Motion-Oriented Attention for a Social Gaze Robot Behavior

Mihaela Sorostinean, François Ferland, Thi-Hai-Ha Dang, and Adriana Tapus

Computer Vision and Robotics Lab, Computer Science and System Engineering Department, ENSTA-ParisTech, Palaiseau, France
{firstname.lastname}@ensta-paristech.fr

Abstract. Various studies have shown that human visual attention is generally attracted by motion in the field of view. In order to embody this kind of social behavior in a robot, its gaze should focus on key points in its environment, such as objects or humans moving. In this paper, we have developed a social natural attention system and we explore the perception of people while interacting with a robot in three different situations: one where the robot has a totally random gaze behavior, one where its gaze is fixed on the person in the interaction, and one where its gaze behavior adapts to the motion-based environmental context. We conducted an online survey and an on-site experiment with the Meka robot so as to evaluate people's perception towards these three types of gaze. Our results show that motion-oriented gaze can help to make the robot more engaging and more natural to people.

1 Introduction

During the last decade, a lot of research has focused on examining the effect of a robot's gaze while interacting with a human, since it is known that gaze plays an important role in human-human communication [1,2]. It has been found that enriching a robot with a human-like gaze behavior helps it to be perceived as more intelligent and social [3]. Furthermore, the attention that the robot is showing to the person in the interaction has great influence in the way a person understands the robot's messages [4].

A possible way to achieve more natural and human-like behavior in interaction tasks is to design a distributed robot attention system that allows the robot to be distracted from the main interaction by external events. In a human-human interaction setting, such events could include another person entering the room or the noise of an object hitting the ground. While such events should not disturb the interaction and communication process, it is entirely expected that people would momentarily shift their attention through gaze to the external disturbing factors. We believe that robots should behave similarly in order to exhibit natural, human-like behavior. Previous works suggest that attention disturbance can encourage people to make the communication with a robot more social, and adaptable attention enables the robot to be accepted as an intentional and proactive communication agent [5]. Furthermore, participants in an

M. Beetz et al. (Eds.): ICSR 2014, LNAI 8755, pp. 310–319, 2014.

online survey showing videos of a receptionist robot believed that it was more human-like when it exhibited random gazes in conjunction with person-directed gaze [6]. In [7], a multimodal, saliency-based bottom-up attention system has been used on the iCub humanoid robot to drive its exploratory behavior. The system guided the gaze of the robot to visually stimulating zones (e.g. moving targets, brightly-colored features) and localized sound sources, while inhibition and habituation mechanisms prevented the robot from having its attention fixed at a single target for a long period of time.

The purpose of this research is to discover how a robot embodied with a human-like visual attention system is perceived in a scenario of interaction with a person. The Meka robot plays the role of a receptionist, and interacts with people that come to ask some specific information regarding a meeting with a professor. During the interaction, the robot displays three types of gazing behavior: randomly gazing at people or objects in the background, a fixed gaze at the person in the interaction, and a motion-oriented gaze that focuses both on the person standing in front of the robot and on other possible targets in the background. Our interest is to find out in which of these three scenarios people would feel more comfortable to interact with the robot, and how much the gaze behavior of the robot affects the interaction. Impressions from people will be collected from both a third- and first-person perspective, first from an online video survey and then with an on-site experiment with the Meka robot.

2 Testbench Configuration

The experiments presented in this work have been conducted with the Meka humanoid robot (see Fig. 1). It has been designed to work in human-centered environments. The robot features compliant force control throughout its body, durable and strong hands, and an omnidirectional base with a prismatic lift. The head is a 7 Degrees-of-Freedom (DOF) robotic active vision head with high resolution FireWire cameras in each eye, integrated DSP controllers, and zero-backlash Harmonic Drive gearheads in the neck. Designed for a wide range of expressive postures, it is a platform particularly well suited for researchers interested in human-robot interaction and social robotics. Additionally, we use a separate webcam to augment the peripheral field of view of the robot. When the robot is focusing on a person, his/her body fills an important area of the Meka's eyes cameras field of view, hence rendering the tracking of the background motion difficult. We also use the microphone of this webcam for speech recognition purposes.

