



Preference-Based People-Aware Navigation for Telepresence Robots

Alberto Bacchin¹ · Gloria Beraldo^{1,2} · Jun Miura³ · Emanuele Menegatti¹

Accepted: 19 March 2024
© The Author(s) 2024

Abstract

This work proposes an innovative people-aware navigation for telepresence robots in a populated environment based on the estimated inclination of people to interact and the context information. The main novelty of the proposed *people-aware shared intelligence* is the ability to fuse the remote operator's commands with the probability of person-robot interaction—from both the operator driving the robot and the people around it—and translate it into semi-autonomous approaching and avoiding behaviors that are not coded a priori but rather dynamically emerge according to the current context-awareness. Experiments involved 45 healthy participants who evaluated the proposed approach on a real robot. Three conditions have been tested: (a) the new *people-aware shared intelligence*; (b) a *shared intelligence* system integrated with the standard ROS social navigation layers and; (c) a direct teleoperation (i.e., no robot's intelligence). Results from our *people-aware shared intelligence* system have shown that the robot's social behaviors were in line with the expectations of the participants in terms of comfort, naturalness, and sociability and coherent with the findings from previous studies. Furthermore, the proposed system has facilitated the social interaction between the remote operator and the surrounding people, making the robot more proactive and without affecting navigation performance.

Keywords People-aware navigation · People approaching · Human–robot interaction · Social signaling understanding · Context awareness

1 Introduction

During the COVID-19 pandemic, telepresence robots have regained importance as a tool to assist humans remotely and provide alternative communication channels to keep them in contact. On the one side, telepresence robots are expected to

implement the commands delivered by the remote user, on the other, they should behave in a social manner with the people around. However, traditional navigation algorithms appear not appropriate to be used in uncontrolled environments populated by people. Indeed, such algorithms just optimize the robot's movements toward a target position by treating humans as mere dynamic obstacles. As a consequence, the social rules respected by people during the interaction are not considered with the possible risk of invading the personal spaces [1].

To tackle the aforementioned problems, in recent years, the concept of *people-aware* or *social navigation* has been introduced to allow the robot to navigate safely and socially behave with humans. According to [2], three are the main goals of *social navigation*: (i) the *comfort* as the absence of annoyance for humans in interaction with robots, (ii) the *naturalness* as the similarity between human and robot low-level behavior and (iii) the *sociability* as the compliance to explicit high-level social conventions. Nevertheless, most of the previous literature on social navigation put effort into the appropriate human–robot distances during the interaction by using as reference the metrics from the previous anthropological studies such as proxemics [1], and F-formations

Alberto Bacchin and Gloria Beraldo contributed equally to this work.

✉ Gloria Beraldo
gloriaberaldo@gmail.com

Alberto Bacchin
bacchinalb@dei.unipd.it

Jun Miura
jun.miura@tut.jp

Emanuele Menegatti
emg@dei.unipd.it

¹ Department of Information Engineering, University of Padova, Padua, Italy

² Institute of Cognitive Sciences and Technologies, National Research Council of Italy, Rome, Italy

³ Department of Computer Science and Engineering, Toyohashi University of Technology, Toyohashi, Japan

[3]. Indeed, the state-of-the-art social navigation algorithms have mainly focused on *people-avoidance* and respecting *social spaces*, but have neglected the interaction aspects—such as the people’s interest in starting the interaction before approaching—that are fundamental for telepresence applications.

For instance, imagine using telepresence robots for monitoring person/people at home (e.g., older people), nursing homes, health facilities, and hospitals, that need social interaction and company during the day [4–7]. In such scenarios, it has been already demonstrated that caregivers and families can support and reduce social isolation even if remotely (i.e., through the robot) thanks to a more interactive communication channel with respect to phones. The caregiver or family members act as the operator, while the person alone or in a group is the user in the same environment as the robot. In these contexts, different situations can happen: (a) the robot should avoid a person who does not wish to interact; (b) human–robot interaction with a person who would like to start a conversation with the remote operator; (c) the remote operator would like to start a conversation with a target person/group of people who are not interacting with the robot in the first instance (e.g., the operator would like to monitor the actions of a person) and teleoperate the robot in his/her/their proximity through directional commands. In these situations, given possible delays in communication due to the internet connection and a limited view of the surrounding environment, it is also important to facilitate the teleoperation for the operator by relying on the robot’s ability to contextualize the situation and manage the operator inputs, hence to semi-autonomously behave properly for both interacting or avoiding people according to the specific situations.

In this work, to the best of our knowledge, we propose the first *people-aware* system in teleoperation that manages person-robot interaction from both the operator and the other people based on their will to interact and translates them into semi-autonomous approaching and avoiding behaviors that are not coded a priori. Exploiting the social signaling contextualization through the *shared intelligence* paradigm [8], the robot is able to provide proactive social behaviors that support the operator during the interaction with other humans, as pointed out by our experiments. Moreover, the proposed framework is designed to be operated with a very simple interface (i.e., based on sending left/right directional commands) that enables people without previous knowledge to easily use such a system in telepresence applications. Globally, the *people-aware shared intelligence* framework takes into account human intentions and promotes social-compliant behaviors without affecting the standard navigation capabilities and facilitating the bi-directional interaction (i.e., from the operator towards the other people and vice versa).

1.1 Related Work

Over the years, several approaches for achieving social navigation have been proposed. Some approaches try to extend the traditional reactive navigation algorithms by introducing other constraints for managing social interactions. For instance, the most explored techniques are based on the Artificial Potential Field, like in [9] and Social Force Model, as in [10]. Both rely on the idea that several forces are exerted on the robot—the attractive force generated by the targets and the repulsive derived by the obstacles—and from the resulting sum the robot’s speed is computed as happens for instance in the studies [11–13]. Although these methods are simple to implement and efficiently computed in real-time, they suffer from the presence of oscillatory behaviors in the robot’s trajectories due to local minima. Moreover, previous studies based on Artificial Potential Fields have already demonstrated the robot might move too close to people [14]. Other approaches have focused on estimating in advance the next people’s positions rather than adopting pure reactive navigation such as [15, 16] to choose the robot’s motion. The main drawback of these methods is associated with the way of predicting each person’s trajectory—independently of the others—that can cause frequent robot’s stops related to the people-people interaction [17]. More recent studies have overcome this limitation, including [18–20], but many computation resources are needed and the performance can be affected by the predictions bias. Similarly, specific training hardware is required when using *data-driven strategies* that typically find the best *policy* to simulate human behaviors using features from human trajectories gathered in simulation and/or in the real world. Supervised learning approaches need to collect and label several samples (e.g., time consuming), while deep reinforcement learning strategies make the agent learn to navigate sociably according to a rewarded function properly created to penalize the undesirable robot’s motion [21, 22]. In between, a recent solution that does not require any additional overhead for the user is proposed by Bacchin et al. [23] namely a simple and light learning approach based on a genetic algorithm optimized to find the best configuration of the parameters behind the standard *ROS navigation stack* while the robot is disturbed by people and trained to implement social person avoidance. Nevertheless, such approaches often depend on an intensive training process and suffer from a lack of interpretability in the results. Furthermore, in the case of pre-trained algorithms using simulators, there is the additional challenge of accurately simulating and modeling human behaviors. Certainly, formalising people and robot behaviors according to pre-defined and strict rules allow the robot to take explainable decisions. For instance, Singamaneni et al. in [24] have proposed a deliberative planner tuned according to the specific human–robot scenario, in [25] Gaussian Mixture Models

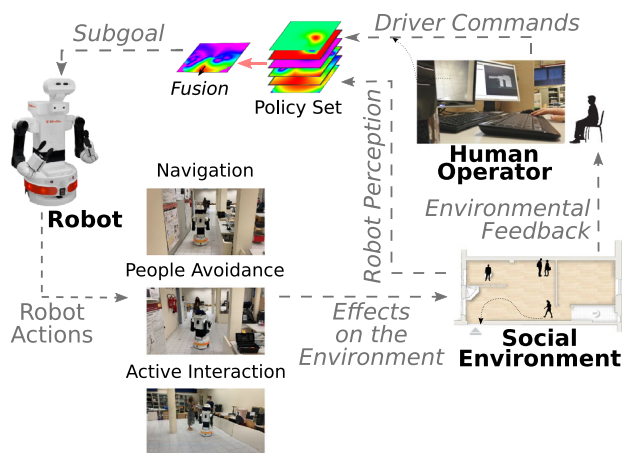


Fig. 1 An illustrative scheme of the main components behind the proposed *people-aware shared intelligence*. The set of *policies* creates probabilistic representations of the robot's context awareness based on the robot's perception, the estimated people will to interact and the operator's commands. Fusing these information, a navigation *subgoal* is computed and provided in input to the standard ROS *navigation stack*. The dynamical update of the *subgoal* allows achieving people-aware behaviors (e.g., safe navigation, people avoidance, approaching people). (Color figure online)

(GMM) are exploited to classify different people behaviors and select a trajectory with a high social score. However, modeling suitable people's behaviors can be very hard and inadequate to generalise in multiple scenarios.

Our system is designed to keep the *proactive* and real-time component with the additional introduction of information about the people's motion prediction to prevent robot's oscillations and take into account the next human actions. Indeed, with respect to the previous approaches, given our innovative purpose of fusing the estimated will to interact of the people with the robot's perception and the operator's commands, it is necessary to combine the current context awareness with the estimation of the situation in the near future to avoid abrupt robot's motions and involuntary stop. Consequently, the robot dynamically interprets the social signals from the surrounding people and the operator's commands without relying on models provided in input or learned from data, which are in addition difficult to create.

1.2 Overview & Contribution

By combining both the remote user's commands and sensor output, our system aims to interpret the situation in order to (a) deviate from people who avoid the robot; (b) approach them when they would like to interact. This scenario adds further complexity. First, modeling the appropriate robot's behaviors is more challenging since they are both dependent on the operator's commands and the motions of the surrounding humans. Second, we face the further challenge of socially approaching people when they are inclined to interact with

the robot and vice versa. Only a few studies in the literature have formulated theories related to the suitable approaching poses for the robot. For instance, Truong and Ngo [26] have presented the complex model called Dynamic Social Zone (DSZ) according to the people's position, orientation, motion, robot's field of view and group relationships, even if they have not considered the real inclination of people to interact. Moreover, previous works of [27, 28] have studied how the robot just estimates the most suitable approaching pose without considering the real people's will to interact. Third, the topic is difficult to study since the tests can be done properly only with real human subjects in both roles (people in the environment vs. remote user), but most previous works are validated only in a simulated environment.

To achieve our objective, we broaden the modular framework proposed in [8] that aims to create an innovative interaction strategy, called *shared intelligence*, derived from the equal contribution of the human's commands and the robot's perception in the decision process. The previous work has shown that treating different sources of information that influence the choice of a navigation *subgoal* (e.g., temporal destination for the robot) by *policies* can lead to the robot's ability to modify or ignore the operator's commands and prevent collisions during a traditional navigation task. In this study, we further expand on this idea by considering the presence of people in the same environment and their intention to interact during a socially compliant task. Therefore, the main novelty of this work is the ability to fuse the remote user's commands with the estimation of the will to interact of the people surrounding the robot to: (a) support the robot's navigation respecting the social conventions; (b) approach people who would like to interact; (c) avoid humans when they are not inclined to the interaction. Hence, we introduce new techniques in the previous system to take advantage of: (a) the estimation of the future positions of all the people around the robot as in the *predictive* algorithms and (b) a formulation of the expected social behaviors (e.g., avoid not target person, approach a target person, safe navigation without collisions) using *policies* to provide an initial guess to the robot about the expected motion. A schematic representation of the system is shown in Fig. 1. Intuitively, the robot perception, the estimated people will to interact and the operator's commands are managed by a set of *policies* that returns probabilistic maps around the robot's position. The fusion of the probabilistic maps is used to compute the robot's navigation *subgoal*. By dynamically updating the *probabilistic maps* according to the context and hence the *subgoal*, socially compliant behaviors are generated (e.g., safe navigation, people avoidance, approaching people). For clarity, it is fundamental to specify that the robot equipped with the proposed system lacks inherent knowledge of any global goal. The global goal solely resides in the user's cognition. If necessary, the user can alter the semi-autonomous robot's behaviors by sending

high-level inputs that guide the robot toward the intended destinations.

