If Looks Could Kill: Humanoid Robots Play a Gaze-based Social Game with Humans

Oskar Palinko, Alessandra Sciutti, Yujin Wakita, Yoshio Matsumoto, Giulio Sandini

Abstract— **Gaze plays an important role in everyday communication between humans. Eyes are not only used to perceive information during interaction, but also to control it. Humanoid robots on the other hand are not yet very proficient in understanding and using gaze. In our study we enabled two humanoid robots to perceive and exert gaze actions. We then performed a pilot experiment with the two android robots playing the "Wink Murder" game with human players. We demonstrate that the designed framework allows the robots to complete the game successfully, validating the efficacy of our gaze tracking system. Moreover, human participants exhibited a rich variety of natural behaviors in the game, suggesting that it could represent a valid scenario for a more in-depth investigation of human-humanoid interaction.**

I. Introduction

Gaze is one of the basic ways people communicate with each other. Often times gaze interaction is not explicit but rather implicit: we are not always aware that besides acquiring visual information (sensing) we are also transmitting information with our eyes (acting). An observer could for example tell what we are interested in by noticing where we look.

Robots, especially humanoid robots, could become more natural by being able to understand the gaze of humans. This way they could become more aware of the intention of people during interaction. Due to their resemblance to people, humanoids might even be expected to understand the gaze of others, as this is something humans are able to do naturally.

The general problem we are addressing is that human-humanoid interaction is not as natural as human-human interaction. We hypothesize that naturalness could be improved by exploiting gaze reading by the robot. Our goal was to improve our gaze tracking algorithms and to design a gaze tracking scenario in which we could employ this eye tracking solution. The requirements for this scenario were: a) to include multiple humans and multiple robots and b) to be engaging enough for human so they would not become bored quickly. These criteria would ensure a wealth of diverse gazing behavior. We decided that a social game would be the ideal scenario which would fulfill the requirements. Therefore we chose the so-called "Wink Murder" party game as the basis for our experiments. This game can be played by multiple humans and multiple robots, thus it can produce complex gazing behavior. This game also employs gaze detection as well as gaze actions (selection with gaze, mutual gaze) which is an added benefit. For the most part the game does not require verbal communication, thus it is easy to port it to different cultural groups. To the best of our knowledge, such a complex game with multiple humans and multiple humanoid robots was not yet implemented anywhere, thus using this setup could generate novel results, as we also show in our pilot experiment. Our main goals were to create a natural gazing scenario to validate our gaze tracking system and to explore human gazing behavior in interaction with androids.

* Research supported by the European Project CODEFROR (PIRSES-2013-612555).

Oskar Palinko, Alessandra Sciutti and Giulio Sandini are with the Robotics, Brain and Cognitive Sciences Department, Istituto Italiano di Tecnologia, Genova, Italy (e-mail: [firstname].[lastname]@iit.it).

Yujin Wakita and Yoshio Matsumoto are with the Robot Innovation Research Center, National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan (e-mail: yoshio.matsumoto@aist.go.jp).

II. Background

As mentioned before gaze recognition is particularly interesting for humanoid robots, as it is a natural human ability that might be expected from human-like robots. There are two aspects of gaze behavior considering robots: gaze generation and gaze understanding. The first approach deals with how robots should display their own gazing. A good overview of this field can be found in [1]. Our work focuses both on humans reading the robot's gaze but more importantly on how robots could understand people's gaze in order to improve their natural behavior. In this field there is a number of observed phenomena which are closely related to observing the eye: mutual gaze [2], joint attention [3], gaze aversion [4], etc. Mutual gaze happens when the gaze of two agents interlock, which has a special importance in human communication. Joint attention is when two people focus their on attention on a single object, which might be an important interaction cue. Gaze aversion happens when the agent diverts its gaze to gain time in a conversation.

Many authors in the field of human-robot interaction have substituted eye gaze with its closest proxy, head orientation, because of technical limitations [5]. Gaze was deemed difficult to calculate or just not available to the authors [6]. Indeed, often times the head and eye orientations are in line, but not always. The eyes definitely carry additional information that head direction does not possess [7].