In our implementation of motion-oriented attention, which uses the Robot Operating System (ROS), the robot can alternate between two tracking sources for its gaze behavior:

- **OpenNI Tracking** To track the person interacting with the robot, we use an ASUS Xtion Live Pro, integrated into the torso of Meka, and the OpenNI2 ROS Package for full skeleton frames tracking. The head frame of the person,

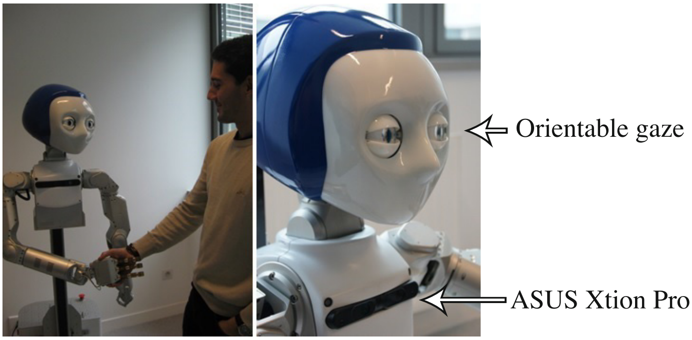

Fig. 1. The Meka humanoid robot

once it is detected, is transformed into the frame of the head of the robot. The origin of this transformed frame is continuously sent to Gaze Control.
- **Motion Tracking** Background motion detection is performed on the webcam camera. Based on [8], semi-dense point tracking is performed to detect and localize motion to the left of the robot. When motion is detected over an empirically-determined threshold quantity, a target is sent to Gaze Control. As the external webcam cannot provide depth information, this target is generated at a fixed, predetermined distance of 2 meters. While this might result in a slight disparity between the gaze target and the actual motion in the scene, this is generally imperceptible from the interacting person's point of view, since the webcam is relatively close to the robot.

Gaze Control generates smooth motion for 6 actuators of the head of the robot: neck pan and tilt, head tilt, and eyes pan (both eyes individually) and tilt. Motion Tracking has priority over OpenNI Tracking. As long as there is no motion in the field of view of the webcam, the robot is looking at the interacting person. However, as soon as motion is detected, the robot focuses on this new target for up to 3 seconds before returning control to OpenNI Tracking. Additionally, speech recognition as well as speech synthesis is performed through an Internet connection by the Google Chrome API.

3 Experimental Design

Figure 2 shows a schematic view of the experimental setup and a screenshot from the video used in the online survey. The Meka robot is positioned behind a tall and narrow desk. The prismatic lift of the robot sets it at a standing height of 1.75 m. The webcam serving the dual purpose of voice capture and motion detection is set on this desk. The person starts the experiment by standing approximately 2 meters from the desk, in front of the robot, for at most 10 seconds, to ensure proper detection by OpenNI. When ready, the experimenter signals the person the beginning of the experiment, and the person walks to the desk and greets the robot (by saying "Hi" or "Hello"). The Meka robot then asks the participant details about the professor's name and the purpose of his/her visit. Thereafter, it provides directions about how to reach the professor's office. The length of one interaction is approximately 1 minute. During this interaction, a person passes in the background for two round trips with the purpose of distracting the robot's attention for at least 4 times. A video (based on sequences from the online survey) showing the interaction with the robot in the three conditions is publicly available online[1].

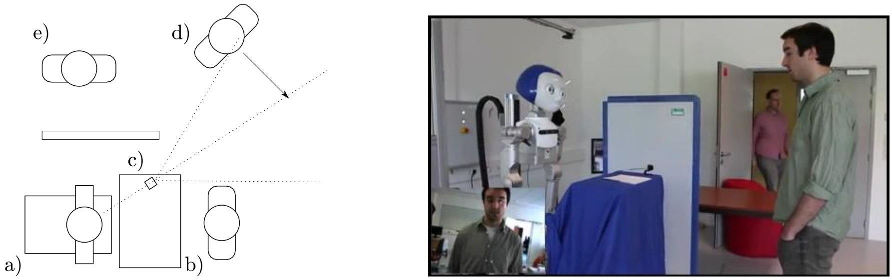

Fig. 2. Experimental setup: (a) the Meka robot, (b) the interacting person, (c) the peripheral view webcam, (d) a person passing in the background, (e) the experimenter. A partition screen prevents the experimenter from accidentally entering the robot's field of view. The robot interacts with the person as someone else is about to enter the robot's field of view. The bottom-left part of the screenshot shows the point of view from the eyes of the Meka robot.

