Playing Charades with a Robot: Collecting a Large Dataset of Human Gestures Through HRI

Jan de Wit*, Bram Willemsen, Mirjam de Haas, Emiel Krahmer, Paul Vogt,
Marije Merckens, Reinjet Oostdijk, Chani Savelberg, Sabine Verdult and Pieter Wolfert
Tilburg center for Cognition and Communication, Tilburg University
Tilburg, the Netherlands
Email: *j.m.s.dewit@uvt.nl

***Abstract*—This work documents a playful human-robot interaction, in the form of a game of charades, through which a humanoid robot is able to learn how to produce and recognize gestures by interacting with human participants. We describe an extensive dataset of gesture recordings, which can be used for future research into gestures, specifically for human-robot interaction applications.**

***Index Terms*—Robot learning, Human-robot interaction, Gesture recognition, Robot motion**

I. Introduction

The ability to produce and recognize non-verbal communication, such as gestures, facilitates understanding between humans and robots, and results in more engaging interactions [1]. Previous work [2] has also shown that a robot's use of iconic gestures [3] is beneficial to second language learning. By enabling the robot to learn these gestures from demonstration [4], we avoid the need to manually design and program them, thereby removing the influence of the designer's frame of reference. The resulting motions could potentially be perceived as more human-like, because they are based on recordings of human gestures that are automatically mapped onto the robot. In this work, we present a dataset of recorded gestures for 35 different objects, which was gathered through a game of charades with a robot.

II. Approach

A. Procedure

After completing a practice round, the robot started the game by performing a gesture from its set of examples, previously recorded from other participants. The participant was then shown a picture of the item that the robot tried to enact, along with three incorrect answers, on the tablet (see Figure 1, left). If the participant guessed incorrectly, the robot performed a gesture for the same object once more for another guess. Then, the roles were reversed and the participant was shown an object on the screen, which they then described using an upper-body gesture (Figure 1, right). The robot tried to recognize the object that was portrayed, and if guessed incorrectly the participant was asked to perform a gesture for the object again for a second attempt. To provide additional insight into the robot's confidence when guessing, the participant was shown a top five of answer candidates, from which the robot picked the top one for its guess. Each game session lasted five rounds of the robot and participant taking turns guessing, covering ten objects — five performed by the robot, five by the participant — out of a total set of 35, which included animals, static objects (furniture, buildings), tools (e.g., cup, book, toothbrush), musical instruments, and vehicles.

Fig. 1. Left: The participant correctly guesses a gesture performed by the robot (*glasses*). Right: The participant is performing a gesture for the robot to guess (*ball*).

B. Implementation

Gesture recognition was implemented by extracting the *gist* of the gesture, inspired by the work of Cabrera and Wachs [5]. This gist was then compared to the complete set of previously recorded gestures using a k-nearest neighbors approach to find the object that was most likely depicted by the current gesture. Hierarchical clustering was used to group similar gestures for each object, and after the participant guessed an object, the weights of these clusters and individual gestures within clusters were increased or lowered based on whether the answer was correct or not. When choosing a gesture to perform, the robot would either explore a new sample (40%), or exploit the cluster and sample with the highest weight (60%). Previously recorded gestures were mapped to the robot's accepted input format for performing motions by calculating the various joint angles that the NAO robot accepts from the joint positions of the participant that were recorded by the Kinect camera. Three gestures for each of the 35 objects were performed by the researcher and added to the system as an initial set for recognition and production. The system was deployed, with an identical setup, at two locations: a science museum that is mostly visited by children and teenagers, with

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No. 688014.

their parents, and a music festival where most visitors were adults. All recorded data were cleared between the two events, so that the robot would have to start learning from scratch again.

III. Descriptives

The system ran for fourteen days at the science museum, and for three days at the music festival. Table I shows the demographics and number of gestures gathered from each location.

TABLE I
Description of Datasets

	Science museum	Music festival
Participants	294	116
Gender	147 Male 141 Female 6 Unknown	49 Male 67 Female
Average age (years)	12.8 (SD = 10.7) 10 unknown	28.3 (SD = 8.7) 2 unknown
Countries	26	4 (1 unknown)
Number of gestures	2,524	1,000

The recorded gestures were stored in the form of a CSV file containing the 3D coordinates of the participant's tracked joint positions, sampled at approximately 30 frames per second from the Kinect camera, as well as a movie file containing a 2D render of the gesture (Figure 2). Furthermore, gestures can be linked to participants and their demographic information by a unique identifier.