Furthermore a number of authors did use actual eye tracking devices to analyze the humans' gaze, most of them used head-mounted systems. Broz et al. looked at human-human interaction for designing more natural human-robot interaction [8]. Sciutti et al. studied how humans follow movements of the robot and compared it to following human movements [9]. Although, as mentioned before, these systems provide high accuracy results, their drawback is in the fact that they are obtrusive for everyday interaction, thus cannot be used in HRI studies where the focus is on the naturalness of communication.

A few groups looked at using remote eye tracking systems built into humanoid robots. Matsumoto and Zelinsky described a design that allowed the estimation of the gaze using a remote system [10]. Ido et al. implemented this system in the HRP2 humanoid robot [11]. It appears that this approach was not used extensively to study human-humanoid interaction. Palinko et al. implemented a remote gaze tracker on the iCub humanoids robot and used it in a collaborative tower building task [12]. Our scenario is based on this implementation of the gaze tracking system.

Considering engaging social interaction scenarios with humanoid robots and multiple humans the research field is less populated. Vazquez et al. used the "chest of drawers" robot, Chester, to play the Mafia game, but only in a Wizard of Oz fashion [13]. Leite et al. used the Keepon mini-robot to explore individual versus group HRI [14]. They found that multiple robots could be beneficial for social interaction with children and that behaviors and social signals can be significantly different when interacting in groups rather than individually with a robot. Crick and Scassellati used a small robot-controlled toy truck to facilitate playing social games between human participants [15].

More specifically looking only at using humanoid robots, Sheiki employed the miniature humanoid Nao as an art exhibit guide who interacted with two humans at the same time [16]. The robot was able to manage a conversation by knowing the head pose of the humans. Mutlu researched humanoid robot (either Asimo or Robovie) communicating verbally and using gaze with two human participants in a storytelling and a tourist agent scenario [1]. In one of his scenarios the humans liked the robot more when Robovie paid more attention to them.

Looking at scenarios where robots played a social game with a human participant, Gori et al. implemented a gesture recognition game which was played between an iCub humanoid robot and one human participant [17]. Both of them needed to memorize sequences of hand gestures and repeat them. Bentivegna et al. taught a humanoid robot to play air hockey against a human, but they didn't report on a naïve human subject study using this setup [18]. Ito and Tani used a small Sony humanoid playing an imitation game to explore joint attention and turn taking [19].

To the best of our knowledge, there has not been yet a study which would employ more than one human subject and more than one humanoid robot in a social game scenario, previous to our work.

III. Approach

This section will describe the android platform, gaze tracking on the robots, the rules of the game as well as the details of the experimental setup.

A. *The android robots*

In our approach we used the Actroid F humanoid robot platform. These robots are quite suitable for our research purpose because of their very human-like appearance and behavior, see Figure 1.

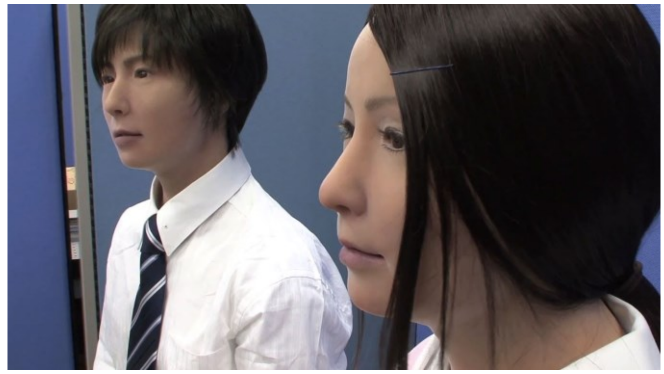

Figure 1. The two Actroid F humanoid robots.

These robots' skin is made of silicon rubber. The faces are made as a verbatim copy of an unnamed person's face. The male and female versions both have the same facial features. They are made different by applying facial hair, wigs and cosmetics. The robots have 13 degrees of freedom: one in the right arm, one in the torso and the rest in the neck and head. The robots can exert facial expressions like happiness, anger, surprise, etc. by moving their lips and eyebrows. The robots are pneumatically driven. The degrees of freedom are actuated using air powered servos. The servos are controlled in position from a remote computer. Due to the usage of pneumatic actuation, the robots are physically compliant, thus can produce human-like movement.

