

Using Socially Assistive Robots In Speech-Language Therapy For Children With Language Impairments

Micol Spitale¹, Silvia Silleresi^{1,2}, Franca Garzotto¹ and Maja J Matarić³

¹Department of Electronics, Information, and Bioengineering, Politecnico di Milano, Via Golgi 39, Milan, 20133, Italy.

²Department of Psychology, University of Milano Bicocca, Piazza Dell'Ateneo Nuovo 1, Milan, 20126, Italy.

³Computer Science Department, University of Southern California, 3710 McClintock Ave, Los Angeles, 90089, California, USA.

Contributing authors: micol.spitale@polimi.it; silvia.silleresi@unimib.it;
franca.garzotto@polimi.it; mataric@usc.edu;

Abstract

Socially assistive robots (SARs) have been shown to be promising therapy tools for children with primary or co-occurring language impairments (e.g., developmental language disorder and autism spectrum disorder), but only a few studies have explored the use of SARs in speech-language therapies. This work sought to address the following research goals: (1) explore the potential of using SAR for training linguistic skills of children with language impairments, targeting specific aspects of language and measuring their linguistic improvements in speech-language therapy; (2) explore children's facial cues during SAR-supported speech-language therapy; and (3) collect therapist perspectives on using SARs in speech-language therapy after having experienced it. Toward these goals, we conducted an 8-week between-subjects study involving 20 children with language impairments and 6 speech-language therapists who conducted the SAR-supported therapy. Children were randomly assigned to either a physical SAR or a virtual SAR condition; both provided the same language impairment therapy. We collected linguistic activity scores, video recordings, therapist questionnaires, and group interview data. The study results show that: (i) the study participants' overall linguistic skills improved significantly in both conditions; (ii) participants who were engaged with the physical SAR (measured based on gaze direction and head position) were more likely to demonstrate linguistic skill improvements and had a significantly higher numbers of speech occurrences in the child-robot-therapist triads with the physical SAR; (iii) therapists reported skepticism about SAR efficacy in this context but believed that SAR could be beneficial for keeping children engaged, motivated, and positive during speech-language therapy.

Keywords: Socially Assistive Robots, Language Impairments, Speech-Language Therapy, Children

1 Introduction

Speech and language disorders are among the most prevalent childhood conditions worldwide, with an estimated overall prevalence rate of 7.5-8% [7, 34]. For some children, such as those with Developmental Language Disorder (DLD), difficulties in language skills constitute the primary domain of impairment. DLD is a neurodevelopmental disorder that involves specific difficulties in mastering aspects of language (e.g., word/sentence structure, aka *morphosyntax*) independent from any kind of intellectual, sensory, or neurological impairment. For other conditions, language impairments constitute a co-occurring factor, such as in Autism Spectrum Disorder (ASD). ASD is a neurodevelopmental condition characterized by impairments in social communication and interaction and the presence of restricted and repetitive behaviors; language impairments may be associated with the primary ASD diagnosis (aka ASD-LI) [52]. Several studies in psycholinguistics have used standardized or experimental tasks that evaluate morphosyntax to investigate the phenotypic similarities among individuals with DLD and ASD-LI, generating an ongoing debate regarding the overlap of their language (dis)abilities [41, 59, 81]. For the purposes of this paper, we assume that DLD and ASD-LI have shared needs for improved language skills [70, 77], and we refer to children with either DLD or ASD-LI diagnoses as *children with language impairments or disorders*.

In recent years, researchers and therapists have both suggested that existing tools used in research and clinical practice often lack engagement and are not tailored to the specific needs of children with language disorders [40, 82]. Past studies have pointed out the potential of adopting innovative technologies into speech-language therapy. Specifically, Socially Assistive Robots (SARs) [28, 37, 80] have emerged as promising tools for supporting children with special needs (e.g., ASD, DLD) in enhancing their social and communication skills [15, 21]. SARs have been shown to have positive impacts on child engagement, joint attention, and turn-taking [17, 21, 46, 65, 66]. However, to date, only a few studies have investigated the use of SAR in the context of speech-language interventions to address morphosyntactic structure. Consequently, this work explores the potential of



Fig. 1 A child interacting with the QT robot with the support of a therapist during a speech-language therapy session.

SAR for training comprehension and production linguistic skills - specifically morphosyntactic ones - of children with language impairments. This work sought to understand the use of SAR in speech-language therapy, in particular exploring i) children's linguistic improvements using a SAR, ii) children's behavior (e.g., facial cues, speech occurrences) during interactions with a SAR, and iii) therapist perspectives on introducing SARs into speech-language therapy.

This paper pursues the following research questions:

- **RQ1.** To what extent do children improve their morphosyntactic linguistic skills from speech-language therapy/training with a socially assistive robot?
- **RQ2.** To what extent do SARs promote facial expressions, eye gaze, and speech in children during their speech-language therapy?
- **RQ3.** To what extent do therapists who have experienced SAR believe in its efficacy and would use SARs in their speech-language therapy practice to train children with language impairments?

To explore these questions, we conducted a longitudinal 8-week between-subject user study involving 20 children with language impairments and 6 speech-language therapists. Children received training of their morphosyntactic skills

through linguistic activities in one of two conditions: with a physical QT robot [1] or with a virtual QT robot. We collected study data from automatic activity logs (i.e., linguistic score obtained by children), video and audio recordings (see Figure 1), questionnaires, and a group interview with the therapists. The contributions of this work are threefold, demonstrating that:

1. The child participants in the study significantly improved their linguistic skills in both conditions: with a physical and virtual SAR);
2. The child participants in the physical robot condition who were engaged in the interaction with the physical SAR (measured by gaze direction and head position) were more likely to show linguistic skill improvements and had significantly more speech occurrences in child-robot-therapist triads;
3. Therapists were skeptical about the adoption of SARs for improving skills of children with language impairments; however, they acknowledged that physical SARs could be beneficial for keeping children engaged, motivated, and positive during therapy.

This paper is structured as follows. Section 2 presents related work on socially assistive robots for children with language impairments. Section 3 describes the hypotheses, the study design, the systems used, the participant recruitment, the linguistic activities used to train comprehension and production skills, and the procedure, measurement, and analyses. Section 4 reports on the results of the study, and Section 5 discusses them. Section 6 concludes the paper.

2 Related Work

Socially assistive robotics (SAR) aims at assisting people with special needs, through social rather than physical support [2, 37, 54]. In recent years, SAR has been studied extensively in the context of training social and communication skills for children, in particular in clinical practice for individuals with autism (e.g., [19, 24, 53, 57, 65, 68]). SARs have the potential to improve child engagement while also allowing therapists to deliver more interactive sessions [29, 32]. The following sections overview past work into SAR-supported speech-language therapies, linguistic interactions, and related interventions, into the role of the robot's

embodiment, and into therapist involvement in such interventions.

2.1 Socially Assistive Robots for Children with Language Impairments

Past studies explored SARs as means of enhancing linguistic skills of children with language impairments, especially in the context of speech-language therapy, but no work to date has focused on morphosyntactic structures. Shimaya et al. [67] investigated how a humanoid SAR could promote verbalization in three teens on the autism spectrum. Participants demonstrated desirable non-echolalic reactions toward discussions about challenges of human relationships. Estévez et al. [33] explored the use of a Nao robot in speech therapy interventions with five children with language disorders. The results suggested that the robot could promote attention, motivation, and readiness to learn. Lee and Hyun [49] investigated the use of a robot companion to promote linguistic interactions by children with language disorders. The study involved four children with autism and suggested that children learned to initiate conversations with the robot and expressed emotions. Egado-García et al. [31] performed a case study to develop adaptive behaviors for a Nao robot used in speech-therapy sessions. The results showed that the robot could play a positive, motivating role in several speech-therapy activities. Robles-Bykbaev et al. [60] presented a system that provided decision support for planning therapy sessions, and a robotic assistant SPELTRA for motivating children with communication disorders to engage in therapeutic activities, along with a module for creating clusters of patients with similar needs and profiles. The system was validated in two phases: the first (N=111) evaluated the robot's appearance and functionality with typically developing youth, and the second (N=70) collected interaction data between the robot and children with communication disorders. The results showed that participants of both studies felt confident and comfortable in interacting with SPELTRA, and capable of carrying out the therapeutic activities.

