Research Laboratory. Adult participants were recruited online via the university’s research participant recruitment portal and included students from the university, university employees, and people unaffiliated with the university. Participants gave informed consent prior to participation and were debriefed extensively.

Study 1: Turing Testing with Echoborgs

Aims

In outlining the logic of his imitation game, Turing (1950) argued that “there was little point in trying to make a “thinking machine” more human by dressing it up in such artificial flesh” (p. 434) and made a clear distinction between what he thought of as the physical (likeness) and intellectual (functional) capacities of humans. However, this distinction has been criticized (Harnad, 2000); perceiving the salient bodily characteristics of other entities is fundamental to how humans infer the subjective states (or lack thereof) of said entities, be they real or unreal in reality (Graziano, 2013). To explore this tension, our first study investigated a Turing Test scenario wherein participants were asked to determine which of two shadowed interlocutors was truly human and which was a chat bot. Furthermore, we sought to determine whether a chat bot voiced by a human shadower would be perceived as more human-like than the same bot communicating via text.

Shadowers and Subjects

Two female graduate students (both aged 23) were trained as speech shadowers. Eighty-two participants (42 female, mean age = 28.93, SD = 12.05) were randomly assigned into pairs within one of two experimental conditions: Text Interface ($n = 21$) and Echoborg ($n = 20$). One participant within each pair was randomly selected to function as the Turing Test interrogator while the second participant was designated as the human interlocutor. In all pairs, participants were both unfamiliar with one another and unaware of the other's role in the study.

Procedure

From the interaction room, the researcher instructed the interrogator that the study involved using a text-based instant messaging client (Pidgin) to simultaneously communicate with two anonymous interlocutors, one of whom was a chat bot (Cleverbot). The interrogator's computer showed two separate text-input windows, one that delivered to “Interlocutor A,” and another that delivered to “Interlocutor B.” The interrogator was told that following 10-min of conversation they would be asked which of these two interlocutors they believed was the real human. Meanwhile, in a separate room, a research assistant instructed the human interlocutor that the study involved holding a 10-min conversation with a stranger and that their task was to simply respond to messages that appeared on a computer screen. The human interlocutor was thus blind to the fact that they were engaged in a Turing Test. Both the interrogator and the human interlocutor were informed that they were free to discuss any topic during the interaction so long as nothing was vulgar.

Text Interface Condition

Once instruction was complete, the researcher relocated to a third room (the source room) where they monitored the interaction using a computer. Messages that the interrogator typed to Interlocutor A were routed to the researcher, who input the received text into Cleverbot and routed Cleverbot's response back through the instant messaging client to the interrogator. Messages the interrogator sent to Interlocutor B, meanwhile, were routed to

the human interlocutor's computer, and the human interlocutor directly responded in text via the instant messaging client.

Echoborg Condition

The interrogator was further instructed that though they would type messages to Interlocutor A and Interlocutor B via the instant messaging client, the responses of these two interlocutors would be spoken aloud by two speech shadowers. The two speech shadowers, with shadowing equipment, entered the room, sat side-by-side facing the interrogator at a distance of roughly six feet, and it was made known to the interrogator which shadower would reproduce the words of Interlocutor A and which would reproduce the words of Interlocutor B (shadowers alternated between trials in terms of the interlocutor they were paired to). The interrogator was informed that the shadowers would speak solely words they received from their respective sources and that at no point during the interaction would the shadowers speak self-authored thoughts. Furthermore, the interrogator was informed that both interlocutors would only respond to typed messages and that nothing the interrogator spoke aloud would be responded to.

Following these instructions, the researcher relocated to the source room. As in the Text Interface condition, messages that the interrogator sent to Interlocutor A were routed to the researcher's computer where they were input by the researcher into Cleverbot. Instead of routing Cleverbot's responses back to the interrogator through the instant messaging client, however, the researcher spoke Cleverbot's responses into a microphone which relayed to the speech shadower paired to Interlocutor A, thus allowing them to hear and repeat Cleverbot's words to the interrogator. Similarly, the human interlocutor's typed responses were routed to the researcher's computer (rather than directly to the interrogator), allowing the researcher to speak these messages into a separate microphone which relayed to the shadower paired to Interlocutor B (see Figure 2).

Stock Responses

Cleverbot's response formats are not programmed; Cleverbot references past conversations it has held with people over the internet when generating a reply to a given user input (Carpenter, 2015). Unlike other bots, therefore, Cleverbot has no consistent identity. Its strength lies in its ability to learn unique ways of responding. We decided, however, that in order to establish consistency between experimental trials, three stock responses would be supplied in both conditions to the interrogator in lieu of a response generated by Cleverbot. Each time the interrogator inquired as to the name of Interlocutor A, the standard response “My name is Kim” was supplied to the interrogator. In response to questions as to what Interlocutor A's occupation was, the response “I'm a psychology student here” was supplied. Finally, in response to questions concerning where Interlocutor A was from, the response “I'm from London” was given.

Measures

Following the interaction, the interrogator indicated on a questionnaire which of the two interlocutors (A or B) they believed was the real human and indicated along a 10-point scale how confident they were that they had made the correct identification (1: not at

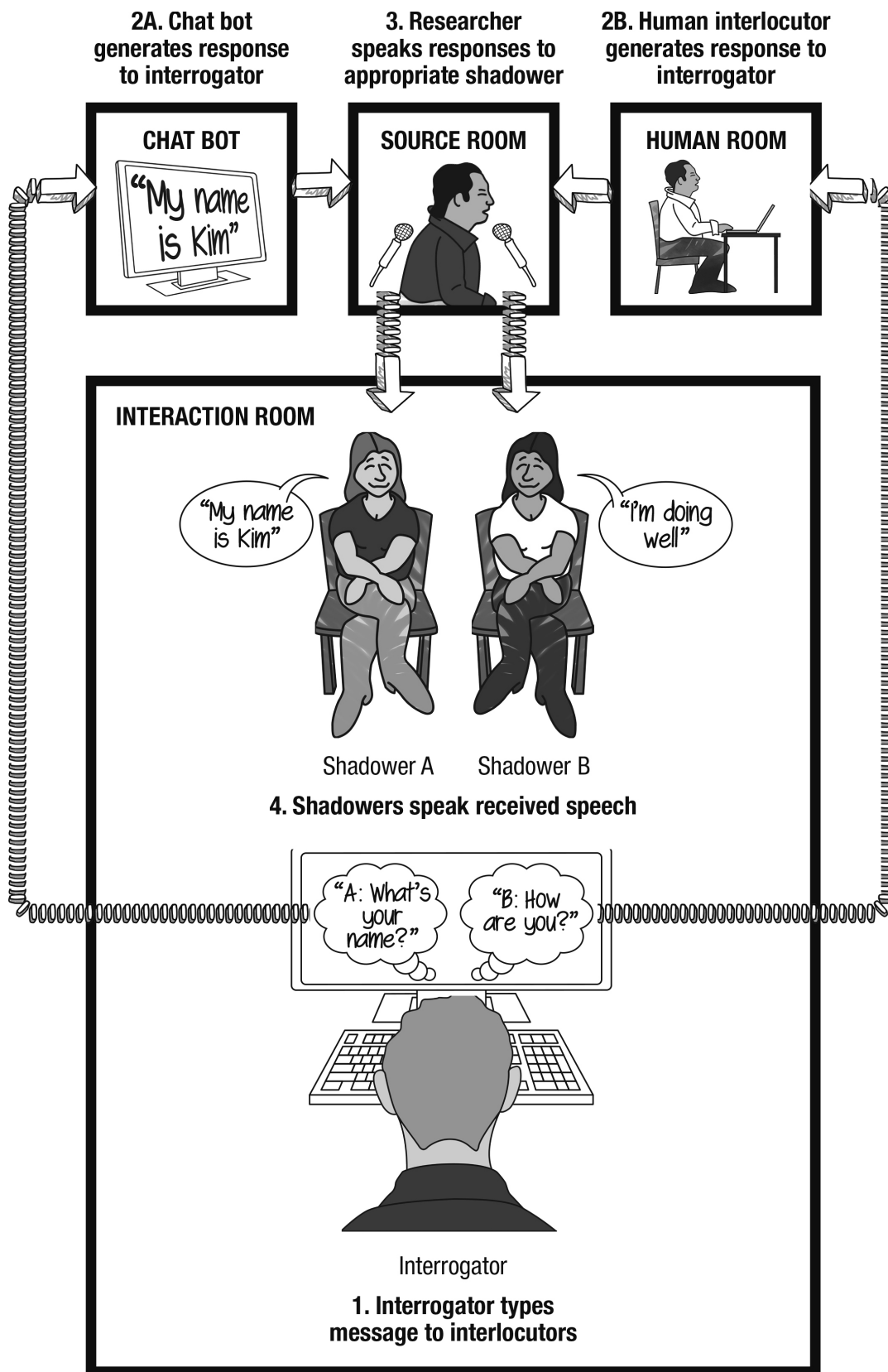


FIGURE 2 | Illustration of a Turing Test scenario involving speech shadowing. This figure visually depicts the Echoborg condition in Study 1.

all confident; 10: highly confident). Interrogators also rated each interlocutor along a 10-point scale in terms of how human-like they seemed (1: seemed very mechanical and computer-like; 10: seemed very human-like).

