Exploring Non-Verbal Strategies for Initiating an HRI*

Francesco Vigni[1] [0000−0001−9918−8485] and Silvia Rossi[1,2] [0000−0002−3379−1756]

(1) Interdepartmental Center for Advances in Robotic Surgery - ICAROS
University of Naples Federico II, Italy
(2) Department of Electrical Engineering and Information Technologies - DIETI
University of Naples Federico II, Italy
{francesco.vigni, silvia.rossi} @unina.it

Abstract. The growing deployment of robots in social contexts implies the need to model their behaviour as social agents. In this context, the way a robot approaches a user and eventually engages in an interaction is a crucial aspect to take into account for the acceptance of these tools. In this work, we explore how the approaching policy and gaze behaviours can influence the perceived intention to interact before the interaction starts. The conducted user study highlights the importance of the robot's gaze behaviour when approaching a human with respect to its approaching behaviour. In particular, if the robot moves in the surroundings of a human, even not straightforward in their direction, but locks the gaze at them, the intention to interact is recognised clearer and faster with respect to the direct approaching of the user but with an adverse gaze.

Keywords: Approach Policy · Social Behaviour · Gaze · Proxemics

1 Introduction

A typical first encounter with a robot and the subsequent interaction provide important insights to the user on the robot's capabilities. When the latter presents humanoid characteristics, the interaction can elicit social expectations [6, 30]. However, if during the interaction the robot fails to comply with social norms, dissatisfaction can arise [27]. To tackle these issues, research is pushing toward the definition of *natural* Human-Robot Interaction (nHRI) [26], which can enable robots to display and perceive all the modalities used by humans in face-to-face interactions.

The spontaneity of Human-Human Interactions (HHIs) is inherently unconstrained and strongly depends on the nonverbal cues employed by the participants in the interaction. These cues can vary with the participant, the context, and also their willingness in taking part in the interaction. However, it becomes

* This work has been supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 955778.

very hard to keep spontaneity in the nonverbal channel of the encounter if one of the participants is a social robot. The question "How does a social HRI begin, and what factors trigger it?" becomes increasingly important and embeds how the nonverbal behaviours of a robot are perceived by a user.

In this work, we start from the assumption that two participants (a human and a robot) about to start a social HRI continuously exchange social signals in a bidirectional manner that could lead to initiating the interaction. In this sense, the behaviour of both (potential) participants of an HRI can lead to initiating or avoiding the interaction at will. If one of the (potential) participants in an HRI is constrained, the behaviour of the partner is considered responsible for initiating or avoiding, the interaction. This claim is inspired by the work presented in [13], in which authors model an HRI in a multimodal fashion and consider the bidirectional component of the communication channels. Similar policies happen among humans. For instance, if we would like to prevent an interaction to start while walking in a public space, we can use an aversive gaze or deviate our path to avoid a specific person. When translating this behaviour to HRI, we can exploit the anthropomorphism of a robot to design behaviours that can purposefully convey the intention to start or avoid an interaction.

To investigate these concepts, we conducted a user study where the approach policy and gaze of a humanoid robot (Pepper) approaching a standing user were manipulated to convey various degrees of intention to initiate the interaction. Our results show that, given a robot approaching a user, the robot's gaze conveys the social intention quicker and clearer than its motion toward the user.

2 Background and Related Work

With the increasing deployment of robots in social contexts, it is relevant to understand how the interactions evolve. Among the well-known metrics used in HRI, engagement evaluation is crucial as it allows estimating the process or state (according to the chosen definition) of the interaction [16].

Engagement in HRI is a widely studied topic that has its roots in human sciences and exploits recent technologies for building effective and robust tools. In a key work in the field, authors define engagement in HRI as "the process by which individuals involved in an interaction start, maintain, and end their perceived connection to one another" [23]. In a similar fashion, we model the interaction in three separate phases namely: approaching, interacting and terminating. In the first phase, the user might infer the intention to initiate an interaction of the robot but no HRI is ongoing. In the second phase the HRI is occurring, and in the last phase, the social cues of the participants convey that the interaction is terminating.

