Project INSIDE: Towards Autonomous Semi-Unstructured Human-Robot Social Interaction in Autism Therapy

Francisco S. Melo[a,b,*], Alberto Sardinha[a,b], David Belo[d], Marta Couto[c], Miguel Faria[a,b], Anabela Farias[c], Hugo Gambôa[d], Cátia Jesus[c], Mithun Kinarullathil[b], Pedro Lima[b], Luís Luz[b], André Mateus[b], Isabel Melo[c], Plinio Moreno[b], Daniel Osório[d], Ana Paiva[a,b], Jhielson Pimentel[b], João Rodrigues[d], Pedro Sequeira[e], Rubén Solera-Ureña[a], Miguel Vasco[a,b], Manuela Veloso[f], Rodrigo Ventura[b]

[a] *INESC-ID, Lisbon, Portugal*
[b] *Instituto Superior Técnico, University of Lisbon, Portugal*
[c] *Hospital Garcia de Orta, Almada, Portugal*
[d] *LIBPhys, Universidade Nova de Lisboa, Caparica, Portugal*
[e] *Northeastern University, Boston, MA 02115, USA*
[f] *Carnegie Mellon University, Pittsburgh, PA 15213, USA*

Abstract

This paper describes the INSIDE system, a networked robot system designed to allow the use of mobile robots as active players in the therapy of children with autism spectrum disorders (ASD). While a significant volume of work has explored the impact of robots in ASD therapy, most such work comprises remotely operated robots and/or well-structured interaction dynamics. In contrast, the INSIDE system allows for *complex, semi-unstructured interaction* in ASD therapy while featuring a *fully autonomous robot*. In this paper we describe the hardware and software infrastructure that supports such rich form of interaction, as well as the design methodology that guided the development of the INSIDE system. We also present some results on the use of our system both in pilot and in a long-term study comprising multiple therapy sessions with children at Hospital Garcia de Orta, in Portugal, highlighting the robustness and autonomy of the system as a whole.

1. Introduction

Recent years have witnessed an increasing number of applications of robotic technology in health and assisted living. Robots of all shapes and forms are

*Corresponding author.
Email addresses: `fmelo@inesc-id.pt` (Francisco S. Melo)

currently used both for high-precision medical interventions [93] and physical and cognitive therapy [32, 50, 18, 87].

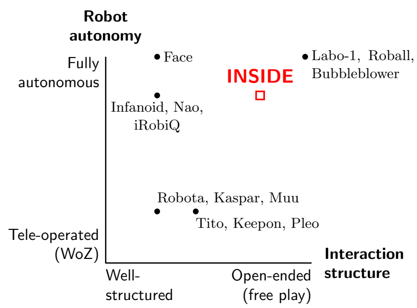
Of particular interest to this paper is the use of robotic technology in autism therapy—see, for example, the survey work of Scassellati et al. [80]. Autism Spectrum Disorders (ASD), as defined by the DSM-V, consist of persistent deficits in social communication and social interaction across multiple contexts [2]. Such deficits include difficulties in social-emotional reciprocity, nonverbal communication, and developing, maintaining, and understanding relationships. Autism is a *spectrum* of conditions, and while there are characteristics common to all ASD individuals, the exact degree to which each individual is affected by this condition varies greatly [9]. Such enormous variability among individuals poses challenges in terms of therapeutical approaches, as ASD requires a wide range of interventions in order to help the different individuals in the best possible way [25].

There is significant controversy regarding the actual prevalence of ASD. Official data from the World Health Organization estimates that 1 in 160 children suffers from some autism spectrum disorder [92]. ASD has social, emotional and economic costs for the autistic individuals, their families and the community [88]. These costs are not limited to childhood: autistic individuals have several impairments and difficulties throughout their adult life, even when they receive successful interventions during childhood [8]. Nevertheless, it is generally accepted that an early and adequate intervention yields a more favorable outcome [79].

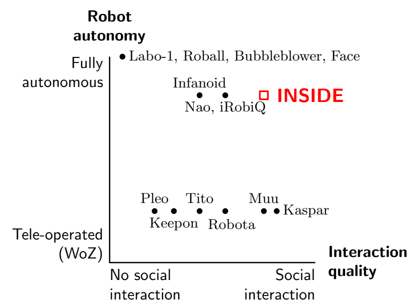
ASD children show little interest in social interaction and instead prefer to interact with objects [24, 84]. Such preference is particularly evident in their fascination by computers, tablets and other electronic devices [23]. For example, it has been reported that even those ASD children that usually interact very little with human therapists are willing to engage with robots [19, 59, 89]. The willingness of ASD children to (socially) interact with robots may, in part, be explained by the predictability and simplicity of their social behavior, when compared with that of human partners [79]. The use of robotic technology may, therefore, provide an important tool to develop novel therapeutic approaches in which children have fun while engaging in a social interaction, something that is typically difficult for ASD children [84]. Several studies report that children with autism in effect create affective bonds with social robots [43, 48].

1.1. Robots in ASD therapy

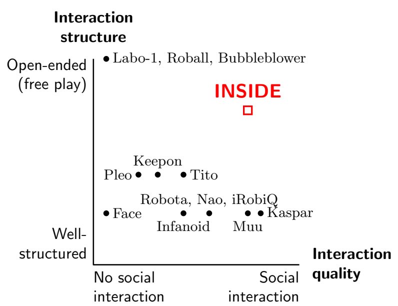
As mentioned above, several projects have explored the use of robots in ASD therapy. Notable examples include Aurora [22] or, more recently, the DREAM project [28]. These projects feature a variety of robot platforms that interact with children during therapy sessions, typically by engaging in some form of joint activity such as an imitation game or other collaborative task. The studies conducted in the context of these projects report promising results: during their interaction with the robots, ASD children are able to exhibit different forms of social behavior, such as joint attention, eye gaze, spontaneous imitation and increased engagement in tasks after interaction [9, 22, 68]. Such success attests



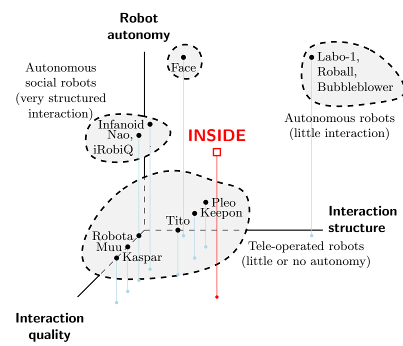
(a) Robot autonomy versus interaction structure.



(b) Robot autonomy versus interaction quality.



(c) Interaction structure versus interaction quality.



(d) Positioning of INSIDE with respect to existing work.

Figure 1: Landscape of robot use in ASD therapy in terms of robot autonomy, interaction quality, and interaction structure. We include a representative (and not comprehensive) list of works identified by the robotic platforms used: Bubbleblower [29, 30], Face [67], Infanoid [46], iRobiQ [94], Kaspar [22, 74], Keepon [48, 47], Labo-1 [89], Muu [52], Nao [33, 28], Pleo [42, 43], the Roball toy [58, 59], Robota [12, 20], and Tito [27].

to the need for further exploration of the potential impact of robot technology in the therapy of children with autism spectrum disorders.

To gain a clearer understanding on the use of robots in ASD therapy, it is educative to carefully consider the interaction between robot, children and therapists. We look at such interaction from three complementary dimensions:[1]

[1] It is interesting to draw a parallel between the three dimensions outlined herein and the discussion in the work of Scassellati et al. [80]. Scassellati et al. discuss the use of robots in autism therapy from a broader perspective, considering aspects of *robot appearance*, *human-robot interaction*, and *evaluation*. Human-robot interaction is further broken down into *targeted behavior*, *role of the robot*, and *robot autonomy*. Although the correspondence is not exact, it is possible to equate our dimension of "interaction quality" with Scassellati et al.'s "role of the robot", and "interaction structure" with "targeted behavior".

- *Interaction quality*, which roughly describes how rich the social interaction between the robot and the children is. One end of the spectrum corresponds to robots that are little more than toys, providing little to no social interaction. The other end of the spectrum corresponds to highly interactive robotic platforms that allow for rich, multimodal forms of social interaction, including dialogue, joint manipulation, etc.

- *Interaction structure*, which roughly describes the type of activities in which the interaction takes place. One end of the spectrum corresponds to highly structured activities, that progress according to a strict script in which the roles and configuration of robot and humans are well defined. The other end of the spectrum corresponds to unstructured activities such as free-play.

- *Robot autonomy*, which describes whether the robot is remotely controlled/tele-operated or fully autonomous.

Figure 1 depicts the landscape of robot use in ASD therapy across the different dimensions outlined above. We include a (non-exhaustive) selection of representative works featuring a wide variety of robotic platforms. As outlined in the diagram, existing work can be roughly clustered into three major groups, that we consider separately.

◇

The larger cluster—marked as "Tele-operated robots"—corresponds to those works adopting a *Wizard-of-Oz* approach [40], in which the robot is remotely controlled by a human operator (the "Wizard") who remains "behind the curtain".[2] The Wizard-of-Oz (WoZ) approach is very attractive from a research perspective: since it relies on human perception, it readily circumvents the fundamental perceptual challenges that plague any autonomous robot; it also allows robots (via tele-operation) to exhibit a wide range of social behaviors and engage in rich interactions, effectively facilitating the study of child-robot interaction in ASD therapy.

It is not surprising, then, that works featuring tele-operated robots exist that feature a wide variety of interaction modalities, ranging from very simple interactions—in which the robot merely exhibits some form of emotional response to the child—to significantly more involved interplay, where robot and child are involved in imitation and turn-taking games. Examples of the former include work with the Keepon [48, 47] and Pleo robots [42, 43]; a representative example of the latter is the work done in the context of the Aurora project using the Kaspar robot [22, 74].

However, in scenarios where the interaction of the child with the robot is mediated by a therapist (as seen in Fig. 13), a WoZ setup may demand two or

[2]Since the robot operator is out of the sight of the participants in the experiment, the robot appears autonomous during the interaction.

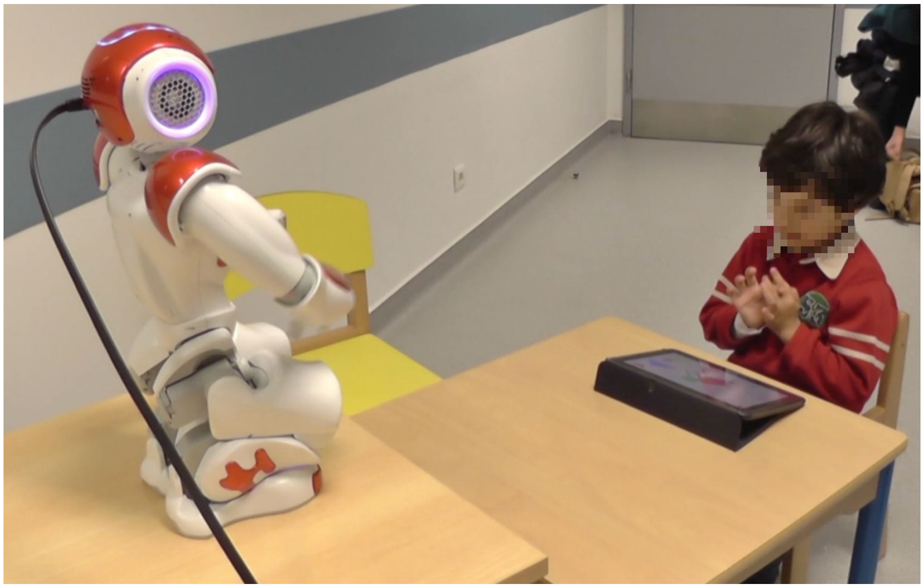


Figure 2: Structured interaction between a child and the Nao robot. The robot and the child play a turn-taking Tangram game [10].

more therapists (besides the robot technical staff). As interactions become increasingly complex, the burden imposed on the human operators also increases, requiring them to process an increasing number of inputs from the interaction and handling an increasingly complex repertoire of actions. Eventually, the robot may need multiple human operators to engage a child in a rich, multimodal interaction. Such a strong dependence on human operators renders WoZ approaches to autism therapy unaffordable in the long term [86, 80, 28].

As for works featuring autonomous robots, we can identify two additional clusters that greatly differ in the way the child interacts with the robot. In one cluster—marked only with "autonomous robots"—we include those works featuring robots such as Labo-1 [89], Roball [58, 59] or the Bubbleblower [29, 30]. These robotic platforms are endowed with very simple reactive behaviors, which allow only for the simplest form of interaction. The children interact with these robots during free play (in an unstructured interaction), but the robots behave as little more than sophisticated toys, exhibiting no social interaction.

In the other cluster—marked as "autonomous *social* robots"—we include works that use robotic platforms such as Infanoid [46], Nao [33, 28] or iRobiQ/CARO [94]. For example, Yun et al. [94] describe a robot that autonomously interacts with a child with ASD during a therapy session. During the therapy sessions, the child and therapist are siting facing the robot and take turns interacting with the robot. The activity consists of several rounds where the child is asked to make eye contact or interpret emotional expressions, either with the robot or with the therapist. The robot autonomously identifies both the child and the therapist, and the activity proceeds.

Scenarios such as the one just described feature autonomous robots capable of exhibiting social interaction capabilities, but rely on very structured interactions, both in terms of the type of activity available to the participants and even in the way they are configured with respect to one another. For example, in most such works the child sits in front of the robot and the interaction follows a strict script (see Fig. 2).

◇

The INSIDE system sets itself apart from the clusters identified above. Unlike the works identified in the first cluster—of the "tele-operated robots"—the robot used in our system acts in an fully autonomous manner. Moreover, unlike the works identified in the second cluster—of the "autonomous robots"—our robot exhibits social interaction capabilities. As an example, the robot promotes and explains the different therapeutical activities, provides feedback on the task—such as encouragement or reinforcement—and asks for assistance. Finally, it also differs significantly from the works in the third cluster—of the "autonomous social robots"—as it accommodates a rich set of activities in which the child is allowed to freely move around the therapy room.