B. *Cameras*

The robots have built-in cameras in both of their eyes. These are fixed focus CMOS devices (NCM13-J) with a horizontal field of view of about 45 degrees. They can provide images of VGA resolution (640x480 pixels) at 30 frames per second or at SXGA resolution (1280x1024) at 10 frames per second. The cameras fit right in the pupil opening of the robots' eyes.

To be able to do gaze tracking of humans at an interaction distance of about 80-100cm higher resolution images are of great benefit, due to the small apparent size of the humans' eyes in the camera images. Unfortunately, when used in SXGA resolution, the robot cameras only provide 10 frames per second with considerable motion blurring due to extended exposure times needed for the higher resolution. Also the robot cameras are located behind a plastic cover of the eyeball, which creates additional blurring of the image. Even though ideally we would have wanted to use these cameras in the robot's eyes, because of the above technical reasons we opted for an outside webcam to provide the images, namely the Logicool C920 (equivalent to the Logitech C920 outside of Japan). These cameras have a very wide horizontal field of view (~80 degrees), with a full HD resolution (1920x1080) providing 30 frames per second due to its built in H.264 hardware encoding.

C. *Gaze tracking*

Gaze tracking of the human subjects was done using the approach we developed in [12]. We introduced a couple of modifications to the original algorithm. Namely, for tracking the corners of the eyes instead the algorithm implemented in the *dlib* library [20] and described in [21], we opted for

CLM [22]. This was done to speed up the execution as *dlib* is computationally expensive and we needed to run the whole algorithm from a single computer. Since we already use CLM for head pose tracking we just used its eye corner detections too, even though they appear to be less precise than *dlib's*. Differently than in [12] but similarly as in [23] we performed the CLM algorithm only on a cutout of the whole image where the face was found in the previous frame. As we were tracking two faces this time, if we lost one of them, we reverted to search for the face on either the left or right half of the image. This windowing also sped up the approach but performing it twice (for two subjects) slowed it down compared to our previous solution [23]. In the end the whole algorithm was performing at around 10 frames per second, which proved to be a usable speed.

As an additional improvement of our gaze tracker, we improved our iris center detection approach, see Figure 2. First we started off with a greyscale image of the eye area. Then we performed blurring of the image using a uniform filter with a width of the estimated iris radius. Iris radius was estimated by assuming an inter-pupillar distance of 65mm and an iris diameter of 12mm [24]. We selected the darkest point of the blurred image as the initial guess of the iris center. Then we performed directional Sobel edge detection to find the outline of the iris, knowing that the transition from the iris to the sclera needs to be a dark to light change. Once the edges were found we applied an own version of the circular Hough transformation with the previously estimated iris radius size to refine the center point recognition.

D. Rules of the game

The original Wink Murder party game has several variations of gameplay. We opted for one which was the easiest for implementation and later data analysis. The practical minimum for playing this game is 4 participants. With 3 participants the game ends very quick, so it does not provide much data to be analyzed, nor it is fun for the players. At the start of the game the players are dealt one card each. Only one of the cards contains an 'X' which designates the "murderer" (also called "villain"). The card should not be revealed to anyone else other than the player who drew it. The other three players who get blank cards are the "innocents" who are also acting as "detectives" at the same time. In the original game the villain "murders" the innocents by winking at them, but since the Actroid F is not able to wink (both eyelids are controlled via one actuator), we needed to find a different action. Blinking was considered but rejected, because human players cannot be expected to withhold their natural blinks for the purpose of the game. Thus we opted for a quick raising and lowering of both eyebrows as a "murder action". This can be performed by both the human and android participants easily without interfering with normal behavior.

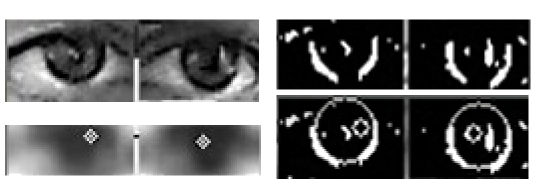

Figure 2. Upper left: original image of eyes; lower left: blurred image; upper right: directional Sobel edge detection; lower right: circular Hough transformation.