Metrics like engagement [16] and visual focus of attention (VFoA) [3] can be used to track how the interaction unfolds. For example, the authors in [5] present a dataset (EU-HRI dataset) containing a rich set of data streams obtained from a 56 days experiment performed in the wild, in which a (static) humanoid robot Pepper was used to interact with users and record data from its sensors. In that work, a distance threshold from the fixed standing robot defines the logic for considering the user as part of the HRI. Hence, engagement evaluation is deployed as a metric “during” the interaction while does not provide information on a social interaction that is about to start. Indeed, among humans, social intelligence can be used to understand a partner that seeks an interaction or that would like to terminate it [25]. Social robots should be able to employ, or at least mimic, similar social intelligence. When an interaction with an anthropomorphic robot is about to start, Kendon’s model [11] can be used to define the social robot skills [2] and greetings behaviours [9, 28].

Gaze [1, 24], proxemics [14] and body movements [21] are among the nonverbal features that can be interpreted as social signals and can be used to convey the robot’s intentions. The interpersonal space, relative body pose and mutual gaze [18] can be used to capture a snapshot of the evolution of an HRI. The way these variables develop over time can give us more insights into the dynamics of the interaction. Yet, an orchestrated employment of these in a multimodal fashion is expected to improve natural HRIs [4, 10].

The approaching phase in a social HRI provides a first impression that can be used to deduct social intentions. Research highlights that proxemics and the robot’s body motions in this phase are pivotal for the users’ perception of the robot’s intention [10, 14, 20]. The way a robot approaches a human can be interpreted by the latter in different ways [19]. A fast movement towards the human might elicit fear and discomfort [12]. On the other hand, if the robot approaches the human too slowly, the latter might not understand its intention to interact. Moreover, if the robot embeds anthropomorphic features, the motion of each body part can also affect the interaction. In particular, gaze can be manipulated to convey positive or negative robot’s mental states and intentions during an interaction [1]. Yet, a robot that stares a human during a social interaction is not positively perceived [29]. In [22], authors developed a model that predicts the walking behaviour of a person in the proximity of the robot, plans a path towards them and finally conveys the intention to start a conversation in a nonverbal fashion. In contrast to the design choices of [22], this work constrains the movement of the humans and focuses on how various nonverbal features of the robot can influence the perceived intention to interact only during the approaching phase. The main idea is that, despite the perception strategies that are currently in use in anthropomorphic robots, users might infer the robot’s social intentions by its body expressions from afar.

3 Methods

A 2x2 experimental design was implemented, to investigate the effects of gaze and approach policy of a robot navigating toward a user in a hall. The controlled variables modulate the nonverbal behaviour of the robot and can be interpreted by the users as social signals [4]. For the robot’s gaze, two conditions are chosen by employing a *social* or an *adverse* gaze during the approach phase. In this sense, a *social* gaze consists in the robot employing a face-directed gaze [1] during the approach phase, while an *adverse* gaze consists in the robot gazing the location that is specular to the human's face with respect to the robot's path. Hence, the robot is gazing an empty location in the hall. We assume this last condition to be perceived as asocial as Normoyle et al. [15] link similar behaviours to low trust.

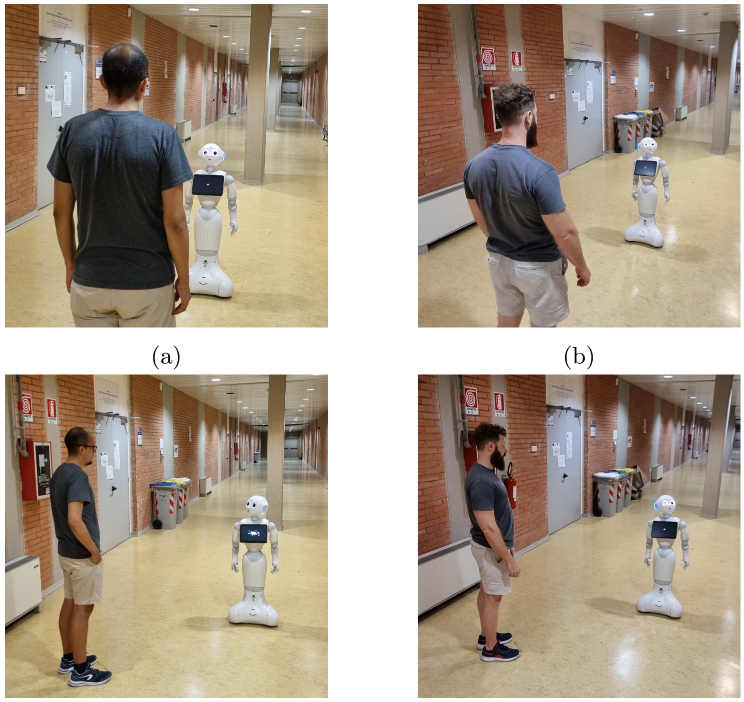
The approach policy is controlled by the position of the standing user in space (front vs side approach). The user can either stand in front of the robot or with a lateral offset from the trajectory at the end of the approaching phase ($30°$ clockwise, see Figure 1). Given the symmetry of the scenario, only one lateral condition is implemented. In both conditions, the user is instructed to face the robot. These two configurations are expected to well mimic a spontaneous encounter with a robot in a public space while still fitting with the social space range defined in [17]. Table 1 shows an overview of the experimental conditions. The conditions are named according to their controlled variables where SF refers to social front, AF to adverse front, SS to social side and AS to adverse side.