1.2. Contributions

In this paper, we report the work conducted in the context of the project INSIDE (`www.project-inside.pt`) towards the development of a networked robot system that can be used in a wide range of therapeutical activities involving children with autism spectrum disorders. Two key concerns drove the development of the system and set it apart from other existing platforms used in robot-enhanced therapy:

- ASTRO, the robot used in the context of the project, should be able to socially interact with the children and be fully autonomous during a therapy session;
- Within the goals and activities of the therapy session, the child-robot interaction should be as unconstrained as possible, allowing for the children to freely move and express herself.

The INSIDE system therefore addresses a gap that can be identified from the previous discussion: our system should enable *autonomous social interaction*, while allowing relatively *unconstrained activity* by the child. Our key contribution can thus be summarized as the development of a networked robot system that allows for social and semi-unstructured interaction while featuring a fully autonomous robot. In particular, the system was developed with the following key features:

- *The robot interacts with a child that is allowed to freely move around the room.* As observed before, such freedom contrasts with most recent work featuring ASD therapy with autonomous social robots. The fact that the child is allowed to freely move around the room poses significant challenges in terms of perception. For example, during the activities we need to monitor the whole space of the room and autonomously distinguish the child from the other agents present in the room (the therapist, the robot, and the care-takers).

- *The child performs a number of different activities during a therapy session*, such as: (i) finding hidden balls in the room, which actually requires the child to move around the room; (ii) solving a geometric puzzle; (iii) play a turn-taking game with the robot; (iv) assisting the robot in moving across the room.[3]

- *The robot has a key role in the therapy session.* During a therapy session, the robot invites the child to perform activities, explains the activities to the child, and provides encouragement with positive reinforcements. In this process, the therapist assists the child, sometimes clarifying the robot's explanations, providing additional reinforcement or aiding in physical aspects of the task; for example, removing the obstacles for the robot to move to the next activity. Figure 13 illustrates a therapist-mediated interaction between a child and a robot at the end of a therapy session conducted in the context of INSIDE. In the background one can see the room where the activities took place.

- *The robot is fully autonomous.* In no aspect of the interaction is the robot controlled by the therapist in the room or the operators outside the room. For safety concerns—given the size of the robot and the unstructured nature of the interaction—the robot is nevertheless monitored by human operators outside the therapy room. These operators have the possibility to take control of the robot if necessary but, as reported in Section 5, the robot was able to operate autonomously during a long-term study that took place in real-world therapy sessions held at Hospital Garcia de Orta, in Portugal.

The INSIDE system thus pushes the state-of-the-art along two directions: networked robot systems for human-robot interaction and potential use of robots in ASD therapy.

Networked robotic systems for human-robot interaction. In literature, robots that are dependable enough to operate autonomously during long periods of time, involving interaction with humans, are not yet often seen. Several successful examples in the literature feature interaction scenarios under routine

[3]The objects in the room are arranged so as to purposefully block the robot's access to a part of the room. The robot autonomously detects the obstacle and, after determining that it cannot go around it, asks for the child's assistance in removing the obstacle.

situations [35, 57]. The interaction with humans involves either elderly or children which are keen to cooperate with the robot through multi-modal interfaces (e.g., speech, touch screens). In other examples, the operation period is neither too long nor too challenging, as it does not include features such as tracking people within an environment or human-aware navigation [4, 26].

The scenario considered in our paper features a robot system that can behave autonomously during a full therapy session, during which it tracks a child moving about a large room (where parents also sit and a therapist also moves around). The robot explains the activities and encourages and challenges the child to improve its performance in the different games. It recognizes words spoken by the therapist, interpreting them in the context of current tasks, and detects successful actions by the child (e.g., in completing a puzzle, or finding a given number of balls hidden all around the room). Given the complexity of the challenges posed by children with ASD, integrating all these functionalities in an autonomous robot system is, by itself, a novelty that pushes the state of the art in robot technology.

Use of robots in ASD therapy. Several projects have explored the use of robots in ASD therapy, as surveyed above. These projects feature a variety of robot platforms that interact with children during therapy sessions, typically by engaging in some form of joint activity such as an imitation game or other collaborative task. These studies can be roughly grouped into two categories:

- One category comprises studies that present unconstrained and unstructured interactions to observe the reactions and the interaction that arises between ASD children and different robots. Examples include the works of Dautenhahn [19], Michaud and Théberge-Turmel [59], and Robins et al. [73].

- A second category comprises studies in which there is a structured and constrained interaction aiming at improving specific skills such as joint attention, imitation, or recognition of emotional expression.[4] Examples include the works of Barakova et al. [3], Bekele et al. [8], Boccanfuso and O'Kane [13], Brok and Barakova [14], Duquette et al. [27], Liu et al. [49], Pierno et al. [65], Simut et al. [83], Srinivasan et al. [84], and Tapus et al. [85].

To be an effective tool for therapy, a robot needs to be autonomous and versatile. Autonomous to alleviate the human cost associated with its operation; and versatile to be of use in activities of different types, aimed at improving different skills. To our knowledge, our work pioneers in the use of an autonomous robot capable of conducting a small therapy session—from the initial greeting to the final goodbye—in which multiple and diverse tasks are performed that address

[4] A notable exception was presented by Clabaugh et al. [17], in which an assistive robot is used at home, for a month.

specific impairments of ASD children, that acts like a social agent and engages in rich social interaction with the children.

2. The INSIDE Intervention Scenarios

In this section we provide a brief overview of the interaction scenarios in INSIDE, discussing the therapeutic goals and outlining the role of the robot. As will soon become apparent, the interaction scenarios considered pose a number of technological challenges—in terms of both perception, cognition and actuation of the robot—that lie at the core of the design options of the INSIDE system.

2.1. Therapeutic Approach: DIR/Floortime

Children with ASD have a range of occupational performance problems (i.e., difficulties in completing everyday activities) and sensory issues that interfere with their full participation in school, home, and community activities [15]. Regardless of their philosophy, most ASD therapies have a common goal: improve the quality of life of children with ASD and their families. Occupational therapy tries to achieve this goal by promoting the social participation of children in natural contexts, developing children's social competence and improving their engagement and participation [60, 7].

Occupational therapists working with ASD children frequently adopt the DIR/Floortime$^{\text{TM}}$ model, developed by the U.S. child psychiatrist Stanley Greenspan (1941-2010) and his colleagues. This model, a semi-structured intervention, was designed to improve social-emotional growth in children with ASD via interactive play activities individually designed to enhance the child's play and social participation [15].

DIR/Floortime$^{\text{TM}}$ focuses on relationships, social skills, meaningful spontaneous use of language and communication, and integrated understanding of human development. The integrated model of human development includes interaction with caregivers and the environment, biological, motor and sensory differences, and the child's functional emotional developmental capacities [62].

Several studies have demonstrated that social engagement directly affects important behaviors like language, requesting, greeting, joint attention and imitation, even when these behaviors are not specifically targeted by the intervention program [76, 56].

Greenspan [34] described six functional emotional developmental levels. The developmental capacities are essential for spontaneous and empathic relationships as well as for the mastery of academic and life skills. The model also takes into account the individual differences regarding sensory processing and modulation that interfere with the child's capacity to plan and sequence actions or ideas. Together, developmental levels and individual differences provide the goals for working with ASD children.

2.2. Tasks and Therapeutical Goals

The core concept that underlies much of the research in INSIDE is the concept of *symbiotic autonomy*. In order to develop an autonomous robot platform, capable of interacting with human agents in the context of a given task, the robot must be able to act in situations in which neither the robot nor the human are able to fully complete such task by themselves without the assistance of the other, due to inevitable limitations of the agents or the design of the task. This concept of symbiotic autonomy lends itself quite naturally to the scenario of therapy with children with ASD. Indeed, many of the activities employed in traditional therapy already take into account a scheme of symbiotic autonomy and are developed in order to focus on the behavior deficiencies of children with ASD, such as the difficulty in addressing help requests or in asking for help.

In designing robot-mediated therapy sessions, we require a set of activities that (1) have therapeutic goals (i.e., activities that train skills that are commonly impaired in ASD children); and (2) are arranged in such a manner that allows the robot to have an active role in the session, as an alternative to being just a fun toy.

With this in mind we created seven activities that address key impairments of children with ASD and/or simulate common social interactions. The activities and their goals are as follows:

- *"Say hello"*: Once the child enters the room, the robot says hello, introduces itself and asks for the child's name. This first moment represents a well-known social routine: the child meets someone new, a social agent, and they say hello to each other. It also allows the child to get acquainted with the robot and presents the robot as a social partner that can communicate with the child.

- *"Ball game"*: In this second activity the child must retrieve 8 colored balls that were hidden in the room. Once the child retrieves a ball, he/she must place it in the robot's basket and then go search for another ball. The activity trains the ability of the children to follow simple instructions (search for hidden balls in the room) and their attention (they must retrieve 8 balls). Because the task is fairly easy and it is something commonly trained with children during early interventions, it works as a good icebreaker, allowing children to do something that they have trained before. The positive feedback provided by the robot helps to keep the children focused on the task, mimicking what therapists typically do in a therapy session.

- *"Obstacle activity"*: The robot wants to reach a table that is placed in a corner of the room. While moving towards the table, the robot's path is blocked by an obstacle; the robot then asks the child for help. The child must remove the obstacle so that the robot can reach the table. This activity aims to improve social reciprocity and empathy. Understanding other's mental states is a significant impairment of ASD children [71, 5] and therefore understanding that others need help can be difficult. In order to successfully complete the task, the child needs to comprehend that the robot wants to reach the table and is incapable of removing the obstacle, which requires the child's ability to understand the robot's perspective. It is one of the most challenging tasks within the session.

- *"Puzzle"*: Once the robot is able to reach the table, it invites the child to assemble a puzzle. In this activity, it is the robot's turn to help the child. Initially, all pieces but one are placed next to the puzzle. For the child to complete the puzzle, he/she must ask the robot for help. When the penultimate piece is on the puzzle, the robot points out that there is still a piece missing and encourages the child to ask for help. One consequence of the social communication impairments in ASD children is the lack of ability to make requests, which is invaluable for social interaction and one of the deficits that peers and family members perceive as a significant limitation in social-communicative behavior [90].

- *"Tangram"*: This activity uses a turn-taking Tangram game previously developed in the work of Bernardo et al. [10]. Turn-taking is so ubiquitous in human interactions and so deeply embedded in common-sense, that it is a largely unconscious process that is, in most cases, extremely difficult to accurately describe [81]. One of the most recognizable purposes of turn-taking is to regulate human conversations. Being such an omnipresent phenomenon, training turn-taking in ASD children is extremely important. Turn-taking in a game follows rules that can be well-defined a priori, unlike what happens with conversations. In this case, each turn corresponds to one piece of a puzzle. Once the player places the piece in the correct position, the turn changes and it is time for the other player. During the game, in one of the robot's turns, the robot will ask for help placing a piece.

- *"Blocks activity"*: Another important piece of social interaction is empathy and reciprocity, which are connected to the ability to understand others state of mind. Inspired by the method used by Plötner et al. [69], we created the "blocks activity" where the therapist knocks down a tower of blocks and waits to see if the child spontaneously helps her to rebuild the tower. This is the only task where the robot has a more passive role. If the child does not help the therapist, the robot will encourage the child to help.

- *"Say goodbye"*: When the tasks are completed the robot says goodbye and tries to leave the room asking the child to open the door. This provides another opportunity for the child to help the robot and is once again a representation of a common social routine, as the robot thanks the child and says goodbye before leaving the room.

2.3. The Role of the Robot

We foresee several advantages in robot-assisted therapy for ASD children. Robots may allow us to develop a therapy more focused on the children's interests and abilities, resulting in motivating and pleasurable interactions that comprise an overall positive experience for ASD children [84]. In other words, technology may allow us to create a therapeutic setting where children have fun while engaging in an interaction, something that is typically difficult for ASD children.

As mentioned in Section 1, ASD is characterized by difficulties in making sense of the social world; on the other hand, ASD children often show ease in understanding the physical world and object-related interactions [6, 44, 45]. A robot is an object that can behave like a social partner, which can be a perfect bridge between the physical and the social world. Using a robot also allows for the embodied characteristics of face-to-face human interactions [49], without all the implicit rules and cues that regulate social interactions between humans and that are so difficult for ASD children to read.

Previous studies demonstrate that it is possible to use robots to improve some common impairments of children with ASD. Robots can be used to create turn-taking and imitation games that teach basic social skills; or as social interaction mediators and shared-attention objects to encourage interactions [21, 72, 47, 66]. However, none of these studies have used a fully autonomous robot that interact with ASD children in unconstrained activities, whereby children can freely move and express themselves.

However, to build a networked robot system that is able to participate in the activities described in Section 2.2, several key technological challenges must be addressed:

- *Perception*: Perceiving people and objects in the room is one of the key challenges in the development of the INSIDE system. A reliable perception system is crucial for the robot to interact with the child in the therapy sessions. For example, the system must be able to detect the child's position and robot's pose in order to: (i) move the robot close to the child; (ii) enable the robot to guide the child to the next activity; and (iii) make the robot look at the child while talking.

- *Robot motion and head rotation*: Another important impairment of ASD children mentioned above is related to the *mind-reading mechanism*, also known as *theory of mind* (TOM) [64, 5]. Before TOM emerges, certain precursors need to be consolidated—namely the eye direction detection (EDD) that involves identifying the eyes and being able to detect whether or not they are looking at us. This mechanism allows dyadic representations that serve as a base for the development of the shared attention mechanism (SAM). Our robot has important features—such as motion and a head rotation—which aim to address such mechanisms and can be a helpful tool in development of TOM.

- *Decision-making*: As discussed before, the therapeutical scenario considered in our work is challenging, since the child and robot may freely move around while engaging in activities within a room. To tackle this challenge from the perspective of the robot's behavior, the decision-making system has to take into account the current state of the environment, based on the input from the perception system, and plan the next actions of the robot. For example, if the robot is currently playing the puzzle activity and detects that one more piece was placed in the right location, it should decide to smile and reinforce the child.