Results

In the Text Interface condition, 21 out of 21 interrogators correctly identified Interlocutor B as being the real human, compared to 18 out of 20 interrogators in the Echoborg condition, a non-significant difference, $z = 1.49$, $p = 0.14$ (two-tailed). There was no significant difference between conditions in terms of how confident interrogators were with regard to their answers, with interrogators in the Text Interface condition reporting an average confidence of 7.67 (SD = 2.61) and interrogators in the Echoborg condition reporting an average confidence of 7.55 (SD = 1.70), $t(39) = 1.68$, SE = 0.69, $p = 0.87$.

Human-likeness ratings were compared using a repeated measures analysis of variance, with *Condition* (Text Interface vs. Echoborg) treated as a between-subjects factor and *Interlocutor* (Interlocutor A vs. Interlocutor B) treated as a within-subjects factor. There was a significant main effect of *Interlocutor* showing that Interlocutor B was perceived as significantly more human-like than Interlocutor A in both conditions, $F(1,39) = 130.87$, $r = 0.88$, $p < 0.001$. There was also a significant interaction between *Condition* and *Interlocutor*, $F(1,39) = 7.23$, $r = 0.40$, $p < 0.05$. Independent samples means tests showed that the average human-likeness rating of Interlocutor A in the Text Interface condition ($M = 2.14$, SD = 1.15) was significantly less than the average rating in the Echoborg condition ($M = 4.05$, SD = 2.42), $t(39) = -3.25$, SE = 0.59, $p < 0.01$. Meanwhile, the average human-likeness rating of Interlocutor B in the Text Interface condition ($M = 8.76$, SD = 1.51) was not significantly different from the average rating in the Echoborg condition ($M = 8.15$, SD = 1.46), $t(39) = 1.32$, SE = 0.46, $p = 0.20$.

Discussion

The interface (human body vs. text) engaged by the interrogator made no statistically significant difference in terms of their ability to discern which interlocutor was the real human. The chat bot, however, was perceived by interrogators as significantly more human-like when being shadowed by a person compared to when simply communicating via text. This contrasted with the fact that how human-like human interlocutors seemed to participants did not depend on whether their words were voiced by a speech shadower. This suggests that as the quality of an interlocutor's discourse capacity improves (i.e., becomes more human) in Turing Test scenarios, the role the interface plays in eliciting judgments about human-likeness declines.

Study 2: A Human Imitating a Chat Bot?

Aims

Study 2 investigated whether attributing human agency to an interlocutor is increasingly determined by the nature of the interface as the words spoken by the interlocutor provide less definitive evidence. We designed a scenario wherein participants encountered an interlocutor and had to determine whether the

interlocutor was (a) a person communicating words that had been generated by a chat bot, or (b) a person merely imitating a chat bot, but nonetheless speaking self-authored words (the former option always being true). The point here was to see whether or not the interface participants encountered (human body vs. text) influenced whether they thought their interlocutor was producing self-authored words or, alternatively, those of a machine. The framing of the scenario leads participants to expect that the communication offered by their interlocutor will be abnormal, thus the conversational limitations of chat bots are not a liability as they are in standard Turing Test scenarios. By design, participants must form an attribution regarding the communicative agency of their interlocutor under conditions of ambiguity.

Research on perceptual salience suggests that people will deem causal what is salient to them in the absence of equally salient alternative explanations (Jones and Nisbett, 1972; Taylor and Fiske, 1975). Dual process information evaluation theories propose that when a person evaluates the communication and behavior of others, stimulus ambiguity increases reliance on heuristic cues (e.g., appearance) at the expense of more thoughtful situational evaluation (Sager and Schofield, 1980; Devine, 1989; Chen and Chaiken, 1999). We extrapolated from this research that when faced with an ambiguous situation in which one's interlocutor was either truly speaking words generated by a chat bot or merely pretending to be one, the interface (and thereby the heuristic cues) salient to the participant would determine how they attributed authorship to the words they encountered. We therefore hypothesized that those who encountered an echoborg would be more likely to see their interlocutor as producing self-authored words (imitating a chat bot) compared to those who encountered an interlocutor through a text interface.

Shadowers and Subjects

A female graduate student (aged 30) was trained to perform as a speech shadower. Fifty-eight adult participants (35 female; mean age = 25.19, SD = 9.08) were randomly assigned to one of two conditions: Echoborg ($n = 28$) and Text Interface ($n = 30$).

Procedure

As with Study 1, Cleverbot, as well as the three stock responses described above, were used in all trials.

The participant was led to an interaction room and instructed by the researcher that the study involved holding a 10-min conversation with an interlocutor who was either (a) communicating solely words that had been generated by a chat bot program (at no point speaking anything self-authored), or (b) simply imitating a chat bot program, but producing self-authored words nonetheless. The researcher ensured that the distinction between these scenarios was clear to the participant and gave the further instruction that the participant would be asked following the interaction which of the two scenarios they believed to have been the case. The participant was informed that they were free to discuss anything they liked with their interlocutor so long they refrained from vulgarity.

Unlike Study 1, which had participants send messages to their interlocutors via an instant messaging client, Study 2 featured participants speaking aloud to their interlocutor as they would during

any other face-to-face encounter, thereby increasing the mundane realism of the scenario. The apparatus for this type of interaction, however, required a means of inputting the participant's spoken words into the chat bot in the form of text. As we deemed speech-to-text software to be insufficient for our purposes (being too slow and inaccurate), we settled on a procedure wherein the researcher (from an adjacent room) acted as the chat bot's ears and speed typed the participant's words into the chat bot as they were being spoken, paraphrasing when necessary for particularly verbose turns. This can be conceptualized as a minimal technological dependency format of the echoborg method (as opposed to a full technological dependency format which would place acoustic perception solely on technology). Although a minimal technological dependency format adds an additional human element to the communication loop, it ensures that accurate representations of interactants' words are processed by the conversational agent.

Text Interface Condition

The participant was seated in front of a computer screen which displayed a blank instant messaging client chat window. The participant was instructed that they were to address their interlocutor by speaking aloud and that their interlocutor would respond via text readable in the chat window. Once instruction was complete, the researcher left the interaction room and returned to the adjacent source room. From the source room, the researcher overheard words spoken by the participant via a covert wireless microphone and speed typed them into Cleverbot's text-input window. Cleverbot's responses were then sent through the instant messaging client to the participant's screen in the interaction room (see **Figure 3**).

Echoborg Condition

The participant was instructed that as soon as the researcher left the interaction room their interlocutor would enter and sit facing the participant (at a distance of roughly six feet). The participant was not made aware of the fact that their interlocutor would be wearing an earpiece and receiving messages via radio, and the cyranoid apparatus was not visible to the participant. The researcher then left the interaction room and returned to the adjacent source room while the shadower entered the interaction room and sat across from the participant. The researcher listened to the words of the participant via a covert wireless microphone, speed typed them into Cleverbot's text-input window, and subsequently spoke Cleverbot's responses into a microphone which relayed to the shadower's inner-ear monitor.

Measures

Following the interaction, the participant indicated on a questionnaire whether they thought their interlocutor had truly been producing words generated by a chat bot program or whether their interlocutor was simply imitating a chat bot.

Results

Of the 30 participants in the Text Interface condition, 11 stated following the interaction that they believed their interlocutor was simply imitating a chat bot compared to 22 of 28 participants in the Echoborg condition. A binary logistic regression model

showed these proportions to be significantly different from one another, $OR = 6.33$, $b = 1.85$, $SE = 0.60$, $p < 0.01$ (indicating that the odds of a participant in the Echoborg condition deciding their interlocutor was imitating a chat bot were 6.33 times greater than the odds of a participant in the Text Interface condition coming to the same conclusion).

To gain a sense of the audio latency dynamics of echoborg interactions involving minimal technological dependency, we randomly selected four trials from the Echoborg condition and measured the time between the conclusion of each interactant-utterance and the commencement of the echoborg's subsequent response. The average latency was 5.15 s ($SD = 3.04$ s).

