Gesture-Bot: Design and Evaluation of Simple Gestures of a Do-it-yourself Telepresence Robot for Remote Communication

Yibo Wang[†]
School of Design
Shanghai Jiao Tong University
Shanghai, China
wangyibo828@sjtu.edu.cn

Chenwei Zhang[†]
Department of Electrical and Electronic Engineering
The University of Hong Kong
Hong Kong SAR
u3584554@connect.hku.hk

Heqiao Wang
City University of Hong Kong
Hong Kong SAR
heqiaow28@gmail.com

Shuya Lu
University of Arts London
London, United Kingdom
slu27529@gmail.com

Ray LC
School of Creative Media
City University of Hong Kong
Hong Kong SAR
LC@raylc.org

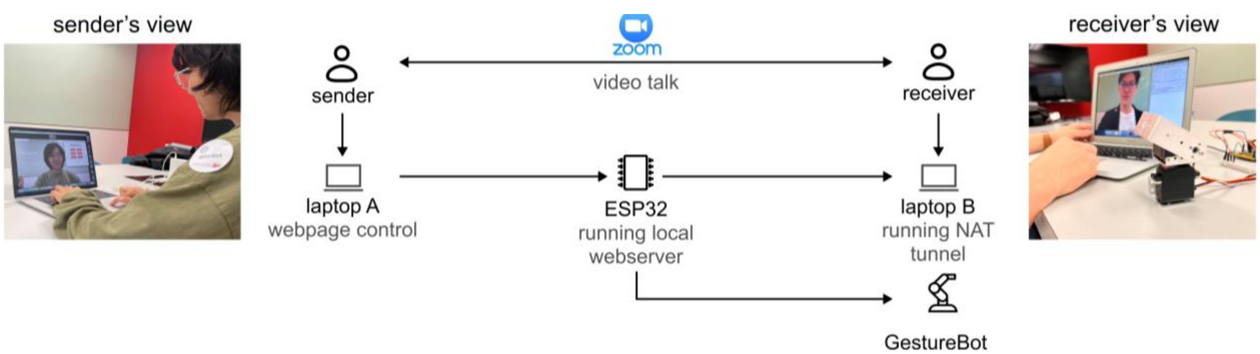

Figure 1: The communication flow when using the Gesture-Bot system. (Left) Sender of movement commands to the Gesture-Bot through a web interface. (Right) Location of the Gesture-Bot with the participant also on the call.

ABSTRACT

Current video conferencing technology allows participants to communicate virtually over distance, but users lack the sense of presence due to the absence of physical cues for interaction. We proposes the design of Gesture-Bot, a DIY telepresence robot that performs pan-and-tilt gestures in the presence of a receiver during video chat. It is command via the web by a remote sender. We conducted a workshop to design and evaluate pan-and-tilt movements with 26 participants in two separate communication scenarios to examine the flow of communication in physical-interactions-assisted remote chat. According to the data collected from the questionnaire and the post-experiment interview, the Gesture-Bot system has shown its potential in assisting remote communications while aspects such as the outlook, the method of controlling the robot, the set of gestures are to be improved in the future.

CCS CONCEPTS

• Human-centered computing →Interaction design →Empirical studies in interaction design

KEYWORDS

Remote Communication, Telepresence, DIY Robotics, Gestures

ACM Reference format:

Yibo Wang, Chenwei Zhang, Heqiao Wang, Shuya Lu and RAY LC. 2023. Gesture-Bot: Design and Evaluation of Simple Gestures of a Do-it-yourself Telepresence Robot for Remote Communication. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI'23 Companion), March13-16, 2023, Stockholm, Sweden.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/ 3568294.3580051

[†]Two authors contributed equally to the paper.

HRI '23 Companion, March 13–16, 2023, Stockholm, Sweden.