In summary, our contributions can be summarized as follows:

- we introduce people-avoidance capabilities in a *shared intelligence system* to support the remote user in navigating in populated environments;
- we propose a solution to approaching people when they are willing to start an interaction or vice-versa when the remote user is interested in the interaction, without the need to trigger the interaction with explicit commands;
- we propose a system that continuously fuses both user's commands, robot perception, and estimated people intentions;
- we validate our system with a real robot, involving more than 40 people in the experiments.

The rest of the paper is organized as follows. Section 2 explains the details of the proposed *people-aware shared intelligence* approach by focusing on the new *social policies*. Section 3 describes the robotic platform, the experimental setup and the examined modalities exploited to test the proposed system. Section 4 is dedicated to present the results in terms of quantitative metrics and answers to a questionnaire about the experience in real-world experiments. The results are discussed with respect to other state-of-the-art studies in Sect. 5 and, finally, Sect. 6 concludes the paper.

2 Shared People-Aware Navigation System

In our system, the generation of *people-aware navigation system* is achieved by the fusion of *policies* related to the operator's commands and the robot's perception to determine a temporary position, called *subgoal*, that the robot has to reach. In line with [8], each *policy* handles a specific kind of information influencing the choice of the *subgoal* such as the direction provided by the operator, the proximity to possible targets, the distance from the obstacles, etc. For a better understanding, a *policy* is modeled as a decision function that receives a specific input and returns a probabilistic grid defined in the area around the robot under the vector $\mathbf{x} = (x, y)$. By fusing all of them in output by the *policies*, a new probabilistic grid is achieved that contains the joint probability of the multiple events influencing the choice of the *subgoal*. Indeed, the *subgoal* is simply computed as the position with the highest probability in the fusion *probabilistic grid*. Since the system presented in [8] is designed only for a safe navigation, it includes the following *policies*:

- *Obstacle-avoidance* that generates probabilistic grids where the probability to set the *subgoal* in one position is proportional to the closest distance of the obstacles. The

probability is forced to zero if the position is occupied by an obstacle.;

- *Distance* that assigns higher probability to the positions inside the preferred range of *subgoal* distances given in input;
- *Direction* that favors the positions around the current robot direction to avoid abrupt directional changes.
- *User input* that attributes higher probability to the zones in the direction chosen by the operator via a discrete input.

Herein, instead, we have proposed new *policies* for making the robot exhibit *social behaviors* that are oriented both to the interaction with people (e.g., the robot autonomously stops in front of the target people for the interaction); and the *people-avoidance* in accordance to the *social norms* (e.g., respecting the social spaces while navigating). Given the aim of including *social behaviors* in the previous system, we have also proposed a different version of the *User Input policy*, managing the user's commands, to allow a continuous interaction with the robot (e.g., both in the time and in the space). Indeed, we hypothesized that a finer control of the robot is more appropriate than the original discrete modality of interaction proposed in [8] when dealing with dynamic targets in unstructured environments. Finally, we have modified the strategy for computing the *subgoal* in order to allow the autonomous stop of the robot when interacting with target people.

Illustrative examples of the application of the *policies* behind the system to situations of interest are shown in Fig. 2. In the middle, the resulting probabilistic grids by each *policy* are represented, and on the right the fusion of them as well. The red areas are the most probable to set the *subgoal*. The white arrow on the fusion indicates the *subgoal* chosen for the robot in each situation.

A detailed formulation of the new *policies* and the computation of the *subgoal* will be described in the next subsections.

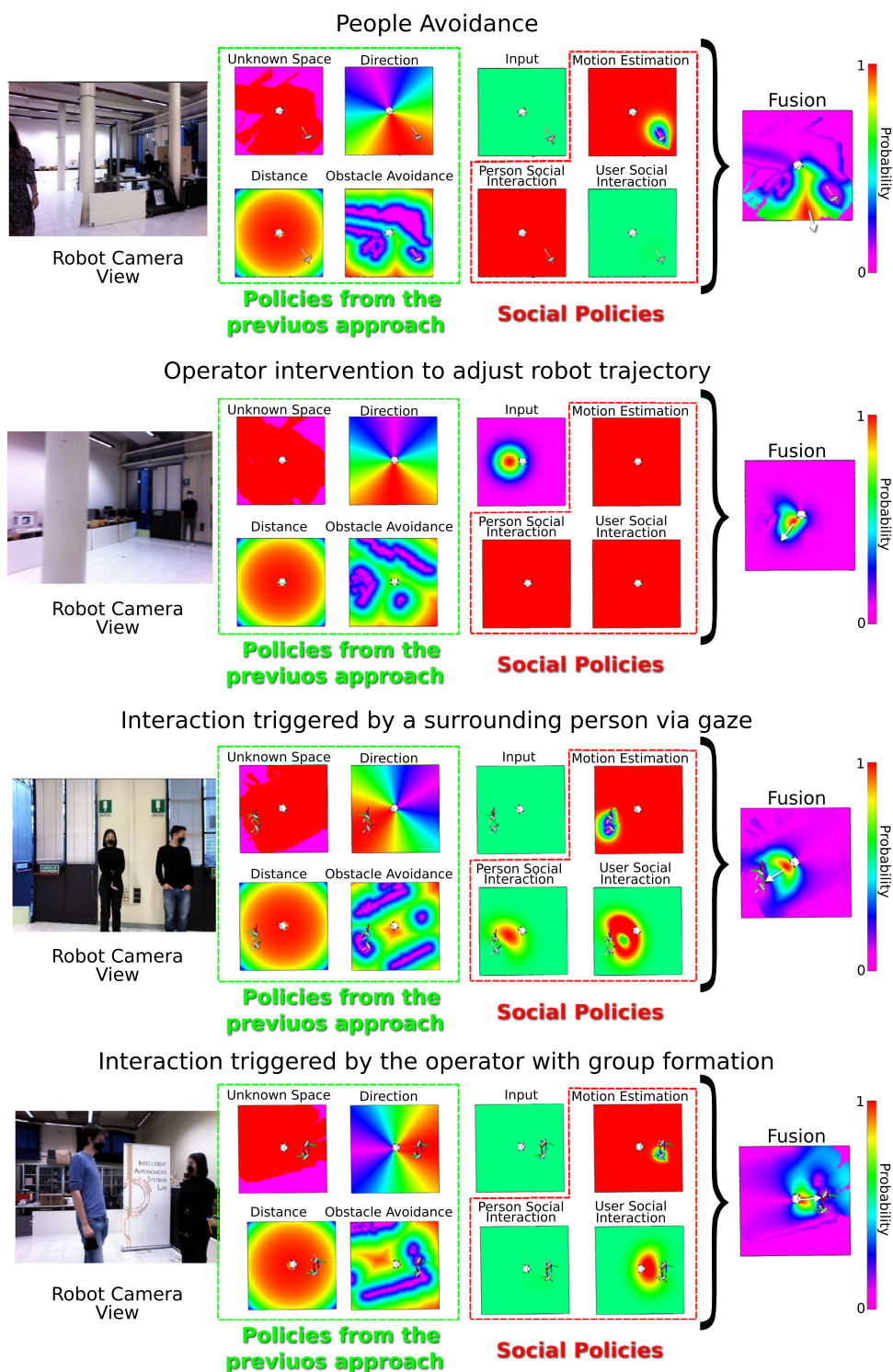
2.1 The Social Policies

To cope with the challenges related to the social navigation depicted in [2], we have introduced three new social *policies*: the *Motion Prediction*, the *Person Social Interaction* and *User Social Interaction*.

2.1.1 The Motion Prediction Policy

The *Motion Prediction Policy* estimates the people's positions with respect to the robot in the near future. The aim of this *policy* is to filter those positions out from the choice of the *subgoal* to make the robot implement *people-avoidance* functionalities respecting the social spaces. We have designed the *Motion Prediction policy* in order to introduce some socially-compliant behaviors such as preventing

Fig. 2 The picture shows the application of the proposed framework in the following situations: *people-avoidance* in a corridor, *user's command* to turn at a crossroad, the *human-robot interaction* triggered by the surrounding people via the gaze and by the operator at a couple of talking people (e.g., when the respect of group social behavior is required). On the left, the robot camera's view (e.g., the feedback for the operator) is reported. In the middle, all the probability grids from *policies* are shown. Stylized 3D models are used to show the detected people (in white). Finally, on the right, the resulting distribution by fusing the *policies* is represented. The white arrow represents the current *subgoal*



the robot to cut the street off. Inspired by the work of [12], we have considered the estimated people's speed in the computation of the resulting probability distribution. Given $\Theta_{MP} = \{(p_i, \gamma_i, v_i), i \in [1, N]\}$ where $p_i = (x_i, y_i)$ indicates the position, γ_i the orientation and $v_i = (\dot{x}_i, \dot{y}_i)$ the velocity of the person i (all are provided by a people tracker,

see Appendix A.2), the *Motion Prediction policy* is modeled as follows:

$$P^{MP}(\mathbf{x}, \Theta^{MP}) = \prod_{i=1}^N f_{motion}(\mathbf{x}, p_i, v_i) \cdot f_{turn}(\mathbf{x}, p_i, \gamma_i) \tag{1}$$

Each person $i \in [1, N]$ contributes to a local minimum in the probability distribution P^{MP} . $f_{motion}(\mathbf{x}, p_i, v_i)$, a bivariate Gaussian distribution which models the future position of the person i and the related uncertainty. To predict the future position, we used the following stochastic process

$$p_i(t + dt) = p_i(t) + v_i(t) \cdot dt + \epsilon_t \quad (2)$$

where $\epsilon_t \sim \mathcal{N}(0, R_t)$ is the noise mode and $dt = 1$ s is the fixed forward time to make the prediction. Thus, the expected future position is $\mu_i = p_i + v_i \cdot dt$ and the covariance matrix is $\Sigma_i = \Sigma^{p_i} + \Sigma^{v_i}$ where Σ^{p_i} and Σ^{v_i} are respectively the co-variance matrices representing the confidence on the pose and velocity estimations of each person returned from the detector. Finally, we get

$$f_{motion}(\mathbf{x}, p_i, v_i) = \mathcal{N}(\mu_i, \Sigma_i; \mathbf{x}) \quad (3)$$

where \mathbf{x} indicates the variable of the Gaussian distribution.

$f_{turn}(\mathbf{x}, p_i, \gamma_i)$ instead is modeling the fact that a person can turn and change direction, which is not considered in the linear model in Eq. 2. We hypothesize that the person likely maintains his/her current position in the nearest future, while she/he may decide to change direction afterward. The distribution is still based on a Gaussian distribution

$$f_{turn}(\mathbf{x}, p_i, \gamma_i) = \mathcal{N}(\gamma_i, \sigma_{tr}^2; \theta) \quad (4)$$

but in this case, the probability decreases with the angular distance $\theta = \text{atan2}(y - y_i, x - x_i)$ from the current motion direction γ_i .

2.1.2 Social Interaction Policies

This subsection aims to illustrate the two *Social Interaction policies* included in our system: the *Person Social Interaction* and the *User Social Interaction*. The first aims to estimate the will of interacting from surrounding people with the remote user, through the robot, based on non-verbal cues. The latter infers the remote user's intention of interacting with specific people predicted by the system. In both *policies*, we focus on gaze as a social cue triggering the interaction, since it has been successfully applied in different human-robot interaction works [29]. Indeed, as demonstrated in other studies, humans commonly catch the attention of a person by looking at him/her. For instance, in [30], the authors have shown that the direct gaze captures the attention of people, and in [31], have demonstrated that gaze is the most important cue to gather people's attention, independently of other non-verbal signals. Based on these premises, we have assumed that a person interested in interacting with the robot at least looks toward it in the *Person Social Interaction*, while the user lets

the robot point toward the target person in the *User Social Interaction* if desires to engage in interaction (see Sect. 4.5).