Discussion

Our results indicate that under conditions of ambiguity wherein the source of an interlocutor's verbal agency is unclear, the interface substantially affects whether one attributes human agency to the words one's interlocutor produces. Participants who communicated with a chat bot via a text interface were significantly more likely to see their interlocutor as actually producing words generated by a chat bot compared to those who encountered the same chat bot but through a human shadower. The results from this study corroborate the notion that the cyranic illusion is robust in circumstances involving extreme source-shadower incongruity: people are biased toward perceiving an echoborg as an autonomous person.

Our findings suggest that it is relatively easy to get a chat bot to be perceived as an autonomous human if one is free to manipulate the contextual frame (i.e., the social psychological context of the interaction). An ostensibly simple suggestion from the experimenter (i.e., that an interlocutor might be a human imitating a chat bot) can shift the entire contextual frame, fundamentally altering attributions of agency. Indeed, whenever it is claimed a certain bot has "passed the Turing Test" or some variant of Turing's game, it usually has less to do with advances in conversational agent technology and more to do with shifting the contextual frame (e.g., when the chat bot Eugene Goostman—a bot that poses as a 13-year-old Ukrainian boy with limited English skills and general knowledge—was declared as having successfully fooled 33% of interrogators in a Turing Test in 2014; You, 2015). This, however, raises a fundamental question: within what contextual frame *should* participants encounter chat bots when we evaluate them? Arguably, the most important frame is the most common, namely, the everyday assumption that our interlocutors are human, just like us.

Study 3: Can Covert Echoborgs Pass as Human in the Everyday Contextual Frame?

Aims

Study 3 examined people's impressions following their conversing with an agent who, unbeknownst to them, produced solely the words of a chat bot. We aimed to gauge whether or not being shadowed by a human improved a chat bot's ability to pass as an actual person within the everyday contextual frame (i.e., under the conditions of a generic social encounter wherein it is assumed an interlocutor is an ordinary human). The concept of "passing"

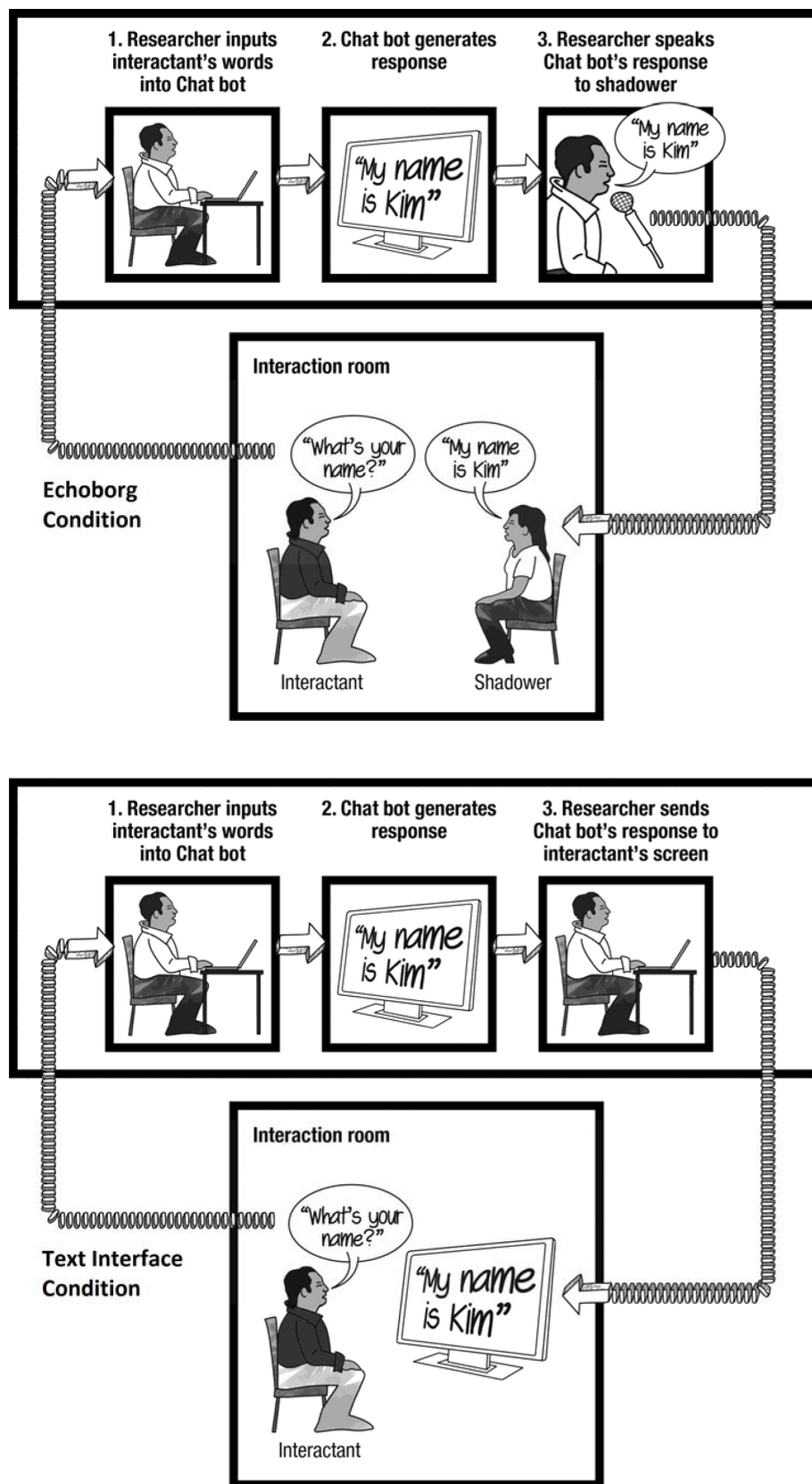


FIGURE 3 | Illustration of interaction scenarios in Study 2 and Study 3.

within such a frame comes from the sociological and social psychological traditions that explore the mechanisms through which people manage identities in order to be accepted as a member of a particular group (Goffman, 1963; Renfrow, 2004; Khanna and Johnson, 2010). For example, the anthropomorphic androids in Dick's (1968) novel *Do Androids Dream of Electric Sheep?* were able to pass as human so long as they concealed their true nature, took part in mundane human activities, and avoided the scrutiny of bounty hunters. The speech shadower in an echoborg is essentially a human mask placed over the peripherals one normally associates with computer systems. From a static third-person point of view, therefore, echoborgs appear to be autonomous human beings and nothing more, raising the question as to whether or not despite their communicative deficiencies people still sense that echoborgs are ordinary people. We predicted that research participants would not leave an interaction with a covert echoborg with the impression of having communicated with something non-human, whereas interacting with a covert chat bot through a text interface would leave participants with a strong impression of having encountered machine intelligence of some sort.

This study also investigated perceptual phenomena associated with the uncanny valley, namely how human-like, eerie, and familiar a covert echoborg interlocutor would seem to those with whom they communicated, and whether or not people would be comfortable in the presence of a covert echoborg. Mori's (1970) original hypothesis suggested that "subtle deviations from human appearance and behavior create an unnerving effect" (MacDorman and Ishiguro, 2006b, p. 299), and our goal was to gauge people's reaction to an interlocutor that was human in all respects but for the fact that a conversational agent determined the words they spoke.

Shadows and Subjects

A female graduate student (aged 23) was trained to perform as a speech shadower. Forty-one adult participants (26 female; mean age = 24.12, SD = 7.59) were randomly assigned to one of two conditions: Echoborg ($n = 20$) and Text Interface ($n = 21$).

Procedure

In addition to Cleverbot, two other chat bots were used in this study: Mitsuku (winner of the 2013 Loebner Prize) and Rose (winner of the 2014 Loebner Prize). In the Echoborg condition, Cleverbot and Rose were each assigned to speak with seven participants while Mitsuku spoke with six participants. In the Text Interface condition, Cleverbot, Rose, and Mitsuku each spoke with seven participants. During Cleverbot trials, the stock responses used in the prior two studies were employed.

The participant was instructed that the study concerned how strangers conversed when speaking for the first time, that it involved simply holding a 10-min conversation with another research participant, and that they were free to decide on topics for discussion so long as vulgarity was avoided. The researcher made no mention of chat bots or of anything related to artificial intelligence. Furthermore, the participant was given no indication that their interlocutor would behave non-autonomously or abnormally. The aim was to invoke the everyday contextual frame, in so far as that can be done within an experimental setting.

This study used the same minimal technological dependency apparatus and procedure as in Study 2. In the Text Interface condition the participant spoke aloud to their interlocutor while their interlocutor's responses were shown in text on a computer screen. In the Echoborg condition the participant encountered a human shadower face-to-face.

