GAIPS tech. rep. series, n. GAIPS-TR-001-18, March 2018

INTELLIGENT AGENTS AND SYNTHETIC CHARACTERS GROUP

Project INSIDE

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

Francisco S. Melo, Alberto Sardinha, Miguel Faria, Ana Paiva, Ruben Solera and Miguel Vasco
INESC-ID, Lisboa, Portugal

Mithun Kinarullathil, Pedro Lima, Luís Luz, André Mateus, Jhielson Pimentel and Rodrigo Ventura
Instituto Superior Técnico, Lisbon, Portugal

Marta Couto, Anabela Farias and Isabel Melo
Child Development Center, Hospital Garcia de Orta, Almada, Portugal

Hugo Gambôa
Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Almada, Portugal

David Belo, D. Noronha Osório and J. Almeida Rodrigues
Lab. Instrumentação, Engenharia Biomédica e Física da Radiação, Universidade Nova de Lisboa, Almada, Portugal

Pedro Sequeira
Northeastern University, Boston, MA 02115, USA

Manuela Veloso
Carnegie Mellon University, Pittsburgh, PA 15213, USA

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. The INSIDE system is the first in which a fully autonomous mobile robot actively engages children with ASD during therapy in a semi-unstructured interaction. 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. Finally, we describe some preliminary results concerning the use of the INSIDE system in the actual therapy with children at Hospital Garcia de Orta, in Portugal.

Contents

GAIPS/INESC-ID
TagusPark, Edifício IST
Av. Prof. Dr. Cavaco Silva
2780-990 Porto Salvo
Portugal

Tel.: +351 214 233 508
Fax: +351 214 233 290
http://gaips.inesc-id.pt/

Corresponding author:
Francisco S. Melo
E-mail: fmelo@inesc-id.pt
http://gaips.inesc-id.pt/~fmelo/

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 currently used both for high-precision medical interventions [82] and physical and cognitive therapy [17, 31, 45, 77].

Of particular interest to this paper is the use of robotic technology in autism therapy (see, for example, the survey in [70]). Autism Spectrum Disorders (ASD), as defined by the DSM-V [2], consist of persistent deficits in social communication and social interaction across multiple contexts. 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 large 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 [81] estimates that 1 in 160 children suffers from some autism spectrum disorder. ASD has social, emotional and economic costs for the autistic individuals, their families and the community [78]. 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 [7]. Nevertheless, it is generally accepted that an early and adequate intervention yields a more favorable outcome [69].

ASD children show little interest in social interaction and instead prefer to interact with objects [24, 75]. 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, 51, 79]. The willingness of ASD children to (socially) interact with robots may, in part, be explained by the predictability and simplicity of their social behaviour, when compared with that of human partners [69]. 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 [75]. Several studies report that children with autism in effect create affective bonds with social robots [37, 42].

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 [27]. 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 behaviour, such as joint attention, eye gaze, spontaneous imitation and increased engagement in tasks after interaction [9, 22, 59]. Such success attests to the need for further exploring 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]

- *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 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 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.

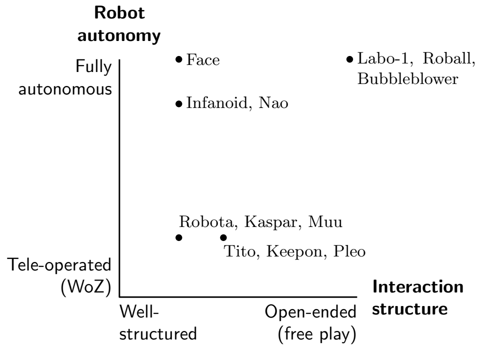
Figure 1 outlines the landscape of robot use in ASD therapy across the different dimensions outlined above. We include a (non-exhaustive) selection of representative works featuring a wide variety of robotic platforms. As outlined in the diagram, existing work can be roughly clustered into three groups,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

The larger cluster corresponds to those works adopting a *Wizard-of-Oz* approach [35], 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 behaviours and engage in rich interactions, effectively facilitating the study of child-robot interaction in ASD therapy.

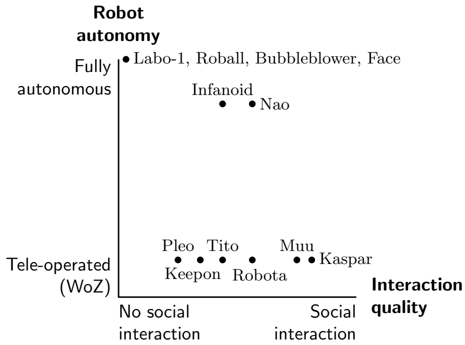
It is not surprising, then, that works featuring tele-operated robots exist that feature a wide variety of interaction modalities, ranging from very simple interactions—in which the robot merely exhibits some form of emotional response to the child—to

[1]It is interesting to draw a parallel between the three dimensions outlined herein and the discussion in the work of Scassellati et al. [70]. 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 behaviour*, *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 behaviour".

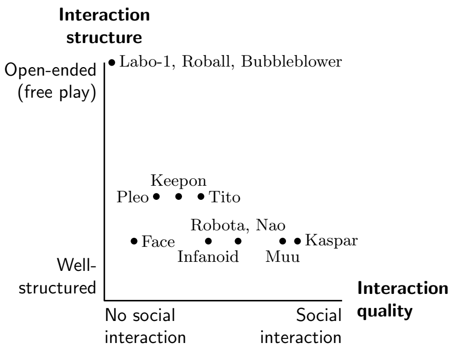
[2]Since the robot operator is out of the sight of the participants in the experiment, the robot appears autonomous during the interaction.



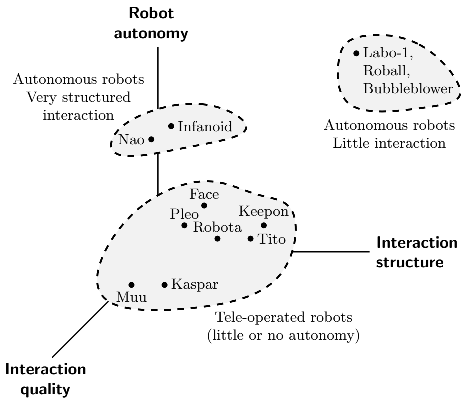
(a) Robot autonomy versus interaction structure.



(b) Robot autonomy versus interaction quality.



(c) Interaction structure versus interaction quality.