In order to investigate people's perception of the gaze of the robot, we chose to compare three types of behavior. People participating in the experiment were exposed to three corresponding experimental conditions:

- **C1-Random** Random Gaze Attention Robot. The robot looks randomly at the background while talking with the person. The gaze target changed every 6 seconds. This condition was used to highlight if any motion, regardless of frequency or targeting, was perceived as natural.
- **C2-Fixed** Fixed Gaze Attention Robot. The robot is embodied with fixed gaze attention, so its gaze is fixed on the face of the person standing in front of it.
- **C3-Distributed** Motion-oriented Gaze Robot. The robot has a distributed gaze behavior, looking both at the person in the interaction and at other moving targets in the background in a human-like manner.

The experiment was conducted in two identical phases with different sets of participants. For the first phase of the experiment, we created an online questionnaire in which we integrated three videos that illustrate the three gaze behaviors and a set of questions that allowed us to evaluate the people's perception

[1] `http://perso.ensta-paristech.fr/~tapus/eng/media/videos/icsr2014.mp4`

about the behavior of the robot and its attention system. The second phase of the experiment was conducted in the laboratory with a face-to-face interaction between the robot and the participants. For both experimental stages the participants filled-out a pre-experiment general questionnaire with 4 items (age, gender, background, and robotics knowledge) that provided us with demographic information. After each condition, the participants answered 9 questions (with a total of 17 sub-items), all presented on a 7-point Likert scale. To avoid the priming effect of knowing the content of the questionnaire after the first condition, the order in which the three conditions appeared was randomized for each participant. An additional set of 3 questions (with a total of 5 sub-items) were asked at the end of the whole experiment to compare all three conditions. The following questions were relevant to the study in this paper:

- **Q1.** Was the robot's behavior human-like?
- **Q2.** Did its overall bodily behavior contributed to its human-likeness?
- **Q3.** Did its gaze movement contributed to its human-likeness?
- **Q4.** Did its head movement contributed to its human-likeness?
- **Q5.** Was the robot embodied with attention?
- **Q6.** Was the robot expressive?
- **Q7.** Did the robot appeared intelligent?
- **Q8.** Was the interaction engaging?
- **Q9.** Which condition was the most social?
- **Q10.** Which condition was the most natural?
- **Q11.** With the robot of which condition would you prefer interact with?

From the participants' responses to these questions, we wanted to test the following hypotheses:

- **H1.** The motion-oriented gaze robot (C3-Distributed) is perceived as more natural and more human-like than the two other types of gaze (C1-Random and C2-Fixed).
- **H2.** The motion-oriented gaze robot (C3-Distributed) appears more attentive to its environment than the two other types of gaze (C1-Random and C2-Fixed).
- **H3.** The motion-oriented gaze robot (C3-Distributed) appears more expressive to the person in interaction than the two other type of gaze do (C1-Random and C2-Fixed).

4 Experimental Results

4.1 Online Survey

The online questionnaire was completed by 69 individuals (45 male, 24 female, aged 19 to 74, average 30, 66.7% from Romania, 31.9% from Canada, 1.4% from France). To analyze the data from the online questionnaire, a single factor analysis of variance (ANOVA) test was conducted between the values rated on a 7-points Likert scale for each two pairs of random, fixed, and motion-oriented gaze behaviors. Figure 3 shows significant results from the questionnaire (results from Q7 and Q8 were not significant).

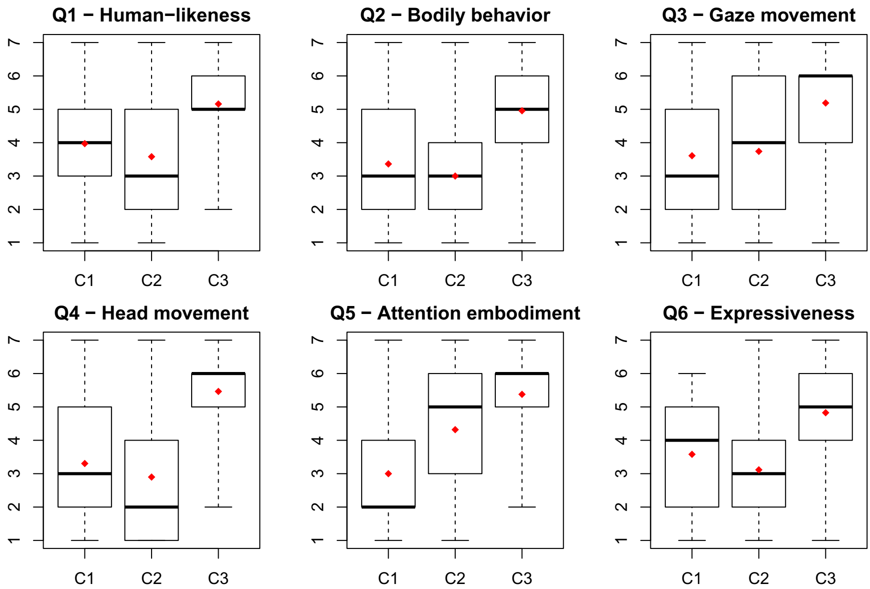

Fig. 3. Online survey Likert-scale results for each condition: C1-Random, C2-Fixed and C3-Distributed. The red dots represent the means.