Overall, previous studies establish that SARs are promising tools for promoting linguistic interactions and eliciting motivation and engagement

of children with language impairments during speech-language therapy.

2.2 Socially Assistive Robots vs. Virtual Agents for Children with Special Needs

Within the assistive technologies literature, not only socially assistive robots but also virtual agents (embodied conversational agents [14]) have shown promising results in providing social support for users with special needs. The success of socially assistive robots is demonstrated in children's preferences towards robots relative to humans [3, 36, 72, 74] or computer-based agents [75]. For example, Fachantidis et al. [36] analyzed interactions of four children with autism and an agent (a person or robot) that guided activities to improve knowledge and comprehension of emotions and recognition of facial expressions, association with social situations, and empathy. The results showed that children who struggled with human-human relationships comfortably approached and talked to a robot. Our past work (Spitale et al. [75]) compared a physical robot with a virtual agent in a speech-language therapy intervention; the results showed that children with DLD preferred the physical robot over the virtual agent; they were more engaged, motivated, and found the robot smarter than the virtual agent.

Embodied conversational agents have been shown to be effective in supporting task-based skill learning by children because they can keep children focused on the task and less distracted by the agent itself. Past works have developed and evaluated many embodied conversational agents in educational and learning contexts [39, 84]. These agents generally operate within a controlled interaction space that helps the user maintain attention during task-based activities. Anzalone et al. [4] argue that SARs can engage users in multi-modal ways, unlike embodied conversational agents, serious games, or other software agents because of their own physical presence in the real world. SARs have been shown to promote joint attention, turn-taking and verbal initialization, enhancing social and communication skills of children [20, 64, 65]. Although virtual agents have been shown to be efficient in promoting language skills in children with language

impairments, socially assistive robots can offer physicality-based interactions that are especially important for children on the autism spectrum [64].

Both SARs and virtual agents can promote the improvements of skills for children with special needs (e.g., autism and DLD), however the literature supports children's preference for SARs over virtual agents in therapeutic social interaction contexts.

2.3 Therapists' Perspectives

Past works have explored the use of robotic assistants for speech-language therapists. Robles-Bykbaev et al. [62] presents RAMSES, a robotic assistant, and a mobile support environment for assisting speech-language pathologists in speech-language therapy. Their results demonstrate the possibility of automating activities involved in speech-language therapy, allowing therapists to perform their activities in various locations and in a way that is comfortable for users with language disorders. Caldwell Marín et al. [13] also developed of a robotic platform for assist speech-language pathologists and showed that the platform could become a useful tool during speech-language therapeutic interventions. Robles-Bykbaev et al. [61] presented a low-cost robotic assistant to support activities during speech-language therapy sessions that was capable of registering patient information, results of therapeutic sessions, and providing remote support for reinforcing activities at home. A pilot study validated the system in 73 therapeutic sessions with 29 children with cerebral palsy. The results showed that children adapted very quickly to the robotic assistant and demonstrated significant improvement in language training.

Despite the promising results of using robots in speech-language therapy, many people are skeptical or even opposed to their use in real therapy contexts [23]. A recent European survey [35] showed that only the 26% of participants were comfortable with "having a robot to provide services and companionship when infirm or elderly" or "having a medical operation performed on them by a robot." Taheri et al. [79] investigated the use of a Nao robot to promote music therapy for children with autism; the therapists involved in the study expressed skepticism about using the robot but acknowledged that the robot could be used

as a facilitator in interventions. Conti et al. [22] explored the perception of practitioners and future professionals (students) in adopting robots into their therapeutic sessions. Their results showed that practitioners were skeptical and perceived the assistive robot as an expensive and limited tool. According to Scassellati et al. [64], this is due to the limited involvement of therapists in the study design process. Overall, despite the promising potential for SARs in the therapy context, some therapists still seem skeptical.

Based on the state-of-the-art in the three areas—SARs for speech-language therapy, virtual agents vs. SARs in therapy, and therapist perspectives—we designed our study to explore SARs for training morphosyntactic linguistic skills (the rules that determine the morphological and syntactic relationship between linguistic forms) of children with language impairments during SAR-supported speech-language therapy administered by therapists.

3 Study Method

We conducted a study to explore the use of SARs in speech-language therapy to i) improve children’s linguistic performance (i.e., the ability to understand and produce a specific linguistic structure), ii) understand what children’s behaviors (e.g., facial cues and speech occurrences) SARs can promote, and iii) gather therapists’ perspectives on the introduction of SARs into therapy.

Due to the COVID-19 pandemic, speech-language therapists were the only ones allowed to interact with the children receiving speech-language therapy; this presented an opportunity to naturally involve the therapists in the study. Consequently, therapists conducted the study, providing the child participants with all the needed support, tools, and materials and also receiving first-hand experience with SARs in the therapy context. The therapists included in this study were not involved in the design phase of the study activities. Instead, the design of the activities was led by experts in psycholinguistics who had previous experience designing speech-language activities for this target group.

3.1 Study Design

We designed a longitudinal study involving 20 children with language impairment and 6 therapists in a European speech-language therapy center. The study lasted 8 weeks (1 session/week) and included: pre-intervention assessment (1 session), intervention (6 sessions), and a post-intervention assessment (1 session) and a post-intervention survey and group interview with the therapists (1 session).

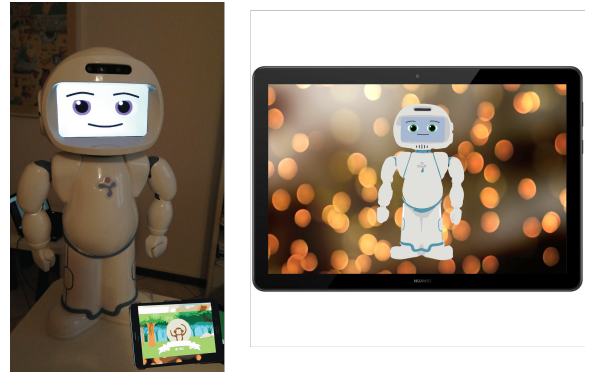


Fig. 2 Left: the QT robot (SAR condition). Right: the virtual QT robot (V-SAR condition).

We conducted a longitudinal between-subjects study (inspired by [8, 20, 47]) with the embodiment of the robot / agent as the independent variable, resulting in two conditions (see Figure 2):

- **Physical Socially Assistive Robot (SAR):** children interact with a QT robot [1] via either their speech and/or a tablet;
- **Virtual Socially Assistive Robot (V-SAR):** children interact with a virtual QT robot via either their speech and/or a tablet.

During the pre- and post-intervention assessment sessions, all participants performed the same activities on a laptop without any support from the robot (physical or virtual).

3.2 Hypotheses

Grounded in our prior work [75], we developed the following hypotheses about the use of SARs in training morphosyntactic skills of children with language disorders:

- **HP1.a:** Child participants will improve their linguistic skills after an intervention with either agent (SAR or V-SAR).
- **HP1.b:** Child participants will improve their linguistic skills more by interacting with the SAR than with the V-SAR (see Section 2.1 and 2.2).
- **HP2.a:** Child participants who engage with the robot will improve more after the intervention than those who do not, as measured by their facial cues (head position and gaze direction).
- **HP2.b:** Child participants in the SAR condition will communicate more (i.e., number of speech occurrences) in the triad interaction (child-robot-therapist) than children in the V-SAR condition (see Section 2.2).
- **HP3:** Therapists will be more skeptical about using the robot for training children’s skills in speech-language therapy (as stated by Looije et al. [51], see Section 2.3) than as a companion.