- *Robot design*: Klin et al. [45] demonstrate how ASD children seem to orient towards non-social contingencies, relying more on audio-visual synchrony than on biological motion. Unlike what happens with humans, audio-visual synchrony can be manipulated in a robot, directing the child's attention to the desired places—for example, to the eye region of the face, or to the robot's pouch, in the case of the ball game. While designing the robot for the INSIDE project, a team of therapists, doctors and researchers included important features (e.g., an LCD for facial expressions, an illuminated pouch in the robot's front with an RFID reader to detect balls) aiming at attracting/focusing the child's attention and improving the interaction possibilities within the therapy sessions.

- *Social interaction*: As mentioned above, one key feature of the robot is an LCD display in the head that is capable of expressing several emotions (e.g., happiness, sadness) in order to improve the interaction with the child.

- *Symbiotic interaction*: Finally, while designing the activities and developing the robot system, we took into account the fact that the robot is not able to perform all the actions in the environment. For example, it is unable to open a door or removing an obstacle in the room. Such limitations, if identified by the robot, provide an excellent opportunity for the robot to ask for the child's help.

As emphasized in Section 1, in all the above the robot should perform in a fully autonomous manner, both in terms of perception and in terms of actuation. The following section describes in detail the architecture of the networked robot system, highlighting the design decisions and the key technological features of the final system.

3. The INSIDE System Architecture

This section goes over the INSIDE system, discussing how its design meets the needs of the intervention scenarios described in Section 2.

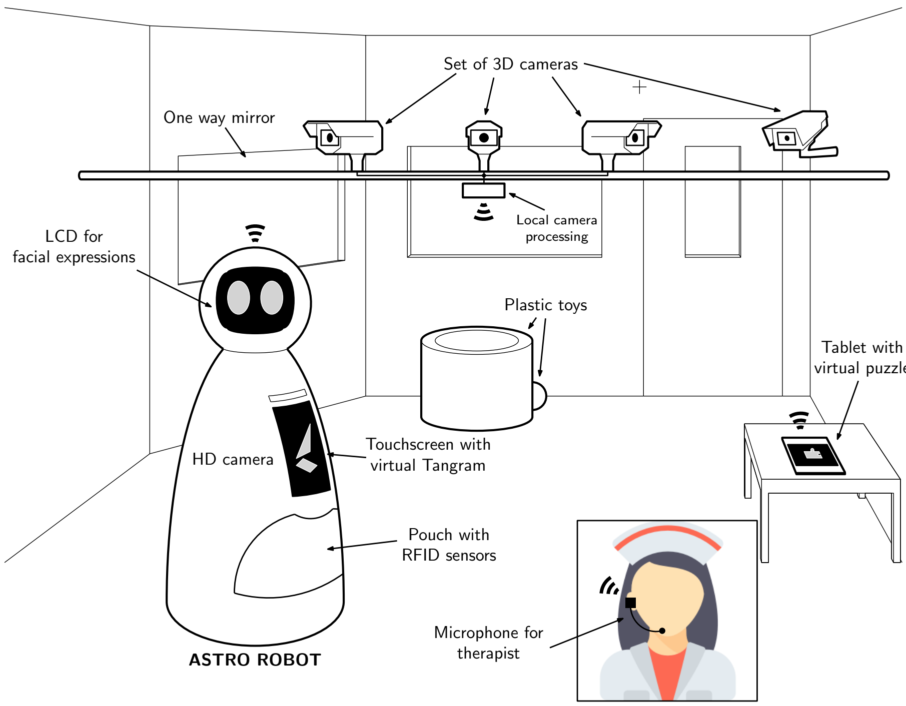


Figure 3: Diagram illustrating the main components of the hardware setup used in INSIDE.

3.1. System overview

In order to address the different challenges posed to the system and alluded to in Section 2, it was necessary to develop a hardware infrastructure that supports the different types of interaction planned for the therapy sessions. Additionally, it was also necessary to design a software architecture that is able to seamlessly integrate the different functionalities required of such a system.

Both hardware and software were designed following an *adjustable autonomy approach*, departing from initial mock-up studies, aimed at identifying key requirements for the system, until the final deployment of the full system, in which human intervention is reduced to monitorization. The adjustable autonomy approach ensured a smooth transition from an initial Wizard-of-Oz (WoZ) paradigm—in which a human operator is fully in charge of perception and robot operation—until the final setup, where perception is automated and the robot is fully autonomous. At the same time, it permitted an adequate design and tuning of the interaction between the robot and the child during therapy sessions, by means of the restricted perception Wizard-of-Oz methodology detailed in Section 4.

The overall hardware setup is illustrated in Fig. 3. It consists of a networked

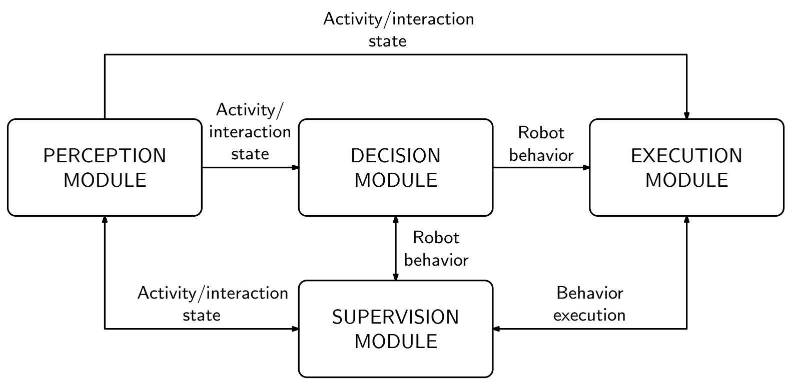


Figure 4: Outline of the main blocks in the software architecture.

robot system built on top of ROS.[5] The network comprises a set of 3D cameras mounted on the walls/ceiling of the therapy room, ensuring complete visual coverage of the room. The stream from each camera is processed locally, to avoid overloading the network. As a whole, the camera network provides the necessary information regarding the human activity in the space (including child detection, etc.). A tablet connected to the network is used to run the puzzle activity, and a remote microphone is used to process the speech of the therapist interacting with the child and robot (mostly used for keyword detection). The remaining activities depend almost exclusively on the robot's onboard sensors.

The robot, named ASTRO, is a mobile platform designed for multi-modal human-robot interaction. It includes a set of lasers, used for autonomous navigation and obstacle detection. It also includes an LCD in the rotating head that is used to animate facial expressions and speech acts. It also includes a touch-screen in the front, used for touch interactions (as featured in the Tangram game and puzzle). The robot's casket also includes a removable pouch covered by an RFID sensor, which can be used to detect when specific objects are placed in the robot's pouch—such as the balls in the ball game.

It is worth noting that the final hardware infrastructure adopted in INSIDE is general-purpose, and can easily be adapted for other scenarios of human-robot interaction. Similarly, the software architecture supporting the interaction was built to allow new tasks to be easily configured and deployed in the environment.

The software relies on a hierarchical architecture, outlined in Fig. 4. The system comprises four major components, namely

The perception module responsible for processing the information arriving from the different sensors in the network.

The decision module responsible for deciding, at each moment, whether to continue with the current activity or move to another activity. It is also

[5] Robot Operating System, see `http://www.ros.org/`.

responsible for selecting and triggering the different robot behaviors, as a function of the current state of the activity and the interaction with the child.

The execution module responsible for executing the behaviors triggered by the decision module and, in fact, executing the actions of the robot.

The supervision module which conveys a channel that allows human users to monitor the whole activity. The supervision module provides human operators with the ability to correct both the robot's perceptions and the robot's actions. Given the sensitivity of the application, security and ethical concerns require constant access to the robot, which is ensured via the supervision module.

The high-level interaction between the different modules ensures that the robot is able to go through the therapy session autonomously and robustly.

3.2. The perception module

The perception module is responsible for acquiring and processing all the information acquired by the sensors and providing the robot (namely, the decision module) the necessary information to perform the activities programmed for the therapy sessions. In particular, in light of the activities outlined in Section 2, the robot should be able to

- Know where it is;
- Know where the child is and what she is doing;
- Detect when the child satisfactorily replies to the robot's interpellation (such as responding with her name when asked);
- Know the state of the current activity.

INSIDE's system was designed taking into consideration such need for information, and the perception module closely reflects the structure of the activities in the INSIDE scenarios. In particular, its hierarchical block structure ensures the necessary flexibility to add or remove activities.

In INSIDE, the perception module receives, as input, the raw data from the different sensors in the environment and the robot, and is responsible for processing such data into meaningful information regarding the state of the robot, the child and of the interaction. Such information (the *state*) also incorporates any corrective feedback provided by the supervision module (more on this ahead), and will drive the decision and execution modules.

It is possible to identify several major blocks in the perception module. In the continuation, we discuss each of these blocks in detail.

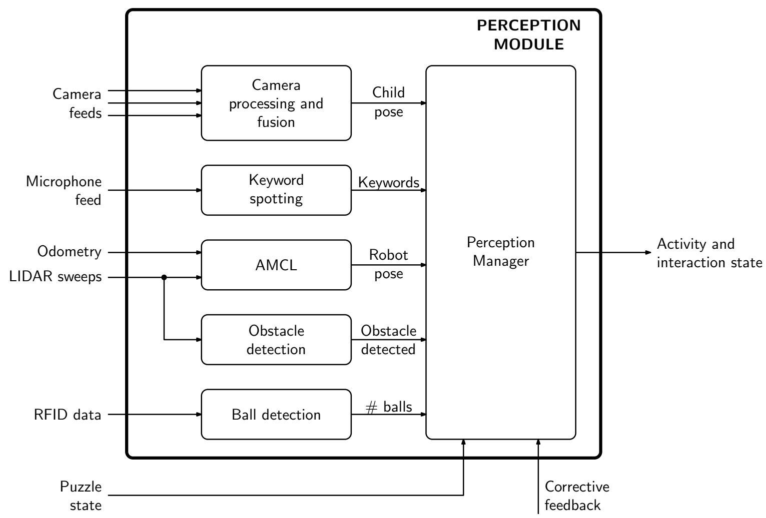


Figure 5: Detailed outline of the perception module.

The camera processing block. This block receives the feeds provided by the 3D cameras in the environment (in our setting, four Microsoft Kinects for Xbox One attached to the walls and ceiling).[6] These feeds are processed locally by a dedicated computer (we use one Intel NUC5i7RYH computer for each camera) to extract the skeleton information from all the people in the room.[7]

Since children have smaller body frame than adults, we use a simple threshold to eliminate detected skeletons corresponding to adults, and the robot's position to avoid confusing the child with the robot. To fuse the detections from multiple cameras, we use the *near-neighbour joint probability data association algorithm*, a Kalman Filter with a constant velocity model implemented in the ROS package `bayes_people_tracking`.[8] The output of the camera processing block is the pose of the child.

The keyword spotting block. In order to have an interaction between child and robot as natural as possible, it would be desirable for the robot to have the ability to process and interpret the child's spoken utterances. Following our adjustable autonomy methodology, we used the initial WoZ sessions to conduct extensive recordings of complete sessions using a number of microphones placed on the environment and on the robot.

These recordings evidenced a number of fundamental technical difficulties in terms of automatic speech recognition: multiple, moving overlapping speakers

[6]The cameras were calibrated at the time of deployment of the system using the standard Xbox calibration procedure.

[7]Th Kinect version used is able to track the skeleton—i.e., the pose of the body joints and links between them—of up to six people simultaneously.

[8]`https://github.com/CentralLabFacilities/bayes\_people\_tracker`

(children and adults), distant speech, room reverberation, different noise sources (e.g. robot's motors), microphone distortion, and mechanical vibrations, among others. In addition to all the technical difficulties, we also observed that the ASD children's utterances during the therapy sessions were scarce, limited, with low intensity and very short duration (monosyllables). Therefore, obtaining a dataset from which speech recognition could be tuned was deemed unfeasible.

In alternative, we opted by having the therapist coordinating the session wear a wireless microphone, with which it can provide verbal inputs to drive the behavior of the robot in specific situations. The use of a close-up earset microphone alleviates most of the technical difficulties identified before and provides an easily configurable and flexible mechanism to provide execution feedback to the system. In particular, we resort to the therapist input to ensure that the robot reacts adequately to the children's responses to some of the robot's interpellations.

To this purpose, the perception module integrates a keyword spotting system (KWS) based on the AUDIMUS automatic speech recognition engine [54, 55]. AUDIMUS is a hybrid speech recognition system that combines the temporal modeling capabilities of a hidden Markov models with the pattern discriminative classification capabilities of multi-layer perceptrons. For the purposes of INSIDE, AUDIMUS uses a specific equally-likely 1-gram language model formed by all the possible target keywords and a competing speech filler model [1]. At any time, the output of the keyword spotting block is a "list" with the keywords detected in the present utterance (if there was one).

The ACML and obstacle detection blocks. As depicted in Fig. 3, the ASTRO is equipped with a front LIDAR which scans a wide region in front of the robot, on a plane parallel to the ground. The LIDAR scan is used to estimate the position of the robot as well as to detect humans, obstacles and other objects.

In order to navigate the space, the AMCL block has available a map of the environment, constructed from LIDAR scans collected prior to the therapy sessions. We then use Monte-Carlo localization [31] to track the position of the robot, using both the odometry and LIDAR data. In our system, we use the AMCL algorithm implemented as the ROS `acml` package.[9] The ACML algorithm uses a particle filter to track the posterior distribution over possible positions of the robot, given the motion model and the sensor readings observed.

The LIDAR data is also used to detect obstacles, by counting the number of detected reflection points inside a pre-defined area in front of the robot. Obstruction is considered to occur when the number of points inside the box is above a pre-defined threshold.