Validation of hypothesis H1. The hypothesis that the motion-oriented gaze robot (C3-Distributed) would be perceived as more natural and human-like was strongly supported in the online questionnaire. The robot embodied with a motion-oriented gaze behavior was significantly more appreciated relatively to the random gaze and to the fixed gaze for its human-like behavior ($F(2, 204) = 20.18, p < 0.001$, Tukey's Honestly Significant Difference (HSD) comparison revealed significant difference between C1-Random and C3-Distributed ($p < 0.001$) and C2-Fixed and C3-Distributed ($p < 0.001$)). This was also the case for its overall bodily behavior ($F(2, 204) = 24.49, p < 0.001$, Tukey's HSD test results: significant difference between C1-Random and C3-Distributed ($p < 0.001$) and between C2-Fixed and C3-Distributed ($p < 0.001$)), its gaze movement ($F(2, 204) = 15.68, p < 0.001$, Tukey's HSD test results: significant difference between C1-Random and C3-Distributed ($p < 0.001$) and between C2-Fixed and C3-Distributed ($p < 0.001$)), and head movement ($F(2, 204) = 44.31, p < 0.001$, Tukey's HSD test results: significant difference between C1-Random and C3-Distributed ($p < 0.001$) and between C2-Fixed and C3-Distributed ($p < 0.001$)). In Q9, 50.73% of the participants rated the social-attentive robot with motion-oriented attention as more social than the robot embodied with fixed attention (36.23%) or the robot embodied with random gaze behavior (13.04%, $\chi_2 = 35.46$, $p < 0.001$). Moreover, in Q10, 82.6% of participants thought that the social-attentive robot was more natural than the other two robot gaze behaviors (random-7.25% and fixed-10.1%, C1-Random and C2-Fixed, respectively, $\chi_2 = 14.22$, $p < 0.001$).

Validation of hypothesis H2. Our prediction that the participants would rate the robot that looked both at the person in the interaction and at the movement in the background (C3-Distributed) as more attentive was also supported. The participants attributed significantly higher scores for the attention with which the robot in this scenario was embodied, in regard to the other two types of behavior, fixed and random ($F(2, 204) = 32.03, p < 0.001$, Tukey's HSD test results: significant difference between C1-Random and C3-Distributed ($p < 0.001$), between C2-Fixed and C3-Distributed ($p = 0.001$) and also between C1-Random and C2-Fixed ($p < 0.001$)).

Validation of hypothesis H3. Regarding the expressiveness between the three robot behaviors, a significant difference was observed for the motion-oriented gaze robot, as ANOVA returns ($F(2, 204) = 15.36, p < 0.001$, Tukey's HSD test results: significant difference between C1-Random and C3-Distributed ($p < 0.001$) and between C2-Fixed and C3-Distributed ($p < 0.001$)). Furthermore, the participants also noted that the interaction with the motion-oriented gaze robot appeared more engaging ($F(2, 204) = 33.66, p < 0.001$, Tukey's HSD test results: significant difference between C1-Random and C3-Distributed ($p < 0.001$) and between C2-Fixed and C3-Distributed ($p < 0.001$)).

Finally, 66.67% of the participants declared in Q11 that they would prefer to interact with the distributed motion-oriented gaze robot (C3-Distributed), compared to 24.6% and 8.7% for C2-Fixed and C1-Random, respectively ($\chi^2 = 72.74$, $p < 0.001$).