Since we assume that the interaction occurs similarly in both cases, we used the same distribution to model both the *policies*:

$$P^{SI}(\mathbf{x}, \Theta^{SI}) = 1 - \prod_{i=1}^N (1 - P_i^{SI}(\mathbf{x}, \Theta_i^{SI})) \quad (5)$$

For each person $i \in [1, N]$ detected in the scene, we can define:

$$P_i^{SI}(\mathbf{x}, \Theta_i^{SI}) = w_i^{SI}(p_i, t) \cdot f_{space}(\mathbf{x}, p_i) \cdot f_{gaze}(\mathbf{x}, p_i, \alpha_i) \quad (6)$$

where $\Theta_i^{SI} = (p_i, \alpha_i, t)$ and $\Theta^{SI} = \{\Theta_i^{SI}, \forall i \in [1, N]\}$. The first component $f_{space}(\mathbf{x}, p_i)$ models the interaction space as a ring according to Hall's theory of proxemics [1]. Formally, it is achieved as a difference between Gaussians:

$$f_{space}(\mathbf{x}, p_i) = \mathcal{N}(p_i, \Sigma_{dM}; \mathbf{x}) - \mathcal{N}(p_i, \Sigma_{dm}; \mathbf{x}) \quad (7)$$

with $\Sigma_{dM} = I_2(\sigma_{dM}^2)$ and $\Sigma_{dm} = I_2(\sigma_{dm}^2)$. The width of the ring is controlled by the variances σ_{dM}^2 and σ_{dm}^2 . Since in formal circumstances people usually keep distances between 1 and 1.5 m during a voice conversation as stated by [32], we set $\sigma_{dm}^2 = 0.75$ and $\sigma_{dM}^2 = 1.0$ to ensure that the probability to interact in the interval [1.0, 1.5] m is ≥ 0.9 . The result is normalized afterward to obtain a proper probability distribution.

The second component $f_{gaze}(\mathbf{x}, p_i, \alpha_i)$ selects the portion of the aforementioned interaction space where the interaction is expected to take place according to the gaze estimation (see Appendix A.2). Specifically, it is shaped as:

$$f_{gaze}(\mathbf{x}, p_i, \alpha_i) = 1 - \mathcal{N}(\alpha_i, \sigma_{ang}^2; \theta) \quad (8)$$

where similarly to Eq. 4, $\theta = \text{atan2}(y - y_i, x - x_i)$ while α_i has a different meaning in the two *policies*. In the *Person Social Interaction policy*, it measures the current gazing direction of the person i , which is estimated by a *gaze detector*. Hence, the larger the angular distance from the gazing direction is, the lower the probability is. In the *User Social Interaction policy*, α_i indicates the direction connecting the segment between the robot and the position of the person i . In accordance with [33] which investigates how people arrange in space while interacting, we assume that the area of interest for human-robot interaction is placed between -90° and 90° with respect to the person orientation. Therefore, we set $\sigma_{ang} = 45^\circ$ to ensure the probability of interacting in that area is greater than 95%.

Finally, $w_i^{SI}(p_i, t)$ is a weighting function that differently behaves in the two *policies*. In the *Person Social Interaction*

policy, it represents the growing interest of a person to interact while he/she is looking towards the target, the robot in this case. For this reason, it is just a time-dependent exponential weight, as modeled in Eq. A1 (see Appendix A), that grows when the person p is looking towards the robot and decreases otherwise. Thanks to the transient phase introduced by the exponential rise/decay, we can filter quick glances out towards different directions. Rise/fall time is empirically set to 3.0 s. In the *User Social Interaction policy*, $w_i^{SI}(p_i, t)$ takes into account several cues related to proxemics, the estimation of the “robot gaze” and a time factor based on the persistence of the robot’s heading towards a person. Specifically, $w_i^{SI}(p_i, t)$ is achieved as the product of the following three factors, which vary between 0 and 1:

- a distance factor $w^d(p_i)$ which models the fact that closer persons are more likely interaction targets.
- a direction factor $w^{dir}(\mathbf{x})$ representing the “robot gaze” (abbreviated with *rg*) i.e., the area where the remote user aims at, that we represented with a Gaussian:

$$w^{dir}(\mathbf{x}) = 1 - \mathcal{N}(\gamma_r, \sigma_{rg}^2; \theta) \quad (9)$$

where $\theta = \text{atan2}(y, x)$ and γ_r is the current robot orientation. To tune σ_{rg} , we took inspiration from the human vision model. A recent research of [34] defined the effective visual field as the region where the discrimination of a simple figure can still be accomplished in a short period. According to this study, the effective visual field extends within 15° of eccentricity, so we set $\sigma_{rg} = 18^\circ$ to be more robust against possible oscillations in the robot motion.

- a time-dependent exponential weight $w^T(t)$ (see Appendix A) used to filter those situations where the robot quickly glances at somebody, or conversely when the robot tries interacting but its heading oscillates due to its motion. The *rise_time* of $w^T(t)$ is proportional to the distance between the person i and the robot. We suppose the closer the robot is to the person, the more probable interaction will happen, and coherently the peak of the exponential in w_d will grow. Instead, when the person i is not spotted by the robot gaze or it is outside the Hall’s Social Space (i.e., $d \geq 3.6$ m), $w^T(t)$ falls in *fall_time* s. The *fall_time* is set to 3 s to give time to the driver to adjust eventual unexpected oscillations.

2.2 The User Input Policy

The *User Input policy* handles inputs delivered by the human to correct the current robot’s behavior. In our system, such inputs correspond to continuous and sustained streams of directional commands in the left and right directions. It is also possible that the user does not deliver commands. Start and stop commands are also supported by the proposed system

simply to activate/deactivate the semi-autonomous navigation based on the fusion of the *policies*.

If the user is satisfied with the current robot’s trajectory, s/he is not required to intervene. The user can modify the robot’s trajectory using the input interface to trigger the *User Input policy* via directional inputs only when necessary.

The new version of the *User Input policy* creates an exponential distribution:

$$P^{UI}(\mathbf{x}, \Theta^{UI}) = w^{UI}(t) \cdot e^{-\frac{d_{UI}^2(\mathbf{x}, A_P)}{2\sigma_{UI}}} \quad (10)$$

where $\Theta^{UI} = (A_P, t)$ and $d_{UI}(\mathbf{x}, A_P)$ is the Euclidean distance measured from an application point $A_P = (x_{A_P}, y_{A_P})$. While such an application point was fixed in the previous version of the system, herein, it can move inside a semicircle with radius $R = 1$ m centered in the robot’s position (i.e., -90° and 90° from the robot’s position according to the user’s inputs), and it is computed as:

$$\begin{pmatrix} x_{A_P} \\ y_{A_P} \end{pmatrix} = R \begin{pmatrix} \cos \theta(t, dir) \\ \sin \theta(t, dir) \end{pmatrix} \quad (11)$$

where $\theta(t, dir)$ depends on the input stream as follows:

$$\theta(t + \Delta t, dir) = \theta(t) + dir \cdot \omega_0 \Delta t + \alpha_0 \Delta t^2 \quad (12)$$

where Δt is the time interval measured between the two last consecutive commands of the same type in the input stream, $\omega_0 = 0.05$ rad/s and $\alpha_0 = 0.05$ rad/s² are respectively the initial angular velocity and acceleration, $dir \in \{-1, 0, 1\}$ indicates the current directional command that can be respectively right, no command and left. Time is reset when a discontinuity in the input stream (e.g., a change in the class of the user’s input) is detected.

It is worth mentioning that, when no commands are delivered, the application point remains steady, but the intensity of the peak starts decreasing according to a time-dependent exponential weight $w_{CUI}(t)$ (see Appendix A), that we added in the newest version. $w_{CUI}(t)$ can also filter spurious commands out through the introduction of a transitory phase. If the distribution is completely suppressed—i.e., $p_{UI} < \epsilon \forall (x, y)$ -, $\theta(t, dir)$ is set to zero, so the position of the application point is re-initialized in front of the robot.

2.3 Fusion and Subgoal Update

The fusion strategy is a fundamental step to create a representation of the environment that is consistent with all the *policies*.

As introduced in Sect. 2, *policies* are essentially probabilistic grids. In particular, each location $\mathbf{x} = (x, y)$ indicates

the probability that the position \mathbf{x} is suitable to place a navigation *subgoal*, according to a specific policy P . In other words, $P(\mathbf{x}_1) \simeq 1$ means that \mathbf{x}_1 is a good navigation *subgoal*, $P(\mathbf{x}_2) \simeq 0.5$ means that $P(\mathbf{x}_2)$ is neither a good nor bad location since placing or not placing a *subgoal* in \mathbf{x}_2 are equally probable events. Finally, $P(\mathbf{x}_3) \simeq 0$ means that, surely, $P(\mathbf{x}_3)$ is not a suitable position for a *subgoal*, e.g. high probability of collision. For instance, the *Obstacle Avoidance policy* assigns low probability to the locations where obstacles are detected—these are like forbidden areas—, while high probability (i.e., close to 1) is assigned to locations far from an obstacle since they are suitable for placing the *subgoal*; an area in the middle between a circumscribed obstacle and the free spaces just nearby has a probability around 0.5 as neither good nor forbidden. Another example is the *Social policies*. High probability is given to locations where is likely to start an interaction, while the rest is set to 0.5. Note that there are no forbidden areas with low probability in this case since *Social policies* do not deliver such information.

In our system, we hypothesize that the fusion should assign the same weight to the *policies* related to the user's input and the robot's perception. In this way, we can handle both the situation where the robot's perception must overcome the user (e.g., a wrong input towards an obstacle) or, on the contrary, when the user must overcome the robot (e.g., wrong behavior), without authority switching. Considering this aspect and in line with the previous version of the system from [8], the fusion is the joint probability of all the simultaneous events modeled by the *policies* and it is easily computed as the element-wise product of them. However, in the future, such weights can be optimised (e.g., via a learning process) and/or tuned in practical settings in order to customize the robot's behaviors and favor the contributions of specific *policies* on the other ones.

Once we have obtained the fusion of all the *policies*, we can extract the *subgoal*. The *subgoal* is computed as the position with the highest probability in the fusion (i.e., the maximum). In the case of multiple maxima, we take one at random.

To make the system reactive to the dynamic motion of the surrounding people and changes in the environment, the *subgoal* S_t is updated with a fixed frequency (e.g., in our case set to 5 Hz in accordance with the system proposed by [35]) rather than at the occurrence of specific events as in [8]. Then, S_t is forwarded to the navigation module when it is far enough from the previous one, i.e. when $\|S_t - S_{t-1}\| < d_{th}$. This strategy allows the robot to autonomously stop in front of the target people (e.g., the position of the *subgoal* is not modified), and re-start moving when its context-awareness is significantly changed (e.g., thanks to the arrival of the user's commands or the people's disappearance). Therefore, the operator is not required to explicitly stop the robot to interact with a person, the proposed system can understand

when and where to stop interpreting the context information encoded in the fusion of the *policies*.

3 Materials and Methods of the Feasibility Study

3.1 Participants

This study involved 45 participants (S1–S45, 26.2 ± 8.3 years old, 23 female), 3 of them repeated the experiment with different roles (12 of them were asked to teleoperate the robot). Eight people have already experience with real robots, 24 among them have at least some theoretical knowledge in robotics, but none of them has previously used a mobile robot as required in this study. All participants voluntarily accepted to take part in the experiments and signed a written informed consent in accordance with the principles of the Declaration of Helsinki.

3.2 Robotic Platform

We used TIAGO++¹ from PAL Robotics (see Fig. 1) as a robotic platform for this study. It is composed of a differential drive base (diameter 0.54 m) and a humanoid upper body. It is equipped with a 2D laser range sensor on the front for obstacles detections. The robot's head integrates an Orbbec Astra RGB-D camera which outputs a $640 \times 480 @ 30$ fps video stream for people detection. Due to the lag observed in the video stream provided by the robot camera, we mounted on its head an Xtion Pro camera characterized by $1280 \times 1024 @ 30$ fps as resolution, to provide visual feedback about the current situation to the operator. The onboard PC is equipped with an Intel Core i7 (Haswell) CPU, 16 GB of RAM. We also used some external PCs: a desktop PC (Intel Core i7-7700 CPU, 16 GB of RAM) connected through Wi-Fi and a laptop (Intel Core i9-8950HK, 16 GB of RAM, NVIDIA GTX 1650 GPU) connected via Ethernet to the robot to run perception nodes (e.g., people detector).

3.3 Experimental Setup

To evaluate the proposed system, participants were required to teleoperate a mobile robot in a shared fashion (i.e., semi-autonomously), meaning that the robot trajectory depends both on the human input and the processing of the contextual information by the *policies* presented in Sect. 2.

The navigation task tested during the experiment was designed to assess the multiple features of our *shared*

¹ <https://pal-robotics.com/robots/tiago/>.

intelligence system: the traditional *obstacle-avoidance* capabilities, the *people-avoidance*, and the *social functionalities* including the estimation of the person's intention to interact with the robot and the robot's interaction with a group of people.