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

Figure 1: **PLima: I can not see the INSIDE project in the diagram of Fig 1d, so how do we position the project wrt the others?** 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 [28, 29], Face [58], Infanoid [40], Kaspar [22, 64], Keepon [41, 42], Labo-1 [79], Muu [46], Nao [27, 33], Pleo [36, 37], the Roball toy [50, 51], Robota [13, 20], and Tito [26].

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 [41, 42] and Pleo robots [36, 37]; a representative example of the latter is the work done in the context of the Aurora project using the Kaspar robot [22, 64].

However, in scenarios where the interaction of the child with the robot is mediated by a therapist (as seen in Fig. 13 for an example), a WoZ setup may require two or 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 require multiple human operators to engage a child in a rich, multimodal interaction.

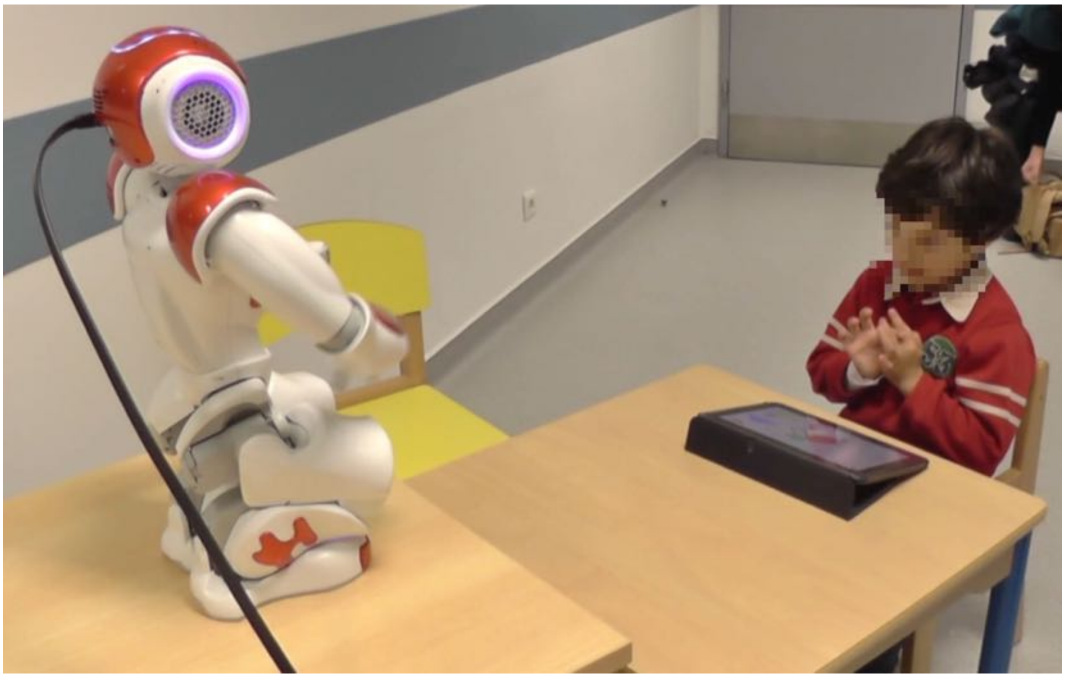


Figure 2: Structured interaction between a child and the Nao robot. The robot and the child play a turn-taking Tangram game [11].

Such a strong dependence on human operators renders WoZ approaches to autism therapy unaffordable in the long term [27, 70, 76].

As for works featuring autonomous robots, we can identify two clusters that greatly differ on the way the child interacts with the robot. In one cluster we include those works featuring robots such as Labo-1 [79], Roball [50, 51] or the Bubbleblower [28, 29]. These robotic platforms are endowed with very simple reactive behaviours, 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 contrast, works using robotic platforms such as Infanoid [40] or Nao [27, 33] rely on very structured interactions, in which the child usually sits in front of the robot and the interaction follows a strict script—see Fig. 2 for an example.

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 fully autonomous during the 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 itself.

The INSIDE system therefore addresses a gap that can be identified in the previous discussion: our system should enable autonomous social interaction, while allowing relatively unconstrained activity by the child. 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.

This paper describes the architecture of the networked robot system designed in the context of INSIDE to support such interaction between robot and child during a therapy session. We discuss the therapeutical considerations that drove the design of the system, and how such considerations influenced the technology in and layout of the system. We detail the design methodology adopted to build the robot's behaviour and report some preliminary results on the use of the system as a whole in therapy sessions at Hospital Garcia de Orta, in Almada, Portugal. We conclude by discussing how the ideas behind the INSIDE system can be ported to other scenarios involving rich human-robot interaction.

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 [14]. 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 [6, 53].

Occupational therapists working with ASD children frequently adopt the DIR/Floortime™ 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 [14].

DIR/Floortime™ 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 [54].

Several studies have demonstrated that social engagement directly affects important behaviours like language, requesting, greeting, joint attention and imitation; even when these behaviours are not specifically targeted by the intervention program [49, 67].

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 of the INSIDE project. Indeed, many of the activities employed in traditional therapy already take into account the scheme of symbiotic autonomy and are developed in order to focus on the behaviour deficiencies of children with ASD, such as the difficulty in addressing help requests or in asking for help. To create robot-mediated therapy sessions we need a set of activities that (1) have therapeutic goals (i.e., goals 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 coloured 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 gets stuck and asks the child for help. The child must remove the block 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 [4, 61] 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 block 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. At this activitity, 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 behaviour [80].

- *"Tangram"*: This activity uses a game previously developed by Bernardo et al. [10] that has a turn-taking mode. 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 [71]. 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 has rules that can be very well defined a priori, unlike what happens with conversations. In this case, each turn corresponds to one piece of a puzzle. Once the player sets the piece in the correct place, the turn changes and it is time for the other player. During the game, on one of the robot's turns, the robot will ask for help.

- *"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 Plotner et al. [60], 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 can foresee some advantages of Robot-assisted therapy for ASD children. Robots may allow us to develop a therapy focused on the children's interests and abilities, resulting in motivating and pleasurable interactions that comprise an overall positive experience for ASD children [74]. In other words, technology may allow us to create a therapeutic setting where children have fun while engaging in an interaction which is typically difficult for ASD children.

ASD is characterized by difficulties in making sense of the social world, but autistic children often show ease in understanding the physical world and object related interactions [5, 38, 39]. 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 [44], 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, 41, 57, 62]. 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.