In the context of this work, we define engagement as: “the process by which [a participant and a robot] establish, maintain, and end their perceived connection” [69].

3.3 Systems

We provided the therapists with a LuxAI QT robot [1], two tablets, a laptop, and a camera. QT is a commercially available humanoid-like tabletop robot (Figure 2) equipped with an RGB-D camera, an array microphone, speaker, and a total of 8 degrees of freedom: 2 in the neck, 2 in each shoulder, and 1 in each elbow. The robot’s perception and action were fully autonomous, and its decision-making was based on pre-scripted responses. The robot/agent was able to transcribe a participant’s speech using real-time speech-to-text, then used natural language processing to check for correctness, and then determined if to move ahead with the training or request that the participant try again.

We used Amazon Web Service (AWS) Polly’s Justin voice (as in other SAR research, e.g., [73, 78]) for both robot agents (SAR and V-SAR). We also used Amazon Polly visemes for synchronizing the robot agents’ mouth positions with the spoken voice, again for both robot agents (SAR and V-SAR). To enhance QT’s expressiveness, we worked with collaborators from psycholinguistics, who advised the design of small movements

for the robot’s head (e.g., nodding when children answered correctly) and arms (e.g., greeting the children at the beginning of the interaction). Both agents used Google Speech-to-Text to transcribe the child participant’s speech and used Amazon Lex Web service to respond verbally. We used the HARMONI [76] framework to compose the human-robot interaction.

The robot and tablets were connected to the WiFi network for access to cloud services, specifically AWS for text-to-speech, and Google speech-to-text. The laptop was used for the pre- and post-intervention assessments, to avoid any contamination of data with the tablet child participants used. We video-recorded all interaction sessions.

3.4 Participants

Table 1 Assignments of therapists and child participants to the two study conditions

Therapist ID	N of children in SAR group	N of children in V-SAR group
T1	1	1
T2	3	3
T3	2	2
T4	1	2
T5	2	1
T6	1	1
Total	10	10

We recruited a total of 41 children and 9 speech-language therapists as volunteers from a speech-language therapeutic center in Italy using the following inclusion criteria for the child participants:

- aged 6 to 12 years;
- diagnosed with developmental language disorder (DLD) or autism spectrum disorder and co-occurring language impairment (ASD-LI);
- monolingual (Italian);
- spontaneous language production: mean length of utterance (MLU) of at least 2.5 ¹.

All children participants in the study were already attending therapeutic sessions at the speech-language center before the beginning of the study.

¹MLU is a measure of children’s linguistic productivity. It is computed by taking 100 child utterances and dividing the number of morphemes, i.e., the smallest meaningful lexical item, by the total number of utterances. A higher MLU indicates a higher degree of linguistic ability.

We excluded children (3) whose scores on the morphosyntactic structures (clitics and passives, defined in Section 3.5 [30]) were greater than 80%, those who did not attend all the study sessions (15), and those who were quarantined due to COVID-19 (3). This resulted in 20 children being included in the study, 14 males and 6 females, aged 6-11 ($M=8.2$, $SD=1.36$); 11 were diagnosed with ASD-LI and 9 with DLD. Because of the exclusion of the 3 child participants, 3 of the 9 therapists did not proceed with the study because their patients were excluded from it.

Due to practical constraints, child participants could not be equally distributed among therapists. Table 1 shows the therapist-child assignments. In the rest of the paper, we use the terms “therapist” and “participant” (child). Participants were randomly assigned to the two conditions. In the SAR condition, 4 had a diagnosis of ASD-LI and 6 had a diagnosis of DLD; in the V-SAR condition, 6 had a diagnosis of ASD-LI and 4 had a diagnosis of DLD. This research was approved by the Politecnico di Milano Research Ethics Board; all therapists and participants were uncompensated volunteers.

3.5 Linguistic Activities

We provided therapists with 16 linguistic activities to assess (6 activities) and train (10 activities) the participants’ comprehension and production skills on two morphosyntactic structures that are known to be difficult for (Italian) children with language impairment: clitic pronouns [5, 58] and passive sentences [30, 50]. Clitic pronouns are monosyllabic words that must accompany a verb and express the gender, number, and case of the object of a transitive verb (e.g., “Maria la lava” = “Mary washes *her*”). Passive clauses are sentences where the object of an active sentence is promoted to be the subject of the passive construction (e.g., “La bambina ‘e lavata da Maria” = “The girl *is washed* by Mary”).

3.6 Assessment

Six activities were used in the pre- and post-intervention assessments of the participants’ linguistic comprehension and production skills, as follows.

3.6.1 Comprehension

Based on guidance from psycholinguistics experts, we created two sets (1 for pre, 1 for post) of assessment picture selection tasks (adapted from [43]) for each structure (4 tasks total). Participants listened to a sentence containing the target linguistic structure (e.g., a passive clause) while being presented a set (2 or more) of pictures. They were then asked to select (by pointing to the laptop screen) the picture that best represented the sentence they had just heard from a set of pictures displayed on the screen. The therapists then used a computer mouse to click on the selective picture. No feedback was given.

3.6.2 Production

We used a Sentence Repetition (SR) test (1 for pre, 1 for post), controlling for sentence complexity, lexical access, and number of words/syllables (adapted from [75]), targeting passives and clitics. The participants listened to a sentence and then were asked to repeat the sentence verbatim. As above, no feedback was given after repeating the sentence.





3.7 Training

Ten activities were developed for robot/agent-supported training of linguistic comprehension and production skills. The role of the robot/agent was to present the training activities, i.e., to show pictures and describe the scene for each activity. Psycholinguistics expert members of the research group recommended the use of the popular storytelling format because linguistic context (as in storytelling) boosts language skill learning and syntactic comprehension [45]. Therefore, participants were presented with a series stories; for comprehension, a story composed of several sentences and a picture to match to a heard sentence, and for production, a sentence from story to repeat verbatim.

3.7.1 Comprehension

Two stories were created to train passive clauses and clitic pronouns. Each story was composed of three episodes (6 activities total); in each, a picture was shown on the tablet screen that depicted the main scene, and then the robot/agent told a sentence while the tablet screen displayed three

Table 2 A comprehension training activity: “Philippe the horse is searching for his hat inside the fence.”

Main scene	Options	Main scene after choice
Inside the fence there is a pig		
Inside the fence there is also a cow		

pictures, one of which was consistent with the spoken sentence. The participants were asked to click on the tablet screen to select the picture that best matched the spoken sentence; if they did so correctly, the robot/agent added the picture to the storyboard (as shown in Table 2) and continued with telling the story; if they chose incorrectly, the robot/agent asked the participant to try again, after repeating the same sentence of the story, until the participant selected the correct picture.

3.7.2 Production

We adapted the Sequential Order Subtest [83] to the story-telling task by using two sets of four pictures for two different stories targeting passive clauses and clitic pronouns. The robot/agent told a story composed of four scenes. As the story was being told, one picture at a time was shown. At the end of the story, the tablet screen showed all four pictures in order, and the robot/agent asked the participant to retell the story as closely to the original as possible (ideally verbatim) while using the pictures as prompts. The robot/agent asked to the participant to repeat the story if they missed any key parts of the story (based on participant’s keywords).

3.8 Questionnaires

At the end of the 6-session intervention, the therapists completed questionnaires that assessed their experience with SAR and V-SAR and their perspectives on using SAR in therapeutic intervention (as in [71]). The questionnaires consisted

of: Adoption of Technology (AoT) [63] (evaluating willingness to adopt a technology), Quebec User Evaluation of Satisfaction with Assistive Technology (QUEST) [27] (measuring the level of satisfaction with a technology), and System Usability Scale (SUS) [11] (assessing the usability of a system). The therapists completed the questionnaires for each of the two versions of SAR they worked with during the intervention. Additionally, they completed a specialized Likert scale (1-5, Strongly disagree - Strongly agree) questionnaire we designed that collected their perspectives about the engagement, likability, and usability of the SARs, based on [44].