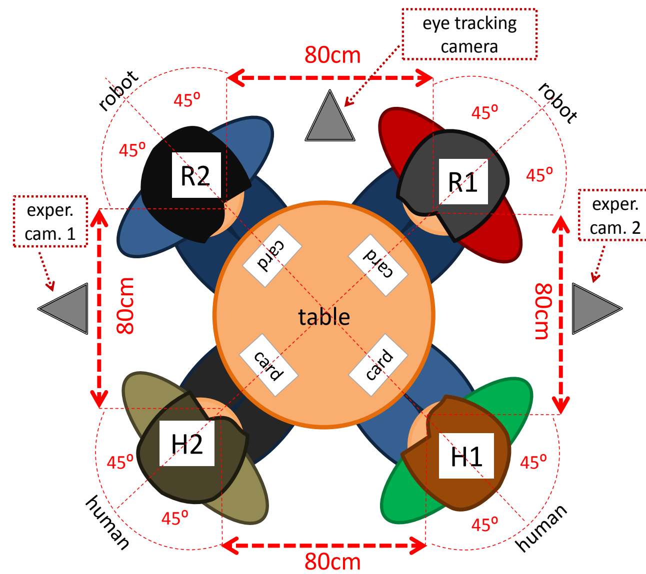

Figure 3. Game setup.

The villain can "murder" an innocent only when there is mutual gaze between them, of course. Once an innocent is "murdered" he needs to continue playing the game for a few more seconds, so that other players can't figure out the identity of the villain because of a sudden reaction. After this time the player needs to pitch his head downwards, signaling that he is out of the game. An innocent can win the game if she witnesses "an act of murder" in which case she is obliged to call out the villain. If she makes a wrong accusation, the game is ended and the accuser loses one point. The game practically ends when there is only the villain and one innocent left, as in this case the latter cannot perpetuate the game by not gazing at the villain.

E. Experiment design

The game is played by two androids (R1 and R2) and two human participants (H1 and H2). One experimental session consists of 12 games played in a pseudo-random order, making sure that each player is the villain 3 times. The physical setup of the experiment can be seen in Figure 3. A photo of the setup can be seen in Figure 4.

The physical location of the androids and humans stays the same throughout the experiment, thus R1 is always the female android, while R2 is the male and the humans don't

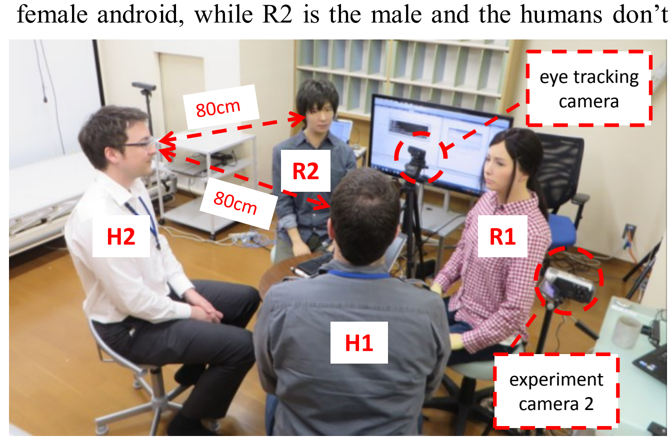

Figure 4. Photo of the experimental setup.

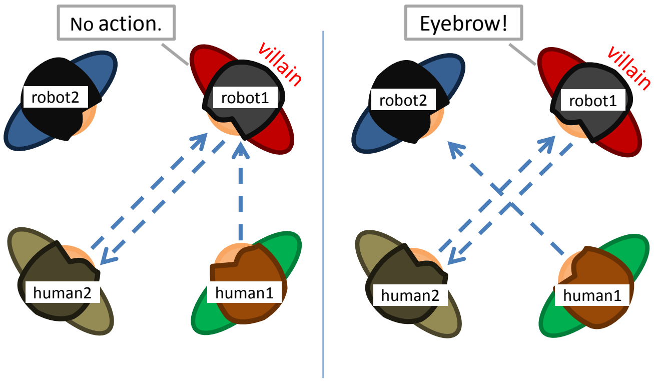

Figure 5. Android strategy as villain.