4.2 Robot Real-World Experiments

For the second phase, 21 participants (17 male, 4 female, aged 20 to 57, average 28.3) have been recruited from the university. 12 participants were from France, 2 from Romania, and one from each of these countries: Algeria, Argentina, Canada, China, Italy, Morocco, and Tunisia. Participants received a written copy of the script for the interaction and were asked to follow it as much as possible. They were allowed to refer to this script while interacting with the robot, but were told in advance that the dialogue was not evaluated in this experiment, and that they should focus on the overall behavior of the robot instead of the quality of the speech recognition or the answers received from it, as speech recognition and the dialogue are not central parts of this work. In case of any speech recognition errors, the operator could manually select the answers (in a manner that could not be perceived by the participant) so that the interaction would not be slowed down. As with the online survey result, we used a one-way within groups ANOVA test to analyze the significance of the participants' responses on a 7-points Likert scale. Figure 4 presents significant results from the questionnaire (results for Q2 to Q4 were not significant).

Validation of hypothesis H1. A significant difference was discovered in the human-like perception of the robot between the first and the third condition (ANOVA

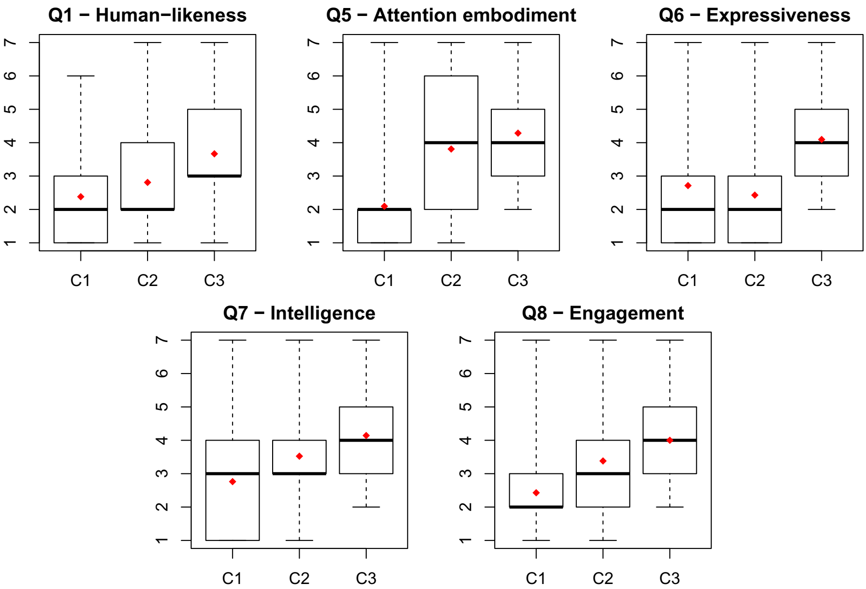

Fig. 4. Live trials Likert-scale results for each condition: C1-Random, C2-Fixed and C3-Distributed. The red dots represent the means.

test: $F(2, 60) = 3.19, p = 0.04$, Tukey's HSD comparison: $p = 0.04$ between C1-Random and C3-Distributed). In Q9, while 52.38% of the participants appreciated the robot with motion-oriented gaze (C3-Distributed) as natural, compared to 23.80% for fixed gaze (C2-Fixed) and 19.01% for random gaze (C1-Random), these results were not significant ($\chi^2 = 4.26$, $p = 0.11$). However, the robot with motion-oriented gaze was significantly perceived in Q10 as more social than the robot embodied with the other two behaviors (52.38% -C3-Distributed, 33.30% -C2-Fixed, and 9.52% -C1-Random, $\chi^2 = 6.07$, $p = 0.04$, one of the participants perceived no natural or social behavior in any of the robots).

Validation of hypothesis H2. The ANOVA test showed a significant difference among the three gaze behaviors in the attention level that participants perceived in each robot ($F(2, 60) = 10.12, p < 0.001$). Tukey's HSD comparisons indicated a significant difference between the robot embodied with random gaze behavior (C1-Random) and the robot embodied with motion-oriented gaze behavior (C3-Distributed, $p < 0.001$) and between the fixed and random behavior (C1-Random and C2-Fixed, $p = 0.003$), but no significant difference was observed between the fixed and the motion-oriented gaze behavior (C2-Fixed and C3-Distributed). Moreover, the robot in C3-Distributed was also considered more intelligent with respect to the one in C1-Random (ANOVA test: $F(2, 60) = 3.43, p = 0.038$, Tukey's HSD comparison: $p = 0.029$ between C1-Random and C3-Distributed).

Validation of hypothesis H3. Our hypothesis that the robot with motion-oriented gaze appears more expressive was supported by the real-world experiment as well (ANOVA test: $F(2, 60) = 5.98, p = 0.004$, Tukey's HSD comparison: significant difference between C1-Random and C3-Distributed ($p = 0.02$) and between C1-Random and C2-Fixed ($p = 0.005$)). Furthermore, the ANOVA test revealed a significant difference on the way the participants perceived the interactions as engaging ($F(2, 60) = 5.19, p = 0.008$), and a Tukey's HSD comparison showed a difference between C1-Random and C3-Distributed ($p = 0.006$).