3.9 Procedure

The study lasted 8 weeks, and consisted of: 3 preparatory team meetings, a pre-intervention assessment session, 6 speech-language training sessions, and a post-intervention assessment session with children, and a group interview meeting with therapists (see Figure 3).

The three 2-hour preparatory meetings with all the therapists and researchers were held in person, following all COVID-19 safety protocols. The first meeting explained the purpose of the study, the study design and protocol, and the theoretical background and reasons for investigating the use of technologies in speech-language therapy. The meeting also trained the therapists on how to use the robot (SAR), agent (V-SAR), and tablet so they could conduct the study without any in-person support. Three researchers and nine therapists (3 of whom were later excluded after their patients did not meet the study inclusion criteria, as described in Section 3.4) participated. The therapists received study user manuals and were asked them to read the materials in time for the second meeting and prepare any questions.

The second meeting addressed all therapist questions, discussed precise inclusion criteria, and gave a demonstration of the technical setup for both conditions. The third meeting was focused on training therapists on how to set up and use the SAR and V-SAR systems, and included three practice run-throughs.

In the pre-intervention assessment session, the participants used a laptop app we created for the study in order to avoid using any technologies that are part of the study conditions (tablet,

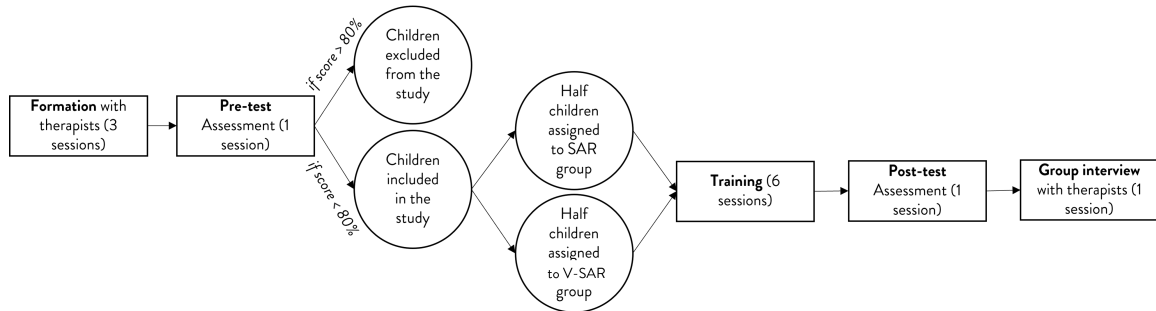


Fig. 3 Study timeline; sessions were weekly.

robot). The assessment was performed in the participants' usual therapy room, at a table, with a therapist who sat next to them and assisted them as necessary without intervening. After the pre-intervention assessment, the participants who met the inclusion criteria (Section 3.4) were randomly assigned to one of the two study conditions (SAR or V-SAR).

In the intervention sessions, therapists followed pre-scripted instructions that involved:

1. Plugging in the robot that then automatically switched on (SAR condition) or powering on the tablet that then displayed the virtual robot (V-SAR condition);
2. Powering on the tablet with the speech and language training activities;
3. Placing the camera on the table.

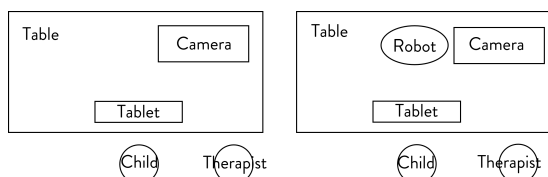


Fig. 4 Setting for the study conditions: V-SAR (left) and SAR (right).

For each intervention session, participants entered the session room, and their therapists sat next to them (as in the pre-intervention assessment, see Figure 4). Figure 1 shows a participant from the SAR condition interacting with the robot during a session. As the end of the study, we conducted a post-intervention assessment following the same procedure as in the pre-intervention assessment.



Fig. 5 Group interview with 5 therapists (one of the therapist was not present when the photo was taken).

After the intervention and the group interview, the therapists completed the AoT, QUEST, and SUS questionnaires and the questionnaire on engagement, likability, and usability of the SAR and V-SAR.

The group interview with the therapists lasted 90 minutes, and followed a semi-structured protocol: we prepared questions for the therapists, and encouraged them to freely express their opinions and asked unscripted follow-up questions. Figure 5 shows the group interview setting. While we are aware of the possibility of incurring social desirability bias [56], we chose the group interview because synergies among therapists may yield more significant insights and raise multiple perspectives than gathering opinions individually [48]. During the interview, we asked therapists to provide their opinion on the usage, engagement, likability, benefits, and challenges of both SAR and V-SAR, and to recall meaningful episodes regarding the child participants' interactions with SAR and V-SAR.

3.10 Data Collection

We collected a large battery of heterogeneous data from this study. We video- and audio-recorded all intervention sessions and the post-intervention group interview with the therapists. For the questionnaires, we used Google Forms to collect therapists' responses. In addition to the video and audio recordings of all intervention sessions, and the various questionnaires, we also collected activity logs of all sessions that included the number of correct and incorrect answers to the story-related questions, types of incorrect answers, time required to give an answer to any questions, and the duration of each training activity. For speech production, we also collected the transcripts of the participants' utterances.

3.11 Analyses

We statistically analyzed the data using IBM SPSS [38] and R [26]. Our sample was not normally distributed, so we adopted non-parametric statistical tests (e.g., Wilcoxon tests). For the activity logs, we defined a *linguistic score* (ls) as the percentage of correct answers with respect to the total number of tasks (scenes) in each linguistic activity:

$$ls_i^t = \frac{n_{correcttasks_i^t}}{n_{totaltasks_i^t}} \quad (1)$$

where i is the participant ID and t is the linguistic activity. Our main *dependent variable* was *linguistic performance* (lp), computed as:

$$lp_i^a = ls_{post-intervention_i}^a - ls_{pre-intervention_i}^a \quad (2)$$

where i refers to the participant ID, and ls is the number of correct answers for each linguistic activity a .

We used OpenFace [6] to extract head pose, facial expressions, and gaze direction. We used pyannote [10] to extract auditory features (e.g., number of utterances, number of speaker changes). OpenFace visual features were extracted frame-by-frame and depended on the length of each recorded video. We pre-processed the data to remove null and constant features. We represented visual features as a fixed-length vector of the following statistical attributes for time-series data:

mean, median, standard deviation, and autocorrelation (lag 1 second). We then normalized the data (with Min-Max scaler) to cluster facial cues, and applied Principal Component Analysis (PCA) to reduce the dimensionality of the features. We then applied K-means clustering on the principal components to examine the participants' facial cues.

To extract auditory features, we exploited the pyannote library², an open-source toolkit for speaker diarization based on the PyTorch framework. We used their pre-trained models to detect voice activity, speaker change, and overlapped speech.

We analyzed the questionnaires (extracted as csv files from Google Forms) via Microsoft Excel. For the group interview, we transcribed the session using Google API for Automatic Speech Recognition and adopted a bottom-up thematic analysis approach [9], identifying patterns or themes within qualitative data, inferring them *a posteriori*.

4 Study Results

4.1 RQ1: Participants' Linguistic Improvements

We performed three non-parametric Mann-Whitney Wilcoxon tests—one for clitics in comprehension, one for passives in comprehension, and one for both clitics and passives in production—to compare pre- and post-intervention scores (ls , see Equation 1). For all three cases, the Wilcoxon tests with continuity correction revealed that there was no significant difference between the scores obtained pre- and post-intervention. This finding is perhaps not surprising since our sample was very heterogeneous in terms of participant age, diagnosis, and response to training sessions. We therefore performed a detailed analysis for each linguistic skill by computing the *linguistic performance* as defined in Equation 2.