switch places. As Figure 3 depicts, the four participants are seated at a round table 90 degrees apart. They are all facing the center of the table. The approximate distance between the faces of the neighboring participants is around 80cm. A folded piece of paper is located in front of each participant. These are used to designate the villain. They are pseudo-shuffled and replaced after each game by the experimenter. The eye tracking camera is mounted on a tripod and placed between the two androids facing the two humans. Its position was chosen so that both humans appear in its field of view, as close as possible to the players. Since the C920 webcam is wide field of view, it can be placed close to the setup, thus providing better angles for eye tracking of the humans, see Figure 4. The TV screen seen in the same figure is only there for debugging purposes and is turned away for the actual experiments. The experimenter sits behind this screen, thus not interfering with the flow of the games. The experimenter sees the video feed from the webcam on the screen, so he can control the start and end of each game.

The whole experiment was recorded with two separate camcorders for analysis purposes ("experiment camera 1" and 2 in Figure 3).

F. Android head and eye movements

During each game the androids were randomly moving their gaze with independent timing. The fixations were uniformly distributed between 2 and 4 seconds. The only three focuses of the robots' attention were at the face of the three other participants. They changed their gaze either by rotating their eyes only (-45, 0, 45 degrees) or in combination with head rotation. In this case the robots turned their heads to -22.5 or 22.5 degrees.

G. Android strategy as villain

When one of the androids was the villain, we provided it with some potential advantages over the human players. First, as an outside eye tracking camera was used, the robot knew the gaze direction of both humans at the same time, regardless of where it was looking. This was done due to the fact that humans can also detect eyebrow movements even using their peripheral vision, thus we wanted to be at least as good as humans. Second, the non-villain robot did not call out the villain robot even if it saw it "murdering" someone. This was done due to the fact that our camera was not wide fielded enough to track all three other subjects, so we made the untrackable subject (the other robot) passive. The robots were also not programmed to eliminate the non-villain robot.

The main strategy of the villain android started with classifying the humans' gaze into three categories: looking at either of the other three participants. Once this was known based on the gaze tracking algorithm, the robot acted as follows: if the robot was looking at a human who was looking back at it, it would exert the eyebrow movement if the other human was looking somewhere else, see Figure 5. If both humans were looking at the villain-robot, then it did not perform the action, in order not to reveal itself. The androids did not look at a human on purpose, but were following a random gaze and performed the "kill" action only when the above conditions were met. This was done so that the games would not end too quickly.

H. Android strategy as innocent/detective

When one of the humans ended up being the villain, again, the robots were programmed to know that it was not the other robot, as we could not track three subjects. On the other hand, even though they were able to track both humans, each robot was programmed to detect eyebrow movements only on humans whom they were looking at. E.g. if one of the humans was trying to "kill" the other one and the eye tracking camera detected this, the action would go undetected if the robots were looking at each other at this time. The eyebrow movement itself was visually detected using the CLM algorithm. We compared any potential eyebrow movement with the movement of the mouth, as reference. If the mouth did not move much but the eyebrows did, it was declared as a "kill action". Unfortunately this was ambiguous with head pitch movements, so we needed to find a threshold which would minimize both false positives and false negatives. We did so by testing our system on a couple of pre-pilot subjects. Once all conditions were met for "kill" detection, the robot called out "Stop! You are the murderer" while looking at the person who was detected to be the culprit. See the accompanying video for examples of these actions.

I. Software component setup

The system's software components and their connections are shown in Figure 6. The webcam's image is fed to the CLM face feature and head pose detector. The algorithm expects to see two faces. Once it recognizes them, tracking

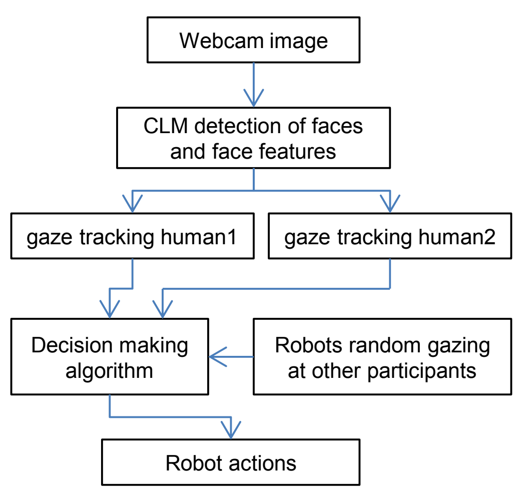

Figure 6. Software components.