ACM ISBN 978-1-4503-9970-8/23/03...$15.00
DOI: https://doi.org/10.1145/3568294.3580051

1 INTRODUCTION

The prevalence of smart devices, coupled with health crises, has increased the use of video talk technology, which offers real-time visibility of users and their surroundings, for remote communication in video conferencing with others [11]. However, virtual events like online video talk do not perform satisfactorily at giving participants a sense of social presence [3]. Unlike in-person gatherings where participants can utilize body movements and gestures to naturally express their feelings as well as exchange nonverbal information, current video talk technology does not support such physical interactions. Also, the low-bandwidth networks can cause latency and reduce video talk quality, which makes it difficult for users to recognize nonverbal information from facial expressions and gestures [8]. Therefore, participants in video talks often lack a sense of involvement and engagement. The loss of nonverbal information also increases the likelihood of misunderstanding and misinterpretation [15]. There are existing works that apply virtual characters to assist distant communications [10]. However, previous studies showed that physical interactions are favored over virtual ones [14].

Previous work in robot-mediated distant communications have included the use of robotic avatars for establishing physical and social presence to others [1], and projection of human face onto sculptural displays for self awareness aims [6]. In mobile robotic telepresence (MRP), a human shaped device consisting of an LCD screen, a microphone, a web camera, speakers, and a mobile robotic base with wheels is employed [5]. Such robots have been useful for distant couples, as it was found that the inclusion of physical agents in communication supports natural interactions akin in-person communication, promoting connection between partners [16]. However, most of the existing telepresence designs, such as virtual reality glasses or humanoid robots, are expensive and oversized for everyday use [2].

As an experimental approach to fill this gap, we focused on the socialized human robotics interactions [9] using gestural game-like interactions [7], and designed Gesture-Bot, a robot that is smaller and of lower cost compared to MRP robots [4]. Gesture-Bot only takes up the space of up to one cubic decimeter, similar to a desktop action figure. By performing pan-and-tilt gestures under control of participants during video talks, Gesture-Bots can simulate physical interactions. Gesture-Bot's mechanism is easy for users to understand and assemble, thus making it an accessible robot for everyone. We designed experiments and conducted an online workshop with participants recruited from university. Based on the data and user reflections, we empirically evaluated the extent to which Gesture-Bot can help communicate human intentions and emotions through its pan-and-tilt movements and the extent to which users can accurately decode the messages conveyed by such movements controlled by other users.

2 DESIGN

We designed an easy-to-DIY prototype at affordable cost. The robotic body consists of a pan-and-tilt head, driven by medium size servos (DS3115) after taking into consideration size, noise, and power needs. Microprocessor ESP32 is used for motor control.