An initial pilot study is performed to validate the designed conditions and to improve the survey. The outcome of it allowed us to rephrase some unclear entries in the survey and optimize the experimental protocol. Then a user study is conducted with a within-subject design in which each participant is exposed to the four experimental conditions. We recruited 26 participants in total, which would allow us to detect an effect size of $d = 0.25$ with .80 power at an alpha level of .05 (calculated using the G*Power software [7]).

Table 1: Overview of the experimental conditions.

	Social	***Adverse***
Front	Condition 1 (SF)	Condition 2 (AF)
Side	Condition 3 (SS)	Condition 4 (AS)

Most of the participants (93%) have already interacted with a robot prior to this experiment and are in average 28.3 years old ($SD = 8.74$). The experiment was carried out in a hall with controlled illumination and no outside distraction. Only the robot, the participant and the experimenter were present in the hall throughout the duration of the experiment (ca. 20 minutes). The robot performs a frontal straight segment of $4m$ (*approaching phase* $0 - 14s$), stops at $1.7m$ from the user, gazes at them and says a short greeting sentence (*interacting phase* $14 - 24s$) and finally goes back to the starting position (*terminating phase* $24 - 51s$). It is important to notice that despite the position of the participant (frontally or with the lateral offset), once the robot has finished the approaching phase, it regulates its face towards them before playing the greeting sentence. Figures 1a, 1b, 1c, 1d shows snapshots of the experimental conditions SF, AF, SS and AS, respectively. The initial distance of $4m$ as well as the distance during the interacting phase of $1.7m$ are selected according to [8]. We are aware that



(a) (b) (c) (d)

Fig. 1: Snapshots of the four experimental conditions at the end of the approaching phase.

proxemics can differ among participants, however, this work focuses on the role of gaze and proximity as drivers for initiating a social HRI and the selected values are reasonable for the goal. Each participant is instructed to stand still throughout the experiment, observe the robot's behaviour and ask themself the question:

– (1) "Would the robot like to start an interaction with me?".

Participants are instructed to say "yes" as soon as they can answer the question (1) positively. At this point, the experimenter halts a stopwatch that keeps track of the time passed between the initial motion of the robot and the time when the participant responded to question (1) with "yes". If the robot can reach the interacting phase location and the participant still has not answered positively to (1), we consider the stopwatch to halt at the end of the interacting phase ($t = 24s$). With this logic, the stopwatch values are taken into account only if belong to the approaching phase and are saturated if answered during the interacting phase or later. Bearing this in mind, the multimodal interaction that unfolds during the interacting phase and its effects on the question (1) are out of the scope of this work. The interpersonal space together with the gaze directed to the participant and a short spoken sentence, allow us to consider that participants surely perceive the intention to interact with the robot during the interacting phase. The controlled variables (gaze and approach policy) differ across the conditions only during the approaching phase. The conditions are administered to the user in a pseudo-randomized fashion to mitigate order effects. After being exposed to each condition the participants answered a brief post-interaction survey comprising of the following 5-point Likert scale entries. In italic are shown the corresponding keywords used in Figure 3.

1. The robot's behaviour is *social*.
2. The robot would like to *interact* with me.
3. I would feel *comfortable* of encountering this robot in a social context.
4. I like the *quality* of the robot.
5. I quickly understood when the robot wanted to *start* the interaction.
6. I quickly understood when the robot wanted to *finish* the interaction.

The response could range between 1 (I fully disagree) to 5 (I fully agree). Finally, after experiencing the four conditions the experimenter tells the participant the goal of the research in detail.

4 Result

The responses to question (1) are collected in terms of seconds (time between the start of the robot motion and the keyword pronounced by the participants).