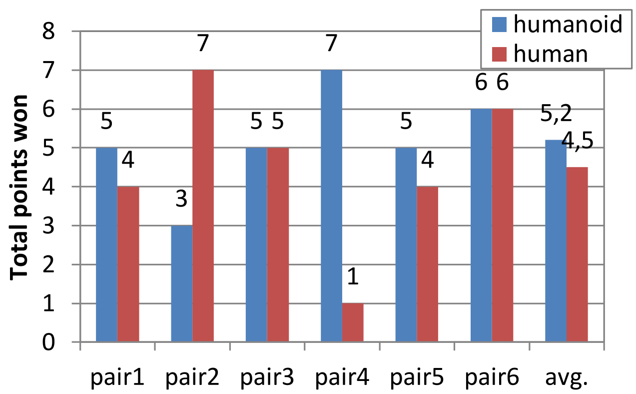

Figure 7. Total points won in experiments by humans and humanoids.

is continued on small areas where the faces are detected in the previous frame. Gaze detection relies on the head pose and eye corner detection of CLM. Once the gaze of both humans is detected it is fed to the decision making algorithm, which acts according to the description under subheadings 2.G and 2.H, above. The robots are randomly looking at either of the three other participants as mentioned before, which is also an input to decision making. Robot actions (eyebrow movement if villain and calling out a villain when innocent) are generated as the output of decision making.

IV. Pilot Experiment Results

The pilot experiment was completed by people from our lab: 12 subjects in 6 pairs. Four of them were female while 8 were male. Five pairs of subjects were Caucasians while one pair was Japanese.

Overall results of the pilot experiments can be found in Figure 7 while details are in Table 1. The total results show that robots won a grand total of 31 games, while humans won 27 games. On average, robots won 5.2 while humans won 4.5 games. Looking at totals for each pair, it can be seen that three experiments were won by robots, one by humans and two experiments were tied. In Table 1 columns represent pairs of subjects, while rows show outcomes of games. The numerator in the results is the actual number of games won, while the denominator represents how many occasions the agent had to win the game. In the last row, humans had twice as many opportunities to win by detection, because the robots could detect only humans (by design), while humans could detect both robots and the other human as the "murderer".

Table 1. Pilot experiment results: [number_of_wins]/[number_of_opportunities_to_win]

	Pair1	Pair2	Pair3	Pair4	Pair5	Pair6
Robot won by "murder"	4/6	1/6	4/6	6/6	3/6	3/6
Human won by "murder"	0/6	1/6	1/6	1/6	0/6	0/6
Robot won by detection	1/6	2/6	1/6	1/6	2/6	3/6
Human won by detection	3/12	7/12	4/12	0/12	4/12	6/12

According to post-experimental discussion with participants, it was revealed that subjects felt quite engaged in the experiment and thought it was fun. One pair reported that after the experiment they thought the robots were "much smarter" compared to what they thought before seeing the robots perform. Only Pair2 reported noticing that the robots "are working together" while others did not notice this even after being asked about it.

It is interesting to notice also the fact that only the pair which involved Japanese participants (Pair4) did not call out anyone (either human or robot) to be the villain. All other pairs did so at least 3 times.