Ball detection module and puzzle state information. In order to keep track of the state of activities such as the ball game and the puzzle, the perception module handles two additional sources of information. First, the readings from

[9] `http://wiki.ros.org/amcl`

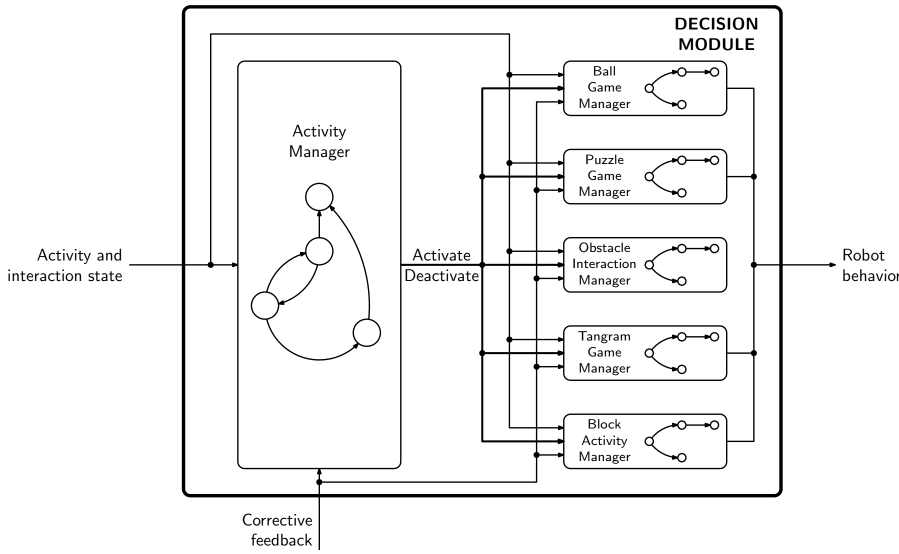


Figure 6: Outline of the decision module. The blocks on the right correspond to *interaction managers*, responsible for defining the behavior of the robot in each activity. The set of interaction managers indicated are merely illustrative.

the RFID sensor on the robot's pouch is processed by a ball counter, which keeps track of which balls have been placed in the pouch, ensuring that the robot is aware of the current state of the task and provides intermediate reinforcement to the child as the task proceeds.

Additionally, the perception module also manages the state of the puzzle activity, provided directly from the tablet where the activity takes place. Together with the keyword mechanism, the ball detection block and puzzle information allow the robot to keep track of the state of the different activities and respond adequately.

Perception Manager. Finally, the perception module includes one additional block, dubbed the *perception manager*. The perception manager is responsible for managing all processed perceptions, incorporating any corrective feedback provided by the supervision module and then forwarding the most up-to-date state information to the decision and execution modules.

3.3. Decision module

The decision module is responsible for parsing the activity and interaction information provided by the perception module and decide the robot's behavior accordingly. In particular, the decision module should be able to determine when to switch between tasks, as well as determining what the robot should do during the activities, as a response to the children's behavior.

As such, the decision module also exhibits a hierarchical structure, outlined in Fig. 6. At the higher level, an *activity manager* tracks the progress of the current activity and determines, as a function of the child's response, the time to switch between activities. The decision-making process can rely on a predefined finite-state machine or a more sophisticated decision-theoretic policy, using for example, the ROS `mdm` package.[10] Therefore, at each moment, the activity manager will activate one activity and deactivate all other activities, depending on the perceptual information provided by the perception module and any feedback provided by the supervision module.

Associated with each individual activity is, in turn, an *interaction manager*, responsible for determining, in the context of that particular activity, what behavior the robot should exhibit at each moment. Such interaction managers can be seen as specialized versions of the activity manager, holding their own decision process (using also a finite-state machine or pre-computer policy). The addition of new activities can be done by simply designing a new interaction manager for that activity. It is worth noting, in particular, that the current system already supports activities involving third-party applications (as is the case of the puzzle and tangram activities, each corresponding to an application that is run within the system).

It is important to note that not all activities are played in every session. The activities for each therapy session are defined beforehand by the therapist through the consoles associated to the supervision module.

Activity Manager. In its current version, the INSIDE system uses a finite-state machine that was carefully designed to match the outline of the therapy session and address all the interaction situations encountered in the pilots—non-responsive children, sudden changes in activity or activity order, among others. Such finite state machine is computationally light, as it takes as input the state (provided by the perception module) and acts upon it - either by triggering a new activity or by allowing the current interaction manager to conclude.

The use of an authored finite state machine as activity manager for our system was a deliberate choice, motivated by the restrictions of the application scenario. First, the manager ought to be lightweight, avoiding unnecessary computation that could disrupt the interaction by introducing unexpected latencies. Second, the behavior of the robot ought to meet—as much as possible—the desiderata derived both from the therapeutical goals and the situations encountered during the preliminary studies.

Similar considerations would apply, had the activity manager resulted from some decision-theoretic framework, such as a Markov decision process. By treating the information provided by the decision module as state, it would be possible to design a reward function and construct a model that would describe the dynamics of the interaction. Then, using standard decision-theoretic tools, a *policy* (or contingency plan) could be computed that could be deployed in the system as the activity manager. We note, however, that for the reasons outlined above (avoiding latency, etc.), it is better to compute the policy offline (and,

[10] `http://wiki.ros.org/markov_decision_making/`

as new information arrives, eventually update it during the robot's idle time). A policy computed offline is no different from a finite-state machine, and could be used as our activity manager with no change in the overall architecture of the system. Finally, we note that the activity manager plays the role—in our system—to the "Control Layer" found in the ROS `mdm` package, parsing "observations" and responding with "actions".

To conclude, the activity manager in our system is a lightweight component whose sole purpose is to monitor the state (provided by the perception module) and, when necessary, interrupt/launch the different interaction managers. Computationally, the "heavy-lifting" part of the INSIDE system is performed at the perception manager and in designing the actual policy adopted by the activity manager—which we have done offline.

Interaction Managers. Currently, the INSIDE system comprises a total of 7 interaction managers, one for each of the activities planned for the therapy sessions, to know

- *Welcome manager.* This interaction manager seeks to engage the child during the welcome process, providing successively richer prompts depending on the child's response, and concluding by inviting the child to play. These richer prompts are meant to call for the attention of the child when this is unresponsive; starting by just calling the child's name, progressing to add music in parallel with calling the child and finally adding movement.

- *Ball game manager.* This interaction manager provides the necessary explanation of the ball game before inviting the child to play. During the game, it provides incentives and reinforcement to the child when the child finds a ball and successfully places it in the robot's pouch. At the end of the game, it requests the child's assistance to remove the balls from its pouch. If at any time during the activity the child loses interest the robot motivates the child to continue the activity, either by asking to find the balls or to continue removing them from the basket.

- *Obstacle interaction manager.* This interaction manager seeks to enroll the assistance of the child every time the robot finds its way blocked by an obstacle, and is unable to find an alternative path.

- *Puzzle manager.* This interaction manager is similar to the ball game manager, in that it explains to the child the puzzle before inviting her to play. Additionally, it is also responsible for incentivizing the child to ask for assistance with respect to the missing pieces and provide the necessary feedback. During the entire game, this manager is also responsible for keeping the child motivated in the game by either praising the child's performance when it places a piece in the correct place, warn her when it places a piece wrongly or by asking the child to continue playing if there are pieces remaining out of place.

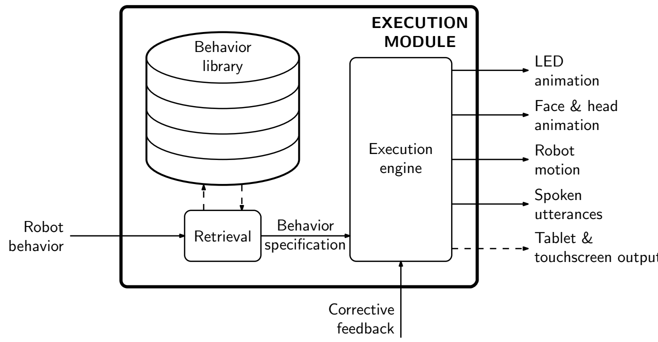


Figure 7: Execution module, responsible for translating the behaviors determined by the decision module into the multimodal output of the robot.

- *Tangram manager.* This interaction manager is similar to the previous one. It explains the Tangram game and the turn-taking play mode. The manager also ensures a positive probability of, during a game, enrolling the child's assistance during the robot's turn to play.

- *Goodbye manager.* This interaction manager conducts the final moments of the session, where robot and child head to the door and say goodbye.

3.4. Execution module

The execution module is responsible for translating the behaviors determined by the decision module into actual robot actions, and its structure is outlined in Fig. 7. In particular, the robot has a set of pre-programmed behaviors—both general and task specific. For our purposes, we define a *behavior* as the composition of the different output modalities supported by the robot, namely LEDs, face animations, head movement, body movement, speech and, in specific activities, the interactive devices (touchscreen and tablet).

An example of a general behavior is the `idle` behavior, in which the robot maintains a neutral facial expression with a slight rhythmic movement that simulates breathing. At the same time, every now and then, the robot performs a small head movement. This idle behavior, used in situations where the robot is expected to play a passive role, was designed to convey a "sensation of living" even during these moments. An example of a task specific behavior is, for example, the `obstacle detected` behavior. When the robot realizes that its path is blocked by an obstacle, it exhibits a sad face while executing small sideway motions, as if looking for a passage. This behavior is only triggered upon the detection of an obstacle and the activation of the "Obstacle activity" (see Fig. 8).

Spoken utterances. A key part of the interaction between the child and the robot relies on the ability of the robot to verbally communicate with the child. Endowing the robot's vocal output with expressiveness and natural emotions

Figure 8: Example of a task specific behavior of the robot. Upon detecting an obstacle, ASTRO makes a sad face and tries to find a way around, before asking the child for assistance. In the background it is visible the whole hardware infrastructure of INSIDE.

is essential to ensure an engaging experience for the children. Therefore, in order to optimize the robot's vocal output, we used the early WoZ studies to evaluate several state-of-the-art text-to-speech engines. In the early stage of the project, we evaluated CereProc,[11] Nuance Vocalizer,[12] Acapela,[13] and the DIXI TTS engine [63]. Such preliminary studies brought to the forefront the limited expressiveness and natural emotion in the speech synthesized by these systems, which led us to opt for pre-recorded human speech for the robot's vocal output. Thus, an extensive set of pre-defined utterances has been recorded in a sound-proof room, which cover all the different activities programmed for the therapy sessions. The general characteristics of the speaker (gender, age, and voice tonality), along with the specific characteristics of each of the recorded utterances (speaking speed, prosody, intonation, expressiveness, emotivity, etc.) were recorded to meet the requirements raised by the medical experts, ensuring an appropriate interaction with the ASD children participating in the studies.

In addition, the AUDIMUS engine used for keyword spotting is employed to perform a forced alignment (phone-level segmentation) of the pre-recorded speech files with their corresponding transcriptions, thus providing the sequences

[11] `https://www.cereproc.com`

[12] `https://www.nuance.com/omni-channel-customer-engagement/voice-and-ivr/text-to-speech.html`

[13] `http://www.acapela-group.com`

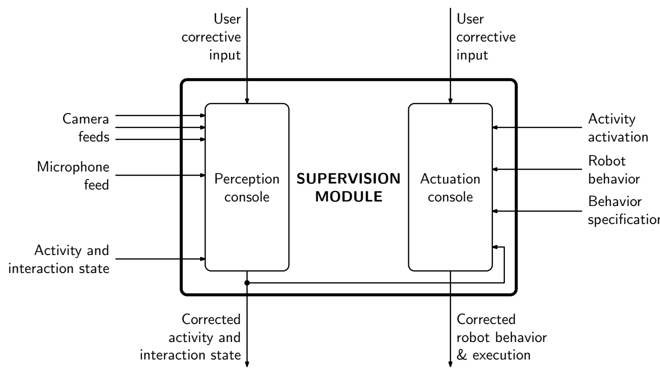


Figure 9: Overview of the supervision module, which provides external users with the necessary tools to take control of the system, if necessary.

of phonemes in each file and their durations. These phonemes and their duration are used by the execution engine to perform speech-lip synchronization.

3.5. Supervision module

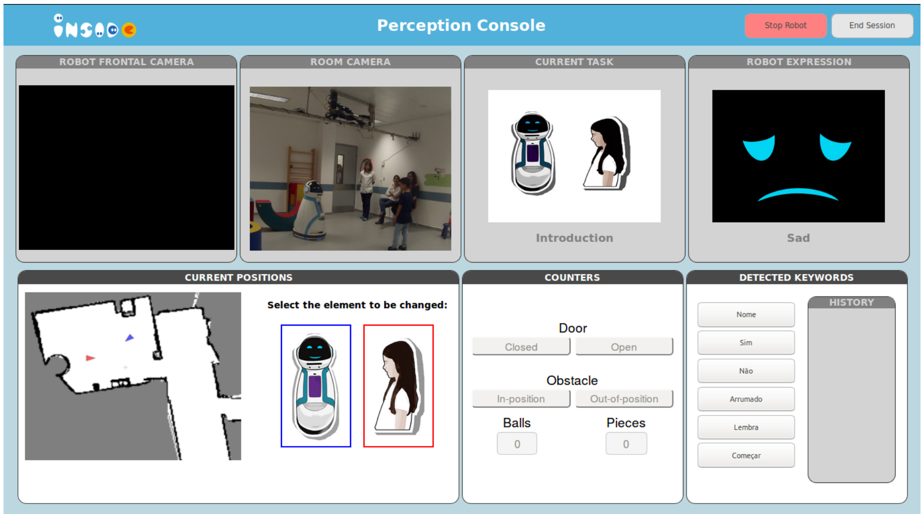
The supervision module provides a backdoor to the perceptual and behavioral elements of the robot, allowing human supervisor to take control of the interaction at any time, should some anomalous circumstance so demand. If no intervention is required, the supervision module has no impact on the system.

The supervision module is outlined in Fig. 9, and comprises two *operator consoles*.