A paired t-test with 95% confidence intervals is performed on the mean of these for each condition. Figure 2 shows the mean time to answer the question (1) and the standard deviation at each condition is shown in terms of error bar length.

In this measure, we found a significant difference between SF ($M = 16.59$, $SD = 6.45$) and AF ($M = 21.06$, $SD = 3.75$), with $t(25) = -2.56$, $p < .05$ and between SF and AS ($M = 22.26$, $SD = 2.06$), $t(25) = -3.76$, $p < .001$.

This shows that participants were able to answer significantly faster to question (1) when the robot employed a social gaze compared to the robot that used an asocial gaze despite its base motion trajectory. Significant difference is found between AF and SS ($M = 12.58, SD = 4.80$) with $t(25) = 7.84$, $p < .001$ and between SS and AS with $t(25) = 8.27$, $p < .001$. Participants took significantly longer to answer question (1) when the robot employed an adverse gaze despite its base motion trajectory. Finally, significant difference is found between SF and SS with $t(25) = 2.20$, $p < .05$. This latter result shows that question (1) was answered significantly faster when the robot employed a social gaze and the user was not in front of the trajectory of the robot.

We could deduct that the base motion trajectory is less relevant than the gaze direction for eliciting the intention to start a social HRI. It is interesting to notice that in Figure 2 only the condition SS obtained a mean time within the

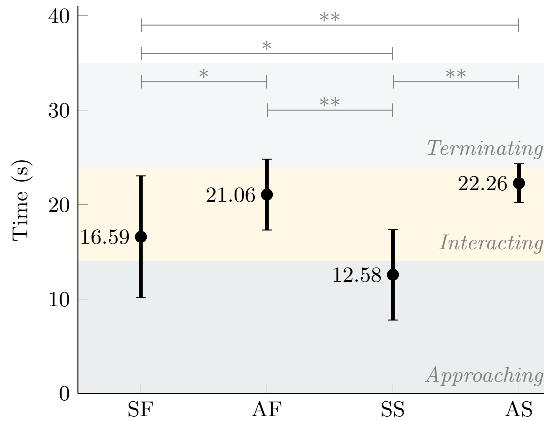


Fig. 2: Responses' means of question (1) per condition. Significant differences between conditions have been indicated with * for $p < .05$ and with ** for $p < .001$.

approaching phase window. Figure 3 shows the mean responses of the survey grouped by conditions with the respective significant differences.

Regarding the questions presented in section 3, a Wilcoxon signed-rank test is performed on the mean responses per each condition. In particular significant differences are found in the *social* question between SF and AS ($T = 55$, $p < .05$), between AF and AS ($T = 48$, $p < .05$) and between SS and AS ($T = 18$, $p < .05$). Regarding the *interact* question significant differences are found between SF and AF ($T = 22$, $p < .05$) and between SF and AS ($T = 20$, $p < .05$). The experiments were designed so that the robot 1) approaches the user in four different ways and 2) terminates the interaction using the same behaviour across all conditions. Surprisingly, no significant difference was found in the *start* question, but a significant difference was found in the *finish* question between SS and AS ($T = 41$, $p < .05$). We suspect that this result is given by the short time allocated to the interaction and the difference in the yaw of the robot head

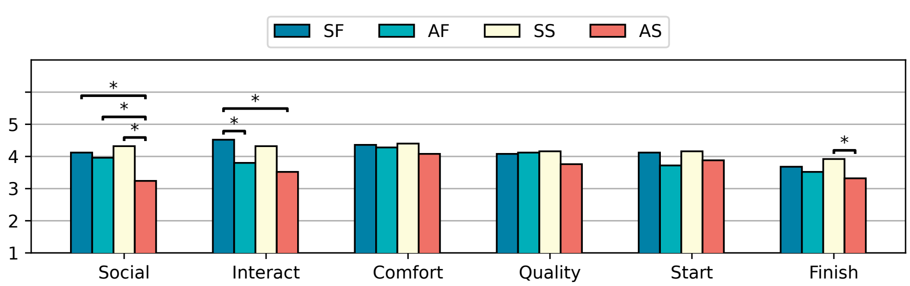


Fig. 3: Average responses of the survey. Significant differences between conditions have been indicated with * for $p < .05$.

between SS and AS. However, further investigation will help us understand this trend.