V. Discussion and Future Work

The results of our pilot experiment show that the proposed gaze tracking system is appropriate to enable a natural interaction in the context of a social game as the Wink Murder. Indeed, the robots managed to win as villains the majority of the times (about 58%), suggesting that they could effectively monitor their human partners' gaze. One reason why humans were not able to win more games as villains can be the fact that the robots were not proficient at detecting the "kill actions" of humans (the raise of the eyebrows), see Section 2.H. This prevented humans from directly winning the games (row 2 of Table 1). On the other side, humans' results are boosted because in some situations they called out the robots when they were trying to "kill" them. Humans thought that the robots were trying to eliminate the other player, i.e. they didn't detect mutual gaze with the robot and that's why they called out the villain (this was the case especially with Pair2 and Pair5). This might have been caused by the fact that in some situations the head movements of the robots were performed somewhat slower than the eye and eyebrow movements. The slower head movement was mostly unpredictable due to the nature of the pneumatic actuation system. In future implementations this technical issue needs to be addressed either by making sure the head turns are executed before the eyebrow action, or by eliminating head movements in favor of only eye movements altogether. The first solution might be problematic because there is no feedback mechanism in the robots which would inform us that the desired position is achieved. The latter solution might not be ideal because performing only eye movements with a still head looks quite unnatural.

As mentioned in the previous sections, the game was set up so that robots have the upper hand in two ways:

- The robots had foveal (sharp) vision of both human subjects at all times. Humans can only have foveal vision on one subject, while the other subject is seen peripherally (blurred).
- The two robots collaborated passively with each other by not calling out each other.

Both these benefits for the robots were caused mostly by technical limitations of our system. Namely, the fact that we had to use an outside camera for gaze tracking, because the in-eye cameras of the robots were not performing well enough for our purposes. In future improvements of our system we plan to fix these deficiencies to create an even playing field for both humans and robots. This might be difficult due to the fact that the space for installation of the camera in the robots' eyeball is quite limited. On the other hand thanks to the recent advancements in miniature camera technology for mobile phone applications, it might be possible to find replacements for the original sensors and optics.

Notwithstanding the current technical limitations, this framework allowed for an engaging interaction which evoked different behaviors and strategies in the participants. For example some people worked on eliminating the other human first and then focused on the robots. Other villains on the other hand tried to reassure their co-player that they were also innocent by not performing the "kill action" right away. For the future, this type of multi-robot multi-human interactive scenario could be exploited to investigate natural coordination and collaboration mechanisms in naturalistic contexts. In particular it will be possible to extend the current understanding of phenomena supporting joint action, as for instance automatic imitation [25], going beyond the simplified, dyadic experimental settings usually adopted.

Since the only Japanese pair of subjects we tested (Pair4) seemed reluctant in calling out others being villains, we wonder if this is a behavior which might be due to cultural differences. Further studies are needed to determine the effect of culture on social games played with robots. Another field of future work could be to compare if an appearance-wise very human-like platform (Actroid F) would be better at playing the Wink Murder game than a humanoid which looks a bit less like a human (e.g. iCub).

VI. Conclusion

Since eye gaze communication is very important in human-human interaction we think humanoid robots should also possess the ability to read others' gaze and communicate their own, to make them more natural, i.e. more human-like. In order to study this ability of gaze reading in human-humanoid interaction, "clever" scenarios need to be devised which would enable natural and diverse gaze behavior of the human participants. It is also needed to keep participants engaged and motivated in completing the experiment task. We think that the Wink Murder social game provides such a platform which could be even further refined. Our main contributions are the improvements of the gaze tracking algorithm and experiment design that allows this game to be played among humans and humanoid robots.

Acknowledgement

The authors would like to thank James Adrien Wright for his help in setting up and testing the game scenario.

References

[1] B. Mutlu, "Designing gaze behavior for humanlike robots," 2009.

[2] N. George and L. Conty, "Facing the gaze of others.," *Neurophysiol. Clin.*, vol. 38, no. 3, pp. 197–207, Jun. 2008.

[3] F. Kaplan and V. V. Hafner, "The challenges of joint attention," *Interact. Stud.*, vol. 7, no. 2, pp. 135–169, 2006.

[4] L.-P. Morency, C. M. Christoudias, and T. Darrell, "Recognizing gaze aversion gestures in embodied conversational discourse," *Proc. 8th Int. Conf. Multimodal interfaces - ICMI '06*, p. 287, 2006.

[5] M. W. Doniec, G. Sun, and B. Scassellati, "Active Learning of Joint Attention," in *Humanoid Robots, 2006 6th IEEE-RAS International Conference on*, 2006, pp. 34–39.

[6] H. Kim, H. Jasso, G. Deák, and J. Triesch, "A robotic model of the development of gaze following," in *2008 IEEE 7th International Conference on Development and Learning, ICDL*, 2008.