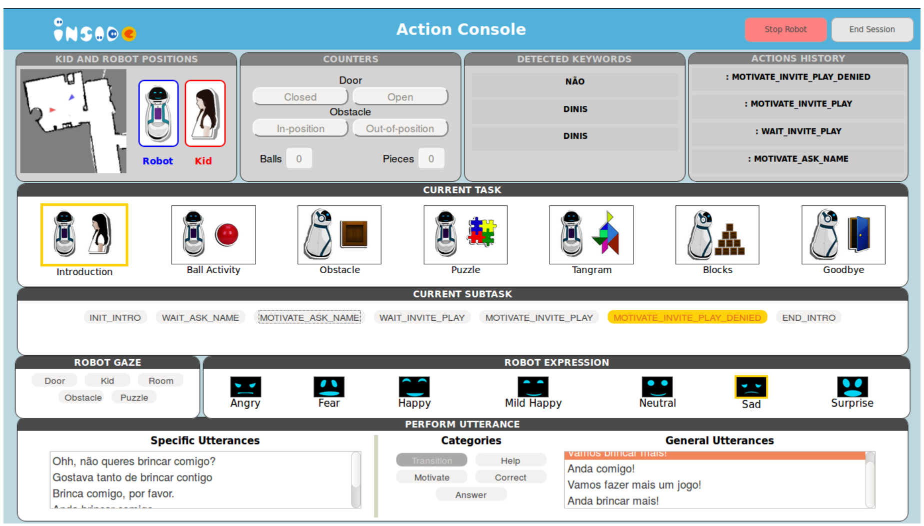
- *The perception console* provides a human operator with direct access to the camera network and microphone feeds. The perception console operator is also positioned behind a one-way mirror, having direct visual perception of the room and the events taking place therein. The goal of the perception console operator is to monitor the perception module, making sure that the sensor data is properly processed and correcting this information whenever problems are encountered. The corrective feedback provided by the operator can later be used to improve the performance of the perception module, as discussed in Section 4.

- *The actuation console* provides a human operator with access to the current (processed) activity and interaction state, as well as to the robot's current decision process. The goal of the actuation console operator is to monitor the decision and execution modules, making sure that the behavior of the robot is adequate to the current situation. The operator of the actuation console must rely only on the information provided by the perception module, according to the restricted perception WoZ methodology described in Section 4.

The supervision module also allows the configuration of several therapy session parameters, such as the number of the session, the name of the child and the



(a) Perception console.



(b) Actuation console.

Figure 10: Screenshots of the two consoles belonging to the supervision module.

activities to be executed in that session. Figure 10 provides a screenshot of the two consoles.

We conclude by noting that the supervision module was crucial to support Wizard-of-Oz experiments in the preliminary studies, as the operators, hidden from the children and supported by a therapist, could check in real-time the situational awareness estimated by the robot system from its sensors, possibly overriding wrong estimates, and similarly check the decisions autonomously

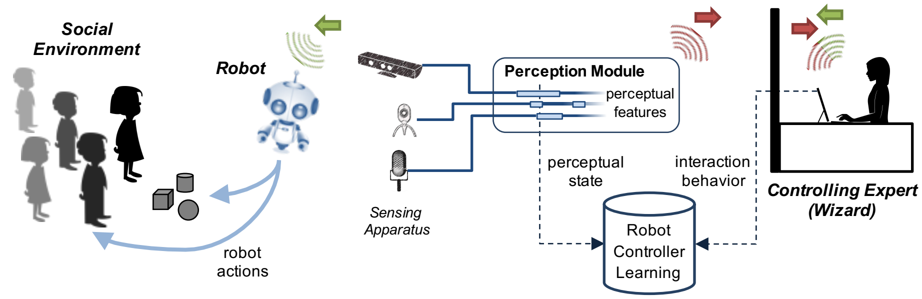


Figure 11: Restricted-perception WoZ methodology. Raw sensor data is filtered by a Perception Module to inform the controlling expert during the studies. Offline, a robot controller algorithm learns a mapping function between the perceived states and the robot behaviors chosen by the wizard. Colored arrows indicate the flow of information during the interaction.

taken by the robot system, possibly overriding them if inappropriate. In the process, the system developers could understand the causes of wrong sensing and wrong decisions and correct them in the next session. Such iterative development process led to a steady progress: while in the initial studies the operators had to intervene frequently, over time they had to intervene less and less until the final, long-term study, in which the robot operated fully autonomously.

4. Social Interaction Design Methodology

As discussed in the previous sections, the purpose of the INSIDE networked robot system is to enable a mobile robot to socially interact with ASD children during therapy sessions in a set of different activities. The social interaction capabilities of the robot are, therefore, determinant to the success of the intervention using the robot. In order to design the interaction of the robot, we adopted a methodology known as *restricted perception Wizard-of-Oz development* [82].

The key idea behind the adopted design methodology is that interactions involving remote operation of the robot—following a WoZ paradigm—actually provide a significant amount of useful information regarding the intended social behavior of the robot. Unfortunately, many perceptual limitations of the robot are often disregarded by giving the Wizard complete access to observations over the interaction, which poses difficulties when automatically extracting the social behavior showcased by the Wizard. We refer to this problem of perceptual mismatch as the *perceptual correspondence problem*: considering that humans and robots have very different sensory capabilities, the same scene may appear very different for the expert and the robot [11]. The practical consequence is that it is often difficult to correctly associate the actions demonstrated by the Wizard with the corresponding perceptual inputs.

To mitigate the perceptual correspondence problem, researchers proposed *immersive remote teleoperation* where the expert is limited to observe the interaction from the robot's perspective, relying exclusively in the robot sensors (e.g., cameras) and actuators [11, 16]. Such technique addresses, to some extent, the perceptual correspondence problem. We argue, however, that even if the Wizard is restricted to perceive everything according to the robot's point-of-view, there is still a significant amount of information that the human extracts from raw sensor feeds that will not be available for the robot to reason upon. We shall refer to such mismatch as the *perceptual-cognitive correspondence*. Consequently, we cannot replicate the Wizards' decision process, making standard teleoperation techniques ill-suited to learn the robot interaction behavior required for social interactions with human users.

To address the aforementioned problems, we instead adopt a *restricted perception Wizard-of-Oz approach*, first proposed in [82], as an methodology to allow a meaningful extraction of behaviors from Wizard interactions. Formally, we define an *interaction strategy* as a mapping between a robot's perceptual state and an interaction behavior from its behavior repertoire. We argue that in order to extract meaningful information that can be used to design an autonomous interaction, the robot's perceptual, behavioral, and cognitive limitations have to be taken into account during the interaction design. As such, the human expert acting as the Wizard should be *restricted* from perceiving everything occurring within the task during the studies.[14]

The adopted experimental procedure is illustrated in Fig. 11. The wizard has access to the a restricted view of the interaction, consisting of the robot's *processed* sensor information, from which it must then develop an appropriate interaction strategy. By enforcing such restricted-perception WoZ interaction, we even out the type and amount of information and the interaction behavior available to both the wizard and the robot—a central tenet of our approach. As a result, we mitigate not only the physical but also the perceptual-cognitive correspondence problem mentioned above.

◇

The design methodology can be broken down in three key steps, summarized in Fig. 12:

- *Data collection*, which includes preliminary studies conducted to gather task-specific data and usually relies on an unrestricted WoZ approach. The data collected in these studies is usually used to test and refine the *task AI*, i.e., the high-level decision process that governs the task-specific behavior of the robot. Such task AI is also used to assist the Wizard during the restricted perception WoZ studies, where the interaction data (perceptions and behaviors) are collected.

- *Behavior extraction*, in which the data collected during the data collection stage is used to design and refine a number of interaction managers—

[14]We henceforth refer to the standard WoZ technique as the *unrestricted WoZ* to denote the capability of the wizard in observing the interaction context.

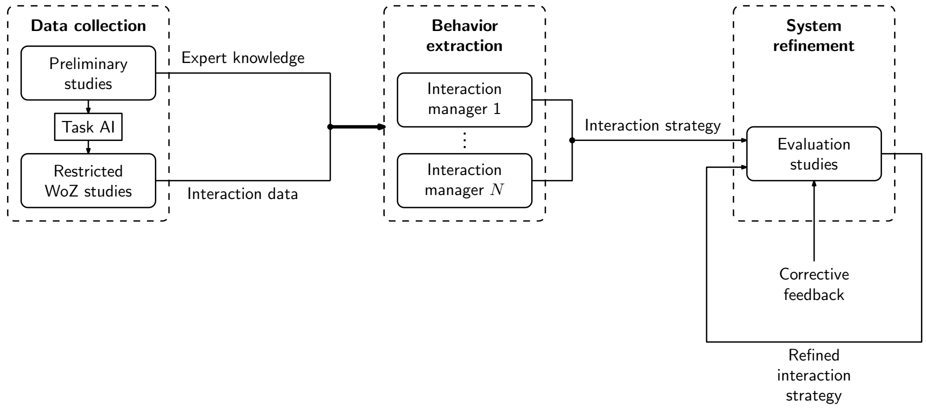


Figure 12: The three steps of the proposed methodology for designing social interaction strategies from restricted-perception Wizard of Oz interactions.

modules that will be responsible for managing the interaction between the robot and the human users in different contexts of the task.

- *System refinement*, in which the robot autonomously interacts in the target scenario. Evaluation studies can be conducted incrementally, allowing for corrective feedback provided by a human supervisor to be used to refine the interaction strategies.

We refer to [82] for further details on the design methodology.

5. Preliminary and long-term studies

In this section, we discuss in further detail the preliminary studies and long-term study ran with the INSIDE system. Our goal in this discussion is to highlight the application of the methodology and to assess the ability of the robot to autonomously perform in a therapy session, as intended. We also provide a brief discussion on how the studies conducted in the context of INSIDE can provide potential insights about the impact of the system in ASD therapy, although such impact falls somewhat outside the scope of this paper.

As discussed in Section 4, we conducted a number of studies involving the use of a mobile robot in the therapy of children with ASD. Such studies were conducted in Hospital Garcia de Orta, in Almada, Portugal, and involved a number of children diagnosed with ASD and followed in the Hospital. In particular,

- We conducted two initial studies involving 5 children, which relied exclusively on a Wizard-of-Oz (WoZ) paradigm. In these studies, the robot was remotely controlled by two operators. One operator was in charge of the robot motion, while the other was in charge of the robot's social behavior.

- We then conducted a restricted perception Wizard-of-Oz study, as described in Section 4. This study involved 6 different children. The robot

Figure 13: Illustration of the preliminary studies. Left: Two wizards control the robot behavior—the operator in front of the computer controls the interaction, while the one standing navigates the robot. The robot top is visible on the one-sided mirror in the background. Right: Therapist mediated interaction at the end of a therapy session. The child brings the robot to the door to say goodbye. Visible in the background is the room where the interaction took place.

behaved autonomously and was monitored by two operators: one was in charge of monitoring the robot's perception and the second was in charge of the robot's actuation. However, there was a significant number of interventions by both operators to correct the perception and the behavior of the robot.

- One final, long-term study in which the robot participated in a total of 121 therapy sessions involving a total of 18 children, which lasted for 4 weeks. During this study, the robot operated autonomously while being monitored by two operators: one was in charge of monitoring the robot's perception and the second was in charge of the robot's actuation. Unlike the previous study, however, the interventions of the operators during the 121 sessions were minimal.

In the continuation, we discuss relevant aspects of the different studies in further detail.

5.1. Preliminary Studies

Preliminary studies were conducted involving 5 boys (ages ranged from 3 years and 9 months to 5 years and 9 months) diagnosed with moderate ASD (according to ADOS-2) in a total of 2 session (session 1, $N = 4$; session 2 $N = 3$). The children were selected from a database for ASD children at the Child Development Center in Hospital Garcia de Orta, Portugal. For the first two sessions, the children were selected based on age and availability.

During this preliminary study, the robot was controlled via a WoZ paradigm. One operator was in charge of controlling the motion of the robot, while a second operator was in charge of controlling the social behavior (speech, facial expressions) of the robot (see Fig. 13). The use of a WoZ paradigm had important

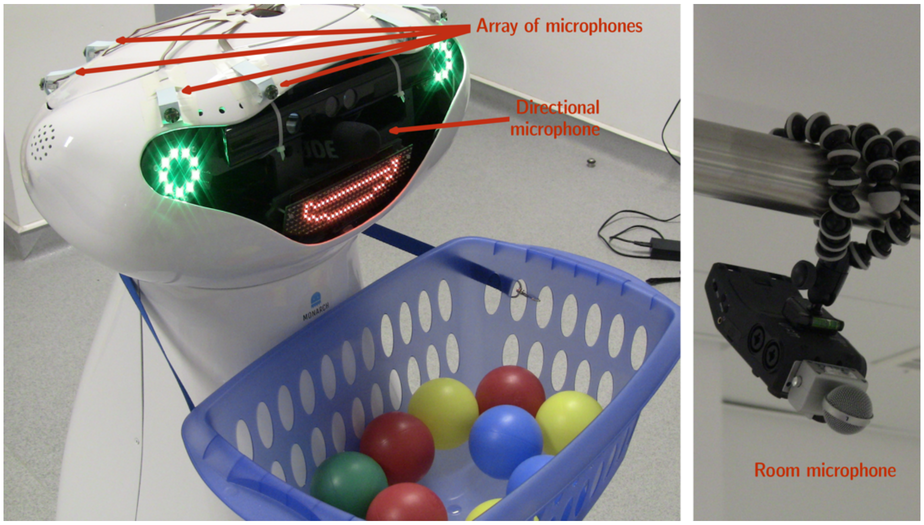


Figure 14: Microphones on the robot and in the room used to capture speech in early pilots.

benefits, namely the ability to experiment with a wide range of utterances and behaviors. For example, in the initial pilot, a comparison was performed between using synthesized speech versus pre-recorded speech. In the former situation, besides a number of pre-selected utterances, the operator had the possibility of typing, at runtime, new utterances that the robot would then voice.

However, while the use of synthesized voice allowed for greater freedom, it was observed that (i) the children were significantly more responsive to the human voice in the pre-recorded speech; (ii) the typing of "new" utterances by the operator introduced a significant lag in the interaction with the child.[15] For these reasons, synthesized speech was dropped in favor of pre-recorded speech.

Similarly, the motion of the robot when controlled by a wizard was very unnatural, and some children were clearly not comfortable when the robot moved. With the automation of the robot motion, the movement of the robot significantly improved, both in terms of fluidity and in terms of the children's response.

The initial studies also were fundamental to define key requirements for our system. Namely, the initial studies allowed us to

- *Realize the perceptual needs for the system.* By observing the interaction between the children, therapist, and robot during these initial studies, we identified the key variables that should drive the behavior of the robot (e.g., child's position, robot's position, state of the current activity, state of the dialogue between child and robot, etc). The perceptual module

[15] As an example, on one occasion a child was going to place a ball in the basket; the wizard decided to type an utterance saying "thank you" but, in the process, the child dropped the ball. Since the utterance had been sent to the robot, it thanked the child anyway, and the child laughed and said "no".

was thus designed to provide the necessary information regarding these variables.