Measures and Post-Interaction Interview

Following the interaction the participant completed a brief questionnaire containing items asking them to indicate on a 10-point scale how human-like (1: very mechanical and computer-like; 10: very human-like), eerie (1: not at all eerie; 10: very eerie), and familiar (1: not at all familiar; 10: very familiar) their interlocutor seemed, as well as how comfortable they felt during the interaction (1: not at all comfortable; 10: very comfortable). Participants were also asked to briefly describe in writing the person they spoke with and what they thought their study was about.

When the questionnaire was completed, the researcher interviewed the participant to gain a sense of their impressions of the interaction and their interlocutor. The participant was asked to describe salient aspects of their interlocutor's personality. In order to ascertain whether the participant had picked up on the fact that they had communicated with a computer program, the researcher asked the participant whether they had suspicions regarding the nature of their interlocutor or about the study generally. Finally, the researcher revealed to the participant the full nature of the interaction and disclosed the purpose of the study.

Results

In the Text Interface condition, 14 of 21 participants (67%) mentioned during their post-interaction interview (prior to the researcher making any allusion to chat bots or anything computer-related) that they felt they had spoken to a computer program or robot. Two participants stated during debriefing that they suspected their interlocutor was a real person acting or using a script. Furthermore, seven participants (33%) explicitly stated in writing on their questionnaires that they believed the purpose of the study was to assess human-computer/human-robot interaction. Of the 14 participants who did not indicate that they thought the purpose of the study involved human-computer interaction, six said that they thought the study concerned how strangers communicated with one another (the stated purpose of the study supplied by the researcher prior to the interaction). Two participants believed the study concerned how people handle abnormal/unexpected situations. Six participants provided unique responses that did not fit into these categories.

Only 3 of 20 participants (15%) in the Echoborg condition stated during their post-interaction interview that they felt as though they had spoken to a computer or robot. Fifteen participants made it clear to the researcher during their interview that they suspected their interlocutor had been acting or giving scripted responses that did not align with their actual persona. Only two participants (10%) indicated in writing on their questionnaires that they believed the purpose of the study was to assess human-computer/human-robot interaction. Of the 18 participants who did not indicate that they thought the study's purpose was to investigate human-computer interaction, only one stated

that they thought the purpose of the study was to investigate communication between strangers. Seven participants believed the purpose of the study related to how people deal with abnormal/unexpected situations (e.g., “how people react when thrown out of their comfort zone” and “how people react to people who do not comply with social norms”). Four participants believed the study’s purpose was to see how people communicated those who were shy/introverted. Three participants stated that they thought the study’s purpose involved how people communicate with those who have a disability such as autism or speech impairment. Four participants provided other unique responses.

We performed a multivariate analysis of variance to see whether *Interface* (Echoborg vs. Text Interface) and *Chat Bot* (Cleverbot vs. Mitsuku vs. Rose) produced effects on participants’ judgments concerning the four questionnaire items that pertained to how familiar, eerie, and human-like their interlocutor seemed as well as how comfortable they felt during the interaction. An initial omnibus test showed a significant effect of *Interface*, $\Lambda = 0.73$, $F(4,34) = 3.18$, $p < 0.05$, $\eta^2 = 0.27$, and a non-significant effect of *Chat Bot*, $\Lambda = 0.74$, $F(8,68) = 1.41$, $p = 0.21$, $\eta^2 = 0.14$. Univariate tests showed a significant effect of *Interface* on how comfortable participants felt during the interaction, $F(1,37) = 10.64$, $p < 0.01$, $\eta^2 = 0.22$, with participants in the Text Interface condition reporting higher levels of comfort ($M = 5.52$, $SD = 2.42$) compared to those in the Echoborg condition ($M = 3.44$, $SD = 2.04$). However, these univariate tests showed non-significant effects of *Interface* with respect to how familiar, $F(1,37) = 1.52$, $p = 0.23$, $\eta^2 = 0.04$, eerie, $F(1,37) = 0.08$, $p = 0.77$, $\eta^2 < 0.01$, and human-like, $F(1,37) = 0.24$, $p = 0.63$, $\eta^2 = 0.01$, interlocutors seemed. In the Text Interface condition, mean scores for familiarity, eeriness, and human-likeness were 3.81 ($SD = 1.89$), 6.19 ($SD = 2.14$), and 2.95 ($SD = 1.63$), respectively, compared to scores of 3.00 ($SD = 2.22$), 6.00 ($SD = 2.00$), and 2.70 ($SD = 1.78$), respectively, within the Echoborg condition.

Two Echoborg condition trials for each chat bot were selected at random and the audio latency was assessed. The average latencies for Cleverbot, Mitsuku, and Rose were 4.43 s ($SD = 2.92$ s), 5.95 s ($SD = 3.98$ s), and 3.96 s ($SD = 3.94$ s), respectively. As each trial made use of the same minimal technological dependency format of interaction, the differences between these latencies can be accounted for by the fact that the chat bots we used differ in terms of the speed at which they generate and return responses.

Discussion

In line with our hypothesis, a majority of participants in the Text Interface condition sensed they were communicating with a chat bot despite being led to believe they would be talking to another research participant while only a small minority of participants in the Echoborg condition came to the same conclusion. These results suggest that a chat bot stands a far greater chance of passing as a human in an everyday contextual frame when being shadowed by a human than when communicating via a text interface. The caveat to these findings, however, is that interactants do not tend to see a person shadowing for a chat bot as genuine. Rather, interactants see such people as deliberately behaving outside of their normal persona. This finding corroborates the general phenomenon observed in Study 2, that people are inclined to perceive

an echoborg as somebody acting but nonetheless speaking self-authored words. We should note, however, that participants’ awareness of being in a laboratory study may have contributed to their suspecting that the persona they encountered was not genuine. Future research may include observational field studies wherein interactants encounter a covert echoborg in real-world social contexts (e.g., a generic social gathering). It is plausible that in such scenarios interactants would be less inclined to form the belief that an echoborg was someone deliberately acting outside of their normal persona.

Although our experiment only considered two types of interfaces as opposed to a continuum of interfaces ranging from the very-human to the very-mechanical, our results contribute a novel finding to the discussion surrounding uncanny valley phenomena. We found evidence that people feel significantly less comfortable speaking to a chat bot through a human speech shadower than they do speaking to the same chat bot through a text interface. General discomfort seemed to derive from the social awkwardness that arose due to the chat bot’s violations of conversational norms. The effect of these violations appears to have been magnified in the Echoborg condition. It is likely that participants in the Echoborg condition held higher expectations about the level of understanding and rapport that would be reached and sustained during the interactions on account of their speaking face-to-face with another human being, for the physical body of the other is laden with social cues that evoke such expectations (Kiesler, 2005). Komatsu and Yamada’s (2011) “adaptation gap” hypothesis suggests that when expectations are not met during interactions with agents (e.g., when the implied social capacity of an agent exceeds that actually experienced by a user), people’s subjective impressions are affected. Accordingly, participants in the Echoborg condition may have felt more uncomfortable compared to their counterparts in the Text Interface condition partly due to their having higher pre-interaction expectations about the quality of interlocution they would experience. What requires further study is the investigation of conditions within which participants are told prior to interacting with either an echoborg or a text interface that their interlocutor will be producing the words of a chat bot. Adding two such conditions to Study 3/s design would allow one to observe whether the body of the other produces effect on feelings of comfort independent of pre-interaction expectations.

General Discussion

We have introduced and demonstrated a new research method, a special type of cyranoid we call an echoborg. Echoborgs make possible interactions with artificial conversational agents that have truly human interfaces. Though an abundance of research has demonstrated various means of embodying machine intelligence in human form, from onscreen embodied conversational agents (e.g., Cassell et al., 2000; Krämer et al., 2009) to 3D agents in immersive virtual environments (e.g., Selvarajah and Richards, 2005; Bailenson et al., 2008) to tangible machine-bodied androids (e.g., Ishiguro and Nishio, 2007; Spexard et al., 2007), the echoborg stands apart from these other methods in that it involves a real, tangible human as the interface.

Study 1 compared a standard text-based version of the Turing Test to an echoborg version and found that although a chat bot's ability to pass a Turing Test was not improved when being shadowed by a human, being shadowed did increase ratings of how human-like the chat bot seemed. This effect of embodiment on human-likeness was unique to chat bot interlocutors, as human interlocutors in these tests were not seen as more human-like when their words were spoken by a human shadower, suggesting that a demonstrated capacity for human-level dialog may override the effect of human embodiment on perceptions of human-likeness in Turing Test contexts. Study 2 showed that in an ambiguous situation wherein participants were told that an interlocutor was either articulating words generated by a chat bot or merely imitating one, participants in a text interface condition were more likely to conclude that they had encountered the words of an actual chat bot than those who encountered an echoborg. The contrast between these two conditions provides evidence for (a) the robustness of the cyranic illusion, and (b) the notion that people's causal attributions align with what is most salient and least ambiguous to them. Study 3 explored the notion of passing and the uncanny valley in an ordinary, everyday contextual frame (i.e., the experimental context attempted to simulate a generic, unscripted, first-time encounter between strangers). Participants engaged with a covert chat bot via either a text interface or an echoborg. When interviewed following these interactions, most of the participants who engaged a text interface suspected they had encountered a chat bot, whereas only a few of the participants who engaged an echoborg held the same suspicion. This suggests that it is possible for a chat bot to pass as fully human given the requisite interface, namely an actual human body, and a suitable contextual frame. This study also found that people were less comfortable speaking to an echoborg than to a text interface.