We involved four people per experiment with different roles: (i) the operator that drives the robot, (ii) one walking person in a corridor, (iii) two static people who, firstly, look in different directions (one gazing at the robot, the other ignoring it), and then move to another position where they talk each other without watching the robot. In detail, the social navigation task is performed in the area illustrated in Fig. 3, where we set three fixed target positions and two *Interaction Stations* that were marked on the floor. In the beginning, the robot is placed in the *S* position. Then, the operator should drive the robot along the corridor where a person *P1* is walking towards it (along a straight line), subsequently to the targets *T1* and *T2*. At this point, the robot should approach the first *Interaction Station* where only person *P3* is gazing at it to communicate her/his desire to interact, while the person *P2* is looking in a different direction (see Fig. 3). If the social navigation task is executed correctly, the robot stops in front of *P3* and the operator is instructed not to send directional commands for around 30 s to simulate a dialogue. After that, people start moving as a natural consequence of the end of the interaction, as illustrated in Fig. 3. The operator is required to send the appropriate left and right commands to reach target *T3* and then the second *Interaction Station*, simultaneously people *P1* and *P3* move there and talk together without gazing at the robot. The latter is expected to approach people *P1* and *P3* per effect of the human's commands. The robot stops for a few seconds for interacting with people, and finally, the operator has to drive the robot back to the target *T1*. During all the social navigation tasks,² the operator does not look directly at the robot, but he/she receives only the robot's camera streaming and position in the environment map as feedback. Although the stop command is supported by our system, in our experiment, we explicitly instructed participants not to use it, since we wanted to test our social *policies* and the capability of the system to autonomously take decisions (e.g., stop near to an interaction target) based on the environmental cues.

Participants were instructed on the task in a familiarization phase in which the experimenter explained the dynamic of the interaction as reported above and they acquired confidence in the system. However, subjects were not asked to follow specific trajectories. The operator was free to send high-level direction commands (i.e., turn left/turn right) or not to the robot at will. The experimenter has only indicated to the

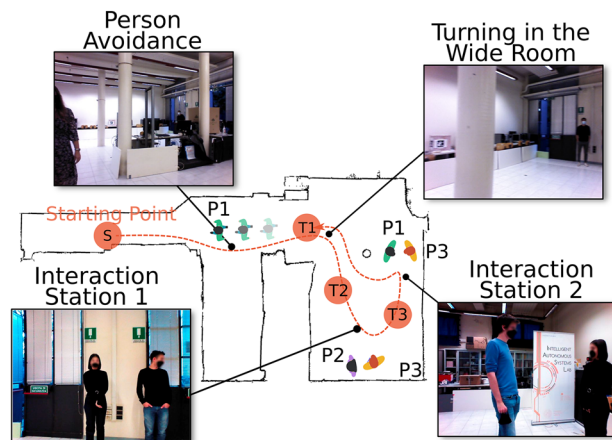


Fig. 3 The experimental setup. The operator is required to teleoperate the robot relying only on the robot's camera streaming from the starting position *S*. A possible robot's trajectory is represented in red. The social navigation task involves three other people *P1–P3* per run. First, *P1* walks towards the robot in the corridor to evaluate the *person-avoidance* ability. Then, we set three target positions *T1–T3* (marked with blue circles) to test the traditional navigation functionalities and two *Interaction stations* for validating the ability of the system to infer the will to interact respectively from the surrounding people and the operator. The task ends when the robot comes back to target *T1*. (Color figure online)

surrounding people when starting to move without providing any information on how to do it.

3.4 Examined Modalities

In this study, we evaluate the performance of our *shared intelligence* system considering the following three modalities, which we named:

- *SocShIn*: the human teleoperates the robot via a 2-class keyboard (turn left vs. turn right). The robot is endowed with the whole *shared intelligence* system described in Sect. 2 to achieve semi-autonomous teleoperation.
- *ShIn+SocLa*: this condition aims to assess the performance of the proposed system vs. an approach available in the literature for social navigation. With this purpose, we focused on *social_navigation_layers*³, the current standard in ROS for social navigation, proposed by [36]. However, to make the two systems comparable, we have kept the basic *policies* related to *obstacle avoidance* and *user's input* using a basic version of the *shared intelligence* system deprived of the new *Social policies* (see Sect. 2A). Hence, in this modality, the *Social policies* are replaced by the *social_navigation_layers*, that were integrated into the ROS navigation stack to achieve social

² Illustrative video: <https://cloud.dei.unipd.it/index.php/s/3YB6YPbiHwCzQHp>.

³ http://wiki.ros.org/social_navigation_layers.

navigation. In this modality, the human controls the robot through the same 2-class keyboard.

- **Joy**: the human directly (i.e., manually) teleoperates the robot namely the operator commands are implemented by the robot without considering the context information and any kind of robot's assistance. In this modality, no *shared intelligence* system is exploited. This condition is used as a reference.

Participants were required to perform two repetitions (i.e., runs) of the social navigation task described in Sect. 3.3 per modality. The testing order of the condition was random to avoid possible biases due to learning/fatigue effects. Overall, the experiment lasted about 1.5 h per participant.

After a total of 12 experiments, we collected 72 runs. We had to discard 12 runs because of failures of the robot's localization (out of the scope of this work), ending up with 60 runs, 20 for each modality.

3.5 Evaluation Methodologies

In this work, we consider the following metrics:

- **navigation_accuracy**: percentage of reached targets. We consider a target reached with a confidence interval of 0.54 m (i.e., the robot footprint diameter).
- **mean_accs**: it measures the average acceleration over the trajectory. The smaller acceleration is, the smoother the trajectory is.
- **concentration_time_ratio**: it is defined as the ratio between the time spent by the operator in delivering input commands to the robot and the total duration of the task.
- **fréchet_dist**: it measures the Fréchet distance [37] between the robot and the person's trajectory during the *people-avoidance* in the corridor, as

$$F(P, Q) = \inf_{\gamma} \max_{t \in [0, 1]} \{d(P(\gamma(t)), Q(\gamma(t)))\} \quad (13)$$

where $\gamma(t)$ is a parametrization of the curves P and Q and d is the Euclidean distance.

- **interaction_accuracy**: percentage of succeeded interactions. We consider an interaction successful when the robot stops at a maximum of 2 m away from the person in accordance with [38] and stays steady for at least 10 s.
- **interaction_social_dist**: it measures the Euclidean distance between the robot and people during the interaction, i.e., when the robot automatically stops near the target person.
- **discomfort_freq**: it measures, in percentage, the number of times the robot violates the Intimate Space, i.e. 0.45 m, when approaching a person.

The first three metrics are associated with navigation performance. We want to evaluate the robot's capacity of reaching some targets, contextualizing the human's commands, and performing smooth trajectories. Then, we focus on the two main functionalities of the proposed system: (i) avoiding people while moving, (ii) approaching people for interaction purposes. The *people-avoidance* capability is measured using the *fréchet_dist*: the larger the distance from the person is, the more comfortable and acceptable the trajectory results. The remaining metrics assess the robot's ability to accomplish social interaction tasks like autonomously approaching the desired person.

Moreover, we administered a questionnaire to participants about their experience and the perception of the human–robot interaction, at the end of each modality. With this purpose, the surveys were different according to the role of the participant in the experiment (e.g., the operator, the one walking in the corridor, and the static interacting people). The set of questions, listed in Table 1, were taken from previous studies in the field of people-aware navigation and teleoperation [8, 39] and adapted to our setup. The respondent was asked to choose where her/his position lies on a 5-point Likert-type (1 = Strongly Disagree, to 5 = Strongly Agree with a given sentence⁴). Finally, once completed the whole experiment, we asked participants which modality they preferred.

Acquired data have been statistically analyzed. A Kolmogorov–Smirnov test was performed to test the normality of each distribution. Given the results of the aforementioned on the data related to the performances of the systems, a One-way ANOVA ($p < 0.05$) was performed tailored by post hoc t -tests with Bonferroni for considering the multiple comparisons (i.e., $p < 0.05/3$) [40].

For the data analysis of the Likert-scale questions based on human evaluation, the Kruskal–Wallis tests were applied (Kolmogorov–Smirnov test $p > 0.05$) tailored by Dunn post hoc tests with Bonferroni correction given the multiple hypotheses tested (i.e., $p < 0.05/\text{number of questions}$).

4 Results

4.1 Navigation Performance

Although this work aims to introduce socially compliant behaviors, it is important the system maintains the traditional navigation performance (e.g., avoid static obstacles, considering the operator's inputs). From this point of view, the experiments were successfully completed by all the par-

⁴ We used different anchor labels for Q2 in *social interaction* questionnaire (i.e. 1 = Too close, 3 = Adequate, 5 = Too far) to have a global vision of people's perception about robot positioning during the interaction.

Table 1 Questionnaires administered per participant's role (↑= higher score is better, ↓ = lower score is better)

(a) Operator (<i>teleoperation</i>)	
Q1	It was easy to control the robot. (↑)
Q2	The robot's behavior seemed natural. (↑)
Q3	The robot's behavior was in line with your intentions. (↑)
Q4	The robot was responsive to commands. (↑)
Q5	You were afraid that the robot would collide with people. (↓)
Q6	The robot made interacting with people easier. (↑)
Q7	You would use the robot in a real context. (↑)
(b) Person walking in the corridor (<i>person-avoidance</i>)	
Q1	Did you feel comfortable moving close to the robot. (↑)
Q2	You felt scared of the robot. (↓)
Q3	You were afraid that the robot would collide with you. (↓)
Q4	The robot has bothered you. (↓)
Q5	You would use this robot in a real context. (↑)
(c) Static person interacting with the robot (<i>social interaction</i>)	
Q1	You felt comfortable interacting with the robot. (↑)
Q2	Rate the distance kept by the robot during the interaction. (See Footnote 4)
Q3	You were afraid that the robot would collide with you. (↓)
Q4	The robot's behavior was in line with your expectations. (↑)
Q5	You felt scared of the robot. (↓)
Q6	You would accept the presence of the robot in a real context. (↑)

ticipants and no collisions happened in the three examined modalities. Figure 4 shows the heat maps of the trajectories performed by the robot. The results are in line with our expectations. In the case of *SocShIn* and *ShIn+SocLa*, there is more variability in the trajectories than the *Joy* due to the attitude of the participants (more in control vs. more robot's autonomy), especially in the less constrained areas (e.g., around the targets *T2* and *T3*). However, a greater number of outliers appear in the *ShIn+SocLa* than in *SocShIn* (e.g., in the area around the *Interaction stations*).

Most of the navigation targets were correctly reached over the runs. We achieved a *navigation_accuracy* equal to $92.86 \pm 11.57\%$, $79.35 \pm 16.23\%$ and $75.0 \pm 19.37\%$ respectively in *Joy*, *SocShIn* and *ShIn+SocLa*. Coherently with the trajectories, missing targets mainly occurred at *T2* and *T3*.

Nevertheless, the trajectories in *SocShIn* and *ShIn+SocLa* result smoother than *Joy* by analysing the average acceleration reported in Fig. 5. This aspect is fundamental for people's comfort and for easing the predictions of the next robot's motion. We found statistical differences among the three distributions (One-way ANOVA, $p_A = 1.1 \times 10^{-13}$), in particular, the acceleration in both *SocShIn* and *ShIn+SocLa* were significantly lower and more constant than *Joy* (i.e., respectively $p_t = 6.07 \times 10^{-9}$ and $p_t = 2.56 \times 10^{-8}$ achieved via post hoc tests).

Finally, since we are focusing on navigation during teleoperation, it is worth mentioning the *concentration_time_ratio* to evaluate the time dedicated by the operator to deliver commands and the level of the robot's autonomy guaranteed by the system. Surely, the *concentration_time_ratio* is 100% in *Joy* because the operator directly teleoperates the robot. The other two conditions reported a score respectively of 26.70% for *SocShIn* and 28.49% for *ShIn+SocLa*, with a slightly reduction in the proposed system.