First, we performed three one-sample Wilcoxon tests, analyzing each condition separately. The tests revealed that, for both SAR and V-SAR conditions, there was a significant improvement in terms of linguistic performance in the comprehension of clitics (V-SAR: $z=2.09$,

²<https://github.com/pyannote/pyannote-audio>

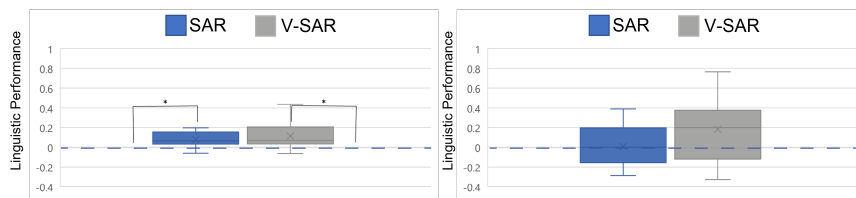


Fig. 6 Comprehension performance for clitics (left) and passives (right) for the SAR and V-SAR conditions. $*p < .05$

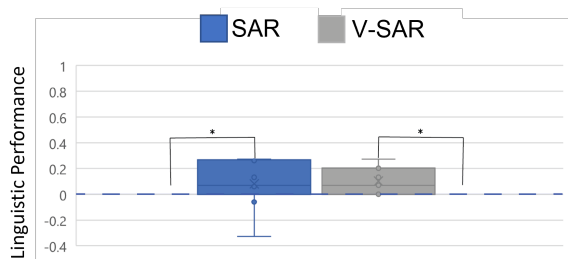


Fig. 7 Production performance for clitics and passives for the SAR and V-SAR conditions. $*p < .05$

$p < .05$; SAR: $z=2.51$, $p < .05$) but not in the comprehension of passives (V-SAR: $z=1.48$, $p = .14$; SAR: $z=1.31$, $p = .20$). *The participants' comprehension skills in clitic pronouns improved significantly in both V-SAR and SAR conditions. The participants' comprehension skills in passives did not improve significantly in either conditions. Therefore, HP1.a is partially supported for comprehension skills.*

We then performed the same tests to evaluate production of clitics and passives. Results showed an improvement in both structures (clitics $p = .022$; passives: $p = .028$). Since both structures were evaluated via the same task, we created a combined score of clitics and passives for the SR task. Thus, the Wilcoxon test results showed that there was a significant improvement (V-SAR: $z=2.09$, $p < .05$; SAR: $z=2.51$, $p < .05$) in production skills on SR in both the SAR and V-SAR conditions. *The participants' production skills in clitics and passives improved significantly in both V-SAR and SAR conditions. Therefore, HP1.a is supported for production skills.*

Second, we evaluated whether a significant difference existed between the SAR and V-SAR conditions. The Mann-Whitney Wilcoxon test showed that there was not a significant difference between the conditions for comprehension of clitics ($p = .05$) and passives ($p = .08$), and also that there was no significant difference for production ($p = .57$). *The linguistic performance for the SAR*

condition was $Mdn=0.06$ in clitics and $Mdn=0$ in passives and for the V-SAR condition was $Mdn=0.06$ in clitics and $Mdn=0.19$ in passives. For production, the SAR condition performance of clitics and passives was $Mdn=0.07$, and the V-SAR condition performance was $Mdn=0.07$. *The participants performed equally well in the SAR and V-SAR conditions for both comprehension and production of clitics and passives. Therefore, HP2.b is not supported.*

4.2 RQ2: Participants' Behaviors

4.2.1 Facial Cues

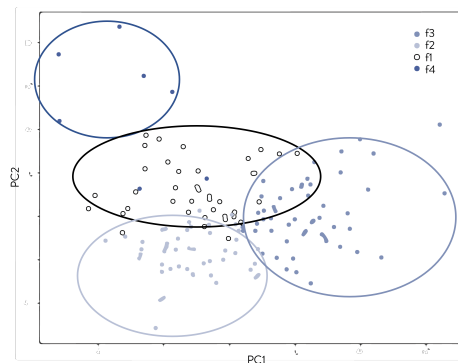


Fig. 8 The four clusters by principal components PC1 and PC2

The facial cues analyzed in this work consisted of head position, gaze direction, and facial action units. The principal component analysis returned three principal components that accounted for 50% of the variance in the dataset. Next, the k-means clustering analysis resulted in $K=6$ according to the inertia score. This yielded 6 clusters; we removed two outlier clusters that included one data point each. Figure 8 shows the clustering based on facial patterns extracted with OpenFace features by highlighting the 4 clusters without outliers.

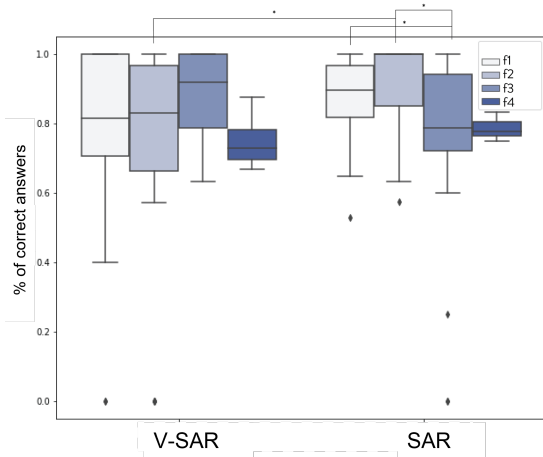


Fig. 9 Linguistic scores (% of correct answers, see Equation 1) for the SAR and V-SAR conditions split into four clusters based on PCA and clustering analyses, $*p < .05$.

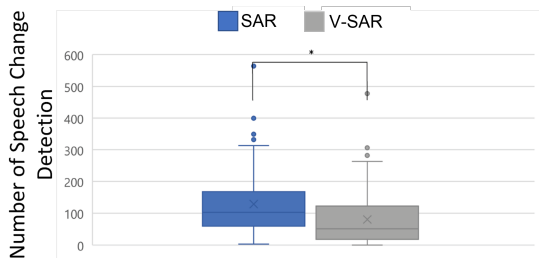


Fig. 10 Auditory features in the SAR and V-SAR conditions: number of speech change detections [continued]. $*p < .05$

We performed a one-way ANOVA to evaluate the linguistic scores among the clusters in each condition with Bonferroni correction (0.05/4) and post-hoc analysis with a t-test to compare the clusters between conditions. The ANOVA revealed that there were statistically significant differences among clusters in the SAR condition ($F(3, 72) = 3.29, p < .05$) and not in the V-SAR condition ($F(3, 72) = 1.19, p = .50$). We ran a t-test post-hoc analysis with Bonferroni correction (0.05/4) to evaluate cluster differences and found that $f3$ was significantly different from $f1$ ($t(148) = -2.19, p < .05$) and $f2$ ($t(148) = -2.44, p < .05$). We also performed t-tests to compare the 4 clusters' correct answers between the V-SAR and SAR conditions. Our results showed that the correct answers in the SAR condition were significantly higher ($t(148) = -2.04, p < .05$) than correct

answers in V-SAR only for $f2$; no significant difference was found in other clusters. Figure 9 shows the scores for each cluster for the two study conditions. We evaluated three most important principal components for each cluster, examining how much each feature contributed to each PC. We only focused on the $f2$ cluster, where correct answers differed between the V-SAR and SAR conditions. Our results showed that the change in gaze direction and head position were the visual features that characterized the most important PCs in the $f2$ cluster, while the facial action units did not contribute significantly to the PCs. This means that participants who were constantly shifting their attention focus between the robot and the screen (where the activity inputs were displayed) were more likely to provide a correct answer. Conversely, participants who did not display a shift in their attention focus (i.e., those who looked at the robot constantly), were distracted by the robot and performed worse on the linguistic tasks. *In the SAR condition, the participants in clusters with significantly higher linguistic scores shifted their attention focus more often (in terms of gaze direction and head positions) than participants in clusters with lower scores. Therefore, HP2.a is supported.*