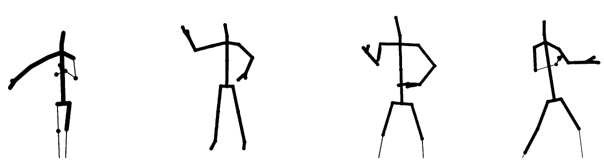

Fig. 2. Four examples of recorded gestures for 'guitar' — first and second are by children, third and fourth by adults.

A. Recognition and Production Performance

After both experiments had finished, we analyzed how the robot's gesture recognition rate developed through time. The results from the science museum are shown in Figure 3. At the music festival, the recognition rate started at 16.37% on the first day, followed by 23.36% and 23.24% on the second and third days. In all cases, the robot performed well above chance (which was approximately 3%). The comprehensibility of the robot's gestures was measured by looking at the number of times participants managed to guess correctly. Figure 3 presents the results from the science museum. During the first day of the music festival, participants managed to guess correctly 50.31% of the time, followed by 51.65% on day two and 50.65% on the last day. This is also above chance (which was 25% for a first attempt, 33% for a second attempt).

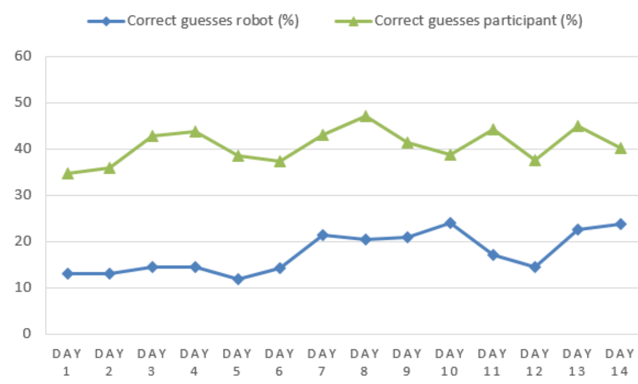

Fig. 3. The robot's and participants' performance (% guessed correctly) during the fourteen days at the science museum.

IV. Conclusion and Discussion

This paper presents an exploratory study where a game of charades was used as a playful method to allow a robot to optimize its own gesture production and recognition abilities. At the same time, an extensive and varied dataset was recorded, to allow future research into gestures, with applications in the field of HRI. We intend to conduct further analyses on the recorded gestures (e.g., which strategies were used, whether these changed between first and second attempts, differences between participant groups), and aim to further improve the robot's ability to recognize and produce gestures. It is difficult to interpret the gesture recognition performance of the system, because existing research tends to work with a smaller set of concepts, and often focuses on detecting a certain predefined gesture, rather than allowing the person performing it to decide on a strategy themselves. However, it does appear that the performance flattens out with a relatively sparse set of data, which can be seen as an indication that we have not reached the maximum potential yet. We would be interested in measuring human performance on recognizing the gestures, to get an idea of the gold standard. The dataset of gesture recordings, as well as the source code of the system will be made publicly available after our further analyses are complete.

References

[1] P. Bremner, A. G. Pipe, C. Melhuish, M. Fraser, and S. Subramanian, "The effects of robot-performed co-verbal gesture on listener behaviour," in *Proceedings of the 11th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*. IEEE, 2011, pp. 458–465.

[2] J. de Wit, T. Schodde, B. Willemsen, K. Bergmann, M. de Haas, S. Kopp, E. Krahmer, and P. Vogt, "The effect of a robot's gestures and adaptive tutoring on children's acquisition of second language vocabularies," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2018, pp. 50–58.

[3] D. McNeill, "So you think gestures are nonverbal?" *Psychological Review*, vol. 92, no. 3, pp. 350–371, 1985.

[4] B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robotics and autonomous systems*, vol. 57, no. 5, pp. 469–483, 2009.

[5] M. E. Cabrera and J. P. Wachs, "A human-centered approach to one-shot gesture learning," *Frontiers in Robotics and AI*, vol. 4, p. 8, 2017.