Implications

Android Science

Drawing from Nunamaker et al.'s (2011) distinction between virtual avatars and embodied conversational agents, in **Figure 4** we visualize a simple two-dimensional matrix differentiating the basic tools available to android science, with one dimension indicating the source of verbal (and potentially non-verbal) agency and the other indicating interface-type. This matrix places the echoborg in relation to current mechanical devices utilized by android researchers (autonomous and tele-operated androids) as well as human beings as experimental subjects. By juxtaposing the field's tools in this manner, we can begin formally distinguishing the unique research questions that lend themselves to each. The fundamental question that each of these tools can be applied to

concerns what happens when the human elements of an interlocutor are removed and replaced by artificial imitations. The unique questions that can be approached via the usage of echoborgs concern how real human bodies (not mere mechanical imitations) fundamentally alter people's perceptions of and interactions with machine intelligence.

In the echoborg paradigm, the communicative limitations of chat bots and other types of conversational agents are not treated as problematic barriers to fluid conversation. Rather, these limitations are directly operationalized; how the human body as an interface mediates the perception of these communicative limitations is what is of interest. We can thus differentiate the echoborg paradigm from the tele-operated android paradigm in the following manner. Tele-operated android research targets the social dynamics between humans and human-like machine interfaces. Given that conversational agents are relatively poor communicators, the tele-operated paradigm cedes speech-interpretation/generation responsibility to a human operator, whose experiences operating an android can also be the subject of inquiry. By contrast, the echoborg paradigm is interested in the social dynamics that emerge when the words artificial systems produce are refracted through actual human bodies during face-to-face interaction.

The affordance which grants the echoborg particular promise as a methodology is that it allows researchers the opportunity to study interactions under conditions wherein people believe they are speaking to an autonomously communicating person. The echoborg can interact covertly (i.e., without interactants expecting that they are communicating with a bot). Of course, chat bots and other conversational agents can be deployed covertly via traditional text interfaces—and many are (e.g., posing as real people in chat rooms, web forums, and social media websites in order to distribute marketing messages and collect user-data; Gianvecchio et al., 2011; Nowak, 2012). But as Study 3 shows, focused interaction with a covert chat bot via a text interface for a sustained period of time is very likely to result in the interactant sensing that that they are not speaking to an actual person. Today's chat bots simply fail to sustain meaningful mixed-initiative dialog, and unless their words are vocalized by a tangible human body, their true nature is quickly exposed.

The Turing Test Paradigm (and Passing)

Over half a century since its conception, the Turing Test paradigm remains a substantial area of interest in artificial intelligence and philosophy of mind. The usefulness of the Turing Test as a technological benchmark, its rules, and what it would mean for a machine to pass such a test (i.e., what, exactly, passing would be evidence of) are issues that have been hotly debated (e.g., Searle, 1980; Copeland, 2000; French, 2000; Harnad, 2000; Chomsky, 2008; Watt, 2008; Proudfoot, 2011). The non-philosophical literature on the Turing Test focuses largely on the technological aspects of candidate conversational agents (e.g., whether they occasionally make spelling mistakes) and the conditions that give rise to increased fooling (e.g., knowing vs. not knowing of the possible presence of a machine intelligence; Saygin and Cicekli, 2002; Gilbert and Forney, 2015). What remains to be explored in sufficient depth are the social psychological dynamics within standard and modified

	SOURCE OF VERBAL (AND POTENTIALLY NONVERBAL) AGENCY:	
	Computer	Human
Android Interface:	Autonomous Android	Tele-Operated Android
Human Interface:	Echoborg	Autonomous Human / Cyranoid

FIGURE 4 | Basic tools of android science.

Turing Test scenarios: causal attributions, identity and power relationships, questions asked and avoided, misunderstandings recognized and repaired, intersubjective achievement, and so on (e.g., Warwick and Shah, 2015). Our position is that the Turing Test is most useful when its orthodox interpretation is relaxed and it is applied not toward assessing the capacities of chat bots *per se*, but toward investigating aspects of human social nature. Indeed, the chat bot itself may be the *least* interesting element within a Turing Test scenario. A chat bot can be made to fool a human interrogator if the expectations of the interrogator are manipulated (e.g., through ambiguous framing). What is interesting is exploring the ways in which the chat bot's utterances interact with the interrogator's expectations, all within a particular contextual frame, so as to produce a social interaction that feels more or less comfortable or human.

In essence, the three studies we have presented are all modified Turing Tests in that they explore passing in one form or another (with Study 1 bearing the closest resemblance to Turing's original concept). What our studies show is how intimately connected passing is to the social psychological framing of an interaction, and how the interface one communicates with affects the meaning of the situation from the point-of-view of interactants. In our own view, the results from Study 3 are at the same time the most profound and the least surprising. Seventeen of 20 people spoke face-to-face with an echoborg in a small room for 10-min and failed to develop even the slightest suspicion that they were interacting with the words of an artificial agent of some kind. They may have seen their interlocutor as strange, introverted, or even acting, but it did not cross their minds that who (or what) they were dealing with was part computer program. This makes sense in light of how we experience mundane human interaction, and implies that, given certain generic social psychological preconditions, an interlocutor's capacity to produce sophisticated or even sensible syntax simply does not factor in to our categorizing them as a human being or as having a "mind." That is to say, rather than taking these results as indicating the sophistication of chat bots, we take these results as indicating the importance of both the body and social psychological framing in social interaction.

Future Research Applications

Creating human-like interfaces that totally override people's awareness that they are interacting with something artificial remains a distant holy grail (Vogele and Bente, 2010). In the interim, however, we can use echoborgs to approximate the conditions of a world in which machines are capable of passing the non-verbal and motor requirements of a Total Turing Test. This opens the doors to a new frontier of human-robot and human-agent interaction research.

Echoborgs can be used to further study uncanny valley phenomena. Most of the literature that has explored the uncanny valley has focused on motor behavior and physical resemblance as independent variables, as well as the effects different levels of participant engagement (passive vs. active) have on perceptions of agents (e.g., von der Pütten et al., 2011). Researchers have also, but to a lesser extent, looked at the role of phonetic quality in relation to the uncanny valley (e.g., Mitchell et al., 2011; Tinwell

et al., 2011). Echoborgs enable us to study uncanny valley phenomena isolating dialogic capacity as an independent variable. Using echoborgs, we can see if an uncanny valley emerges when a spectrum of conversational agents ranging from the very poor (machine-like) to the very advanced (human-like) are communicated through a human speech shadower in unscripted face-to-face interactions.

Another possible avenue of research concerns the use of echoborgs in comparative person perception studies. Experiments can be designed with conditions differentiated in terms of the interface through which participants communicate with a particular conversational agent (text interface, embodied conversational agent, echoborg, and so on). Researchers could then observe how the various interfaces shape aspects of the personality perceived by the participant, from minimal interfaces all the way up to a face-to-face human body.

A particularly enticing possibility for future research involves developing bots that simultaneously dictate words to a shadower while directing elements of the shadower's motor behavior. In the echoborgs we have thus far constructed, the bot supplies the speech shadower with what to say while the shadower retains full control over their non-verbal functioning. We can imagine, however, developing a bot that delivered to the shadower's left ear monitor words to speak while delivering basic behavioral commands (e.g., "smile," "stand up," "extend right hand for handshake") to the shadower's right ear monitor. This would grant the bot greater agency over the echoborg's behavior.

The exciting opportunity opened up by echoborgs more generally is the opportunity to study human-computer interaction under the conditions of face-to-face human-human interaction. The problem for human-computer interaction research in general, and android science in particular, is that humans approach human-computer interaction differently from human-human interaction (as our own research shows). Human-human interaction triggers a huge range of complex phenomena, from identity dynamics to social emotions to basic taken-for-granted assumptions to an incredibly subtle intersubjective orientation to the other (Gillespie and Cornish, 2014). The echoborg method enables us to test conversational agents within face-to-face interaction scenarios, simultaneously pushing AI into a new domain and also to probing the full complexity of the human-human *inter-face*.