4.2.2 Speech Occurrences

We ran three t-tests to evaluate the difference between the conditions in terms of auditory features extracted with the pyannote library: number of overlaps, number of speech activity detections, and number of speech changes for each session. There was a statistically significant difference between the conditions in the number of speech activity detections ($t(148) = -4.42, p < .05$) and the number of speech changes ($t(148) = -3.13, p < .05$); both were significantly higher in the SAR condition than the V-SAR condition. Specifically, the number of overlaps in SAR was $M = 28.71$ with $SD = 25.99$, and in V-SAR has $M = 20.91$ with $SD = 25.34$; the number of speech activity detections in SAR was $M = 67.11$ with $SD = 35.72$, while in V-SAR it was $M = 42.00$ with $SD = 33.24$; the number of speech changes in SAR was $M = 128.17$ with $SD = 99.72$, while in V-SAR it was $M = 80.25$ with $SD = 85.62$. There was no statistically significant difference between the conditions in the

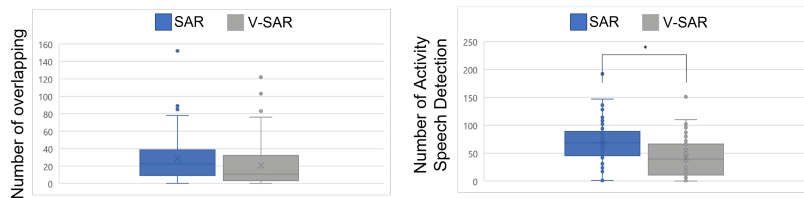


Fig. 11 Auditory features in SAR and V-SAR conditions: number of overlaps, activity detections, and speech change detections. $*p < .05$

number of overlaps ($t(148) = -1.84, p = .067$). Figure 11 plots the auditory features: number of overlaps, number of speech activity detections, and the number of speech changes in both conditions. *The participants in the SAR condition spoke significantly more (number of speech activity detections) than in the V-SAR condition and the number of times speakers–participant, therapist, and robot–spoke (number of speech changes) was significantly higher in the SAR condition than in the V-SAR condition. Therefore, HP2.b was supported for speech.*

4.3 RQ3: Therapists’ Perspectives

From the questionnaire results and the thematic analysis of the therapist group interview, three main topics emerged about the therapists’ perspectives about their experiences with the SAR and V-SAR: i) *usability and adoption* – therapists found both tools easy to learn and easy to use, however they encountered some issues with the SAR, and they reported their frustration in setting it up. Still, therapists were skeptical about the adoption of those technologies in the context of therapeutic sessions for improving children’s communication, social, and autonomy skills; ii) *engagement and likability* – therapists perceived both technologies very engaging and appealing for the children; and iii) *robot benefits* – therapists reported that the robot had beneficial effects on children compared to the paper-based traditional method.

4.3.1 Usability and Adoption

We could not perform any statistical analyses of the therapists’ questionnaires, because of the small sample size ($N=6$). Therefore, we report the results obtained in terms of median, minimum, and maximum scores. Table 3 shows the scores of

the SUS obtained from the therapists. The overall SUS average score was 73 for the SAR and 69 for the V-SAR. The SUS results indicate that the therapists did not have problems with using the SAR and the V-SAR. We would expect that they could have had more issues with using the SAR because of the higher technological complexity and lower familiarity of physical robots, but the data do not demonstrate that. Instead, the therapists were able to use both the SAR and the V-SAR equally well.

The group interview revealed the main issue therapists had during the intervention, which did not emerge in the questionnaire data: the time required for system setup. Specifically, for the SAR condition, therapists had to take the robot out of a box, unwrap it, and place it on the table. Next, they had to connect it to the power supply, and wait for for it to power on. In the mean time, they had to place the video camera on the table and check it field of view, and then power on the tablet and place it in front of the robot on the table. Finally, they had to switch on the mobile router to access to Internet connection for enabling cloud services. Once the robot has “awake”, the therapists invited the child participant into the room, and started the intervention session. The system setup took about 10 minutes, and that time was seen as wasted considering that each therapeutic session lasted about an hour and they had less time with the child.

Tables 4 and 5 report the items of the Adoption of Technology (AoT) and Quebec User Evaluation of Satisfaction with Assistive Technology (QUEST) questionnaires, respectively. We observed that there was no difference between the two conditions (SAR vs. V-SAR), and that therapists were overall quite satisfied with both technologies. However, they still had some doubts about the use of SAR and V-SAR for helping

Table 3 System Usability Scale (SUS): Median (Min; Max) results from therapists answers (6 therapists for SAR and V-SAR). In the SAR condition, the term “robot” refers to the physical robot QT, while in the V-SAR condition the same term refers to the virtual character.

Higher scores are preferred	SAR	V-SAR
I would like to use the robot frequently	3 (2;4)	2.5 (2;4)
I thought the robot was easy to use	4.5 (3;5)	4 (3;5)
The functions of the robot were well integrated	3 (2;4)	2.5 (2;4)
Most people would learn to use the robot quickly	4 (2;5)	3.5 (2;5)
I felt very confident using the robot	4 (3;5)	4 (3;5)
The robot is a tool that would be easy to incorporate into my work routine	4 (4;5)	4 (2;5)
Lower scores are preferred	SAR	V-SAR
I found the robot unnecessarily complex	1 (1;2)	1.5 (1; 3)
I need the support of a technical person to use the robot	2 (2;3)	2.5 (2;3)
There was too much inconsistency in the robot	3 (1;3)	2.5 (2;3)
I found the robot very cumbersome to use	1 (1;1)	1 (1;5)
I needed to learn a lot of things before I could get going with the robot	1 (1;3)	1.5 (1;3)
I sometimes find the robot frustrating to use	2 (1;3)	2 (1;3)

Table 4 Adoption of Technology (AoT): Median (Min; Max) results from therapists answers (for SAR and V-SAR 6 therapists) where higher scores are preferred. In the SAR condition, the term “robot” refers to the physical robot QT, while in the V-SAR condition it refers to the virtual character.

	SAR	V-SAR
The robot helps children stay engaged	4(3;5)	4 (3;5)
The robot helps children stay motivated	4 (2;5)	3 (2;5)
The robot helps children stay positive	4 (2;4)	3.5 (2;4)
The robot helps improve children’ emotional well-being	3 (2;4)	3 (2;4)
The robot helps children with social skills	2 (1;3)	2 (1;3)
The robot helps children with academic skills	3 (2;4)	3 (2;4)
The robot helps children with life skills	2 (2;3)	2 (2;3)
The robot helps children with communication skills	2 (2;4)	2 (2;4)
The robot helps children stay physically active	3 (1;4)	3 (1;4)
The effects of using the robot at school is apparent to others	3(2;4)	3 (2;4)

children with their social, communication, and autonomy life skills (they scored those items as 2 out of 5, see Table 4). *Our results suggest that therapists were skeptical of the introduction of SAR and V-SAR into the speech-language therapy for improving children’s skills. However, they believed that those technologies could help children to stay engaged, motivated, and positive. Therefore HP3 was supported.*

4.3.2 Engagement and Likability

In the group interview, therapists reported that both SAR and V-SAR can help to keep children motivated and engaged, however the robot (SAR) condition was thought to be more attractive for children. This is consistent with the AoT questionnaire results (see Table 4). For example, T4 stated that her patients were excited to play with the robot, and they really liked it. She also thought

Table 5 Quebec User Evaluation of Satisfaction with Assistive Technology (QUEST): Median (Min; Max) results from therapists' answers (for SAR: 6 therapists, for V-SAR: 6 therapists), higher scores are preferred. In the SAR condition, the term "robot" refers to the physical robot QT, while in the V-SAR condition it refers to the virtual character.