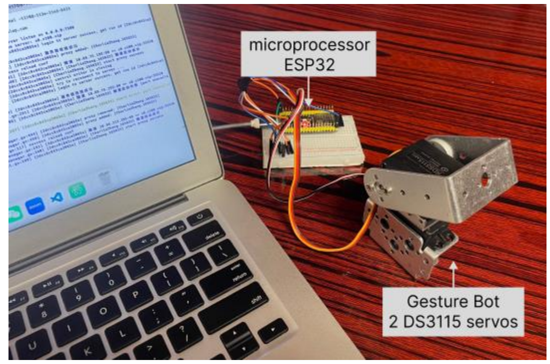

Figure 2: Gesture-Bot platform and its hardware components

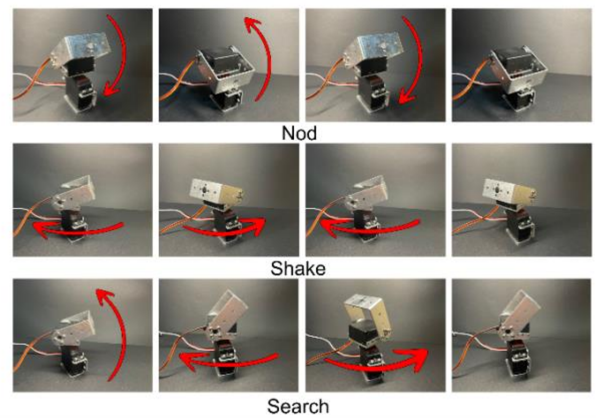

Figure 3: examples of the pre-designed gestures

The prototype only requires users to download the Arduino IDE, upload the code, and activate a NAT-Tunnel software on the market. The workflow of (Figure 2) is: the microprocessor ESP32 will start a local async-web server when connected to local Wi-Fi, prepare the webpage-based user interface, print out the local IP of the UI in the serial monitor and listen to updates. By mapping the local IP of the UI to a public IP address using NAT-Tunnel software, the UI will be accessible globally. After the remote user clicking on the webpage UI, the ESP32 will receive a command and send PWM signals to the servos and the robot will perform a preset gesture corresponding to the action selected. We designed a few pre-set gestures for Gesture-Bot basing on previous work, including gestures like nodding, shaking, and searching [18] (Figure 3). The components of the Gesture-Bot system are easily accessible, and the total cost is below 30 USD.

3 METHODS

To evaluate the effectiveness of Gesture-Bot in assisting video talk, we conducted an online within-subject workshop, which serves as a preliminary step before in-person studies.

We conducted convenience sampling and recruited 26 participants (17 female, 9 male) from institutions who signed up online. To cope with each participants' schedule, 26 participants were divided into two groups with 10 and 16 persons respectively and attended the workshop in two different time slots. The average age of the participants was 20 years old (SD=3), and are Chinese university students. The experiment was conducted in the form of online workshop using Voov Meeting due to Covid-19 and the non-physical nature of our research topic (Figure 4). At the start of the workshop, we presented to all participants the idea of robot aided video talks, the Gesture-Bot setup, and the possible ways to control it. After acknowledging their understanding of the above, a web page link that can be used to remotely control the Gesture-Bot was sent to the participants. All participants were encouraged to control the robot through the webpage while the real-time gestures of the robot were shown on a live webcam.

Figure 4: Participants inventing gestures through a simulated video talk based on the scenario during the workshop

To demonstrate the use of Gesture-Bot, we randomly selected one participant to simulated a video chat. The participant could control the robot as they wished during the demo, and other participants can see the Gesture-Bot in real-time via the webcam.

In the actual workshop, participants were given two different pre-designed semi-structured two-person video conversation scenarios. One scenario focused on quarantine time and the other focused on family emotions. Each scenario included a background story, a main character to control the Gesture-Bot, as well as a supporting character to complete the conversation.

Participants were separated into pairs to facilitate the video conferencing simulation. They were asked to try to empathize with the characters in the scenarios and encouraged to express actively. Then they were asked to use their hands to simulate the pan-and-tilt movements of Gesture-Bot to aid their self-expression. Each simulation lasted around 5 minutes. After the simulation, participants completed a retroactive survey to evaluate Gesture-Bot on a 7-point Likert scale and answered questions in an interview. The items to be evaluated in the questionnaires asked were are given in Figure 5.

The first round of the workshop included 10 participants out of 26. After the procedure, some participants voiced their wish to talk more regarding the design of the platform, the concept of robot-aided communication, and the procedure of the experiment. In the second round of the workshop, we held an additional interview and recorded the feedback. Questions we asked were: "What do you think are the differences between conventional online communication and today's conversation? What do you think are the pros and cons of using this kind of robot during the conversation? Do you remember some of the specific emotional content of the conversation? Etc." Participants were encouraged to talk more about whichever topic they showed more interest in.

4 RESULTS

4.1 Quantitative Findings

In the semi-structured interviews conducted after, the result of the questionnaire was analyzed using R studio and visualized using a heat map and shown in Figure 4. The item with the highest average rating is 1 and 5. The overall result indicates that using gestures to assist conversation could to some extent but not significantly improve the quality of communication. Regarding two related questions (3, 7 and 7, 8), the Wilcoxon test was performed and did not show significance (p>0.05). Thus, there is no significant difference between the user's intended emotion and the actual emotion expressed, and between intention and emotion. The use of the robot system appear to allow users to remember conversations better. Interestingly this is despite the robot having shifted attention away from the conversation for a portion of the participants. This suggests that the robot takes up attention resources of the users but its novelty makes the conversations engaging and memorable.