As an example, it was apparent from the early pilots that the children's voice was too low to be captured by microphones either in the robot or in the room (see Fig. 14 for an example of the microphone setup used in the early pilots to test voice capture). That motivated the use of a head-mounted microphone for the therapist, who repeats some of the answers provided by the children in a way that the system can recognize.

- *Build detailed scripts for the different activities*, which in turn allowed us to prepare the robot to respond to most common situations in a fully autonomous manner. The initial studies allowed us to refine the interaction abilities of the robot, since they already provided a wide range of situations to which the robot could respond. As an example, one child in the initial studies responded "No" when the robot invited them to play. This behavior was used to improve interaction and, in subsequent sessions, the robot was prepared for this reply, to which it would make a sad face and say "I would really love to play with you".

  As another example, we also noticed in the preliminary studies that, when a child was struggling to complete a task (such as removing the obstacle) and the robot kept asking for help, some of the children would lose interest and start exploring the room. However, once the therapist removed the obstacle and the robot moved to a different game, the children looked at the robot and followed it to see the new game. From this observation we endowed the robot with a behavior that, upon detecting that the child moved away from the robot in certain activities (indicating that the child may have lost interest), the robot would actively try to call the child's attention, if necessary by moving to a different task.

All these insights were drawn from the preliminary studies, and were used to improve the system for the long-term study. In any case, and regardless of how well planned the sessions were, there was sometimes unexpected behavior from the children; for example, one of the children asked the robot "are you from Star Wars?". In these situations, the therapist would provide an adequate response to the child and allow the session to continue without any disturbances.

Overall, we noticed that the children enjoyed the sessions. Some tasks were easier for them to understand, like the ball game, whereas others, like helping the robot to remove an obstacle, were more challenging. The fact that the robot was able to play some games appeared to be very important.

5.2. Restricted Perception Wizard of Oz Study

We applied the restricted-perception WoZ methodology described in Section 4 in the context of INSIDE to design the social interaction behaviors of the robot. Specifically, after concluding the preliminary studies described above, and with the task AI in place—i.e., the perception and decision modules designed from the interactions during the preliminary studies—we conducted a restricted-perception WoZ study, in which a Wizard accessing only the processed perceptual data from the robot remotely controlled the interaction between the robot and the child.

The task AI controlled a significant part of the robot decision process (including the robot motion and the sequence of activities conducted during the therapy sessions), allowing the Wizard to focus on the social interaction. Nevertheless, it was always possible for the Wizard to override the AI and modify the behavior of the robot.

During the restricted Wizard of Oz studies, we included a second human supervisor (which was not in contact or within sight of the Wizard operator) which would be able to observe the raw feeds from the cameras in the environment and provide corrective feedback over the perception modules, correcting any perception that is wrongly processed. It is the corrected processed perception that the Wizard operator accesses to control the interaction. The objective behind this approach is to provide a chance to correct and further improve the perception algorithms, and also to ensure that the interaction behavior extracted from the restricted perception Wizard is matched with perceptual information as accurate as possible. This is especially important in complex situations involving very dynamic and unpredictable elements in the environment, as is the case of the INSIDE scenario.

The outcome of the restricted WoZ studies was used to refine the perception and behavior modules—particularly in what social interaction is concerned. The final behaviors implemented in the robot are, therefore, a direct result of the data collected in these preliminary studies and allow the robot to operate in a fully autonomous manner during the final study.

5.3. Long-term Study

In the long-term study, we ran a total of 121 sessions spanning a total of 4 weeks and involving 18 children. The session would last between 15 and 25 minutes of interaction between the robot and the child. Each day comprised, on average, between 6 and 7 sessions with different children. As mentioned before, for safety concerns, two operators continuously monitored the system, during the 121 sessions. One of the operators was responsible for supervising the perception, while the second operator was in charge of supervising the action of the robot.

The operators monitored both anomalous behavior by the robot but also other failures in the system that could compromise the activity, since such failures could render the data collected during the session useless. During the full period of the study, we observed the following interventions by the operator in charge of perception:

- The perception module sometimes confused the child's position with that of the therapist, particularly in situations where the therapist would get down to address the child. In order to minimize the impact of such missed localizations, which influences the gaze and body posture of the robot, the perception operator would sometimes provide corrective feedback, indicating the correct position of the child. On average, the operators provided corrective feedback on 6% of the child localization measures.

- Similarly, the perception module would sometimes miss a keyword, particularly in situations where the therapist would say a keyword while other people in the room were talking, such as the child or the parents. In some of these situations, the therapist repeated the keyword and the system continued autonomously; however, in some other situations the perception operator provided corrective feedback. This occurred on 2% of the keywords uttered by the therapist.

Note that the interventions listed above corresponded to *perceptual corrections* and do not directly address any aspect of the robot's behavior.

Besides the perceptual interventions, there were 6 occasions, within 121 sessions, in which the communication between the tablet, where the puzzle was played, and the rest of the system experienced undue delay. This situation prevented the robot from knowing that the child had already started the puzzle activity. It was only in these situations that the action operator manually provided the robot with this information, thus triggering the corresponding robot behavior.

5.4. Impact in ASD therapy

The long term study described above, besides providing a useful scenario to assess the autonomy and performance of the INSIDE system as a whole, also provides a key opportunity to evaluate the impact of the INSIDE system in the therapy for children with ASD. Although such analysis is out of the scope of this paper, it is nevertheless interesting to consider that the preliminary studies, besides fundamental to refine the system itself, were also extremely useful to define and refine the measures to be collected in the long-term study.

5.4.1. Observation measures

Since the preliminary studies involved only a small sample, we focused on a qualitative analysis as the starting point, preparing the long-term study in which we take a mixed methods approach [37, 38, 51]. In particular, we used the preliminary studies to assess the willingness of the children to interact with a robot, the appropriateness of the activities—whether or not the games, which are used on regular therapy sessions, would be adequate for a session conducted by the robot and moderated by a human therapist—and to obtain a set of measurements that could allow us to evaluate the children's engagement on the long-term study.

One difficulty in defining such measures is that engagement is a ubiquitous concept that crosses many research fields, and it seems that there is no consensus on its definition, even if we only consider definitions of engagement in the context of HRI (see, e.g., the review of Salam and Chetouani [78]). This lack of consensus in the field of HRI with adults is extendable to the definition and measurement of engagement in ASD populations [53, 39, 77], with many studies assessing engagement via external observers [41, 61, 70, 91].

However different the definitions may seem, they all understand engagement as a mental state that is linked to an interaction and has to be inferred, meaning that there is not a test that can accurately tell us whether a person is engaged or not at a given moment. Inferring engagement for adult participants and typically developing children is difficult and complex. For example, Ivaldi et al. [36] mention a very pertinent issue that goes beyond the assessment of eye gaze, speech or individual differences, stating that the estimates of engagement in HRI might be biased by the task engagement. If a child does not like the game that the robot is inviting her to play, than assessments of engagement during that task are probably biased.

The difficulties in studying engagement seem to grow larger as we move to clinical populations, especially populations with developmental disorders that impair social interactions, such as ASD.

For this reason, we opted by inferring engagement in our long term study indirectly, focusing on the overall quality of the sessions. In the preliminary studies we measured

Task completion Rate (TCR). We defined TCR as the percentage of games that the children could complete on their own. All participants were able to complete more than 50% of the tasks.

Given the differences between pilots, comparing TCRs would be unfruitful. However we noted that children with higher TCRs seemed to be paying more attention to the robot, thus understanding what they had to do, and they seemed to be speedier. This observation suggested that time could be a good measure when associated with TCR, and that children that were paying more attention and performing better were more engaged with the robot. When we looked at the TCR and session lengths, we indeed noticed a tendency for shorter sessions to have higher TCRs, but with a small sample it was difficult to draw any definitive conclusions.

Time. One of the earliest observations during the pilots was that shorter sessions appeared to correspond to sessions where the children performed better (higher TCR's, see above). Since our sample was small, we could not definitely establish this fact, but we hypothesized that shorter sessions would correspond to higher completion rates.

Eye gaze. For the first two pilots we used a simple observation grid that only allowed us to evaluate the frequency of eye gaze directed at the robot. However, gaze frequency in itself does not provide information on the percentage of time the children look at the robot, only how many times they gaze at it.

We thus opted by replacing eye gaze frequency with the percentage of time that the children spend looking at the robot, and replaced the observation grid with ELAN® (`http://tla.mpi.nl/tools/tla-tools/elan/`). Eye gaze is extremely difficult to assess in ASD children in general, but is even more so in sessions where both child and robot are allowed to move freely, as was our case. There is a tendency to infer that higher percentages of eye gaze correspond to more engagement and are therefore desirable. However, our preliminary studies showed that a higher percentage of eye gaze directed at the robot does not necessarily reflect a desirable behavior. On the contrary, higher percentages of eye gaze often appear in sessions where the children stare at the robot with their faces nearly touching, which corresponds to behavior that is inadequate in social interactions. This observation raises a challenging but interesting question: how can we evaluate and interpret the quality of eye gaze in future sessions to have a good measure of the children's engagement in social interaction? Should we discount or take into account the staring at the robot as inappropriate/undesirable behavior in social interaction? Or instead, should we consider staring as some form of engagement with the robot?

Speech. ASD entails deficits in communication and language impairments are frequent. For this reason, at an initial stage we opted by considering only the vocalizations directed towards the robot; this option had the advantage of not depending on the quality or even the appropriateness of the vocalization. However, we realized that the data thus collected would not allow us to account for stereotypies and echolalia. Without content, it was impossible to interpret.

We latter decided to carry out content analysis for both the children's speech and the therapist's, including vocal stereotypies and vocalizations that manifest emotional states. This analysis of the children's speech reflected mainly individual differences in language ability, but our observations support that it is relevant to assess progress throughout the long-term study. We noted that most of the childrenâĂŹs speech was directed at the robot, which is a great indication of both engagement with the robot and enjoyment.

Content analysis of the therapist speech also gives an overview of the quality of the sessions. For example, if the therapist needs to repeat the instructions or needs to provide a lot of corrective feedback to the child, then the child probably did not comprehend the instructions and/or was doing something that was not supposed to do (e.g., throwing the balls at the robot instead of placing them on the correct spot).

As a result of the observations above, we defined the overall quality of the session as an interaction of these factors. Finding a measure of engagement that takes into account all these factors allows us to not have to rely on subjective appreciations of independent observers and to have a more consistent measure for all sessions and all participants. Most importantly, the preliminary studies conducted so far have sparked a discussion that led to protocol changes that greatly improved our original protocol.

Figure 15: T-shirt designed to collect physiological signals from the children interacting with the robot. The T-shirt comprises a chest-band and an acquisition module. The chest band is located near the diaphragm of the user, while the module is placed in a small pocket, on the down right position of the T-shirt.

5.4.2. Physiological signals analysis

In addition to the set of measures discussed above—which rely on the observation of the behavior of the children during their interaction with the robot—we also analyzed a posteriori how the interaction affects the children from a physiological perspective, thus providing a more complete understanding of the impact of robots in therapy. The physiological signals are not used to inform the interaction between the children and the robot in any way, but rather to complement the observational measures discussed above regarding the impact that the interaction with the robot had on the child, from a physiological point-of-view.

For this purpose, a T-shirt with wearable sensors has been designed to acquire physiological data. This T-shirt, depicted in Fig. 15, is equipped with:

- A chest band for electrocardiogram (ECG) monitoring. The chest band is elastic and has 2 conductive textile contact zones with a sponge and dual electrode inside.
- An acquisition module developed by PLUX, S.A., which has an integrated accelerometer (ACC) and is connected to the chest band, is used to acquire the ECG. This module transmits the data acquired (both ECG and ACC data) via bluetooth to the connected computer with the OpenSignals software.

The design requirements for the T-shirt were that the child should not feel

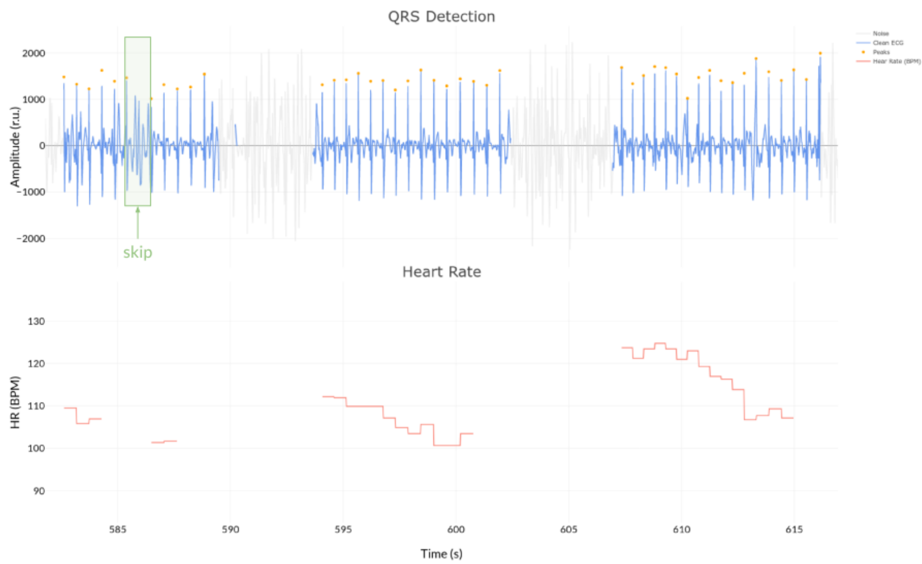


Figure 16: Heart Rate determination in the areas of clean ECG. The areas that are not represented in the Heart Rate plot are either noise or false peak detections.

too uncomfortable, ensuring that wearing it would not compromise the child's performance during the therapy sessions. Once again, the preliminary studies were key in refining the design of both the T-shirt and the wearable device.

In the long-term study, 4 children wore the T-shirt during all sessions. Some sessions were discarded due to the poor quality/corruption of the ECG signals. The corruption of the ECG signals was mainly caused by the child's movements, but also due to stretching, displacement and even the detachment of the ECG sensor from chest-band during the session.