	SAR	V-SAR
The dimensions of the robot are appropriate	3 (2;5)	3.5 (2;5)
The weight of the robot is appropriate	3 (2;5)	3 (2;5)
The appearance of the robot is appropriate	3 (2;4)	3 (2;4)
The voice of the robot is appropriate	3 (2;4)	3.5 (2;4)
The robot behavior is appropriate	3 (2;4)	3 (2;5)
The robot is safe	5 (3;5)	4.5 (3;5)
The robot is durable	3 (3;5)	3.5 (3;5)
The robot is an effective device to assist in education	4 (2;4)	3 (2;4)

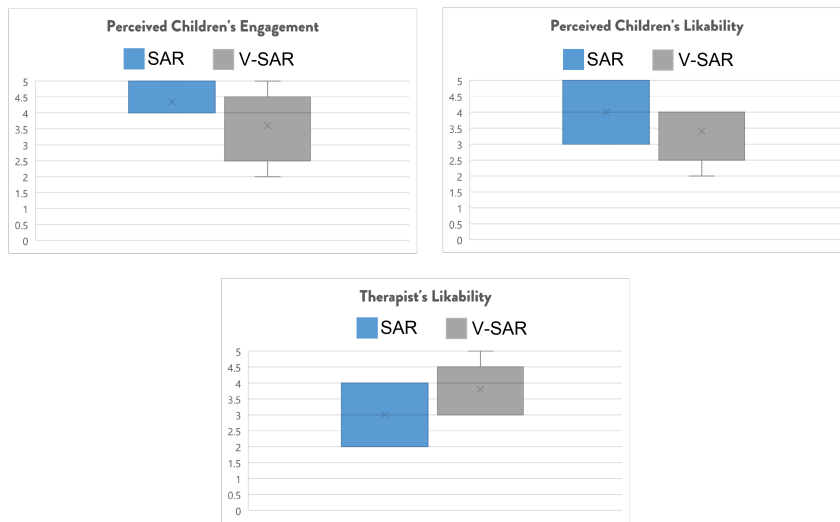


Fig. 12 Therapists' opinions about the participants' engagement and robot's likability, and their likability scoring.

that the robot was more attractive and engaging for her children. T5 reported that her patients were excited to play with the robot at the beginning, but towards the end of the intervention they got bored by playing the same type of activities at each session. She believed that they were really engaged at the beginning because it was a new game with a robot (novelty effect) but that they were not so interested in playing the activities. She believed that children would have preferred to change the activities even if they liked the robot. T2 stated that her patients were sad to not get to play with the robot once the intervention ended, and she reported that C7 asked her: "Where is QT?". Also T5 reported her patients requested to play with the robot. C5 said: "I do want to play

with the robot again." T1 reported that C14 interacted with the V-SAR in a very immersive way, leading C14 to isolate because they wanted only to play alone without the help of the therapist. T1 said: "I felt excluded".

We asked therapists to score how much they thought their patients liked and engaged with the SAR and V-SAR interaction. As shown in Figure 12), therapists seem to like V-SAR more than SAR; their likability scores for SAR were: $Mdn = 3$, $Min = 2$, $Max = 4$, while for V-SAR they were: $Mdn = 4$, $Min = 3$, $Max = 5$. However, the therapists believed that their patients preferred SAR over V-SAR because they perceived that the patients were more engaged with the SAR. Their perception of the participants' engagement was:

for SAR $Mdn = 4$, $Min = 4$, $Max = 5$; for V-SAR $Mdn = 4$, $Min = 2$, $Max = 5$; and their perception of the participants' judgement of agent likability was: for SAR $Mdn = 4$, $Min = 3$, $Max = 5$; for V-SAR $Mdn = 4$, $Min = 2$, $Max = 4$. *Our results show that therapists perceived that the participants were engaged with and preferred SAR over V-SAR.*

4.3.3 Benefits of the Physical Robot

In the group interview, therapists reported several beneficial aspects and also challenges of adopting SAR into the therapeutic context.

First, they stated that the patient's diagnosis impacts on their willingness to interact with a robot. T2 reported that very rigid children (i.e., those with an inflexible mindset) were frightened that they could not predict the robot's behavior. T2 told us that one of his patients (C15) rejected the interaction with the robot at the beginning because of a rigid mindset, and that C15 perceived the robot as unpredictable. T2 suggested having a longer familiarization phase for patients with similar mindsets, to prepare them for the interaction. T4 reported that all of her patients really liked the robot from the very beginning, because they found it very captivating.

Second, therapists also believed that the interaction with the robot allowed children to feel free and not judged. T3 reported that one of her patients worked very well with the robot; the patient did not fear judgment from others while interacting with the robot, and for this reason he particularly liked it. Therapists also thought that there was an important difference between the perceptions the children had of the SAR vs. V-SAR. In the therapists' opinion, the children saw V-SAR on the tablet as just another character in the training activity, while in the case of SAR, children perceived the robot as a tutor. T4 said: "The character on the tablet is just another character in the story activity, they didn't see it as a tutor, as in the case of the robot."

Third, in T4's opinion, the robot worked best in conversational tasks, such as the speech production activities. T4 reported that, after interacting with the robot, one of her patients was able to verbalize the speech rule behind the linguistic structure.

Forth, T4 reported that often the "fragility" of the robot was a physical limitation for the children who wanted to touch the robot. She reported that a child shook hands with the robot whenever he entered the room. T4 suggested that a soft robot (or one that did not break) would be much better for interacting with children.

Finally, therapists reported that, during their therapy sessions, they continued to prefer traditional methods in which they themselves created content over the SAR and V-SAR solutions. T2 and T1 stated that both technologies can be used at home to support therapy practice.

Our findings suggest that therapists believed that SAR can be beneficial for children with non-rigid mindsets, and that children did not feel judged by SAR and were free to express themselves to the robot. However the therapists still preferred traditional therapy methods they controlled over the use of SAR and V-SAR. Therefore HP3 is supported.

5 Discussion

5.1 Study Insights

The results of the study show that participants significantly improved their linguistic skills involving clitic pronouns when trained with both SAR and V-SAR. However, the participants did not show any significant improvement in passive clauses. This may be explained by the fact that clitics are a simpler linguistic structure for children to acquire in general (acquired between 2-4 y.o.), while passives are more complex and require more time to learn (acquired between 4-6 y.o.) [42]. It is likely that participants did not have enough time in our study to improve significantly in the more complex linguistic structures.

We did not find any statistically significant difference between the SAR and V-SAR conditions in terms of participants' linguistic improvement. While a great deal of work has shown significant differences between physical robots and virtual agents (for a review see [28]), some task-based interactions do not demonstrate a difference between agents, such as [25], where participants performed equally well in human and robot conditions.

We analyzed the participants' facial expressions not only in terms of action units, but also in

terms of facial landmark positions and gaze direction. The PCA results showed that gaze and facial landmark positions were most correlated with linguistic improvement in the SAR condition. We observed that participants who improved their linguistic skills most shifted their eye gaze between the robot and tablet many times over the session. This can be interpreted to mean that some participants experienced effective training when the robot acted as a supporter of the linguistic training, while for others the robot drew attention away from the task to itself. These results are in line with the findings of [55], who investigated children’s behavioral patterns in a triad interaction (child-child and robot) in an educational setting. They defined Productive Engagement as leading to a positive learning outcome, as was the case with the participants in the SAR condition of our study.

Finally, we observed that the number of speech changes and interactions were significantly higher in the SAR condition than in the V-SAR condition. This is consistent with other HRI literature, where many studies showed evidence of the potential of using SAR for promoting communication and social skills [67].