4.2 People Avoidance Performance

This section assesses the *people avoidance* capabilities of the system by focusing on the *fréchet_dist*. Figure 6 highlights the distributions resulting from the interaction between the walking person and the teleoperated robot in the corridor (see Fig. 3), compared to Hall's intervals. It is worth noticing that both *SocShIn* and *ShIn+SocLa* significantly perform better than *Joy* as qualitatively emerged from the trajectories (see Fig. 4 there is a more marked deviation in the two conditions than *Joy*). The One-way ANOVA returned a p -value $p_A = 3.6626 \times 10^{-9}$, while the post hoc t -test $p_t = 2.2387 \times 10^{-10}$ between *Joy* and *SocShIn* and $p_t = 2.1699 \times 10^{-7}$ between *Joy* and *ShIn+SocLa*. Although most values belong to the Personal Space, the results are consistent considering the narrow area around the corridor

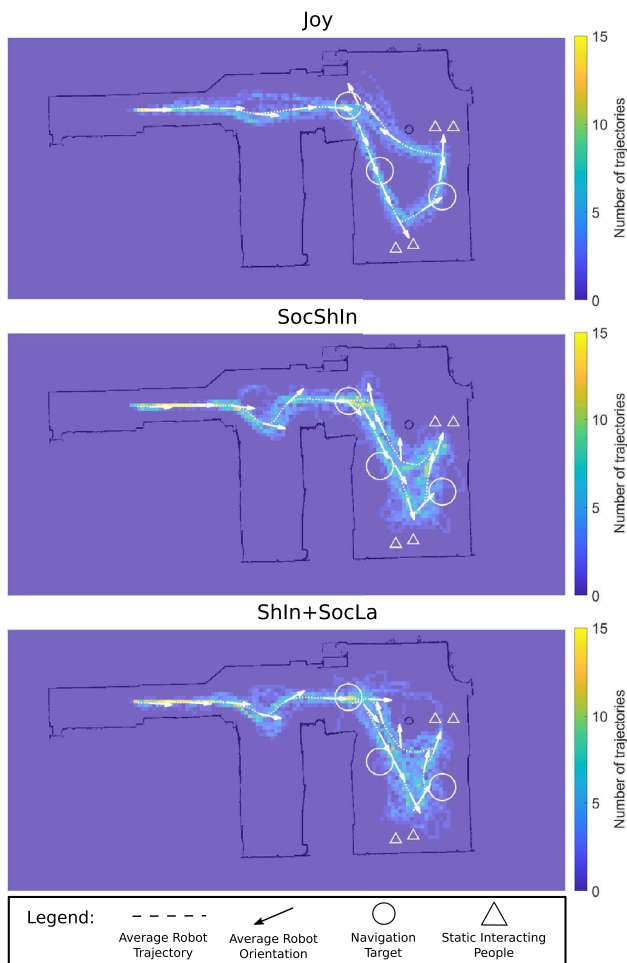


Fig. 4 Heat maps of the trajectories completed by the robot in the three tested modalities. Maps resolution is 15 cm. The color palette ranges from blue (less frequent) to yellow (more frequent). The dashed line represents the average trajectory. (Color figure online)

(i.e., corridor width = 2.20 m, robot diameter = 0.54 m). No significant difference has been found between *SocShIn* and *ShIn+SocLa* ($p_t = 9.6331 \times 10^{-1}$). This result suggests that the system provides comparable *people-avoidance* functionalities with the current ROS standard. However, it is worth highlighting that, the robot never crossed the *Intimate Space* in *SocShIn* as happened with *ShIn+SocLa*. Furthermore, by focusing on the distributions, the variance resulting from the proposed system is less than the one in *ShIn+SocLa*, implying more stability. Summing up, we can state that the performance of *SocShIn* and *ShIn+SocLa* are comparable in terms of *people avoidance*, with a slight advantage for the first one. However, it is worth highlighting that *SocShIn* has the advantage to interact with the other *policies* in the fusion and directly contribute to the merge of the different information behind the choice of the *subgoal*.

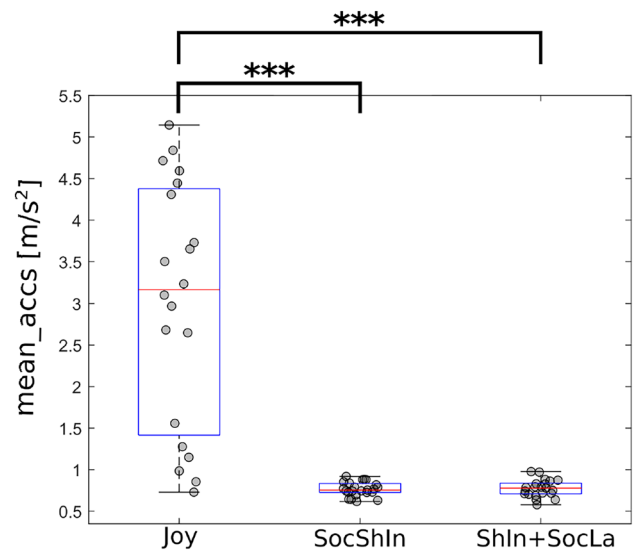


Fig. 5 Distribution of the *mean_accs* per modality. On each box, the red line indicates the median, and the bottom and top edges of each box represent the 25th and 75th percentiles, respectively. Statistically significant differences are reported with One-way ANOVA tailored by *t*-test post hoc tests with Bonferroni correction, *** $p \lll 0.01$. (Color figure online)

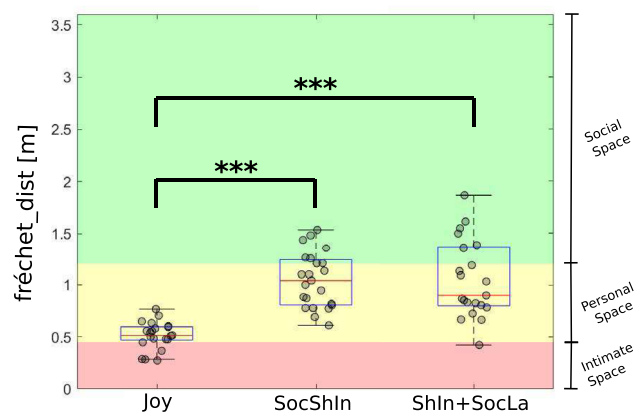


Fig. 6 Distribution of the *fréchet_dist* per modality. On each box, the red line indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. Statistically significant differences are reported with One-way ANOVA tailored by *t*-test post hoc tests with Bonferroni correction, *** $p \lll 0.01$. The colors highlight the Hall spaces [1]. (Color figure online)

4.3 Interaction Performance

It is worth reminding that the most novelty contribution of this work is associated with the robot's capability of autonomously triggering the interaction and inferring the target people. By considering both those aspects in verifying the number of times the robot correctly stopped towards the target people at *Interaction stations*, we achieved a *interaction_accuracy* of 63.04% in *SocShIn* vs. 26.19% in *ShIn+SocLa*. Surely, as expected, the *interaction_accuracy*

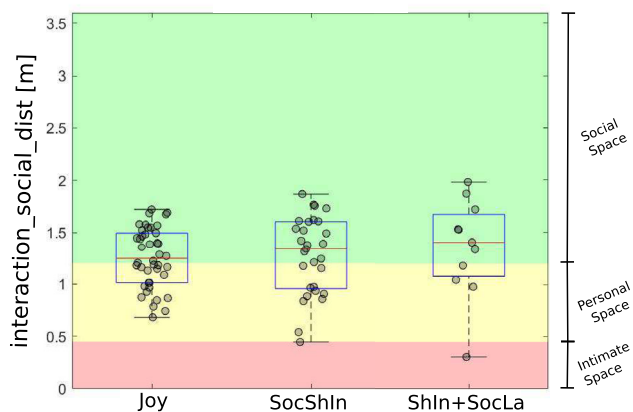


Fig. 7 Distribution of the *interaction_social_dist* per modality. On each box, the red line indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The colors highlight the Hall spaces [1]. (Color figure online)

in *Joy* was 100% because, in this case, the robot stops only per effect of the user's decision (e.g., no assistance from the robot). By analysing the proxemics through the *interaction_social_dist* in the successful interactions that we represent in Fig. 7 with respect to Hall intervals introduced in [1], no significant difference emerged from the One-way ANOVA test ($p_A = 7.3369 \times 10^{-1}$). This outcome suggests the proposed system (i.e., *SocShIn*) is able to keep the expected distance from the target people as it would happen when the operator chooses to stop (i.e., *Joy*). The comparisons with *ShIn+SocLa* might not be relevant for the limited number of correct interactions (i.e., 26.19%). Nevertheless, considering the successful interactions, the *discomfort_freq* achieved in *ShIn+SocLa* appears higher than in *SocShIn*. This might suggest that *ShIn+SocLa* violates the Intimate Space of people more often than the proposed system. Furthermore, coherently with the results shown in Fig. 4, the trajectories tend to be more widespread in *ShIn+SocLa* than in *SocShIn*, suggesting the operator's difficulty in stopping at *Interaction Stations* as also arisen from the *interaction_accuracy*. Moreover, in *ShIn+SocLa*, the robot did not respect the group behavior passing in the middle between the people *P1* and *P3* twice vs. no violations in the other two conditions.

4.4 Human Evaluation

Herein, we analyse the results from the three kinds of questionnaires administered to collect subjective feedback about the *teleoperation*, the *people-avoidance*, and the *social interaction* (see Table 1) in the three modalities. Figure 8 reports the results per typology of participants: the operators (i.e., 12 answers), the people walking in the corridor (e.g., 12 answers from *P1* see Fig. 3) and the static people at the *Interaction stations* (e.g., 36 answers from *P1–P3* see Fig. 3). The left vertical axis refers to the number of answers, while the right

one to the questionnaire score (1 = Strongly Disagree, to 5 = Strongly Agree). The average of the questionnaire scores is marked with a grey circle. Furthermore, we evaluate the questionnaire scores with respect to the distribution of answers. To simplify the visualization, we gathered the questionnaire scores into three options (1–2 = Disagree, 3 = Neutral, 4–5 = Agree), that we represent with different intensities in the colors associated with the modalities.

4.4.1 Teleoperation Questionnaire

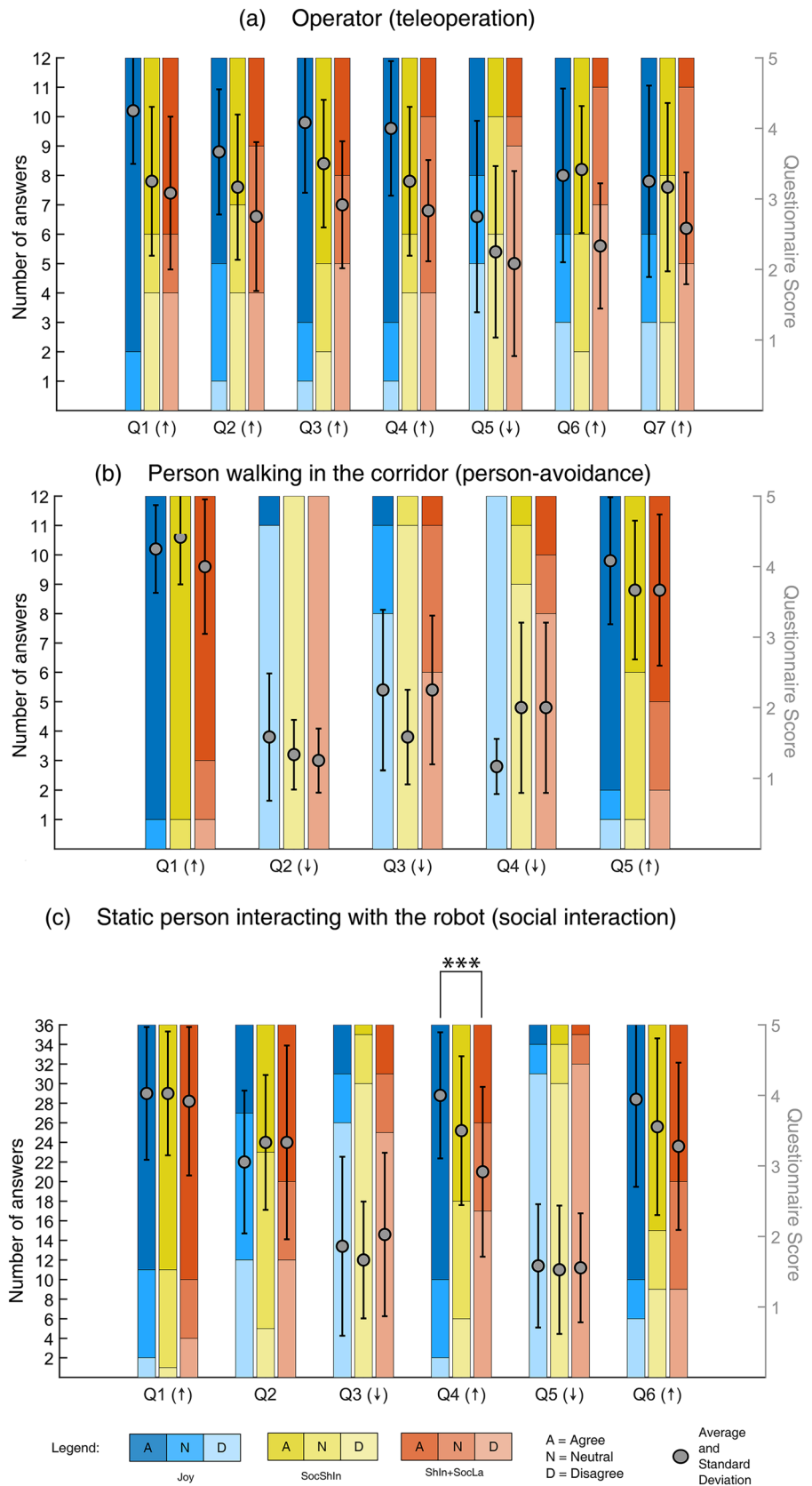
The operators were asked to evaluate their experience focusing on the *teleoperation* and the assistance provided by the robot in the social navigation task, the responsiveness of the systems, and the consistency with the operator's intentions. The results are shown in Fig. 8a.