The presence of highly corrupted data required an opportunistic approach, where the information is retrieved whenever there is a window of opportunity for it. Therefore there is the need to find the areas of interest of the signals acquired, from which can be retrieved indicators. This data can be used in conjunction with observational information by therapists and psychologists to get more insights on the interaction quality factors.

The first approach for the development of an opportunistic model that retrieves relevant information from the ECG signal, was a noise detection clustering algorithm. This has been developed in order to distinguish noise from clean samples of the signal by means of an agglomerative clustering method based on a combination of statistical and morphological features, and the accelerometer data. By combining these features, it has been demonstrated that it is possible to find the areas of the signal that are corrupted with several types of noise and apply more specific and appropriate methods for the de-noising process, while being useful as well for signal reconstruction [75]. The purpose was to select the clean areas of the signal in order to analyze the heart rate variability (HRV).

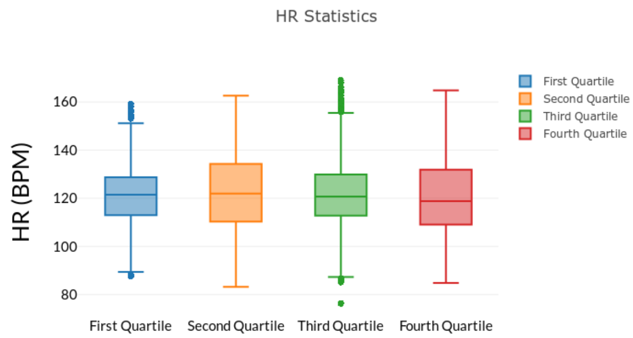


Figure 17: Overall evaluation of Heart rate over the entire set of ECG signals. The box plot here presented shows the heart rate median, interquartile range, maximum and minimum values. These values correspond to quartiles of each session.

The analysis of the ECG has been made after selecting the clusters of the signal that were suitable for peak detection. Figure 16 depicts a small segment of an ECG signal of one of the subjects, in which it is possible to identify the areas selected as clean or noisy. The ECG peaks were detected with the Pan Tompkins algorithm, which is typically used for real-time QRS detection. The areas of the signal that would be considered noise would not be used to calculate the HR. Besides, after calculating the HR, an evaluation would be made in order to identify skips or the appearance of a false peak. These cases can be visualized in Fig. 16.

After separating the clean areas of the signal from the noisy areas, it was possible to evaluate the heart rate over the sessions performed by the children. In this preliminary analysis, the heart rate was averaged for all children and over all sessions, in order to infer the statistical variance of this group. Figure 17 shows this analysis in four quartiles of the session, that is, the average of the heart rate in each quartile of all sessions, for all children. We can conclude that there is no tendency in this case. This result is purely statistical.

The second analysis involved the measure of the heart rate during specific tasks performed by the children during the sessions. In this case four tasks were evaluated: ball game, tangram, obstacle and puzzle. As can be seen in Fig. 18, all children have a higher heart rate during the balls and obstacle tasks.

Further analysis can be made to retrieve more information from the children performance during the tasks. With this deeper analysis, these results could be used by the therapists and psychologists to give them more information on how the child reacted over specific stimuli during the session. This can benefit the management of future sessions, in order to reorder tasks and help in planning more personalized sessions.

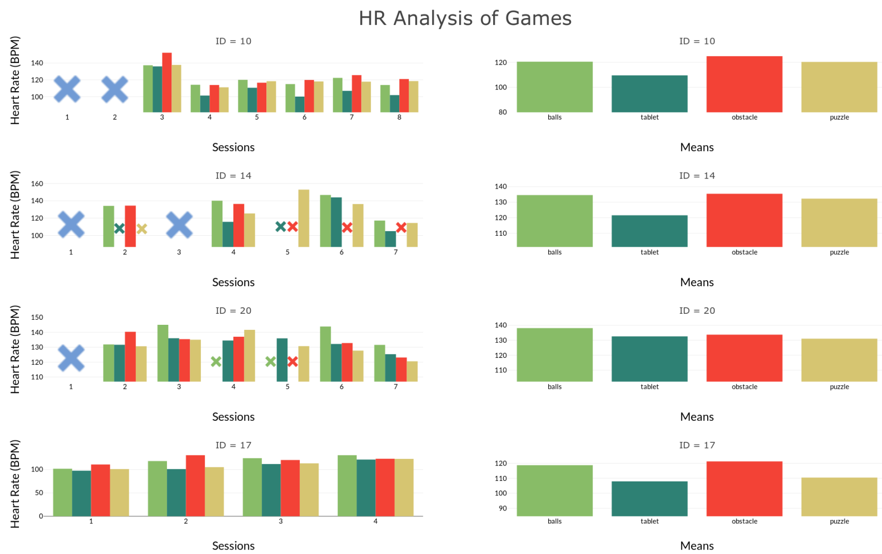


Figure 18: Heart rate evaluation during 4 specific games (balls, tablet, obstacle and puzzle). On the plots of the left are presented the means in each session. In the other hand, the right side shows the average of the heart rate for all games in all sessions performed by a subject. The blue cross indicates that there was not enough clean data to perform this evaluation. The remaining crosses indicate that there was not enough data to perform the evaluation for the corresponding game.

6. Conclusions

We described the work conducted in the context of the project INSIDE towards the development of a networked robot system that can be used in a wide range of therapeutical activities involving children with autism spectrum disorders.

Our system comprises an autonomous mobile robot, ASTRO, which is able to engage in social interaction throughout a therapy session, as the child moves around the room while completing several different activities. The robot plays a key role in the therapy session, as it is the robot's role to invite the child to perform the different activities, explain the activities to the child, and provide encouragement. To our knowledge, our work pioneers the use of an autonomous robot capable of driving a therapy session with multiple activities, acting like a social agent and engaging in rich social interaction with the children in a fully autonomous manner.

We described the methodology behind the development of the INSIDE system, from the preliminary studies in which the robot was tele-operated—following a Wizard-of-Oz paradigm—to the final, long-term study, in which the robot operated autonomously for 4 weeks and 121 actual therapy sessions in Hospital Garcia de Orta, in Portugal. The studies establish the robustness and autonomy of our system and hint into the potential that such a broad-purpose system can have in the therapy of children with ASD.

As future work, we note that the long-term study provided a significant amount of data that is currently being analyzed in terms of the metrics discussed in Section 5, to further assess the potential impact of INSIDE in terms of ASD therapy.

Acknowledgements

This work was partially supported by national funds through the Portuguese Fundação para a Ciência e a Tecnologia under project UID/CEC/50021/2013 (INESC-ID multi-annual funding) and the CMU | Portugal Program and its Information and Communications Technologies Institute, under project CMUP-ERI/HCI/0051/2013.

References

[1] A. Abad, A. Pompili, A. Costa, I. Trancoso, J. Fonseca, G. Leal, L. Farrajota, and I. Martins. Automatic word naming recognition for an on-line aphasia treatment system. *Computer Speech & Language*, 27:1235–1248, 9 2013.

[2] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association, 5th edition, 2013.

[3] E. Barakova, P. Bajracharya, M. Willemsen, T. Lourens, and B. Huskens. Long-term LEGO therapy with humanoid robot for children with ASD. *Expert Systems*, 32(6):698–709, 2015.

[4] E. Barakova, M. de Haas, W. Kuijpers, N. Irigoyen, and A. Betancourt. Socially grounded game strategy enhances bonding and perceived smartness of a humanoid robot. *Connection Science*, 30(1):81–98, 2018.

[5] S. Baron-Cohen. *Mindblindness: An Essay on Autism and Theory of Mind*. MIT Press, 1997.

[6] S. Baron-Cohen. The extreme male brain theory of autism. *Trends in Cognitive Sciences*, 6(6):248–254, 2002.

[7] C. Baum. Participation: Its relationship to occupation and health. *OTJR: Occupation, Participation and Health*, 23(2):46–47, 2003.

[8] E. Bekele, J. Crittendon, A. Swanson, N. Sarkar, and Z. Warren. Pilot clinical application of an adaptive robotic system for young children with autism. *Autism*, 18(5):598–608, 2014.

[9] S. Bernardini, K. Porayska-Pomsta, and T. Smith. ECHOES: An intelligent serious game for fostering social communication in children with autism. *Information Sciences*, 264:41–60, 2014.

[10] B. Bernardo, P. Alves-Oliveira, M. Santos, F.S. Melo, and A. Paiva. An interactive tangram game for children with autism. In *Proc. 16th Int. Conf. Intelligent Virtual Agents*, pages 472–475, 2016.

[11] A. Billard and D. Grollman. Robot learning by demonstration. *Scholarpedia*, 8(12):3824, 2013.

[12] A. Billard, B. Robins, J. Nadel, and K. Dautenhahn. Building Robota, a mini-humanoid robot for the rehabilitation of children with autism. *Assistive Technology*, 19(1):37–49, 2007.

[13] L. Boccanfuso and J. O'Kane. CHARLIE: An adaptive robot design with hand and face tracking for use in autism therapy. *Int. J. Social Robotics*, 3(4):337–347, 2011.

[14] J. Brok and E. Barakova. Engaging autistic children in imitation and turn-taking games with multiagent system of interactive lighting blocks. In *Proc. 9th Int. Conf. Entertainment Computing*, pages 115–126, 2010.

[15] J. Case-Smith and M. Arbesman. Evidence-based review of interventions for autism used in or of relevance to occupational therapy. *American Journal of Occupational Therapy*, 62(4):416–429, 2008.

[16] J. Chen, E. Haas, and M. Barnes. Human performance issues and user interface design for teleoperated robots. *IEEE Trans. Systems, Man, and Cybernetics C*, 37(6):1231–1245, 2007.

[17] C. Clabaugh, D. Becerra, E. Deng, G. Ragusa, and M. Matarić. Month-long, in-home case study of a socially assistive robot for children with autism spectrum disorder. In *Proc. Companion of the 13th ACM/IEEE Int. Conf. Human-Robot Interaction*, pages 87–88, 2018.

[18] A. Cook, K. Adams, J. Volden, N. Harbottle, and C. Harbottle. Using LEGO robots to estimate cognitive ability in children who have severe physical disabilities. *Disability and Rehabilitation: Assistive Technology*, 6 (4):338–346, 2011.

[19] K. Dautenhahn. Design issues on interactive environments for children with autism. In *Proc. 3rd Int. Conf. Disability, Virtual Reality and Associated Technologies*, 2000.

[20] K. Dautenhahn and A. Billard. Games children with autism can play with Robota, a humanoid robotic doll. In *Proc. Cambridge Workshop on Universal Access and Assistive Technology*, pages 179–190, 2002.

[21] K. Dautenhahn and I. Werry. Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmatics & Cognition*, 12(1):1–35, 2004.

[22] K. Dautenhahn, C. Nehaniv, M. Walters, B. Robins, H. Kose-Bagci, N. Mirza, and M. Blow. KASPAR—a minimally expressive humanoid robot for human-robot interaction research. *Applied Bionics and Biomechanics*, 6(3-4):369–397, 2009.

[23] M. Davis, N. Otero, K. Dautenhahn, C. Nehaniv, and S. Powell. Creating a software to promote understanding about narrative in children with autism: Reflecting on the design of feedback and opportunities to reason. In *Proc. 6th IEEE Int. Conf. Development and Learning*, pages 64–69, 2007.

[24] G. Dawson, A. Meltzoff, J. Osterling, J. Rinaldi, and E. Brown. Children with autism fail to orient to naturally occurring social stimuli. *J. Autism and Developmental Disorders*, 28(6):479–485, 1998.

[25] G. Dawson, S. Rogers, J. Munson, M. Smith, J. Winter, J. Greenson, A. Donaldson, and J. Varley. Randomized, controlled trial of an intervention for toddlers with autism: The early-start Denver model. *Pediatrics*, 125 (1):17–23, 2010.

[26] M. de Haas, A. Aroyo, E. Barakova, W. Haselager, and I. Smeekens. The effect of a semi-autonomous robot on children. In *Proc. 8th IEEE Int. Conf. Intelligent Systems*, pages 376–381, 2016.

[27] A. Duquette, F. Michaud, and H. Mercier. Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism. *Autonomous Robots*, 24(2):147–157, 2008.

[28] P. Esteban, P. Baxter, P. Belpaeme, E. Billing, H. Cai, H. Cao, M. Coeckelbergh, C. Costescu, D. David, A.D̃e Beir, Y. Fang, Z. Ju, J. Kennedy, H. Liu, A. Mazel, A. Pandey, K. Richardson, E. Senft, S. Thill, G. Van de Perre, B. Vanderborght, D. Vernon, H. Yu, and T. Ziemke. How to build a supervised autonomous system for robot-enhanced therapy for children with autism spectrum disorder. *Paladyn J. Behavioral Robotics*, 8:18–38, 2017.

[29] D. Feil-Seifer and M. Mataric. B^{3}IA: A control architecture for autonomous robot-assisted behavior intervention for children with autism spectrum disorders. In *Proc. 17th IEEE Int. Symp. Robot and Human Interactive Communication*, pages 328–333, 2008.

[30] D. Feil-Seifer and M. Mataric. Toward socially assistive robotics for augmenting interventions for children with autism spectrum disorders. In *Proc. 11th Int. Symp. Experimental Robotics*, pages 201–210, 2009.

[31] D. Fox. Adapting the sample size in particle filters through KLD-sampling. *International Journal of Robotics Research*, 22(12):985–1003, 2003.

[32] R. Gockley, A. Bruce, J. Forlizzi, M. Michalowski, A. Mundell, S. Rosenthal, B. Sellner, R. Simmons, K. Snipes, A. Schultz, and J. Wang. Designing

robots for long-term social interaction. In *Proc. 2005 IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, pages 1338–1343, 2005.

[33] J. Greczek, E. Kaszubski, A. Atrash, and M. Mataric. Graded cueing feedback in robot-mediated imitation practice for children with autism spectrum disorders. In *Proc. 23rd IEEE Int. Symp. Robot and Human Interactive Communication*, pages 561–566, 2014.

[34] S. Greenspan. *Infancy and early childhood: The practice of clinical assessment and intervention with emotional and developmental challenges.* International Universities Press, Inc., 1992.