To build a networked robot system for the activities in Section 2.2, we present an outline of the key technological challenges of the INSIDE project:

- *Perception*: Perceiving people and objects in the room is one of the key challenges of the INSIDE project. In addition, 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 on the Mindreading mechanism, also known as Theory of Mind (TOM) [4, 56]. 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). The robot in the INSIDE project 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*: Building an autonomous robot for therapeutic purposes is a challenging research problem. In particular, the INSIDE project has a complex setting, whereby the child and robot may freely move around while engaging in activities within a room. To tackle this challenge, 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, the robot is currently playing the puzzle activity and detects that one more piece was placed in the right location, thus it decides to smile and reinforce the child.

- *Robot Design*: Klin et al.[39] 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 basket 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, a basket with an RFID reader to detect balls) with an aim to increase the child's attention and improve the interaction within the therapy sessions.

- *Social Interaction*: A key challenge in INSIDE is to develop a social robot aimed at interacting and playing with ASD children within a therapy session. A 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*: While designing the activities and developing the robot system, we took into account that the robot is not capable of performing all the actions in the environment (e.g., open a door at the end of a therapy session or detecting an obstacle in the room). Hence, the robot has to detect that an action cannot be performed and ask for the child's help in order to complete the task. This is a challenging research problem from both a system development and therapeutic perspective, because children with ASD have difficulties.

The following section describes in detail the architecture of the networked robot system, highlighting the design decisions and the key technological features.

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.

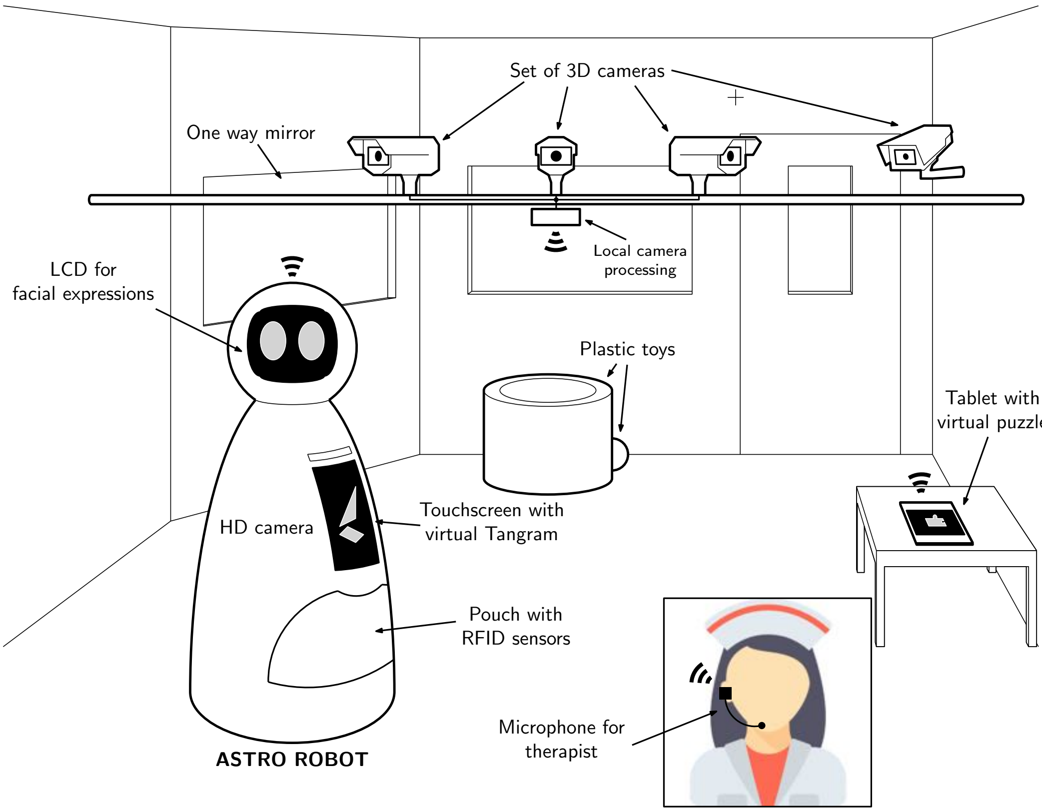


Figure 3: Diagram illustrating the main components of the hardware setup used in INSIDE.

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. 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 robot system built on top of ROS.[3] The network comprises a set of 3D cameras mounted on

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

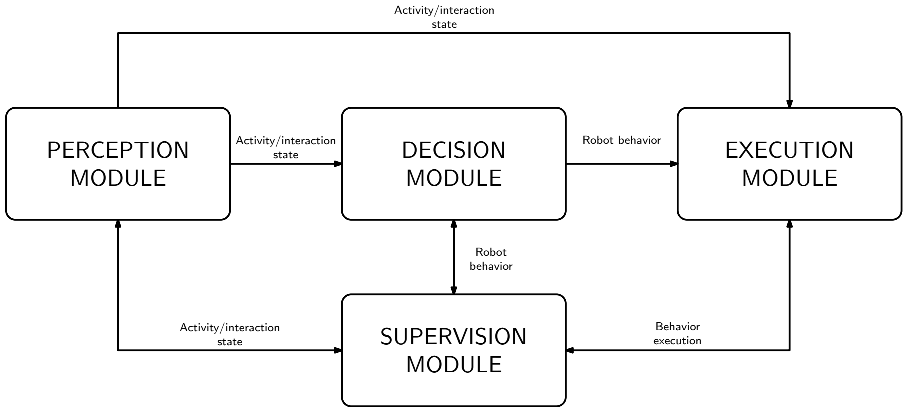


Figure 4: Outline of the main blocks in the software architecture.

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.

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 networks.

The decision module responsible for deciding, at each moment, whether to continue with the current activity or move to another activity. It is also 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. 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, while still allowing anytime human intervention, if the circumstances so dictate.

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.

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). 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.[4]

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

[4]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.