Figure 5: Survey Results. 1: strongly disagree; 7: strongly agree. Understandable: *"The way the robot moves in general makes sense to me."* Shifted Attention: *"Using the robot shifted my attention away from the conversation."* Represent Intention: *"The moving patterns of the robot represented my intentions well."* Represent Human Gesture: *"The robot represents the gestures of a human well."* Remember Conversation: *"The robot interactions allowed me to remember the contents of the conversation."* Frequency of Use: *"I frequently adopt the robot during conversation."* Express Intended Emotion: *"The robot expresses my intended emotions well."* Express Emotion: *"The robot expresses emotion well."*

4.2 Qualitative Findings

The video of the workshop was recorded, and the dialogue audio was translated into English using Voov meeting. By examining through the text line-by-line, we summarize these findings.

In the semi-structured interviews conducted after 2 scenarios were played, participants showed different attitudes toward the robot. A minority of participants (2/16) considered the Gesture-

Bot as redundant under a video call. Of the majority view, a participant said that "*the video delivers my emotions much better in a more detailed way*" while another user regarded the Gesture-Bot as "*barely noticeable*", while the rest of participants indicated that the robot was in some way helpful to the conversation. The following statements were mentioned: "*made the conversation interesting and entertaining*" (3/16); "*helped to deliver one's emotion*" (3/16); "*made the conversation impressive and unforgettable*" (4/16). Additionally, one participant mentioned that "*I would love to use the robot to perform some crazy, exaggerated or even impossible gestures to deliver my subtle emotions.*" and another suggested that the robot could play a role in distant relationships as it provides a sense of "*presence*" and "*companionship*". Negative effects toward the conversation mostly include distraction, while a participant suggested that "*It becomes less distracting as I get used to it.*"

Regarding the design of the robot, a significant part of participants (7/16) desired a robot which can move automatically based on facial expressions or verbal commands, while the opponents (2/16) think it is better to have preset gestures. Some (4/16) proposed that a better visual appearance for the Gesture-Bot.

4.3 Participant Behavior

On video analysis, it was observed that most participants tended to either stop talking when performing gestures or explain their reasons to perform certain gestures.

Although with different amplitude, speed and times of repetition, a great portion of total gestures (23/47) observed were nodding and shaking. Among the observed 23 nodding and shaking gestures, 14 of them expressed simple "confirming" or "denying" to a question. Apart from which, nodding was mostly used to show "carefully listening" to another, and shaking was used to express "disappointed" and "disagree."

It is intriguing that "heads up" and "heads down" almost always indicate positive feelings and negative feelings respectively. To elaborate, the emotions delivered using "heads up" were happy and surprised, and by doing "heads down" the participants tried to express disappointment, frustration, and desperation. One exception is that heads down and up was used to explain "infected" and "recovered" from Covid-19.

Another phenomenon worth noticing is that the speed and time of repetition could indicate the intensity of emotions. For example, a participant expressed "frustration" by "slowly heads down", while in his interpretation "rapidly heads down" meant "desperation." Another participant expressed strong disagreement by "shaking repetitively" while according to him "shaking once" only meant negation. This suggests that the limited movement set of a pan-and-tilt robot led users to adapt, expressing different emotions using the same gesture but varying speed and repetition.

5 DISCUSSIONS AND LIMITATIONS

The workshop focused on how one participant (main character) can use pan-and-tilt gestures to assist distant communication, but the question regarding how the other participant (supporting character) interpret such gestures were unexplored. Also, due to policies under Covid-19, we were unable to conduct off-line experiments, thus the sample size of participants were restricted. Accordingly, one of our future works is to conduct offline experiments where both parts of the conversation have the Gesture-Bot.