We found statistical differences in the ease of controlling the robot (Q1, Kruskal-Wallis test $p_K = 0.015$), consistency of the robot behavior with the operator's intentions (Q3, Kruskal-Wallis test $p_K = 0.021$), responsiveness to operator's commands (Q4, Kruskal-Wallis test $p_K = 0.018$), and the ability of system to facilitate the interaction with people among the modalities (Q6, Kruskal-Wallis test $p_K = 0.026$). However, examining the Dunn post hoc tests, no significant differences were observed between *Joy* and *SocShIn* (Q1–Q7, Dunn post hoc test, $p_D > 0.05$). This result suggests that the semi-autonomous robot's behaviors in *SocShIn* reflected the operator's intentions and the expected reactivity much like when they were under direct human control.

The robot endowed with *ShIn+SocLa* seemed more difficult to control (Q1, Dunn post hoc test, $p_D = 0.025$), less coherent with the operator's expectations (Q3, Dunn post hoc test, $p_D = 0.017$), and less responsive (Q4, Dunn post hoc test, $p_D = 0.015$) than in the *Joy* modality, but such differences were not statistically significant anyway considering the Bonferroni correction.

Analyzing question Q5 about the perception of colliding with people, the scores related to both *SocShIn* and *ShIn+SocLa* are slightly lower than the ones in *Joy*, suggesting that drivers trusted the autonomous *people-avoidance* capabilities of the systems. However, the answers to question Q6 confirmed that it was easier to interact with people by the operator in the *SocShIn* than *ShIn+SocLa* (Q6, Dunn post hoc test, $p_D = 0.045$), even if the difference was not statistically significant by applying the Bonferroni correction. Finally, Q7 shows that the participants would prefer teleoperating the robot in a real-world scenario via *SocShIn* than *ShIn+SocLa* and, almost as much as in *Joy*. Furthermore, 66.5% of them have chosen *SocShIn* as their favorite system. Overall, *SocShIn* achieved greater consensus than *ShIn+SocLa* from the operator side.

Fig. 8 Results from the three questionnaires related to *teleoperation*, the *people-avoidance* and the *social interaction*. The left vertical axis reports the distribution of the answers, while the right one refers to the questionnaire score (1 = Strongly Disagree, to 5 = Strongly Agree). The average and the standard deviation of the questionnaire scores are shown (through a grey circle). The different color intensities in the bars represent the correlation between the distribution of the answers and the questionnaire scores in the three modalities. For this purpose, we converted the questionnaire scores into three options (1–2 = Disagree, 3 = Neutral, 4–5 = Agree) for simplifying the visualization. Statistically significant differences are reported: * $p < 0.05$, ** $p < 0.01$, *** $p < << 0.01$. (Color figure online)



4.4.2 Person-Avoidance Questionnaire

People who walked in the corridor close to the robot were required to assess the robot's motion and respect the social distance. It is worth highlighting that the results from Q1–Q2 in Fig. 8b show the proposed system is perceived as slightly more comfortable than the other modalities and reduces the people's fear of the robot than *Joy*. Consistently, from answers to question Q3, *SocShIn* is considered the safest system since people had the least perception of crashing with the robot. The remaining questions show homogenous scores between *SocShIn* and *ShIn+SocLa* in accordance with the results from the objective metrics.

We did not find significant differences (Q1–Q5, $p_K > 0.05$ Kruskal–Wallis tests) among the modalities in the answers to the person-avoidance questionnaire, suggesting that the three systems were comparable according to the people who walked in the corridor.

4.4.3 Social Interaction Questionnaire

Static people interacting with the robot were requested to judge the robot's stop and the observation of the proxemics rules during the interaction. The results in Fig. 8c report very similar scores in terms of comfort (i.e., Q1) and the level of fear towards the robot among the modalities (i.e., Q5). Considering answers to question Q2 about the distance kept from the robot during the interaction, *SocShIn* is slightly more appropriate than the other two systems. Coherently, responses to question Q3 confirm that in *SocShIn* the robot's behaviors were perceived as the least intimidating (less perception of collision) than the other modalities. Significant differences were found in the scores related to question Q4 about the consistency of the robot's behaviors with the people's intentions (Q4, Kruskal–Wallis test $p_K = 3.184 \times 10^{-4}$) and question Q6 about the acceptance of using the system in a real context (Q6, Kruskal–Wallis test $p_K = 0.043$). The Dunn post hoc test revealed that *ShIn+SocLa* significantly differed from *Joy* in terms of coherence of the robot's behaviors with the surrounded people's intentions (Q4, Dunn post hoc test, $p_D = 1.82 \times 10^{-4}$). In addition, the robot endowed with *ShIn+SocLa* seemed less accepted outside the laboratory (Q6, Dunn post hoc test, $p_D = 0.037$) than in *Joy* modality, but the difference was not statistically significant by applying the Bonferroni correction. The scores related to *SocShIn* were comparable to the ones associated with *Joy* for both the questions (Q4, Q6 Dunn post hoc test, $p_D > 0.05$). Therefore, a higher level of agreement was observed again in favor of *SocShIn* compared to *ShIn+SocLa* among people interacting with the robot.

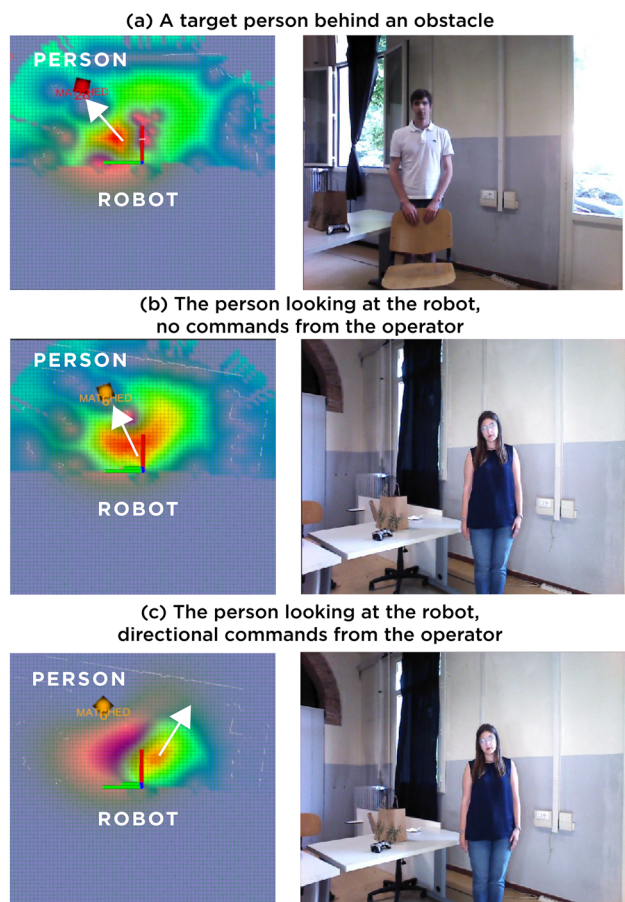


Fig. 9 Experiments showing the functioning of the system in three edge cases: **a** an obstacle is in front of a target person who wishes to interact with the robot; **b** a person stops to the left of the robot and continues looking at it. The areas to the left and right of the robot have similar environment configurations; **c** the operator does not wish to interact with the target person identified by the system in **b** and sends directional commands to avoid the interaction. The fusion of the *policies* is shown on the left, with the *User Social Interaction policy* overlaid in transparency, and the *subgoal* is represented via a white arrow. On the right, the robot's camera image is shown per each situation

4.5 Experiments on Select Edge Cases

The main experiment reported above was designed to test the overall capabilities of the system and in particular the new *social policies*. We have verified the performance in typical situations that can happen with telepresence robots for monitoring people and we focused on the evaluation according to the different roles of the people involved. However, it was not possible to experiment many edge cases due to time constraints. As stated in Sect. 3.4, each experiment took 1.5h on average which is already demanding for volunteers.

Herein, we present supplementary qualitative results achieved from experiments on select edge cases that we performed later to verify the fusion of the *policies* and the *subgoal* resulted in the following edge cases: (a) an obstacle

is placed in front of a target person who wishes to interact with the robot; (b) the robot is in the middle of two areas comparable in terms of environment configuration (e.g., similar positions of obstacles). Then, a person stops to the left of the robot and continues looking at it; (c) in the previous situation, the operator, teleoperating the robot, does not like to interact with the target person identified by the system (such a person continues gazing the robot), and hence, sends directional commands to avoid the interaction. Figure 9 shows the fusion of the *policies* in these situations by highlighting the activation of the *Social Interaction policies* and the corresponding *subgoal* represented via a white arrow.

In the first edge case in Fig. 9a, the *subgoal* is placed in front of the target person and before the obstacle. It is worth noticing that the *Obstacle-avoidance policy* prevents any collision, while the final interaction region is shifted to a suitable location considering both the presence of the person and the obstacle.

As regards the second scenario in Fig. 9b, as expected, the surrounding person that continues looking at the robot over time makes the *Person Social Interaction* activate (e.g., the related peak in the probability distribution can be seen in Fig. 9b). The operator leaves the robot heading towards the same person, and hence, the *User Social Interaction policy* infers the operator would like to interact with the person. Both *Social Interaction policies* are consistent and the *subgoal* is placed close to the target person for starting the interaction. Differently, in the last situation in Fig. 9c, if the operator does not wish to interact with the target person and sends directional commands (i.e., right in the illustrated example), the *User input policy* is activated (e.g., a new peak appears on the right), the robot points towards a different direction as a consequence. Therefore, the *User Social Interaction policy* infers the operator would not like to interact with the person. The fusion is modified accordingly and, the *subgoal* is set on the right, which differs from the previous scenario.

The qualitative results achieved in these experiments on select edge confirm the expected functioning of the proposed system in such edge cases.

5 Discussion

In this paper, we propose a system for achieving social navigation behaviors during teleoperation. The main novelty of this work is to show for the first time the robot's capacity to infer the will to interact from the operator and the surrounding people and then behave consequently. Furthermore, the presented system is also able to manage *people avoidance* behaviors respecting social distances and group formation. One relevant aspect with respect to other previous approaches consists of the way to choose the robots' behaviors. In our system, both traditional navigation and

social behaviors are not coded and activated at the occurrence of specific events, on the contrary, they result from the fusion of the probabilities distribution provided by *policies*, making the system modular and appealing. From our tests involving 45 participants, the system shows socially compliant behaviors coherently with the situations (e.g., avoidance and interaction) and the social norms without affecting the traditional navigation capabilities. In addition, by evaluating the examined modalities, overall, *SocShIn* performs better than *ShIn+SocLa* considering the quantitative metrics and the questionnaire. The robot's trajectories are simpler and easily predictable by the surrounding people who feel more comfortable.