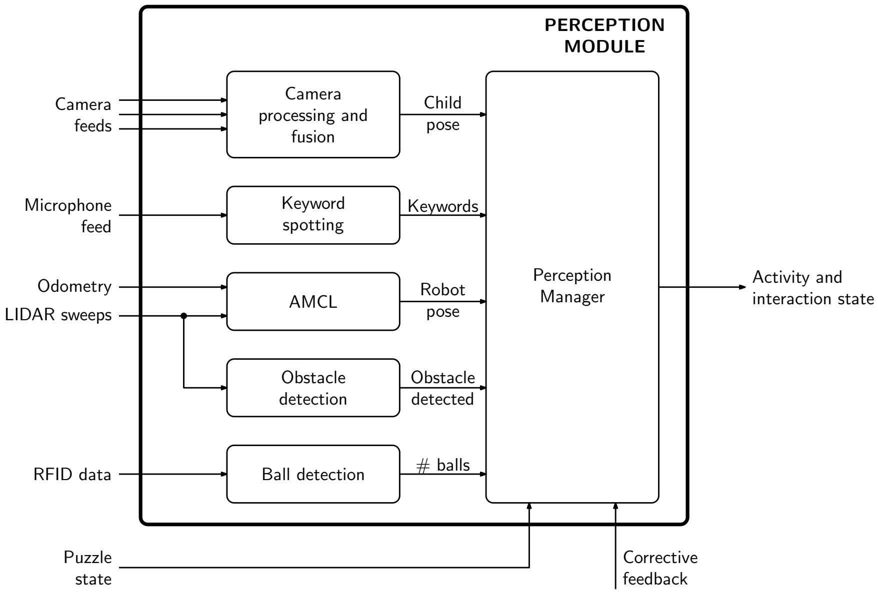


Figure 5: Detailed outline of the perception module.

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`.[5] 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 speech recognition: multiple, moving and different type of simultaneous speakers (children and adults), distant speech, room reverberations, different noise sources (background speech, the robot's motors, etc.), microphone distortions, 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 and very short (monosyllables). Therefore, obtaining a dataset from which speech recognition could be trained 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 ear-set 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,

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

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 [47, 48]. 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]. The output of the keyword spotting block is a "list" of detected keywords.

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 [30] 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.[6] 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 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, incorporate any corrective feedback provided by the supervision module and then forward the most up-to-date state information to the decision and execution modules.

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

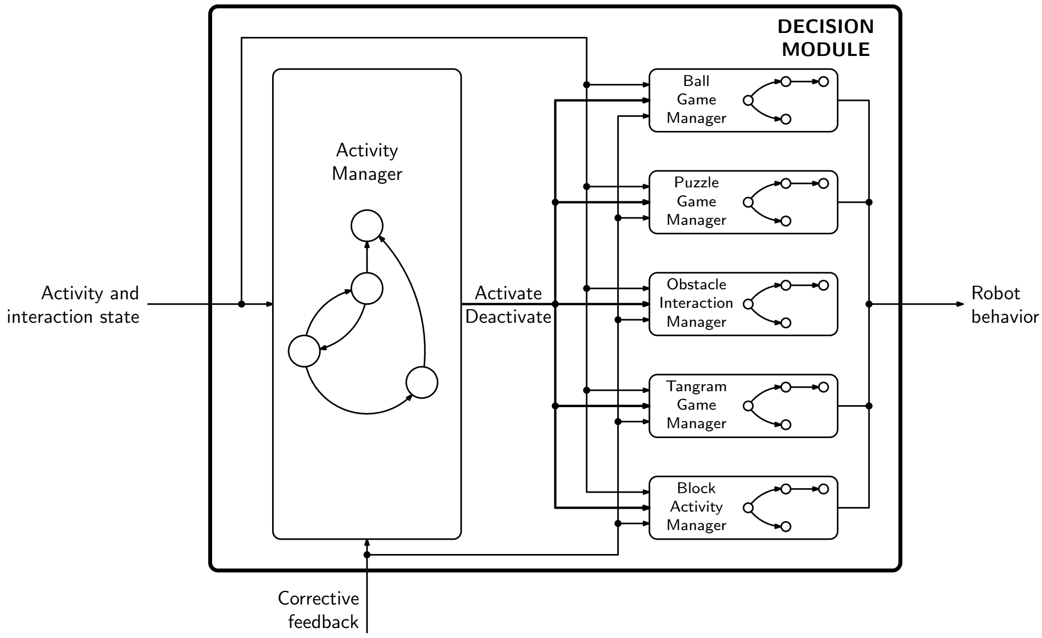


Figure 6: Outline of the decision module. The blocks on the right correspond to *interaction managers*, responsible for defining the behavior of the robot in each activity. The set of interaction managers indicated are merely illustrative.

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 pre-defined finite-state machine or a more sophisticated decision-theoretic policy, using for example, the ROS `mdm` package.[7] 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

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

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).

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.
- *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 as a function of the child's engagement and success. At the end of the game, it requests the child's assistance to remove the balls from its pouch.
- *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.
- *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.

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.

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. 8. In particular, the robot has a set of pre-programmed behaviors—both general and task specific. For our purposes, we define *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

Figure 7: 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.

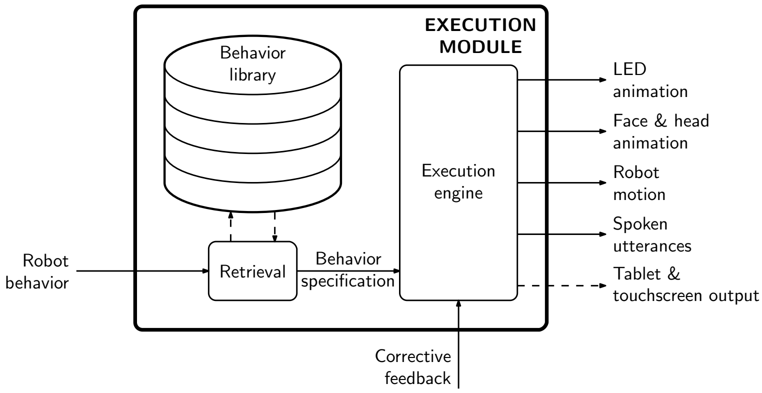


Figure 8: Execution module, responsible to translating the behaviors determined by the decision module into the multimodal output of the robot.

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. 7).