Ethical Considerations

In exploring social contexts involving a covert echoborg, mild deception is required in order to preserve the participant's belief that they are encountering an autonomous person. Careful experimental design (e.g., choice of conversational agents and shadowers, duration of interaction, communicative setting, etc.) and thorough piloting of procedures is strongly recommended so as to render participant distress unlikely. Participants should be exhaustively debriefed to gauge whether or not adjustments need to be made to the research procedures in order to avoid potential negative experiences. As a guideline, the debrief procedure in Study 3 involved asking the participant if they had any concerns regarding the ethics of the study as well as if they would object to a close friend or relative taking part in the same study under the

same conditions. All participants said no to both questions. We can anecdotally report that all of our participants enjoyed taking part in our research, with many expressing positivity toward the echoborg concept during debriefing and linking their experiences with what they had seen in popular science fiction films.

Limitations

Our studies were highly exploratory in nature. As such, various aspects of our investigations could have been more finely controlled. Though best attempts were made to standardize the body language of shadowers across all experimental trials, we did not make specific considerations for controlling certain behaviors (in particular, consistency of eye-contact). Moreover, the identity features of the shadowers (e.g., gender, ethnicity, age, and so on) may have produced unobserved effects on participants. We did not formally investigate such effects as they were not deemed to be of theoretical interest; however, we do acknowledge that questions regarding the relationship between the physical identity of the shadower and the social perception of the echoborg warrant future investigation. Sample sizes in our studies were relatively small due to practical constraints. Had our sample size for Study 3 been larger we might have been able to conduct a comprehensive comparison between the three chat bots used (Cleverbot, Rose, and Mitsuku). Also, we disclose that our choice of chat bots was based on prior familiarity with these programs.

We did not systematically analyze the effects audio latency may have had on participants' experiences. The delay between interactant-utterances and echoborg-responses in the studies that involved participants speaking aloud to an echoborg certainly degraded the mundane realism of interactions to some degree. Minimizing this latency is a major research priority as we continue to refine the echoborg methodology. At the moment we face a

trade-off between speed and accuracy: the use of a speed-typing third party (the minimal technological dependency model) slows the pace at which the conversational agent receives the words spoken by the interactant, yet better guarantees that the agent will process an accurate representation of the interactant's words.

Conclusion

This article has demonstrated the possibility and potential of echoborgs: human-bodied entities whose words (and potentially motor actions) are partially or completely determined by a computer program. Researchers can use echoborgs to study how people interact face-to-face with machine intelligence under the assumption that it is human. This methodology opens up a new paradigm for human-computer interaction research as to date people have interacted with computers, even sophisticated agents and highly lifelike androids, as machines (i.e., as things categorically different from real humans). Pairing a conversational agent with a human being to create an echoborg fundamentally transforms how people perceive and emotionally experience an in person encounter with social technology. Perhaps the most exciting takeaway from this initial examination of echoborgs is that under certain social psychological conditions echoborgs pass as fully autonomous human beings.

Acknowledgments

The authors would like to thank the following people for their contribution to this research: Geetha Reddy, Cristina Roem-mich, Shrabani Naha, Jamie Moss, Silvia Elaluf-Calderwood, Ivan Deschenaux, Alisa Gordon, Steve Bennett, Steve Gaskell, Ly Voo, and Mark Noort.

References

- Abildgaard, J. R., and Scharfe, H. (2012). "A geminoid as a lecturer," in *Social Robotics: Proceedings of the 4th International Conference on Social Robotics (ICSR)*, Chengdu, eds S. S. Ge, O. Khatib, J. Cabibihan, R. Simmons, and M. Williams (Berlin, Germany: Springer), 408–417. doi: 10.1007/978-3-642-34103-8_41
- Bailenson, J. N., Blascovich, J., and Guadagno, R. E. (2008). Self-representations in immersive virtual environments. *J. Appl. Soc. Psychol.* 38, 2673–2690. doi: 10.1111/j.1559-1816.2008.00409.x
- Bailenson, J. N., Swinith, K., Hoyt, C., Persky, S., Dimov, A., and Blascovich, J. (2005). The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence* 14, 379–393. doi: 10.1162/105474605774785235
- Bailly, G. (2003). Close shadowing natural versus synthetic speech. *Int. J. Speech Technol.* 6, 11–19. doi: 10.1023/A:1021091720511
- Bänziger, T., Grandjean, D., and Scherer, K. R. (2009). Emotion recognition from expressions in face, voice, and body: the multimodal emotion recognition test (MERT). *Emotion* 9, 691–704. doi: 10.1037/a0017088
- Becker-Asano, C. (2011). Affective computing combined with android science. *Künstliche Intell.* 25, 245–250. doi: 10.1007/s13218-011-0116-9
- Becker-Asano, C., and Ishiguro, H. (2011). Intercultural differences in decoding facial expressions of the android robot Geminoid F. *J. Artif. Intell. Soft Comput. Res.* 1, 215–231.
- Becker-Asano, C., Ogawa, K., Nishio, S., and Ishiguro, H. (2010). "Exploring the uncanny valley with Geminoid HI-1 in a real-world application," in *Proceedings of the IADIS International Conference on Interfaces and Human Computer Interaction*, Freiburg, ed. K. Blashki (Lisbon: IADIS Press), 121–128.
- Bessho, F., Harada, T., and Kuniyoshi, Y. (2012). "Dialog system using real-time crowdsourcing and Twitter large-scale corpus," in *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, Seoul, eds G. G. Lee, J. Ginzburg, C. Gardent, and A. Stent (Stroudsburg, PA: Association for Computational Linguistics), 227–231.
- Blascovich, J., Loomis, J., Beall, A. C., Swinith, K. R., Hoyt, C. L., and Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychol. Inq.* 13, 103–124. doi: 10.1207/S15327965PLI1302_01
- Carpenter, R. (2015). *Cleverbot [computer program]*. Available at: <http://www.cleverbot.com> [accessed January 26, 2015].
- Cassell, J., Sullivan, J., Prevost, S., and Churchill, E. (eds). (2000). *Embodied Conversational Agents*. Cambridge, MA: MIT Press.
- Chen, S., and Chaiken, S. (1999). "The heuristic-systematic model in its broader context," in *Dual Process Theories in Social Psychology*, eds S. Chaiken and Y. Trope (New York, NY: The Guilford Press), 73–96.
- Chomsky, N. (2008). "Turing on the 'imitation game,'" in *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, eds R. Epstein, G. Roberts, and G. Beber (New York, NY: Springer), 103–106. doi: 10.1007/978-1-4020-6710-5_7
- Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