5 Discussion

Interpersonal distance and gaze are some preceptors of the interaction but cannot always reflect the intention of initiating an interaction. In the scenarios that we build, despite the motion constraint given to the user, we could simulate unstructured spontaneous encounters with a robot. While the robot is approaching a user, its arm gestures and body pose can influence the way the user perceives it [6]. The interactions developed for this study aim to simulate a spontaneous' one that can happen in the wild. Clearly, an unstructured scene is far more complex and the environment can influence the interaction in a positive or negative fashion. For instance, if two users are in a hall without illumination, it would be very hard to establish an interaction. Hence, the social cognition is preceded by the perception of the surroundings.

Figure 2 suggest that the robot's head acts as a stronger social cue with respect to the user's relative body position while approaching a social HRI. Furthermore, in SS the robot gives the impression to actively looking for the user while in FS since the user is already on the robot's trajectory, the intentionality of it is not clearly perceived.

Some participants described the behaviour of the robot during SS as secure and more natural, while other participants suggested that the relative torso pose would have also benefited the study. Additionally, few participants described the behaviour of the robot during SF as aggressive and unnatural. Despite the slight changes among the conditions, our results show that even when the robot is far from the human, its gazing behaviour is read as a social signal and can convey the intention to initiate an interaction.

6 Conclusion

In this work, we investigate the role of nonverbal robot behaviour before a social HRI takes place. We model the time evolution of an HRI with a three-phase representation and conduct a user study to investigate the effects of gaze and approach policy when the interaction has not started yet. Participants are instructed to pronounce the keyword "yes" as soon as they can answer positively to the question "Would the robot like to start an interaction with me?". At this point, a stopwatch records how long it took each participant to pronounce the keyword in each condition. Each participant also fills in a post-interaction survey after experiencing each condition. Results show that participants significantly preferred conditions in which the robot employs a social gaze while approaching them. However, it might be that when the user is directly on the path of the robot the intentionality of the social gaze is perceived as a consequence of the robot's motion.

Future research will investigate the effects of different robot designs and more complex robot paths have on the perceived robot's intentions in similar scenarios. Moreover, we plan to target different age groups of participants to investigate whether their perception of the robot's intention differs.

References

1. Admoni, H., Scassellati, B.: Social eye gaze in human-robot interaction: a review. Journal of Human-Robot Interaction **6**(1), 25–63 (2017)
2. Avelino, J., Garcia-Marques, L., Ventura, R., Bernardino, A.: Break the ice: a survey on socially aware engagement for human–robot first encounters. International Journal of Social Robotics **13**(8), 1851–1877 (2021)
3. Ba, S.O., Odobez, J.M.: Recognizing visual focus of attention from head pose in natural meetings. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **39**(1), 16–33 (2008)
4. Belpaeme, T., Baxter, P., Read, R., Wood, R., Cuayáhuitl, H., Kiefer, B., Racioppa, S., Kruijff-Korbayová, I., Athanasopoulos, G., Enescu, V., et al.: Multimodal child-robot interaction: Building social bonds. Journal of Human-Robot Interaction **1**(2) (2012)
5. Ben-Youssef, A., Clavel, C., Essid, S., Bilac, M., Chamoux, M., Lim, A.: Ue-hri: a new dataset for the study of user engagement in spontaneous human-robot interactions. In: Proceedings of the 19th ACM international conference on multimodal interaction. pp. 464–472 (2017)
6. Breazeal, C., Dautenhahn, K., Kanda, T.: Social robotics. Springer handbook of robotics pp. 1935–1972 (2016)
7. Faul, F., Erdfelder, E., Lang, A.G., Buchner, A.: G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior research methods **39**(2), 175–191 (2007)
8. Hall, E.T.: The hidden dimension, vol. 609. Anchor (1966)
9. Heenan, B., Greenberg, S., Aghel-Manesh, S., Sharlin, E.: Designing social greetings in human robot interaction. In: Proceedings of the 2014 conference on Designing interactive systems. pp. 855–864 (2014)
10. Hong, A., Lunscher, N., Hu, T., Tsuboi, Y., Zhang, X., Franco dos Reis Alves, S., Nejat, G., Benhabib, B.: A multimodal emotional human–robot interaction architecture for social robots engaged in bidirectional communication. IEEE Transactions on Cybernetics **51**(12), 5954–5968 (2021)
11. Kendon, A.: Conducting interaction: Patterns of behavior in focused encounters, vol. 7. CUP Archive (1990)
12. MacArthur, K.R., Stowers, K., Hancock, P.A.: Human-robot interaction: proximity and speed—slowly back away from the robot! In: Advances in human factors in robots and unmanned systems, pp. 365–374. Springer (2017)
13. Maniscalco, U., Storniolo, P., Messina, A.: Bidirectional multi-modal signs of checking human-robot engagement and interaction. International Journal of Social Robotics pp. 1–15 (2022)
14. Neggers, M.M., Cuijpers, R.H., Ruijten, P.A., IJsselsteijn, W.A.: Determining shape and size of personal space of a human when passed by a robot. International Journal of Social Robotics **14**(2), 561–572 (2022)
15. Normoyle, A., Badler, J.B., Fan, T., Badler, N.I., Cassol, V.J., Musse, S.R.: Evaluating perceived trust from procedurally animated gaze. In: Proceedings of Motion on Games, pp. 141–148 (2013)