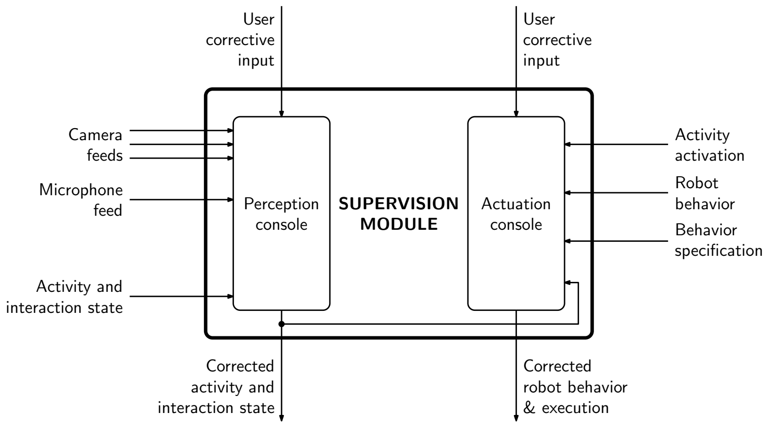


Figure 9: Overview of the supervision module, which provides external users with the necessary tools to take control of the system, if necessary.

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 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. We evaluated CereProc,[8] Nuance Vocalizer,[9] Acapela,[10] and the DIXI TTS engine [55]. 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 (genre, 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 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 is a central element in the INSIDE software architecture, as it provides a mandatory backdoor to the perceptual and behavioral elements of the robot, allowing human supervisor to take control of the interaction at any time,

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

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

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

should circumstances so demand. In particular, if a problem is detected in the robot or the network, the supervision module allows a human operator to intervene and mitigate the potential impact of such problem. 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 will be 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.

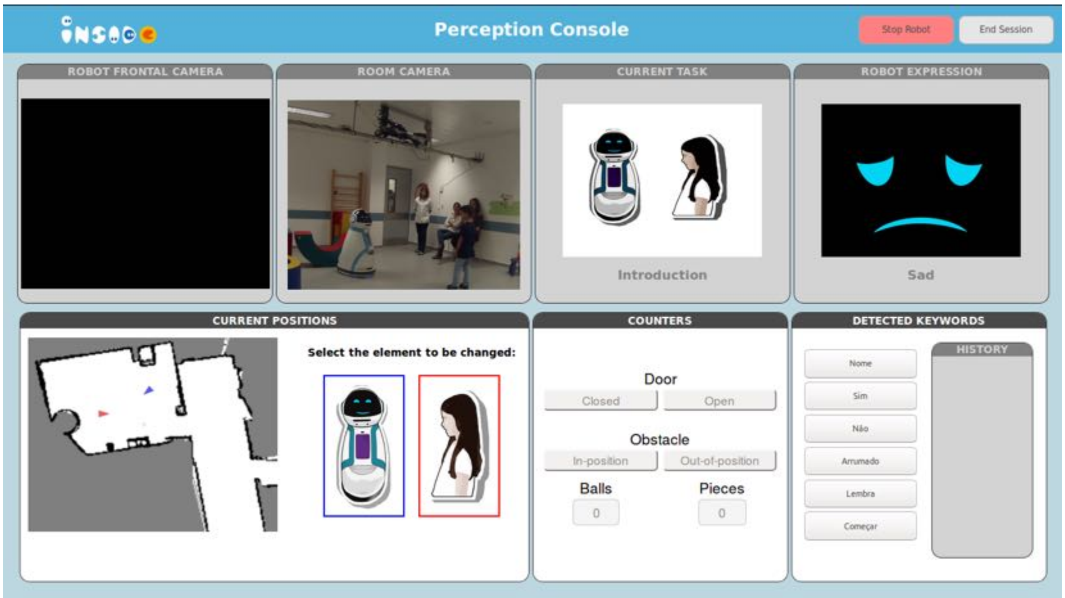
The supervision module also allows the configuration of several therapy session parameters, such as the number of the session, the name of the child and the activities to be executed in that session. Figure 10 provides a screenshot of the two consoles.

4 Social Interaction Design Methodology

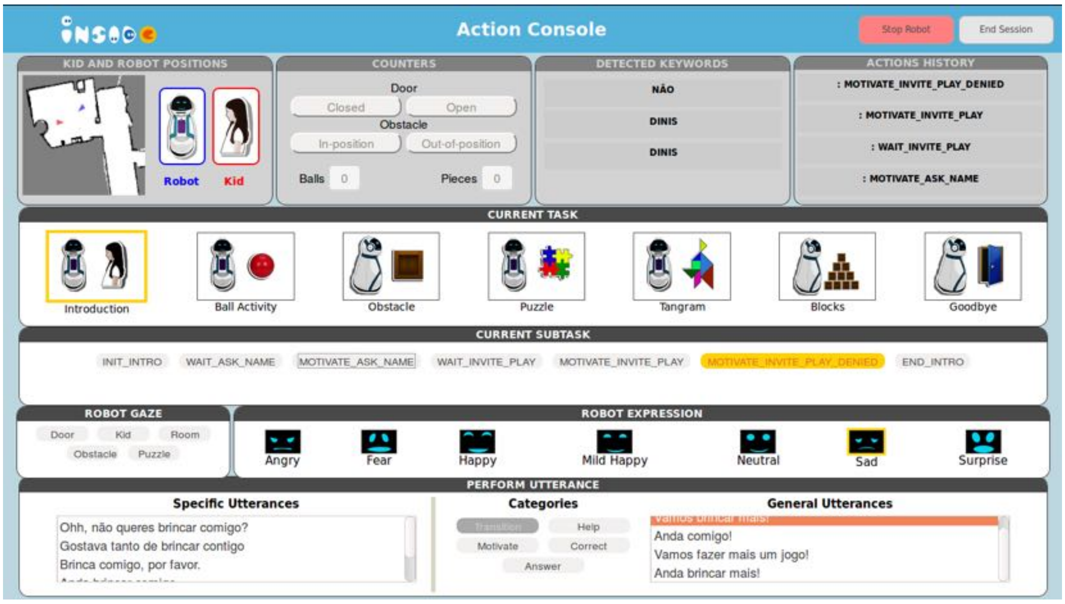
As discussed in the previous sections, the purpose of the INSIDE networked robot system is to enable a mobile robot to socially interact with ASD children during therapy sessions in a set of different and well-defined activities. The social interaction capabilities of the robot are, therefore, determinant 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* [72].

4.1 Restricted Perception Wizard-of-Oz

The key idea behind our 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 behaviour 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 behaviour 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



(a) Perception console.



(b) Actuation console.