Finally, when asked in Q11 which robot they would rather interact with, an equal percentage of the participants (42.85%) selected the robot with motion-oriented attention (C3-Distributed) and the robot with fixed attention (C2-Fixed) while 14.3% preferred the robot with random attention. However, these results are not significant enough to prefer one of the conditions between the three ($\chi_2 = 3.41$, $p < 0.181$). To get some insight into these results, the opinions of the participants to the real-world experiment were taken into account. One of the participants told us that he considered the random behavior natural in a robot receptionist because it made the robot look busy, which is a normal behavior for a real life receptionist. Another participant found that the constant switching of the robot's gaze when a person was passing by tended to distract him from the conversation, which was also annoying.

5 Discussion and Conclusion

In this paper, we described an experiment evaluating embodied, socially attentive motion-oriented gaze control for a humanoid robot interacting as a receptionist. We have shown that a behavior distributing attention between the participant in the interaction and the motion in the background contributes in making a robot seem more natural, human-like, and attentive to various people as supported by an online survey and an on-site experiment. From this experiment we noticed that the percentage of people that considered a robot embodied with attentive behavior as more social in respect to the other two behaviors is approximately the same for both the online survey and the direct interaction with the robot. This leads to the conclusion that a receptionist robot should have a motion-oriented attention behavior in order to be perceived as social during an interaction with a person.

Nevertheless, there were differences between the online and the real-world experiments regarding how natural the robot behavior was perceived in each of the three conditions. It can be seen in the results above that in a direct interaction with the robot the percentage of people that preferred the fixed or the random behavior for the robot are considerably higher in comparison with the online survey. This is mainly due to the influence of the social physical embodied presence of the robot to the participants' perception. Furthermore, while the majority of participants in the online survey considered the robot with motion-oriented gaze as the one they would rather interact with in a receptionist-visitor scenario, that is not the case for the on-site experiment, where an equal percentage of people chose the fixed and the motion-oriented behavior.

From the results of this work, we believe that there are several ways of making social gaze for robots that interact with humans. Instead of expecting to have one unique social gaze for the robot in all kind of situations, an adaptive behavior would be more appropriate. In future work, we plan on studying automatically adaptive behaviors in applications such as tour guide robots and assistive care.

Acknowledgements. We would like to thank the ERASMUS program that funded the internship of Mihaela Sorostinean, Fabio Pardo for allowing us to use his code for speech rendering, and Sean Andrist for participating in the first phase of the experiment.

References

1. Exline, R.V., Gray, D., Schuette, D.: Visual behavior in a dyad as affected by interview content and sex of respondent. In: Tomkins, S., Izzard, C. (eds.) Affect, Cognition and Personality, Springer, New York (1965)
2. Kendon: Some functions of gaze direction in social interaction. Acta Psychologica 26(1), 1–47 (1967)
3. Koelemans, D.: The effect of virtual eye behaviour during human-robot interaction. In: Proceedings of the 18th Twente Student Conference on IT (2013)
4. Mutlu, B., Forlizzi, J., Hodgins, J.: A storytelling robot: Modeling and evaluation of human-like gaze behavior. In: Proceedings of the 6th IEEE-RAS International Conference on Humanoid Robots, pp. 518–523. IEEE (2006)
5. Muhl, C., Nagai, Y.: Does disturbance discourage people from communicating with a robot? In: The 16th IEEE International Symposium on Robot and Human interactive Communication, RO-MAN 2007, pp. 1137–1142. IEEE (2007)
6. Holthaus, P., Pitsch, K., Wachsmuth, S.: How can i help? International Journal of Social Robotics 3(4), 383–393 (2011)
7. Ruesch, J., Lopes, M., Bernardino, A., Hornstein, J., Santos-Victor, J., Pfeifer, R.: Multimodal saliency-based bottom-up attention a framework for the humanoid robot icub. In: Proceedings of the International Conference on Robotics and Automation, pp. 962–967. IEEE (2008)
8. Garrigues, M., Manzanera, A., Bernard, T.M.: Video extruder: a semi-dense point tracker for extracting beams of trajectories in real time. Journal of Real-Time Image Processing, 1–14 (2014)