While the in-the-moment ratings on the questions in the survey were non-significant, participants voiced differing impressions on their experience with the Gesture-Bot. Most of the participants reflected that Gesture-Bot suggests a new way of using robot simulated physical gestures to aid video talk, but a significant portion voiced that maintaining a state of talking and thinking of how to use the robot at the same time, especially when it was their first experience with the Gesture-Bot, could be challenging. However, it's encouraging that the robot led to more memorable and attention-taking conversations despite its limitations.

The participants' behavior during scenarios hints that they had trouble expressing complicated emotion through a limited degree of freedom (DOF), as most of the gestures observed were nodding and shaking. The participants' post-study feedback also includes numerous requirements of using more complicated gestures.

In this study, important aspects of robot design such as anthropomorphic appearance, coloring, and sound were not taken into consideration. However, these functionalities could be of great importance when expressing emotions [12]. As the purpose of this preliminary study was to gather general reactions to use in a specific task, the robot was designed to be as simple as possible. Nevertheless, we believe that the concept of gestures assisted remote communication could be applied in other scenarios, and the configurational concepts of the Gesture-Bot platform can be used to empower a wide range of robotic devices [13]. This study shows an open-source hardware platform for innovators to make their own robot, and [17] is an instance of human-shaped social robots that could be remotely controlled.

Considering the suggestions given by participants, our future work is to develop an automated telepresence robot system with better controls and support for a wider set of gestures.

6 CONCLUSIONS

This work explores the use of gestures performed by a telepresence robot to assist video talk through an easy-to-DIY platform. We designed this robot using simple pan-and-tilt gestures. As a result, the robot system was easy to replicate and low cost. During the workshop, participants reported more memorable conversations due to use of the system, expressed capacity to express emotional content using the robots, proposed scenarios for future applications, and shared feedback regarding the design of the platform. The increased engagement and relatively simple use of the system through a web interface suggest the design of future DIY technologies for gesture-based interactions for creating presence and tangibility in increasingly digitalized communication.

REFERENCES

[1] Aryel Beck, Lola Cañamero, and Kim A. Bard. 2010. Towards an Affect Space for robots to display emotional body language. In 19th International Symposium in Robot and Human Interactive Communication, 464–469. DOI:https://doi.org/10.1109/ROMAN.2010.5598649

[2] John-John Cabibihan, Wing-Chee So, and Soumo Pramanik. 2012. Human-Recognizable Robotic Gestures. IEEE Trans. Auton. Ment. Dev. 4, 4 (2012), 305–314. DOI:https://doi.org/10.1109/TAMD.2012.2208962

[3] Matthew J. Bietz, Nitesh Goyal, Nicole Immorlica, Blair MacIntyre, Andrés Monroy-Hernández, Benjamin C. Pierce, Sean Rintel, and Donghee Yvette Wohn. 2022. Social Presence in Virtual Event Spaces. In CHI Conference on Human Factors in Computing Systems Extended Abstracts, ACM, New Orleans LA USA, 1–5. DOI:https://doi.org/10.1145/3491101.3503713

[4] Maria Karam. A taxonomy of Gestures in Human Computer Interaction. 45.

[5] Annica Kristoffersson, Silvia Coradeschi, and Amy Loutfi. 2013. A Review of Mobile Robotic Telepresence. Adv. Hum.-Comput. Interact. 2013, (2013), 1–17. DOI:https://doi.org/10.1155/2013/902316

[6] RAY LC, Aaliyah Alcibar, Alejandro Baez, and Stefanie Torossian. 2020. Machine Gaze: Self-Identification Through Play With a computer Vision-Based Projection and Robotics System. *Front. Robot. AI* 7:580835. doi: 10.3389/frobt.2020.580835