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

capabilities, the same scene may appear very different for the expert and the robot [12]. 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 [12, 15]. Such technique addresses, to some extent, the perceptual

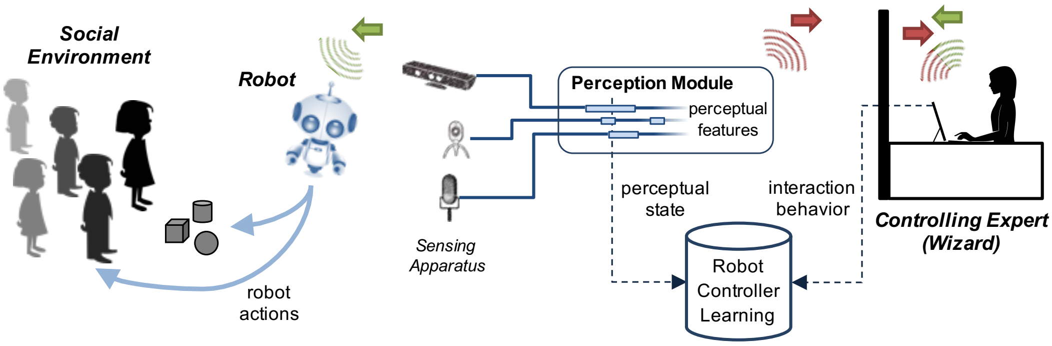


Figure 11: Restricted-perception WoZ methodology. Raw sensor data is filtered by a Perception Module to inform the controlling expert during the studies. Offline, a robot controller algorithm learns a mapping function between the perceived states and the robot behaviours chosen by the wizard. Colored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

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 behaviour 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 [72], as an methodology to allow a meaningful extraction of behaviours from Wizard interactions. Formally, we define an *interaction strategy* as a mapping between a robot's perceptual state and an interaction behaviour from its behaviour repertoire. We argue that in order to extract meaningful information that can be used to design an autonomous interaction, the robot's perceptual, behavioural, 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.[11]

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 behaviour 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.

◇

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

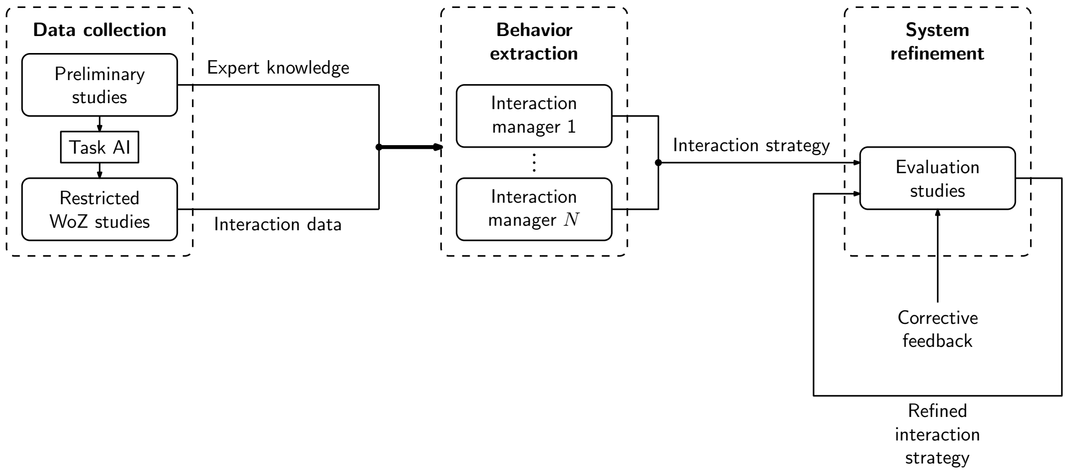


Figure 12: The three steps of the proposed methodology for designing social interaction strategies from restricted-perception Wizard of Oz interactions.

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 behaviour 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 behaviours) are collected.

- *behaviour extraction*, in which the data collected during the data collection stage is used to design and refine a number of interaction managers—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 [72] for further details on the design methodology.

4.2 Restricted Perception Wizard-of-Oz in INSIDE

We applied the restricted-perception WoZ methodology described above in the context of INSIDE to design the social interaction behaviours of the robot. Specifically, we conducted several preliminary studies to assess the children's reactions to the presence of the robot and the behaviour they exhibited during several activities. The sensing apparatus included a camera and omnidirectional microphone placed in

Figure 13: Illustration of the preliminary studies. Left: Two wizards control the robot behaviour—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.

the room and a camera and directional microphone placed inside the robot. The interaction between child and robot was always mediated by a therapist.

Two wizard operators controlled the robot—one was in charge of the interaction behaviour and another for the movement of the robot. Both operators observed the interaction through the feed from the camera, the sound from the omnidirectional microphone, and also through a one-way mirror (see Fig. 13).

As outlined in the methodology of Section 4.1, the preliminary studies were used to define and design the necessary perception and actuation modules for the task AI, before the restricted-perception WoZ experiments could be performed. Specifically, the perception, decision and execution modules detailed in Section 3 were directly informed by the preliminary studies.

With the task AI in place, 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 behaviour 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 behaviour 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.

5 Preliminary Results and Discussion

To evaluate the system and its efficacy in supporting the therapy for children with autism, we needed to obtain a set of measures that can be used to assess the overall engagement of the children, as well as their progress throughout the sessions of the main study. In that context we explored different alternative measures and analysis in such a way that can be used in future studies focusing on symbiotic interactions between children and robots.

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. These preliminary studies allowed us to design a set of measures that will be used for future studies.

Engagement: The the main metrics we considered for the analysis was engagement. So far our results are very encouraging. The children were willing to engage with the robot and all participants completed more than 50% of the tasks. Given that there is no time limit for the session, we found the total length of the session to be a good indicator of the overall quality. Shorter sessions correspond to higher percentages of tasks completed. Observing the available videos we concluded that engagement, although an interesting and relevant measure has to be carefully considered. Engagement of the children with the robot is desirable, but being engaged does not always mean that the children were paying attention to what the robot was saying.

Eye Gaze: Eye gaze is extremely difficult to assess in ASD children in general but more specifically in sessions where child and robot move freely. There is a tendency to infer that higher percentages of eye-gaze correspond to more engagement and are therefore desirable. However our studies so far show that higher percentages of eye gaze directed to the robot do not seem to reflect desirable behaviour. On the contrary, they appear in sessions where children stare at the robot with their faces nearly touching, which corresponds to behaviour that is inadequate in social interactions. Which raises the challenging yet 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 behaviour in social interaction? Or instead, should we consider staring as some form of engagement with the robot?