[7] A. Borji, D. Parks, and L. Itti, "Complementary effects of gaze direction and early saliency in guiding fixations during free viewing.," *J. Vis.*, vol. 14, no. 13, p. 3–, Jan. 2014.

[8] F. Broz and H. Lehmann, "Mutual gaze, personality, and familiarity: Dual eye-tracking during conversation," *RO-MAN, 2012 …*, 2012.

[9] A. Sciutti, A. Bisio, F. Nori, G. Metta, L. Fadiga, and G. Sandini, "Anticipatory gaze in human-robot interactions," in *Gaze in HRI from modeling to communication workshop at the 7th HRI Conference, Boston, Massachusetts, USA*, 2012.

[10] Y. Matsumoto and A. Zelinsky, "An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement," in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, 2000, pp. 499–504.

[11] J. Ido, Y. Matsumoto, T. Ogasawara, and R. Nisimura, "Humanoid with interaction ability using vision and speech information," in *IEEE International Conference on Intelligent Robots and Systems*, 2006, pp. 1316–1321.

[12] O. Palinko, A. Sciutti, F. Rea, and G. Sandini, "Eye Gaze Tracking for a Humanoid Robot," in *IEEE/RAS International Conference of Humanoids Robotics*, 2015.

[13] M. Vazquez, A. Steinfeld, and S. E. Hudson, "Parallel detection of conversational groups of free-standing people and tracking of their lower-body orientation," in *IEEE International Conference on Intelligent Robots and Systems*, 2015, vol. 2015-Decem, pp. 3010–3017.

[14] I. Leite, M. McCoy, D. Ullman, N. Salomons, and B. Scassellati, "Comparing Models of Disengagement in Individual and Group Interactions," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, 2015, pp. 99–105.

[15] C. Crick and B. Scassellati, "Intention-based robot control in social games," in *Proceedings of the Cognitive Society Annual Meeting*, 2009.

[16] S. Sheikhi and J.-M. Odobez, "Combining dynamic head pose–gaze mapping with the robot conversational state for attention recognition in human–robot interactions," *Pattern Recognit. Lett.*, 2014.

[17] I. Gori, S. R. Fanello, G. Metta, and F. Odone, "All gestures you can: A memory game against a humanoid robot," in *IEEE-RAS International Conference on Humanoid Robots*, 2012, pp. 330–336.

[18] D. Bentivegna, A. Ude, C. Atkeson, and G. Cheng, "Humanoid robot learning and game playing using PC-based vision," *Int. Conf. Intell. Robot. Syst.*, vol. 3, pp. 2449–2454, 2002.

[19] I. Masato and J. Tani, "Joint attention between a humanoid robot and users in imitation game," *Proc. Thrid Int. Conf. Dev. Learn.*, 2004.

[20] D. E. King, "Dlib-ml : A Machine Learning Toolkit," *J. Mach. Learn. Res.*, vol. 10, pp. 1755–1758, 2009.

[21] V. Kazemi and J. Sullivan, "One Millisecond Face Alignment with an Ensemble of Regression Trees," in *Computer Vision and Pattern Recognition (CVPR), 2014*, 2014.

[22] T. Baltrusaitis, P. Robinson, and L. P. Morency, "3D Constrained Local Model for rigid and non-rigid facial tracking," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2610–2617.

[23] O. Palinko,F. Rea,G. Sandini, A. Sciutti, "A Robot Reading Human Gaze: Why Eye Tracking Is Better Than Head Tracking for Human-Robot Collaboration," in *Proceedings of 2016 Conference on Intelligent Robots and Systems (IROS)*, 2016.

[24] C. C. Gordon, T. Churchill, C. E. Clauser, B. Bradtmiller, and J. T. McConville, "Anthropometric survey of US army personnel: methods and summary statistics 1988," 1989.

[25] A. Bisio, A. Sciutti, F. Nori, G. Metta, L. Fadiga, G. Sandini & T. Pozzo, "Motor Contagion during Human-Human and Human-Robot Interaction, " *PloS one*, vol. 9, no. 8, e106172, 2014.