[35] N. Hawes, C. Burbridge, F. Jovan, L. Kunze, B. Lacerda, L. Mudrová, J. Young, J. Wyatt, D. Hebesberger, T. Körtner, R. Ambrus, N. Bore, J. Folkesson, P. Jensfelt, L. Beyer, A. Hermans, B. Leibe, A. Aldoma, T. Fäulhammer, M. Zillich, M. Vincze, E. Chinellato, M. Al-Omari, P. Duckworth, Y. Gatsoulis, D. Hogg, A. Cohn, C. Dondrup, J. Fentanes, T. Krajník, J. Santos, T. Duckett, and M. Hanheide. The STRANDS project: Long-term autonomy in everyday environments. *IEEE Robotics & Automation Magazine*, 24(3):146–156, 2017.

[36] S. Ivaldi, S. Lefort, J. Peters, M. Chetouani, J. Provasi, and E. Zibetti. Towards engagement models that consider individual factors in HRI: On the relation of extroversion and negative attitude towards robots to gaze and speech during a human-robot assembly task. *Int. J. Social Robotics*, 9 (1):63–86, 2017.

[37] R. Johnson and A. Onwuegbuzie. Mixed methods research: A research paradigm whose time has come. *Education*, 33(7):14–26, 2004.

[38] R. Johnson, A. Onwuegbuzie, and L. Turner. Toward a definition of mixed methods research. *J. Mixed Methods Res.*, 1(2):112–133, 2007.

[39] D. Keen. Engagement of children with autism in learning. *Australian J. Special Education*, 33:130–140, 2009.

[40] J. Kelley. An iterative design methodology for user-friendly natural language office information applications. *ACM Trans. Information Systems*, 2 (1):26–41, 1984.

[41] C. Kemp, Y. Kishida, M. Carter, and N. Sweller. The effect of activity type on the engagement and interaction of young children with disabilities in inclusive childcare settings. *Early Childhood Research Quarterly*, 28: 134–143, 2013.

[42] E. Kim, R. Paul, F. Shic, and B. Scassellati. Bridging the research gap: Making HRI useful to individuals with autism. *J. Human-Robot Interaction*, 1(1):26–54, 2012.

[43] E. Kim, L. Berkovits, E. Bernier, D. Leyzberg, F. Shic, R. Paul, and B. Scassellati. Social robots as embedded reinforcers of social behavior in children with autism. *J. Autism and Developmental Disorders*, 43(5):1038–1049, 2013.

[44] A. Klin, J. Lang, D. Cicchetti, and F. Volkmar. Brief report: Interrater reliability of clinical diagnosis and DSM-IV criteria for autistic disorder: Results of the DSM-IV autism field trial. *J. Autism and Developmental Disorders*, 30(2):163–167, 2000.

[45] A. Klin, D. Lin, P. Gorrindo, G. Ramsay, and W. Jones. Two-year-olds with autism orient to nonsocial contingencies rather than biological motion. *Nature*, 459(7244):257–261, 2009.

[46] H. Kozima, C. Nakagawa, N. Kawai, D. Kosugi, and Y. Yano. A humanoid in company with children. In *Proc. 4th IEEE/RAS Int. Conf. Humanoid Robots*, pages 470–477, 2004.

[47] H. Kozima, C. Nakagawa, and Y. Yasuda. Interactive robots for communication-care: A case-study in autism therapy. In *Proc. 2005 IEEE Int. Workshop on Robot and Human Interactive Communication*, pages 341–346, 2005.

[48] H. Kozima, M. Michalowski, and C. Nakagawa. Keepon. *Int. J. Social Robotics*, 1(1):3–18, 2009.

[49] C. Liu, K. Conn, N. Sarkar, and W. Stone. Online affect detection and robot behavior adaptation for intervention of children with autism. *IEEE Trans. Robotics*, 24(4):883–896, 2008.

[50] A. Lo, P. Guarino, L. Richards, J. Haselkorn, G. Wittenberg, D. Federman, R. Ringer, T. Wagner, H. Krebs, B. Volpe, C. Bever, D. Bravata, P. Duncan, B. Corn, A. Maffucci, S. Nadeau, S. Conroy, J. Powell, G. Huang, and P. Peduzzi. Robot-assisted therapy for long-term upper-limb impairment after stroke. *New England J. Medicine*, 362(19):1772–1783, 2010.

[51] M. Mark. Mixed and multimethods in predominantly quantitative studies, especially experiments and quasi-experiments. In S. Hesse-Biber and R. Johnson, editors, *Oxford Handbook of Multimethod and Mixed Methods Research Inquiry*, pages 21–41. Oxford University Press, 2015.

[52] N. Matsumoto, H. Fujii, and M. Okada. Minimal design for human-agent communication. *Artificial Life and Robotics*, 10(1):49–54, 2006.

[53] R. McWilliam, D. Bailey, D. Bailey, and M. Wolery. Promoting engagement and mastery. *Teaching Infants and Preschoolers with Disabilities*, 2:230–255, 1992.

[54] H. Meinedo, D. Caseiro, J. Neto, and I. Trancoso. AUDIMUS.MEDIA: A broadcast news speech recognition system for the European Portuguese language. In *Proc. 6th Int. Workshop Computational Processing of the Portuguese Language*, pages 9–17, 2003.

[55] H. Meinedo, A. Abad, T. Pellegrini, J. Neto, and I. Trancoso. The L2F Broadcast News speech recognition system. In *Proc. VI Jornadas en Tecnología del Habla*, pages 93–96, 2010.

[56] J. Mercer. Examining DIR/Floortime™ as a treatment for children with autism spectrum disorders: A review of research and theory. *Research on Social Work Practice*, 27(5):625–635, 2015.

[57] J. Messias, R. Ventura, P. Lima, J. Sequeira, P. Alvito, C. Marques, and P. Carriço. A robotic platform for edutainment activities in a pediatric hospital. In *Proc. 2014 IEEE Int. Conf. Autonomous Robot Systems and Competitions*, pages 193–198, 2014.

[58] F. Michaud and S. Caron. Roball, the rolling robot. *Autonomous Robots*, 12:211–222, 2002.

[59] F. Michaud and C. Théberge-Turmel. Mobile robotic toys and autism: Observations of interaction. In K. Dautenhahn, A. Bond, L. Cañamero, and B. Edmonds, editors, *Socially Intelligent Agents: Creating Relationships with Computers and Robots*, pages 125–132. Kluwer Academic Publishers, 2002.

[60] American occupational therapy association. Occupational therapy practice framework: Domain and process (2nd ed.). *American Journal of occupational therapy practice*, (62):625–683, 2008.

[61] S. Odom. *Widening the circle: Including children with disabilities in preschool programs*. Early Childhood Education Series. Teachers College Press, 2002.

[62] K. Pajareya and K. Nopmaneejumruslers. A pilot randomized controlled trial of DIR/Floortime™ parent training intervention for pre-school children with autistic spectrum disorders. *Autism*, 15(5):563–577, 2011.

[63] S. Paulo, L. Oliveira, C. Mendes, L. Figueira, R. Cassaca, C. Viana, and H. Moniz. DIXI – A generic text-to-speech system for European Portuguese. In *Proc. 8th Int. Conf. Computational Processing of the Portuguese Language*, pages 91–100, 2008.

[64] J. Perner. Theory of mind. In M. Bennett, editor, *Developmental Psychology: Achievements and Prospects*, pages 205–230. Psychology Press, 1999.

[65] A. Pierno, M. Mari, D. Lusher, and U. Castiello. Robotic movement elicits visuomotor priming in children with autism. *Neuropsychologia*, 46(2):448–454, 2008.

[66] G. Pioggia, R. Igliozzi, M. Ferro, A. Ahluwalia, F. Muratori, and D. De Rossi. An android for enhancing social skills and emotion recognition in people with autism. *IEEE Trans. Neural Systems and Rehabilitation Engineering*, 13(4):507–515, 2005.

[67] G. Pioggia, M. Sica, M. Ferro, R. Igliozzi, F. Muratori, A. Ahluwalia, and D. De Rossi. Human-robot interaction in autism: Face, an android-based social therapy. In *Proc. 16th IEEE Int. Symp. Robot and Human Interactive Communication*, pages 605–612, 2007.

[68] A. Piper, E. O'Brien, M. Morris, and T. Winograd. SIDES: A cooperative tabletop computer game for social skills development. In *Proc. 20th Conference on Computer Supported Cooperative Work*, pages 1–10, 2006.

[69] M. Plötner, H. Over, M. Carpenter, and M. Tomasello. Young children show the bystander effect in helping situations. *Psychological Science*, 26 (4):499–506, 2015.

[70] C. Ponitz, S. Rimm-Kaufman, K. Grimm, and T. Curby. Kindergarten classroom quality, behavioral engagement, and reading achievement. *School Psychology Review*, 38:102–120., 2009.

[71] C. Rasga, A. Quelhas, and R. Byrne. How children with autism reason about other's intentions: False-belief and counterfactual inferences. *J. Autism and Developmental Disorders*, 47(6):1806–1817, 2017.

[72] B. Robins, P. Dickerson, P. Stribling, and K. Dautenhahn. Robot-mediated joint attention in children with autism: A case study in robot-human interaction. *Interaction studies*, 5(2):161–198, 2004.

[73] B. Robins, K. Dautenhahn, R. Te Boekhorst, and A. Billard. Robotic assistants in therapy and education of children with autism: Can a small humanoid robot help encourage social interaction skills? *Universal Access in the Information Society*, 4(2):105–120, 2005.

[74] B. Robins, K. Dautenhahn, and P. Dickerson. From isolation to communication: A case study evaluation of robot assisted play for children with autism with a minimally expressive humanoid robot. In *Proc. 2nd Int. Conf. Advances in Computer-Human Interactions*, pages 205–211, 2009.

[75] J. Rodrigues, D. Belo, and H. Gamboa. Noise detection on ECG based on agglomerative clustering of morphological features. *Computers in Biology and Medicine*, 87(1):322–334, 2017.

[76] S. Rogers. Empirically supported comprehensive treatments for young children with autism. *J. Clinical Child Psychology*, 27(2):399–409, 1998.

[77] O. Rudovic, J. Lee, L. Mascarell-Maricic, B. Schuller, and R. Picard. Measuring engagement in robot-assisted autism therapy: A cross-cultural study. *Frontiers in Robotics and AI*, 4, 36, 2017.

[78] H. Salam and M. Chetouani. A multi-level context-based modeling of engagement in human-robot interaction. In *Proc. 11th IEEE Int. Conf. Automatic Face and Gesture Recognition*, pages 1–6, 2015.

[79] B. Scassellati. How social robots will help us to diagnose, treat, and understand autism. In *Proc. 12th Int. Symp. Robotics Research*, pages 552–563, 2007.

[80] B. Scassellati, H. Admoni, and M. Mataric. Robots for use in autism research. *Annual Review of Biomedical Engineering*, 14:275–294, 2012.

[81] E. Schegloff. Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, 29(1):1–63, 2000.

[82] P. Sequeira, P. Alves-Oliveira, T. Ribeiro, E. Di Tullio, S. Petisca, F.S. Melo, G. Castellano, and A. Paiva. Discovering social interaction strategies for robots from restricted-perception Wizard-of-Oz studies. In *Proc. 11th ACM/IEEE Int. Conf. Human Robot Interaction*, pages 197–204, 2016.

[83] R. Simut, J. Vanderfaeillie, A. Peca, G. Van de Perre, and B. Vanderborght. Children with autism spectrum disorders make a fruit salad with Probo, the social robot: An interaction study. *J. Autism and Developmental Disorders*, 46(1):113–126, 2016.

[84] S. Srinivasan, I. Eigsti, L. Neelly, and A. Bhat. The effects of embodied rhythm and robotic interventions on the spontaneous and responsive social attention patterns of children with autism spectrum disorder (ASD): A pilot randomized controlled trial. *Research in Autism Spectrum Disorders*, 27:54–72, 2016.

[85] A. Tapus, A. Peca, A. Aly, C. Pop, L. Jisa, S. Pintea, A. Rusu, and D. David. Children with autism social engagement in interaction with Nao, an imitative robot: A series of single case experiments. *Interaction Studies*, 13(3):315–347, 2012.

[86] S. Thill, C. Pop, T. Belpaeme, T. Ziemke, and B. Vanderborght. Robot-assisted therapy for autism spectrum disorders with (partially) autonomous control: Challenges and outlook. *Paladyn J. Behavioral Robotics*, 3(4):209–217, 2012.

[87] J. Trafton, A. Schultz, D. Perznowski, M. Bugajska, W. Adams, N. Cassimatis, and D. Brock. Children and robots learning to play hide and seek. In *Proc. 1st ACM Conf. Human-Robot interaction*, pages 242–249, 2006.

[88] Z. Warren, Z. Zheng, A. Swanson, E. Bekele, L. Zhang, J. Crittendon, A. Weitlauf, and N. Sarkar. Can robotic interaction improve joint attention skills? *J. Autism and Developmental Disorders*, 45(11):3726–3734, 2015.

[89] I. Werry and K. Dautenhahn. Applying mobile robot technology to the rehabilitation of autistic children. In *Proc. 7th Symp. Intelligent Robotic Systems*, 1999.

[90] B. Wert and J. Neisworth. Effects of video self-modeling on spontaneous requesting in children with autism. *J. Positive Behavior Interventions*, 5 (1):30–34, 2003.

[91] C. Wong and C. Kasari. Play and joint attention of children with autism in the preschool special education classroom. *J. Autism Development Disorders*, 42:2152–2161, 2012.

[92] *Autism Spectrum Disorders.* World Health Organization, http://www.who.int/mediacentre/factsheets/autism-spectrum-disorders/en/.

[93] D. Yates, C. Vaessen, and M. Roupret. From Leonardo to da Vinci: The history of robot-assisted surgery in urology. *British J. Urology International*, 108(11):1708–1713, 2011.

[94] S. Yun, H. Kim, J. Choi, and S. Park. A robot-assisted behavioural intervention system for children with autism spectrum disorders. *Robotics and Autonomous Systems*, 76:58–67, 2016.

There are no conflicts of interest.

Francisco S. Melo (corresponding author)