Speech: Furthermore, the content analysis of the therapist speech also gives an overview of the quality of the sessions. On the other hand, the content analysis

Figure 14: T-shirt designed to collect physiological signals from the children interacting with the robot. The T-shirt comprises ECG and accelerometer sensors and a bluetooth emitter that sends the captured signals to a dedicated computer.

of the child's speech seems to reflect mainly individual differences in language ability. However both measures will be valuable to assess progress through the main study.

It is however interesting to note that most of the children's speech was directed to the robot which is a great indication of engagement with the robot and enjoyment. Most importantly the pilots conducted so far have sparked a discussion that lead to protocol changes that greatly improved our original protocol.

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 will also analyze how the interaction affects the children from a physiological perspective, thus providing a more complete understanding of the impact of robots in therapy.

To this purpose, we built comfortable wearable technology, so that the analysis of the physiological data could be acquired with a T-shirt (see Fig. 14). However, the environment, the conditions in which the tasks are performed and the movements made by the children, are major issues that led to the partial corruption of the electrocardiographic signals acquired. Thus, in order to face these problems, two algorithms were developed for noise detection [65] and signal reconstruction [8].

In the literature, several techniques can be found for signal denoising through

frequency and time domain filtering, like wavelet thresholding [16], and adaptive filtering, such as Bayesian filters [68] and Kalman filters [73]. Other techniques such as independent component analysis (ICA) [83] and empirical mode decomposition (EMD) [43] are typically used as well. These, however, are more appropriate for noise removal. The Noise Detection Clustering algorithm has been created 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, each being more associated with a type of noise and to fine-tune the algorithm's performance. 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 denoising process, while being useful as well for signal reconstruction [65].

The acquired ECGs are heavily noisy and the features are hard, or even impossible to read, due to the mentioned contaminations. Several algorithms explore the cleaner regions, exploiting their features and serving as input to generative machine learning models [3, 18, 52, 66]. These algorithms have the limitation of using a feature extraction method that is specific to a single type of signal, therefore a method that synthesizes accurately biosignals using the raw signal was created. The developed method learns the morphology of the raw signal that is used as input, skipping the feature extraction step, by adopting the Deep Neural Networks (DNN) generative architecture [8]. This work profits from the prediction power of the Recurrent Neural Networks (RNN), more specifically of the Gated Recurrent Units (GRU), in order to generate the samples based on the previous predicted ones in a recursive loop [8, 32].

5.1 Discussion

With the metrics proposed, and an improved protocol and robot, the next steps involve the test with a group of children diagnosed with ASD to participate in the study. Given the immense diversity among ASD children and sampling constraints we are going to use a repeated measures design where children undergo 10 therapy sessions. We will evaluate the children's progress throughout the sessions and finally, generalization. Generalization implies that a child is able to apply what he/she learnt in the therapy setting, to other contexts, for example school or at home. This is extremely difficult to assess, and as such we will conduct an eleventh session without the robot and compare the children's performance on the last session with the robot, to an identical session without the robot and see if they are capable of performing at a similar level.

Following a recommendation from Robins et al. [63], our goal is to promote relationships between autistic individuals and other people. Bonding with a robot is just an intermediate step, but we believe it will have a very high impact in their life.

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] *Diagnostic and statistical manual of mental disorders (DSM-5).* American Psychiatric Association, 2013.

[3] H. Atoui, J. Fayn, and P. Rubel. A neural network approach for patient-specific 12-lead ecg synthesis in patient monitoring environments. In *Computers in Cardiology, 2004*, pages 161–164, Sept 2004. doi: 10.1109/CIC.2004.1442896.

[4] S. Baron-Cohen. *Mindblindness: An essay on autism and theory of mind*. MIT Press, 1997.

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

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

[7] 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.

[8] D. Belo, J. Rodrigues, J.R. Vaz, P. Pezarat-Correia, and H. Gamboa. Biosignals learning and synthesis using deep neural networks. *BioMedical Engineering OnLine*, 16(1):115, Sep 2017.

[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] Alves-Oliveira P. Santos M. Melo F. S. & A. Paiva A. Bernardo, B. Me and you together: A study on collaboration in manipulation tasks. In *In Proc. 16th Int. Conf. Intelligent Virtual Agents*, pages 472–475, 2016.

[11] B. Bernardo, P. Alves-Oliveira, M. Santos, F.S. Melo, and A. Paiva. Me and you together: A study on collaboration in manipulation tasks. In *Proc. 16th Int. Conf. Intelligent Virtual Agents*, pages 472–475, 2016.

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

[13] 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.

[14] 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.

[15] 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.

[16] V. Cherkassky and S. Kilts. Comparison of wavelet thresholding methods for denoising ecg signals. In *Proceedings of the International Conference on Artificial Neural Networks*, ICANN '01, pages 625–629, London, UK, 2001. Springer-Verlag.

[17] 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.

[18] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk. Wavelet-based statistical signal processing using hidden markov models. *IEEE Transactions on Signal Processing*, 46(4):886–902, Apr 1998.

[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] 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.

[27] 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.

[28] D. Feil-Seifer and M. Mataric. B^3IA: 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.

[29] 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.

[30] D. Fox. Adapting the sample size in particle filters through kld-sampling. *International Journal of Robotics Research*, 22, 2003.

[31] 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.

[32] A. Graves. Generating sequences with recurrent neural networks. *CoRR*, abs/1308.0850, 2013. URL `http://arxiv.org/abs/1308.0850`.

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

[36] 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.

[37] 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.

[38] A. Klin, J. Lang, D. Cicchetti, and F.R. Volkmar. Brief report: Interrater reliability of clinical diagnosis and dsm-iv criteria for autistic disorder: Results of the dsm-iv autism field trial. *Journal of autism and Developmental disorders*, 30(2):163–167, 2000.

[39] A. Klin, D. J. 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.

[40] 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.

[41] 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.

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

[43] J. Lee, D.D. McManus, S. Merchant, and K.H. Chon. Automatic motion and noise artifact detection in holter ecg data using empirical mode decomposition and statistical approaches. *IEEE Transactions on Biomedical Engineering*, 59 (6):1499–1506, June 2012.

[44] C. Liu, K. Conn, N. Sarkar, and W. Stone. Online affect detection and robot behavior adaptation for intervention of children with autism. In *IEEE transactions on robotics*, volume 24, pages 883–896, 2008.

[45] 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.

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

[47] H. Meinedo, D. Caseiro, J. Neto, and I. Trancoso. AUDIMUS.MEDIA: A Broadcast News Speech Recognition System for the European Portuguese Language. In N.J. Mamede, I. Trancoso, J. Baptista, and M. das Graças Volpe Nunes, editors, *Computational Processing of the Portuguese Language: 6th International Workshop, PROPOR 2003*, pages 9–17, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.