- Copeland, B. J. (2000). The Turing Test*. *Minds Mach.* 10, 519–539. doi: 10.1023/A:1011285919106
- Corti, K., and Gillespie, A. (2015). Revisiting Milgram's cyranoid method: experimenting with hybrid human agents. *J. Soc. Psychol.* 155, 30–56. doi: 10.1080/00224545.2014.959885
- Dennett, D. (2004). "Can machines think?" in *Alan Turing: Life and Legacy of a Great Thinker*, ed. C. Teuscher (Berlin: Springer), 295–316. doi: 10.1007/978-3-662-05642-4_12
- Devine, P. G. (1989). Stereotypes and prejudice: their automatic and controlled components. *J. Pers. Soc. Psychol.* 56, 5–18. doi: 10.1037/0022-3514.56.1.5
- Dick, P. K. (1968). *Do Androids Dream of Electric Sheep?* New York, NY: Doubleday.
- Dougherty, E. G., and Scharfe, H. (2011). "Initial formation of trust: designing an interaction with Geminoid-DK to promote a positive attitude for cooperation," in *Social Robotics: Proceedings of the 3rd International Conference on Social Robotics (ICSR)*, Amsterdam, eds B. Mutlu, C. Bartneck, J. Ham, V. Evers, and T. Kanda (Berlin: Springer), 95–103. doi: 10.1007/978-3-642-25504-5_10
- Ekman, P. (1992). An argument for basic emotions. *Cogn. Emot.* 6, 169–200. doi: 10.1080/02699939208411068
- French, R. M. (2000). The Turing Test: the first 50 years. *Trends Cogn. Sci.* 4, 115–122. doi: 10.1016/S1364-6613(00)01453-4
- French, R. M. (2012). Moving beyond the Turing Test. *Commun. ACM* 55, 74–77. doi: 10.1145/2380656.2380674
- Friesen, N. (2009). Discursive psychology and educational technology: beyond the cognitive revolution. *Mind Cult. Act.* 16, 103–144. doi: 10.1080/10749030802707861
- Gianvecchio, S., Xie, M., Wu, Z., and Wang, H. (2011). Humans and bots in internet chat: measurement, analysis, and automated classification. *IEEE ACM Trans. Netw.* 19, 1557–1571. doi: 10.1109/TNET.2011.2126591
- Gilbert, R. L., and Forney, A. (2015). Can avatars pass the Turing Test? Intelligent agent perception in a 3D virtual environment. *Int. J. Hum. Comput. Stud.* 73, 30–36. doi: 10.1016/j.ijhcs.2014.08.001
- Gillespie, A., and Cornish, F. (2014). Sensitizing questions: a method to facilitate analyzing the meaning of an utterance. *Integr. Psychol. Behav. Sci.* 48, 435–452. doi: 10.1007/s12124-014-9265-3
- Goffman, E. (1963). *Stigma: Notes on the Management of Spoiled Identity*. London: Penguin Books.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychol. Rev.* 105, 251–279. doi: 10.1037/0033-295X.105.2.251
- Graziano, M. S. A. (2013). *Consciousness and the Social Brain*. Oxford: Oxford University Press.
- Harnad, S. (1991). Other bodies, other minds: a machine incarnation of an old philosophical problem. *Minds Mach.* 1, 43–54.
- Harnad, S. (2000). Minds, machines and Turing: the indistinguishability of indistinguishable. *J. Log. Lang. Inform.* 9, 425–445. doi: 10.1023/A:1008315308862
- Ishiguro, H. (2005). "Android science: toward a new cross-disciplinary framework," in *Proceedings of the 27th Annual Conference of the Cognitive Science Society: Toward Social Mechanisms of Android Science (A CogSci 2005 Workshop)*, Stresa, 18–28.
- Ishiguro, H., and Nishio, S. (2007). Building artificial humans to understand humans. *J. Artif. Organs* 10, 133–142. doi: 10.1007/s10047-007-0381-4
- Jones, E. E., and Nisbett, R. E. (1972). "The actor and the observer: divergent perceptions of the causes of behavior," in *Attribution: Perceiving the Causes of Behavior*, eds E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, and B. Weiner (Morristown, NJ: General Learning Press), 79–94.
- Khanna, N., and Johnson, C. (2010). Passing as black: racial identity work among biracial Americans. *Soc. Psychol. Q.* 73, 380–397. doi: 10.1177/0190272510389014
- Kiesler, S. (2005). "Fostering common ground in human-robot interaction," in *Proceedings of the 14th IEEE International Workshop on Robot and Human Interactive Communication (Ro-Man 2005)*, Nashville, TN, 729–734. doi: 10.1109/roman.2005.1513866
- Komatsu, T., and Yamada, S. (2011). Adaptation gap hypothesis: how differences between users' expected and perceived agent functions affect their subjective impression. *J. Syst. Cybern. Inf.* 9, 67–74.
- Krämer, N. C., Bente, G., Eschenburg, F., and Troitzsch, H. (2009). Embodied conversational agents: research prospects for social psychology and an exemplary study. *Soc. Psychol.* 40, 26–36. doi: 10.1027/1864-9335.40.1.26
- Levinas, E. (1991). *Totality and Infinity: An Essay on Exteriority*. Dordrecht: Kluwer Academic Publishers.
- Linell, P. (2009). *Rethinking Language, Mind, and World Dialogically: Interactional and Contextual Theories of Human Sense-Making*. Charlotte, NC: Information Age Publishing.
- Loebner, H. (2008). "How to hold a Turing Test contest," in *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, eds R. Epstein, G. Roberts, and G. Beber (New York, NY: Springer), 173–179. doi: 10.1007/978-1-4020-6710-5_12
- MacDorman, K. F., and Ishiguro, H. (2006a). Toward social mechanisms of android science. *Interact. Stud.* 7, 289–296. doi: 10.1075/is.7.2.12mac
- MacDorman, K. F., and Ishiguro, H. (2006b). The uncanny advantage of using androids in cognitive and social science research. *Interact. Stud.* 7, 297–337. doi: 10.1075/is.7.3.03mac
- Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature* 244, 522–523. doi: 10.1038/244522a0
- Marslen-Wilson, W. (1985). Speech shadowing and speech comprehension. *Speech Commun.* 4, 55–73. doi: 10.1016/0167-6393(85)90036-6
- Mavridis, N. (2015). A review of verbal and non-verbal human-robot interactive communication. *Rob. Auton. Syst.* 63, 22–35. doi: 10.1016/j.robot.2014.09.031
- Mavridis, N., Petychakis, M., Tsamakos, A., Toulis, P., Emami, S., Kazmi, W., et al. (2010). FaceBots: steps towards enhanced long-term human-robot interaction by utilizing and publishing online social information. *Paladyn* 1, 169–178. doi: 10.2478/s13230-011-0003-y
- Milgram, S. (1974). *Obedience to Authority: An Experimental View*. New York, NY: Harper and Row.
- Milgram, S. (1977). *Cyranic Speech: Pilot Studies* [video prepared for National Science Foundation review panel]. Stanley Milgram Archives (Series II: Studies, Box 26.1). Manuscripts and Archives Department, Sterling Memorial Library, Yale University, New Haven, CT.
- Milgram, S. (2010). "Cyranoids," in *The Individual in a Social World: Essays and Experiments*, ed. T. Blass (London: Pinter and Martin), 402–409.
- Mitchell, R. (2009). "An in your face interface: revisiting cyranoids as a revealing medium for interpersonal interaction," in *Proceedings of the 5th Student Interaction Design Research Conference (SIDeR): Flirting with the Future*, Eindhoven, eds I. H. C. Wouters, F. P. F. Kimman, R. Tieben, S. A. M. Offermans, and H. A. H. Nagtzaam (Eindhoven: Eindhoven University of Technology), 56–59.
- Mitchell, W. J., Szerszen, K. A. Sr., Lu, A. S., Schermerhorn, P. W., Scheutz, M., et al. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception* 2, 10–12. doi: 10.1068/i0415
- Mizumoto, T., Nakadai, K., Yoshida, T., Takeda, R., Otsuka, T., Takahashi, T., et al. (2011). "Design and implementation of selectable sound separation on the Texai telepresence System using HARK," in *Proceedings of the 2011 IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai (New York, NY: IEEE Press), 2130–2137. doi: 10.1109/ICRA.2011.5979849
- Mori, M. (1970). Bukimi no tani. *Energy* 7, 33–35.
- Nakadai, K., Takahashi, T., Okuno, H. G., Nakajima, H., Hasegawa, Y., and Tsujino, H. (2010). Design and implementation of robot audition system "HARK"—open source software for listening to three simultaneous speakers. *Adv. Rob.* 24, 739–761. doi: 10.1163/016918610X493561
- Nishio, S., Ishiguro, H., Anderson, M., and Hagita, N. (2008). "Expressing individuality through teleoperated android: a case study with children," in *Proceedings of the 3rd IASTED International Conference on Human Computer Interaction*, Innsbruck, ed. D. Cunliffe (Anaheim, CA: ACTA Press), 297–302.
- Nishio, S., Ishiguro, H., and Hagita, N. (2007a). Can a teleoperated android represent personal presence? A case study with children. *Psychologia* 50, 330–342. doi: 10.2117/psysoc.2007.330
- Nishio, S., Ishiguro, H., and Hagita, N. (2007b). "Geminoid: teleoperated android of an existing person," in *Humanoid Robots: New Developments*, ed. A. C. D. P. Filho (Vienna: I-Tech), 343–352.
- Nitsch, V., and Popp, M. (2014). Emotions in robot psychology. *Biol. Cybern.* 108, 621–629. doi: 10.1007/s00422-014-0594-6
- Nowak, P. (2012). Deceptibots: when machines go bad. *New Sci.* 214, 45–47. doi: 10.1016/S0262-4079(12)61636-4
- Nunamaker, J. F., Derrick, D. C., Elkins, A. C., Burgoon, J. K., and Patton, M. W. (2011). Embodied conversational agent-based kiosk for automated interviewing. *J. Manage. Inform. Syst.* 28, 17–48. doi: 10.2753/MIS0742-1222280102
- Ogawa, K., Taura, K., Nishio, S., and Ishiguro, H. (2012). "Effect of perspective change in body ownership transfer to teleoperated android robot," in *Proceedings of the 21st IEEE International Symposium on Robot and*