Results comparisons with other social navigation studies may be complex and inappropriate due to different testing conditions and experimental setups. However, it is worth highlighting that our results are consistent in terms of proxemics social rules with the findings from previous studies. For instance, the recent work by Teja et al. [24] presents a tunable human-aware navigation planner with different modes to manage a variety of contexts populated by people. In their experiments, the robot keeps an average minimum distance of 1.29 m from the person in open spaces and 0.66 m and 0.89 m respectively in narrow and pillar corridors, which are in line with the distances in the range [0.61 m, 1.53 m] (1.28 m on average) achieved in our tests (see Fig. 6). Similarly, our results satisfy the constraints found in the study proposed in [41], where different passing distances between the person and the robot in a corridor have been evaluated in terms of acceptability in a setup similar to ours. Specifically, the authors found that people prefer robots to stay out of their intimate space (≤ 0.45 m) when they pass each other in a 2.5 m wide corridor, which always occurs in the *SocShin* modality in our experiments. Furthermore, the minimum robot-distance in *SocShin* is also in line with the real-time results in [14] (i.e., 0.61 m in our vs. 0.56 m in [14] respectively), that already demonstrated to be safer with respect to other state-of-the-art approaches based on social forces (e.g., APF, FTG-SC, SPF-SC).

Several studies in the literature have focused on *person-avoidance* that could be mentioned, however, since the novelty part of this work is based on the robot's prediction to interact in a social manner, herein it is worth discussing the interaction performance. For this purpose, for instance, we notice that our results have been consistent with the findings from Repiso et al. [35] that have proposed a method based on the Social Force Model to enhance the side-by-side navigation. Such a system is designed to accompany and approach walking people, as well as predict the best meeting point considering the group formation and the future target person's position. Although the scenario and the application are different than the one proposed in this paper, most of the *interaction_social_dist* in our experiments belong to the social

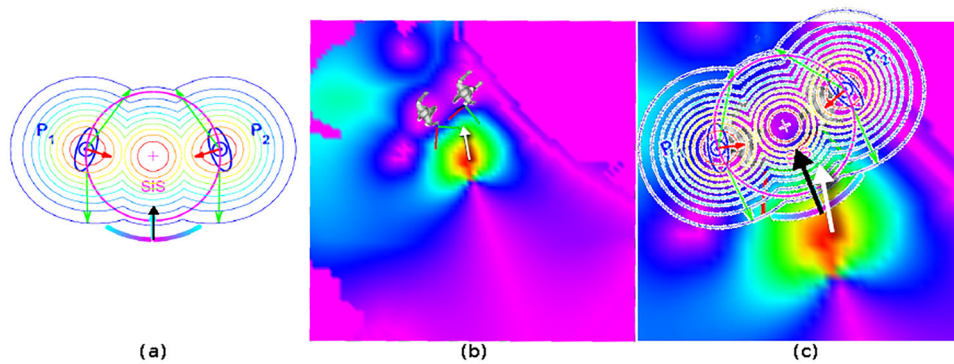


Fig. 10 Qualitative comparison of robot's approaching to social group formation. **a** The dynamic social zone (DSZ) model from [26, 42] showing the approaching pose for the robot with respect to person P1 and person P2. **b** The output fusion from the proposed *SocShIn* when the

robot is approaching two people. **c** The DSZ model in **a** and the resulting fusion in **b** are overlapped. Black arrow and white arrow represent the robot's approaching poses computed respectively with DSZ and *SocShIn*

space (average $d_{our} = 1.279 \pm 0.3748$ m), and precisely to the interval [1.25–2 m] estimated as good performance by [35] according to their validation both in simulation and on the real robot. Similarly, the works from [26, 42] have proposed the concept of the dynamic social zone (DSZ) to represent the space around humans and predict the best approaching robot's pose to people. Among the set of metrics, we have estimated the SDI index from [26] used to evaluate the approach direction of the robot to the humans. On average, we have obtained a value of 0.66 on our data which is coherent with the ones reported by authors in the most similar conditions (i.e., 0.72 when the robot is approaching only one person close to an obstacle, 0.62 in the case of two people), suggesting that in our system, the robot approached humans in the proper position and direction. Another relevant aspect modeled in DSZ is group relationships. Authors in [26, 42] explicitly detect group formations to embed the information in the model managing the social navigation. In our system, although we do not insert any a priori knowledge about group formations, some group relationships arise from the fusion of the *policies*. Figure 10 represents the qualitative comparison between the two systems restricted to the case of two people, showing again small differences.

Finally, it is worth noticing that differently from other studies, in our system, the robot is simultaneously teleoperated by the operator whose commands might lead to less safe human-person distances as observed in *Joy* modality (see Sect. 4.2), but this fact is mediated thanks to the robot's intelligence in the *SocShIn*. Moreover, considering the operator's intervention based on the *concentration_time_ratio*, our results are again consistent with other previous experiments based on *shared control* and *shared autonomy* algorithms. For instance, in [43], participants were required to control a robotic avatar in the lab remotely from their homes through an app. In this context affected by possible net-

work delays, participants let the simulated robot implement social navigation in autonomy without interacting for more than 50% of the entire time. Similarly, in [44], participants have provided high-level goals via a brain-machine interface, less reactive and accurate than the keyboard, to be reached in autonomy by a telepresence robot, and achieved a *concentration_time_ratio* equal to 28% vs. 26.70% (in our *SocShIn*). These findings might open future perspectives of our method to augment the human–robot social interaction in these applications where the robot's intelligence is fundamental to handle the situations when the user cannot interact [8, 45].

6 Conclusion

In this work, we have proposed a *shared intelligence* approach for telepresence robot navigating in environments populated by people. In our system, the robot exhibits the capabilities of: (a) avoiding people, (b) autonomously inferring the intention from the operator and the surrounding people to interact with each other, and in case, (c) approaching people properly for starting the interaction (e.g., a dialogue). To the best of our knowledge, this paper has the following contributions. First, people are not treated as simple obstacles/goals according to the driver's commands as traditionally happens, but both the inclination of the driver and the other people around are factors that equally determine the next robot's behaviors. The former is associated with the driver's commands, the latter is estimated from the people's gaze—in both cases they are not set a priori.

Second, it is the first attempt that such social and teleoperated behaviors result from the fusion of multiple *policies*, representing heterogeneous information combined with the same influence and then, validated in a feasibility study with more than 40 participants.

The tests with the real robots have revealed the presence of satisfactory social-compliant behaviors that are coherent with the expected comfort, naturalness, and sociability principles as reflected in the quantitative metrics and the answers of the participants to the questionnaires. Moreover, the results are also in line with related state-of-the-art studies. The comparison of the proposed system with teleoperation points out a higher smoothness in the robot's trajectories and a safer and more acceptable *people-avoidance*. The introduction of *Social Interaction policies* in the *shared intelligence system* provides better sociability compared with the *shared intelligence system* endowed with *ROS social_navigation_layers*, thanks to the robot's capability of better approaching people.

However, both approaches have some limitations mainly related to the perception and the localization modules (out-of-the-scope of this work). Indeed, it has been assumed to exploit robust people and robot tracking and gaze detection modules to ensure the expected functioning of both systems. For instance, in our lab-setting experiments, we had to discard some runs due to the robot's delocalization, while the perception outputs were consistent. Probably, in more crowded scenarios, the robot's perception based on the onboard sensors should be strengthened with the introduction of additional environment sensors that allow better detecting and tracking both the artificial and the human agents from different perspectives (e.g., top view) and hence to establish coherent socially-compliant behaviours.

In the future, we will further investigate the rising of other high-level social behaviors achieved from the fusion of *policies* in more challenging out-of-the-lab environments where the *policies* can interact in more complex fashions. For instance, during this experimentation, we also observed a sort of *person-following* behavior emerging from our system which requires further evaluations. Moreover, it is worth highlighting that the modularity of the system allows to easily extend the current robot's functionalities by adding new *policies*. Another interesting investigation may include the application of learning-based methods, especially for the fusion of the *policies*, using the data already recorded in our experiments.

Supplementary Information

This article is accompanying by the supplementary video available at <https://cloud.dei.unipd.it/index.php/s/3YB6YPbiHwCzQHP> showing the experimental setup.

Appendix A: Implementation Details

A.1: Time-Dependent Weights

In our system, to avoid abrupt changes, we have modulated the amplitude of the distributions behind the *policies* by a time-dependent exponential weight that is defined as:

$$w_t(t+1) = \begin{cases} w_t(t) \frac{1}{\epsilon} \frac{1}{f^{T_R}} & \text{if rising} \wedge \epsilon < w_t(t) < 1 \\ w_t(t) \epsilon \frac{1}{f^{T_F}} & \text{if falling} \wedge \epsilon < w_t(t) < 1 \\ 0 & \text{if } w_t(t) \leq \epsilon \\ 1 & \text{if } w_t(t) \geq 1 \end{cases} \quad (\text{A1})$$

where f is the update rate of the system, T_R and T_F are the desired rising and falling periods respectively and ϵ is a threshold under that we consider finished the transient.

A.2: Software Frameworks

The system is integrated into Robot Operating System (ROS). The navigation system relies on ROS *navigation stack*⁵, with a TEB local planner and the default parameters. For people detection and tracking, we exploit the SPENCER Multi-Modal People Detection and Tracking Framework from [46] in combination with a PCL detector designed by [47]. Gaze estimation is performed through RT-GENE [48]. Finally, the outputs of *policies* are efficiently stored using Grid Map library from [49].

Acknowledgements This research was partially supported by MIUR (Italian Minister for Education) under the initiative "Departments of Excellence" (Law 232/2016). The authors would like to thank all participants.

Funding Open access funding provided by Consiglio Nazionale Delle Ricerche (CNR) within the CRUI-CARE Agreement.

Data Availability The authors declare that [the/all other] data supporting the findings of this study are available within the article [and its supplementary information files]

Declarations

Conflict of interest The authors declare that they have no Conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence,

⁵ <http://wiki.ros.org/navigation>.

unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Hall ET (1966) The hidden dimension: man's use of space in public and private. The Bodley Head, London
- Kruse T, Pandey AK, Alami R et al (2013) Human-aware robot navigation: a survey. *Robot Auton Syst* 61(12):1726–1743. <https://doi.org/10.1016/j.robot.2013.05.007>
- Kendon A (1990) Conducting interaction: patterns of behavior in focused encounters. Cambridge University Press, Cambridge
- Mois G, Beer JM (2020) The role of healthcare robotics in providing support to older adults: a socio-ecological perspective. *Curr Geriatr Rep*. <https://doi.org/10.1007/s13670-020-00314-w>
- Niemelä M, Van Aerschoot L, Tammela A et al (2021) Towards ethical guidelines of using telepresence robots in residential care. *Int J Soc Robot* 13(3):431–439
- Isabet B, Pino M, Lewis M et al (2021) Social telepresence robots: a narrative review of experiments involving older adults before and during the COVID-19 pandemic. *Int J Environ Res Public Health* 18(7):3597
- Beraldo G, De Benedictis R, Cesta A et al (2023) Toward AI-enabled commercial telepresence robots to combine home care needs and affordability. *IEEE Robot Autom Lett* 8(10):6691–6698. <https://doi.org/10.1109/LRA.2023.3310396>
- Beraldo G, Tonin L, Millán JDR et al (2022) Shared intelligence for robot teleoperation via bmi. *IEEE Trans Hum-Mach Syst* 1:1. <https://doi.org/10.1109/THMS.2021.3137035>
- Koren Y, Borenstein J et al (1991) Potential field methods and their inherent limitations for mobile robot navigation. In: ICRA, pp 1398–1404
- Helbing D, Molnar P (1995) Social force model for pedestrian dynamics. *Phys Rev E* 51(5):4282
- Ferrer G, Garrell A, Sanfeliu A (2013) Robot companion: a social-force based approach with human awareness–navigation in crowded environments. In: 2013 IEEE/RSJ international conference on intelligent robots and systems. IEEE, pp 1688–1694
- Hoshino S, Maki K (2015) Safe and efficient motion planning of multiple mobile robots based on artificial potential for human behavior and robot congestion. *Adv Robot* 29(17):1095–1109. <https://doi.org/10.1080/01691864.2015.1033461>
- Pradhan N, Burg T, Birchfield S (2011) Robot crowd navigation using predictive position fields in the potential function framework. In: Proceedings of the 2011 American control conference, pp 4628–4633. <https://doi.org/10.1109/ACC.2011.5991384>
- Reddy AK, Malviya V, Kala R (2021) Social cues in the autonomous navigation of indoor mobile robots. *Int J Soc Robot* 13(6):1335–1358
- Bennewitz M, Burgard W, Cielniak G et al (2005) Learning motion patterns of people for compliant robot motion. *Int J Robot Res* 24(1):31–48
- Lu DV, Smart WD (2013) Towards more efficient navigation for robots and humans. In: 2013 IEEE/RSJ international conference on intelligent robots and systems. IEEE, pp 1707–1713
- Cheng J, Cheng H, Meng MQH et al (2018) Autonomous navigation by mobile robots in human environments: a survey. In: 2018 IEEE international conference on robotics and biomimetics (ROBIO). IEEE, pp 1981–1986
- Rösman C, Oeljeklaus M, Hoffmann F et al (2017) Online trajectory prediction and planning for social robot navigation. In: 2017 IEEE international conference on advanced intelligent mechatronics (AIM). IEEE, pp 1255–1260
- Vemula A, Muelling K, Oh J (2017) Modeling cooperative navigation in dense human crowds. In: 2017 IEEE international conference on robotics and automation (ICRA). IEEE, pp 1685–1692
- Boldrer M, Antonucci A, Bevilacqua P et al (2022) Multi-agent navigation in human-shared environments: a safe and socially-aware approach. *Robot Auton Syst* 149(103):979
- Chen YF, Everett M, Liu M et al (2017) Socially aware motion planning with deep reinforcement learning. In: 2017 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, pp 1343–1350
- Chen C, Liu Y, Kreiss S et al (2019) Crowd-robot interaction: crowd-aware robot navigation with attention-based deep reinforcement learning. In: 2019 international conference on robotics and automation (ICRA). IEEE, pp 6015–6022
- Bacchin A, Beraldo G, Menegatti E (2021) Learning to plan people-aware trajectories for robot navigation: a genetic algorithm*. In: 2021 European conference on mobile robots (ECMR), pp 1–6. <https://doi.org/10.1109/ECMR50962.2021.9568804>
- Teja Singamaneni P, Favier A, Alami R (2021) Human-aware navigation planner for diverse human–robot interaction contexts. In: 2021 IEEE/RSJ international conference on intelligent robots and systems (IROS), pp 5817–5824. <https://doi.org/10.1109/IROS51168.2021.9636613>
- Sebastian M, Banisetty SB, Feil-Seifer D (2017) Socially-aware navigation planner using models of human-human interaction. In: 2017 26th IEEE international symposium on robot and human interactive communication (RO-MAN), pp 405–410. <https://doi.org/10.1109/ROMAN.2017.8172334>
- Truong XT, Ngo TD (2018) “to approach humans?”: a unified framework for approaching pose prediction and socially aware robot navigation. *IEEE Trans Cogn Dev Syst* 10(3):557–572. <https://doi.org/10.1109/TCDS.2017.2751963>
- Narayanan VK, Spalanzani A, Pasteau F et al (2015) On equitably approaching and joining a group of interacting humans. In: 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, pp 4071–4077
- Yang F, Peters C (2019) App-Istm: data-driven generation of socially acceptable trajectories for approaching small groups of agents. In: Proceedings of the 7th international conference on human-agent interaction, pp 144–152
- Boucher JD, Pattacini U, Lelong A et al (2012) I reach faster when i see you look: gaze effects in human-human and human–robot face-to-face cooperation. *Front Neurobot* 6:3. <https://doi.org/10.3389/fnbot.2012.00003>
- Senju A, Hasegawa T (2005) Direct gaze captures visuospatial attention. *Vis Cogn* 12(1):127–144. <https://doi.org/10.1080/13506280444000157>
- Kuhn G, Tatler BW, Cole GG (2009) You look where i look! effect of gaze cues on overt and covert attention in misdirection. *Vis Cogn* 17(6–7):925–944. <https://doi.org/10.1080/13506280902826775>
- Hall ET (1959) The silent language. Edward Hall, Doubleday Garden City
- Kendon A (2010) Spacing and orientation in co-present interaction. Springer, Berlin, pp 1–15. https://doi.org/10.1007/978-3-642-12397-9_1
- Moniri MM, Luxenburger A, Schuffert W et al (2016) Real-time 3d peripheral view analysis. In: Proceedings of the 26th international conference on artificial reality and telexistence and the 21st eurographics symposium on virtual environments. Eurographics Association, Goslar, DEU, ICAT-EGVE '16, pp 37–44

35. Repiso E, Garrell A, Sanfeliu A (2020) Adaptive side-by-side social robot navigation to approach and interact with people. *Int J Soc Robot*. <https://doi.org/10.1007/s12369-019-00559-2>
36. Lu DV, Hershberger D, Smart WD (2014) Layered costmaps for context-sensitive navigation. In: 2014 IEEE/RSJ international conference on intelligent robots and systems. IEEE, pp 709–715
37. Alt H, Godau M (1995) Computing the Fréchet distance between two polygonal curves. *Int J Comput Geom Appl* 5(01n02):75–91
38. Vinciarelli A, Pantic M, Bourlard H et al (2008) Social signals, their function, and automatic analysis: a survey. In: Proceedings of the 10th international conference on multimodal interfaces. Association for Computing Machinery, New York, NY, USA, ICMI '08, pp 61–68. <https://doi.org/10.1145/1452392.1452405>
39. Weiss A, Bernhaupt R, Tscheligi M et al (2008) A methodological variation for acceptance evaluation of human–robot interaction in public places. In: RO-MAN 2008—the 17th IEEE international symposium on robot and human interactive communication, pp 713–718. <https://doi.org/10.1109/ROMAN.2008.4600751>
40. Field A, Miles J, Field Z (2012) *Discovering statistics using R*. Sage, New York
41. Pacchierotti E, Christensen HI, Jensfelt P (2006) Evaluation of passing distance for social robots. In: ROMAN 2006—the 15th IEEE international symposium on robot and human interactive communication, pp 315–320. <https://doi.org/10.1109/ROMAN.2006.314436>
42. Truong XT, Ngo TD (2016) Dynamic social zone based mobile robot navigation for human comfortable safety in social environments. *Int J Soc Robot* 8:663–684
43. Beraldo G, Koide K, Cesta A et al (2021) Shared autonomy for telepresence robots based on people-aware navigation. In: International conference on intelligent autonomous systems. Springer, pp 109–122
44. Rebsamen B, Guan C, Zhang H et al (2010) A brain controlled wheelchair to navigate in familiar environments. *IEEE Trans Neural Syst Rehabil Eng* 18(6):590–598
45. Beraldo G, Tonin L, Cesta A et al (2021) Brain-driven telepresence robots: A fusion of user's commands with robot's intelligence. In: Baldoni M, Bandini S (eds) *AIxIA 2020 - Advances in Artificial Intelligence*. Springer International Publishing, Cham, pp 235–248
46. Linder T, Breuers S, Leibe B et al (2016) On multi-modal people tracking from mobile platforms in very crowded and dynamic environments. In: 2016 IEEE international conference on robotics and automation (ICRA), pp 5512–5519. <https://doi.org/10.1109/ICRA.2016.7487766>
47. Munaro M, Menegatti E (2014) Fast rgb-d people tracking for service robots. *Auton Robots* 37(3):227–242
48. Fischer T, Chang HJ, Demiris Y (2018) RT-GENE: real-time eye gaze estimation in natural environments. In: European conference on computer vision, pp 339–357
49. Fankhauser P, Hutter M (2016) A universal grid map library: implementation and use case for rough terrain navigation. In: Koubaa A (ed) *Robot operating system (ROS)—the complete reference (Volume 1)*, chap 5. Springer. https://doi.org/10.1007/978-3-319-26054-9_5

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Alberto Bacchin holds a BSc degree in Information Engineering and an MSc in Computer Science Engineering from the University of Padova. Currently, he is pursuing a Ph.D. degree in Information Engineering at the Intelligent and Autonomous Systems Laboratory, Uni-

versity of Padova. His research ranges from Social Robots to Deep Learning techniques, particularly in the context of robotic perception and manipulation for waste sorting applications. In recognition of his dedication to sustainability, he was honored with the Graduation Award BCC Patavina in 2021. Between 2022 and 2023, he served as a Research Assistant at the Nara Institute of Science and Technology.

Gloria Beraldo received the M.Sc. degree cum laude in computer science engineering and the Ph.D. in information engineering with Doctor Europaeus mention from the University of Padova in 2017 and 2021, respectively. She was a visiting researcher at École polytechnique fédérale de Lausanne in 2019 and at Universitat Politècnica de Catalunya under the TERRINet (European Robotics Research Infrastructures) initiative in 2022. She is currently a fixed-term researcher at the National Research Council of Italy. She has been involved in different projects that aim at enhancing robots with AI for providing personalized assistance via human-mediated interactions. She is also a contract professor at the University of Padova for the courses of “Artificial Intelligence” and “Digital Systems”. Her research is focused on designing novel semiautonomous human-robot interactions with particular attention to the case of brain-machine interface-driven robotics devices. She is investigating how to decode high-level user intention and fuse it with the perception of the robot and goal-oriented task planning to achieve advanced forms of human-robot interaction. Her research interests include human-robot interaction, shared control and shared autonomy, telepresence robots, neurorobotics, socially assistive robotics, and intelligent systems. She regularly serves as a reviewer for several top-level conferences and journals on robotics, artificial intelligence and human-robot interaction. She is one of three coordinators of the Italian Group of Artificial Intelligence and Robotics under the Italian Association for Artificial Intelligence. She has been involved in the program committees of different workshops and conferences including IUI, ICAPS and ECAI. She co-organized two special sessions at IEEE ROMAN and she has been served as an associate editor in 2023 and 2024. She is a publicity chair for ICAPS 2024. She is a guest editor in Springer's International Journal of Social Robotics and Frontiers in Neurorobotics. She has received many awards, including the AIxIA Pietro Torasso for her contribution to the field of Artificial Intelligence in the PhD thesis, IEEE Brain BMI Workshop Best Student Paper Award 2020 (Third Place) at SMC2020 Conference, the BEST LBR NOMINEE at HRI 2023.

Jun Miura received the B.Eng. degree in mechanical engineering and the M.Eng. and Dr.-Eng. degrees in information engineering from The University of Tokyo, Japan, in 1984, 1986, and 1989, respectively. From 1989 to 2007, he was with the Department of Computer-Controlled Mechanical Systems, Osaka University, Japan, first as a Research Associate and later as an Associate Professor. From March 1994 to February 1995, he was a Visiting Scientist with the Department of Computer Science, Carnegie Mellon University, USA. Since 2007, he has been a Professor with the Department of Computer Science and Engineering, Toyohashi University of Technology, Japan. To date, he has authored or coauthored more than 260 peer-reviewed scientific papers in the field of robotics and autonomous systems in internationally reputable journals and conferences. He is a member of the IEEE-RA Society, the IEEE-SMC Society, the IEEE-Computer Society, Japanese Society of Mechanical Engineers (JSME), Japanese Society of Artificial Intelligence (JSAI), the Information Processing Society of Japan (IPJS), and the Institute of Electronics, Information, and Communication Engineers (IEICE). He has been a fellow of the Robotics Society of Japan (RSJ), since 2011. He has received numerous awards, including the Best Paper Award from the Robotics Society of Japan in 1997 and the finalist of the Best Paper Award at IEEE

ICRA in 1995 and the Best Service Robotics Paper Award at IEEE ICRA in 2013.

Emanuele Menegatti is a Full Professor of the Dept. of Information Engineering at the University of Padova. His research interests are in the field of Robot Perception. In particular, he is working on Neurorobotics and Collaborative Robotics for Industry 4.0. Menegatti was the European coordinator of the FP7 FoFEU project “Therobot” and local principal investigator for other 5 European Projects. He also served as Project Reviewer for the European Commission in FP7 and H2020. He is teaching the courses of “Intelligent Robotics” and “Data Structures and Programming”. Menegatti has supervised 21 Ph.D. students, 13 post-docs, 60+ master theses and 15+ bachelor theses and co-supervisor of 2 Japanese PhD students. Menegatti is an Associated Editor of the journals: “Robotics and Autonomous Systems” (Elsevier), “Robotica” (Cambridge Press), “IPSI Transactions on Computer Vision and Applications” (Springer), “Frontiers in Neurorobotics” (Frontiers), “International Journal of Advanced Robotic Systems” (Sage). He has served as Associate Editor for IEEE ICRA Conference and IEEE IROS Conference, and IJCAI Conference. He is an author of more than 50 articles in international journals and more than 120

publications in international conferences. Since 2020, he has been president of the International Intelligent Autonomous System Society (IAS-Soc). Menegatti is a founding partner of two startups IT+Robotics srl and EXiMotion srl. He is a co-author of 4 patents.