[48] H. Meinedo, A. Abad, T. Pellegrini, J. Neto, and I. Trancoso. The L2F Broadcast News Speech Recognition System. In *Proc. of the VI Jornadas en Tecnología del Habla and II Iberian SLTech Workshop (FALA 2010)*, pages 93–96, 2010.

[49] 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.

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

[51] 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.

[52] J.C. Nunes and A. Nait-Ali. Hilbert Transform-Based ECG Modeling. *Biomedical Engineering*, pages 36–40, 2005.

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

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

[55] S. Paulo, L.C. Oliveira, C. Mendes, L. Figueira, R. Cassaca, C. Viana, and H. Moniz. Dixi – a generic text-to-speech system for european portuguese. In A. Teixeira, V.L.S. de Lima, L.C. de Oliveira, and P. Quaresma, editors, *Computational Processing of the Portuguese Language: 8th International Conference, PROPOR 2008*, pages 91–100, Berlin, Heidelberg, 2008. Springer.

[56] J. Perner. Theory of mind. *Developmental psychology: Achievements and prospects*, pages 205–230, 1999.

[57] 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. In *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, volume 13, pages 507–515, 2005.

[58] 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.

[59] 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.

[60] 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.

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

[62] 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.

[63] B. Robins, K. Dautenhahn, and J.Dubowski. Robots as isolators or mediators for children with autism a cautionary tale. In *In Procs of the AISB 05 Symposium on Robot Companions.*, 2005.

[64] 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.

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

[66] R. Rodríguez, J. Bila, A. Mexicano, S. Cervantes, R. Ponce, and N.B. Nghien. Hilbert-huang transform and neural networks for electrocardiogram modeling and prediction. In *2014 10th International Conference on Natural Computation (ICNC)*, pages 561–567, Aug 2014.

[67] S. J. Rogers. Empirically supported comprehensive treatments for young children with autism. *Journal of clinical child psychology*, 27(2):399–409, 1998.

[68] R. Sameni, M. Shamsollahi, C. Jutten, and G. Clifford. A nonlinear bayesian filtering frameweork for ecg denoising. *IEEE Transactions of Biomedical Engineering*, 54(12):2172–2185, 2007.

[69] 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.

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

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

[72] 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.

[73] H. Sivaraks and C.A. Ratanamahatana. Robust and accurate anomaly detection in ECG artifacts using time series motif discovery. *Comp. Math. Methods in Medicine*, pages 1–20, May 2015.

[74] 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.

[75] 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.

[76] 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.

[77] 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.

[78] 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.

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

[80] B. Y. Wert and J. T. Neisworth. Effects of video self-modeling on spontaneous requesting in children with autism. . *Journal of positive behavior interventions*, 5(1):30–34, 2003.

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

[82] 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.

[83] H. Yoon, H. Kim, S. Kwon, and K. Park. An automated motion artifact removal algorithm in electrocardiogram based on independent component analysis. In *eTELEMED 2013, The Fifth International Conference on eHealth, Telemedicine and Social Medicine*, pages 15–20, Nice, France, February 2013.

Acknowledgements

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

GAIPS Technical Report Series

This report is part of the GAIPS Technical Report Series. The reports currently available are:

[a] P. Sequeira, F.S. Melo, A. Paiva. Learning by appraising: An emotion-based approach for intrinsic reward design Tech. Rep. GAIPS-TR-001-12, GAIPS/INESC-ID, March 2012.

[b] P. Sequeira, F.S. Melo, A. Paiva. Associative metric for learning in factored MDPs based on classical conditioning. Tech. Rep. GAIPS-TR-002-12, GAIPS/INESC-ID, June 2012.

[c] S. Mascarenhas, R. Prada, A. Paiva. Social importance dynamics: A model for culturally-adaptive agents. Tech. Rep. GAIPS-TR-001-13, GAIPS/INESC-ID, April 2013.

[d] S. Mascarenhas, R. Prada, A. Paiva. Planning culturally-appropriate conversations using adjacency pairs. Tech. Rep. GAIPS-TR-002-13, GAIPS/INESC-ID, April 2013.

[e] P. Sequeira, F.S. Melo, A. Paiva. Emergence of emotion-like signals in learning agents. Tech. Rep. GAIPS-TR-003-13, GAIPS/INESC-ID, October 2013.

[f] S. Mascarenhas, S. Rodrigues, J. Dias, A. Paiva. A process model of empathy for virtual agents. Tech. Rep. GAIPS-TR-004-13, GAIPS/INESC-ID, November 2013.

[g] B. Bernardo. An Interactive Tangram Game For Autistic Children Tech. Rep. GAIPS-TR-001-16, GAIPS/INESC-ID, November 2016.

[h] F. Correia. EMYS: A Social Robot that Plays “Sueca” Tech. Rep. GAIPS-TR-002-16, GAIPS/INESC-ID, November 2016.

[i] J. Campos, J. Guerreiro, B. Ravenet, R. Prada, A. Paiva. Computer supported training of joint investigation teams Tech. Rep. GAIPS-TR-001-17, GAIPS/INESC-ID, May 2017.

[j] J. Campos, J. Guerreiro, B. Ravenet, R. Prada, A. Paiva, E. Jaspaert, G. Vervaeke, G. Nahari, A. Roth, S. Kraus. Designing a training tool for transnational joint investigation teams Tech. Rep. GAIPS-TR-002-17, GAIPS/INESC-ID, May 2017.

[k] F.S. Melo, A. Sardinha, M. Faria, A. Paiva, R. Solera, M. Vasco, M. Kinarullathil, P. Lima, L. Luz, A. Mateus, J. Pimentel, R. Ventura, M. Couto, A. Farias, I. Melo, H. Gambôa, D. Belo, D. Osório, J. Rodrigues, P. Sequeira, M. Veloso Project INSIDE: Towards Autonomous Semi-Unstructured Human-Robot Social Interaction in Autism Therapy Tech. Rep. GAIPS-TR-001-18, GAIPS/INESC-ID, March 2018.

You may obtain any of the GAIPS technical reports from the corresponding author or on the group’s web page (`http://gaips.inesc-id.pt/`).

gaips

INTELLIGENT AGENTS AND SYNTHETIC CHARACTERS GROUP