16. Oertel, C., Castellano, G., Chetouani, M., Nasir, J., Obaid, M., Pelachaud, C., Peters, C.: Engagement in human-agent interaction: An overview. Frontiers in Robotics and AI **7**, 92 (2020)
17. Patompak, P., Jeong, S., Nilkhamhang, I., Chong, N.Y.: Learning proxemics for personalized human–robot social interaction. International Journal of Social Robotics **12**(1), 267–280 (2020)
18. Pereira, A., Oertel, C., Fermoselle, L., Mendelson, J., Gustafson, J.: Effects of different interaction contexts when evaluating gaze models in hri. In: Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction. pp. 131–139 (2020)
19. Rossi, S., Ercolano, G., Raggioli, L., Savino, E., Ruocco, M.: The disappearing robot: An analysis of disengagement and distraction during non-interactive tasks. In: 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). pp. 522–527 (2018)
20. Rossi, S., Staffa, M., Bove, L., Capasso, R., Ercolano, G.: User's personality and activity influence on hri comfortable distances. In: Social Robotics. pp. 167–177. Springer International Publishing, Cham (2017)
21. Sanghvi, J., Castellano, G., Leite, I., Pereira, A., McOwan, P.W., Paiva, A.: Automatic analysis of affective postures and body motion to detect engagement with a game companion. In: Proceedings of the 6th international conference on Human-robot interaction. pp. 305–312 (2011)
22. Satake, S., Kanda, T., Glas, D.F., Imai, M., Ishiguro, H., Hagita, N.: How to approach humans? strategies for social robots to initiate interaction. In: Proceedings of the 4th ACM/IEEE international conference on Human robot interaction. pp. 109–116 (2009)
23. Sidner, C.L., Lee, C., Kidd, C.D., Lesh, N., Rich, C.: Explorations in engagement for humans and robots. Artificial Intelligence **166**(1), 140–164 (2005)
24. Staffa, M., Gregorio, M.D., Giordano, M., Rossi, S.: Can you follow that guy? In: 22th European Symposium on Artificial Neural Networks, ESANN 2014, Bruges, Belgium, April 23-25, 2014 (2014)
25. Sternberg, R.J., Smith, C.: Social intelligence and decoding skills in nonverbal communication. Social cognition **3**(2), 168 (1985)
26. Strazdas, D., Hintz, J., Felßberg, A.M., Al-Hamadi, A.: Robots and wizards: An investigation into natural human–robot interaction. IEEE Access **8**, 207635–207642 (2020)
27. Syrdal, D.S., Dautenhahn, K., Walters, M.L., Koay, K.L.: Sharing spaces with robots in a home scenario-anthropomorphic attributions and their effect on proxemic expectations and evaluations in a live hri trial. In: AAAI fall symposium: AI in Eldercare: new solutions to old problems. pp. 116–123 (2008)
28. Vigni, F., Knoop, E., Prattichizzo, D., Malvezzi, M.: The role of closed-loop hand control in handshaking interactions. IEEE Robotics and Automation Letters **4**(2), 878–885 (2019)
29. Zhang, Y., Beskow, J., Kjellström, H.: Look but don't stare: Mutual gaze interaction in social robots. In: International Conference on Social Robotics. pp. 556–566. Springer (2017)
30. Złotowski, J., Proudfoot, D., Yogeeswaran, K., Bartneck, C.: Anthropomorphism: opportunities and challenges in human–robot interaction. International journal of social robotics **7**(3), 347–360 (2015)