[7] RAY LC, Maurice Benayoun, Permagnus Lindborg, Hongshen Xu, Hin Chung Chan, Ka Man Yip, and Tianyi Zhang. 2022. Power Chess: Robot-to-Robot Nonverbal Emotional Expression Applied to Competitive Play. In 10th International Conference on Digital and Interactive Arts (ARTECH 2021). Association for Computing Machinery, New York, NY, USA, Article 2, 1–11. https://doi.org/10.1145/3483529.3483844

[8] Jialang Victor Li, Max Kreminski, Sean M Fernandes, Anya Osborne, Joshua McVeigh-Schultz, and Katherine Isbister. 2022. Conversation Balance: A Shared VR Visualization to Support Turn-taking in Meetings. In CHI Conference on Human Factors in Computing Systems Extended Abstracts, ACM, New Orleans LA USA, 1–4. DOI:https://doi.org/10.1145/3491101.3519879

[9] Hamed Mahzoon, Ayaka Ueda, Yuichiro Yoshikawa, and Hiroshi Ishiguro. 2022. Effect of robot's vertical body movement on its perceived emotion: A preliminary study on vertical oscillation and transition. PLOS ONE 17, 8 (August 2022), e0271789. DOI:https://doi.org/10.1371/journal.pone.0271789

[10] Kosaku Namikawa, Ippei Suzuki, Ryo Iijima, Sayan Sarcar, and Yoichi Ochiai. 2021. EmojiCam: Emoji-Assisted Video Communication System Leveraging Facial Expressions. In Human-Computer Interaction. Design and User Experience Case Studies, Masaaki Kurosu (ed.). Springer International Publishing, Cham, 611–625. DOI:https://doi.org/10.1007/978-3-030-78468-3_42

[11] Ken-Ichi Okada, Fumihiko Maeda, Yusuke Ichikawaa, and Yutaka Matsushita. 1994. Multiparty videoconferencing at virtual social distance: MAJIC design. In Proceedings of the 1994 ACM conference on Computer supported cooperative work (CSCW '94), Association for Computing Machinery, New York, NY, USA, 385–393. DOI:https://doi.org/10.1145/192844.193054

[12] Sichao Song and Seiji Yamada. 2017. Expressing Emotions Through Color, Sound, and Vibration with an Appearance-Constrained Social Robot. In 2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI, 2–11.

[13] Cesar Vandevelde, Francis Wyffels, Bram Vanderborght, and Jelle Saldien. 2017. Do-It-Yourself Design for Social Robots: An Open-Source Hardware Platform to Encourage Innovation. IEEE Robot. Autom. Mag. 24, 1 (March 2017), 86–94. DOI:https://doi.org/10.1109/MRA.2016.2639059

[14] Joshua Wainer, David Feil-seifer, Dylan Shell, and Maja Mataric. 2006. The role of physical embodiment in human-robot interaction. In ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication, IEEE, Univ. of Hertfordshire, Hatfield, UK, 117–122. DOI:https://doi.org/10.1109/ROMAN.2006.314404

[15] Harald G. Wallbott. 1998. Bodily expression of emotion. Eur. J. Soc. Psychol. 28, 6 (1998), 879–896. DOI:https://doi.org/10.1002/(SICI)1099-0992(1998110)28:6<879::AID-EJSP901>3.0.CO;2-W

[16] Lillian Yang, Carman Neustaedter, and Thecla Schiphorst. 2017. Communicating Through A Telepresence Robot: A Study of Long Distance Relationships. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, ACM, Denver Colorado USA, 3027–3033. DOI:https://doi.org/10.1145/3027063.3053240

[17] Naoki Ise and Takamasa Iio. 2021. Social Robot Encouraging Two Strangers to Talk with Each Other for Their Relationships. In Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, ACM, Boulder CO USA, 144–147. DOI:https://doi.org/10.1145/3434074.3447147

[18] Moretti S, Greco A. Nodding and shaking of the head as simulated approach and avoidance responses. Acta Psychol (Amst). 2020;203: 102988. doi:10.1016/j.actpsy.2019.102988