- Human Interactive Communication (Ro-Man 2012), Paris, 1072–1077. doi: 10.1109/ROMAN.2012.6343891
- Pardo, J. S., Jordan, K., Mallari, R., Scanlon, C., and Lewandowski, E. (2013). Phonetic convergence in shadowed speech: the relation between acoustic and perceptual measures. *J. Mem. Lang.* 69, 183–195. doi: 10.1016/j.jml.2013.06.002
- Perlis, D., Purang, K., and Andersen, C. (1998). Conversational adequacy: mistakes are the essence. *Int. J. Hum. Comput. Stud.* 48, 553–575. doi: 10.1006/ijhc.1997.0181
- Pieraccini, R. (2012). *The Voice in the Machine: Building Computers that Understand Speech*. Cambridge, MA: MIT Press.
- Proudfoot, D. (2011). Anthropomorphism and AI: Turing's much misunderstood imitation game. *Artif. Intell.* 175, 950–957. doi: 10.1016/j.artint.2011.01.006
- Raine, R. (2009). "Making a clever intelligent agent: the theory behind the implementation," in *Proceedings of the IEEE International Conference on Intelligent Computing and Intelligent Systems (ICIS 2009)*, Shanghai, 398–402. doi: 10.1109/ICISYS.2009.5358137
- Ranky, G. N., and Ranky, P. G. (2005). Japanese prototype service robot R&D trends and examples. *Ind. Rob.* 32, 460–464. doi: 10.1108/01439910510629163
- Raudaskoski, P., and Mitchell, R. (2013). "The situated accomplishment (aesthetics) of being a cyranoid," in *Proceedings of the Participatory Innovation Conference (PIN-C 2013)*, Lahti, eds H. Melkas and J. Buur (Lappeenranta: LUT Scientific and Expertise Publications), 126–129.
- Renfrow, D. G. (2004). A cartography of passing in everyday life. *Symb. Interact.* 27, 485–506. doi: 10.1525/si.2004.27.4.485
- Rosenthal-von der Pütten, A. M., and Krämer, N. C. (2014). How design characteristics of robots determine evaluation and uncanny valley related responses. *Comput. Hum. Behav.* 36, 422–439. doi: 10.1016/j.chb.2014.03.066
- Sager, H. A., and Schofield, J. W. (1980). Racial and behavioral cues in black and white children's perceptions of ambiguously aggressive acts. *J. Pers. Soc. Psychol.* 39, 590–598. doi: 10.1037/0022-3514.39.4.590
- Sakamoto, D., Kanda, T., Ono, T., Ishiguro, H., and Hagita, N. (2007). "Android as a telecommunication medium with a human-like presence," in *Proceedings of the 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI'07)*, Arlington, VA, eds A. Schultz, C. Breazeal, T. Fong, and S. Kiesler (New York, NY: ACM), 193–200. doi: 10.1145/1228716.1228743
- Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., and Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Soc. Cogn. Affect. Neurosci.* 7, 413–422. doi: 10.1093/scan/nsr025
- Saygin, A. P., and Cicekli, I. (2002). Pragmatics in human–computer conversations. *J. Pragmat.* 34, 227–258. doi: 10.1016/S0378-2166(02)80001-7
- Schegloff, E. A. (1986). The routine as achievement. *Hum. Stud.* 9, 111–151. doi: 10.1007/BF00148124
- Schumaker, R. P., Ginsburg, M., Chen, H., and Liu, Y. (2007). An evaluation of the chat and knowledge delivery components of a low-level dialog system: the AZ-ALICE experiment. *Decis. Support Syst.* 42, 2236–2246. doi: 10.1016/j.dss.2006.07.001
- Schwitzgebel, R. K., and Taylor, R. W. (1980). Impression formation under conditions of spontaneous and shadowed speech. *J. Soc. Psychol.* 110, 253–263. doi: 10.1080/00224545.1980.9924252
- Searle, J. R. (1980). Minds, brains, and programs. *Behav. Brain Sci.* 3, 417–424. doi: 10.1017/S0140525X00005756
- Selvarajah, K., and Richards, D. (2005). "The use of emotions to create believable agents in a virtual environment," in *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'05)*, Utrecht, eds M. Pechoucek, D. Steiner, and S. Thompson (New York, NY: ACM), 13–20. doi: 10.1145/1082473.1082476
- Seyama, J., and Nagayama, R. S. (2007). The uncanny valley: effect of realism on the impression of artificial human faces. *Presence* 16, 337–351. doi: 10.1162/pres.16.4.337
- Shahri, H. H., and Perlis, D. (2008). "Finding ontological correspondences for a domain-independent natural language dialog agent," in *Proceedings of the 20th Innovative Applications of Artificial Intelligence Conference (AAAI-08)*, Chicago, IL, eds M. Goker and K. Haigh (Palo Alto, CA: AAAI Press), 1685–1692.
- Shockley, K., Sabadini, L., and Fowler, C. A. (2004). Imitation in shadowing words. *Percept. Psychophys.* 66, 422–429. doi: 10.3758/BF03194890
- Spence, C., and Read, L. (2003). Speech shadowing while driving: on the difficulty of splitting attention between eye and ear. *Psychol. Sci.* 14, 251–256. doi: 10.1111/1467-9280.02439
- Spexard, T. P., Hanheide, M., and Sagerer, G. (2007). Human-oriented interaction with an anthropomorphic robot. *IEEE Trans. Rob.* 23, 852–862. doi: 10.1109/TRO.2007.904903
- Straub, I., Nishio, S., and Ishiguro, H. (2010). "Incorporated identity in interaction with a teleoperated android robot: a case study," in *Proceedings of the 19th IEEE International Symposium on Robot and Human Interactive Communication (Ro-Man 2010)*, Viareggio, 119–124. doi: 10.1109/ROMAN.2010.5598695
- Taylor, S. E., and Fiske, S. T. (1975). Point of view and perceptions of causality. *J. Pers. Soc. Psychol.* 32, 439–445. doi: 10.1037/h0077095
- Tinwell, A., Grimshaw, M., and Williams, A. (2011). "Uncanny speech," in *Game Sound Technology and Player Interaction: Concepts and Developments*, ed. M. Grimshaw (Hershey, PA: Information Science Reference), 213–234. doi: 10.4018/978-1-61692-828-5.ch011
- Tomassello, M. (2008). *Origins of Human Communication*. Cambridge, MA: MIT Press.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind* 59, 433–460. doi: 10.1093/mind/LIX.236.433
- Vogele, K., and Bente, G. (2010). "Artificial humans": psychology and neuroscience perspectives on embodiment and nonverbal communication. *Neural Netw.* 23, 1077–1090. doi: 10.1016/j.neunet.2010.06.003
- von der Pütten, A. M., Krämer, N. C., Becker-Asano, C., and Ishiguro, H. (2011). "An android in the field," in *Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI'11)*, Lausanne, eds A. Billard, J. A. Adams, P. Kahn, and J. G. Trafton (New York, NY: ACM), 283–284.
- Wallace, R. (2008). "The anatomy of A.L.I.C.E.," in *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, eds R. Epstein, G. Roberts, and G. Beber (New York, NY: Springer), 181–210. doi: 10.1007/978-1-4020-6710-5_13
- Wallace, R. (2015). *A.L.I.C.E. [computer program]*. Available at: <http://alice.pandorabots.com> [accessed January 26, 2015].
- Warwick, K., and Shah, H. (2015). Human misidentification in Turing Tests. *J. Exp. Theor. Artif. Intell.* 27, 123–135. doi: 10.1080/0952813X.2014.921734
- Watanabe, M., Ogawa, K., and Ishiguro, H. (2014). "Field study: can androids be a social entity in the real world?," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction (HRI'14)*, eds G. Sagerer, M. Imai, T. Belpaeme, and A. Thomaz (New York, NY: ACM), 316–317. doi: 10.1145/2559636.2559811
- Watt, S. (2008). "Can people think? Or machines?," in *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, eds R. Epstein, G. Roberts, and G. Beber (New York, NY: Springer), 301–318. doi: 10.1007/978-1-4020-6710-5_18
- Weizenbaum, J. (1966). Eliza—a computer program for the study of natural language communication between man and machine. *Commun. ACM* 9, 36–45. doi: 10.1145/365153.365168
- Wilcox, B. (2015). *Rose [computer program]*. Available at: <http://brilligunderstanding.com/rosedemo.html> [accessed January 26, 2015].
- Worswick, S. (2015). *Mitsuku [computer program]*. Available at: <http://www.squarebear.co.uk/mitsuku/housebot.htm> [accessed January 26, 2015].
- Yee, N., and Bailenson, J. N. (2007). The Proteus effect: the effect of transformed self-representation on behavior. *Hum. Commun. Res.* 33, 271–290. doi: 10.1111/j.1468-2958.2007.00299.x
- You, J. (2015). Beyond the Turing Test. *Science* 347, 116. doi: 10.1126/science.347.6218.116
- Ziemke, T., and Lindblom, J. (2006). Some methodological issues in android science. *Interact. Stud.* 7, 339–342. doi: 10.1075/is.7.3.05zie
- Zue, V. W., and Glass, J. R. (2000). Conversational interfaces: advances and challenges. *Proc. IEEE* 88, 1166–1180. doi: 10.1109/5.880078

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Corti and Gillespie. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