5.2 Lessons Learned From Therapists

Our results show that children really enjoyed interacting with SAR. Therapists reported that, in their opinion, children preferred the interaction with SAR over V-SAR. Besides enjoying and engaging with the robot during the therapy session, children also asked to play with the physical robot after the end of the intervention; this is a promising results for long-term interventions that go beyond the length of this study (6 weeks). Therapists reported that children perceived the SAR as a companion, and the V-SAR as a character of the training activity. This is consistent with past findings [12, 16].

The development and adoption of robots into therapeutic contexts is still an open-challenge for HRI. In our study, therapists’ perspectives about the introduction of either SAR or V-SAR was clear: they preferred to keep using their traditional methods that gave them autonomy instead of adopting new technologies. Accordingly, they

questioned the efficacy of SAR and V-SAR in helping children with language disorders with their social, communication, and autonomy life skills (per the AoT results). While we strove to make therapists aware of the benefits and challenges of introducing SAR technologies into therapy in the introductory training session, they still maintained some skepticism and prejudice against SAR and V-SAR, and suggested that those solutions may be best for the home context. Many studies (e.g., [51]) reported that both caregivers and parents can be skeptical about a robot’s role. Even when the majority of them acknowledged the effectiveness of SARs in therapeutic interventions, they still had doubts about their use [18]. A possible way of addressing this challenge is to involve therapists (and parents) from the early stage of the research, allowing them to be involved in shaping the technologies toward their full potential.

5.3 Limitations and Future Work

Our study had several limitations. First, due to the highly complex logistics of working with children with special needs, especially during a pandemic, our sample size was small, limiting the generalization power of the results. This is especially challenging because, as is also typical for working with special needs populations, the sample is heterogeneous. Even if we tried to group children with similar linguistic capabilities in the pre-assessment phase, each child is characterized by different needs, and consistent group results are rare [64]. Next, because of the pandemic conditions, therapists had to conduct the study without the help of researchers. This could have lead to some bias, both from possible frustrations of working with novel technologies, and from human error in the administration of activities. For example, therapists could have, perhaps unintentionally, influenced the children’s answers during the study even though we instructed them not to intervene unless it was necessary.

Future work can address more in-depth investigation of children’s behavior during interactions with SAR in a linguistics training context. Additionally, such would could accumulate a dataset to be used for model development and application of real-time interactions between children and SAR or V-SAR in linguistic therapy contexts. Additionally, future work can further analyze the gender

and age data to examine what if any role those attributes had on the findings.

6 Conclusion

This paper explored the use of physical and simulated socially assistive robots (SARs) for supporting training of comprehension and production skills of children with language impairments. A 8-weeks between-subject empirical study was conducted by six therapists and involved 20 children with language impairments (DLD or ASD-LI) randomly assigned to interact with a physical or a virtual robot. Our results confirm that SARs can be effective tools for training language skills in children with language impairments because they promote triadic interactions during speech-language therapy. Although the study results are promising in terms of child speech improvements, therapists reported their skepticism about using SAR for improving skills of children with language impairments, but believed that SAR can keep children engaged, motivated, and positive during speech-language therapy. This work aims to inspire and motivate further work into SAR for speech-language therapy and adoption of SAR technologies in real-world settings.

Citation Diversity Statement

Recent work in several fields of science has identified a bias in citation practices such that papers from women and other minority scholars are under-cited relative to the number of such papers in the field. To heighten the awareness of this problem, we state the distribution of citations in this work: 8.4% have been published by a solely female team, 74.7% by a female/male team, and 16.9% by a solely male team.

Data availability

Data and materials are available upon reasonable request to the authors. With these data, any researcher will be able to run any other type of statistical analysis.

Acknowledgment

This research was supported in part by EIT Digital and IBM Italy (supporting Micol Spitale), in

part by the Politecnico di Milano (supporting Silvia Silleresi, and Franca Garzotto), and in part by the University of Southern California (supporting Maja Matarić). The authors thank the speech-language therapists involved in the study for their help with the recruitment process, empirical study design, and for running the study with a great enthusiasm. The entire research team thanks the study participants.

References

- [1] Qtrobot: Humanoid social robot for research and teaching. URL <http://luxai.com/qtrobot-for-research/>.
- [2] Nida Itrat Abbasi, Micol Spitale, Joanna Anderson, Tamsin Ford, Peter B. Jones, and Hatice Gunes. Can robots help in the evaluation of mental wellbeing in children? an empirical study. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1459–1466, 2022. doi: 10.1109/RO-MAN53752.2022.9900843.
- [3] Angelos Amanatiadis, Vassilis G Kaburlasos, Ch Dardani, and Savvas A Chatzichristofis. Interactive social robots in special education. In *2017 IEEE 7th international conference on consumer electronics-Berlin (ICCE-Berlin)*, pages 126–129. IEEE, 2017.
- [4] Salvatore Maria Anzalone, Jean Xavier, Sofiane Boucenna, Lucia Billeci, Antonio Narzisi, Filippo Muratori, David Cohen, and Mohamed Chetouani. Quantifying patterns of joint attention during human-robot interactions: An application for autism spectrum disorder assessment. *Pattern Recognition Letters*, 118:42–50, 2019.
- [5] Fabrizio Arosio, Chiara Branchini, Lina Barbieri, and Maria Teresa Guasti. Failure to produce direct object clitic pronouns as a clinical marker of sli in school-aged italian speaking children. *Clinical linguistics & phonetics*, 28(9):639–663, 2014.
- [6] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. Openface: an open

- source facial behavior analysis toolkit. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–10. IEEE, 2016.
- [7] Hoffman HJ Black LI, Vahratian A. Communication disorders and use of intervention services among children aged 3–17 years: United states, 2012. *NCHS data brief*, 205:7, 2015.
- [8] Indu P Bodala, Nikhil Churamani, and Hatice Gunes. Teleoperated robot coaching for mindfulness training: A longitudinal study. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, pages 939–944. IEEE, 2021.
- [9] Virginia Braun and Victoria Clarke. Thematic analysis. 2012.
- [10] Hervé Bredin. pyannotate. metrics: A toolkit for reproducible evaluation, diagnostic, and error analysis of speaker diarization systems. In *INTERSPEECH*, pages 3587–3591, 2017.
- [11] John Brooke. Sus: a “quick and dirty” usability. *Usability evaluation in industry*, 189, 1996.
- [12] John-John Cabibihan, Hifza Javed, Marcelo Ang, and Sharifah Mariam Aljunied. Why robots? a survey on the roles and benefits of social robots in the therapy of children with autism. *International journal of social robotics*, 5(4):593–618, 2013.
- [13] Eldon Glen Caldwell Marín, Carlos Andrés Morales, Emilia Solis Sanchez, Miguel Cazorla, and Jose Maria Cañas Plaza. Designing a cyber-physical robotic platform to assist speech-language pathologists. *Assistive Technology*, (just-accepted), 2021.
- [14] Justine Cassell, Timothy Bickmore, Mark Billinghurst, Lee Campbell, Kenny Chang, Hannes Vilhjálmsón, and Hao Yan. Embodiment in conversational interfaces: Rea. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 520–527, 1999.
- [15] Fabio Catania, Micol Spitale, and Franca Garzotto. Conversational agents in therapeutic interventions for neurodevelopmental disorders: A survey. *ACM Comput. Surv.*, aug 2022. ISSN 0360-0300. doi: 10.1145/3564269. URL <https://doi.org/10.1145/3564269>. Just Accepted.
- [16] Thierry Chaminade, David Da Fonseca, Delphine Rosset, Ewald Lutchter, Gordon Cheng, and Christine Deruelle. Fmri study of young adults with autism interacting with a humanoid robot. In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, pages 380–385. IEEE, 2012.
- [17] Nancy Charron, Lundy Lewis, and Michael Craig. A robotic therapy case study: Developing joint attention skills with a student on the autism spectrum. *Journal of Educational Technology Systems*, 46(1):137–148, 2017.
- [18] Carlos A Cifuentes, Maria J Pinto, Nathalia Céspedes, and Marcela Múnera. Social robots in therapy and care. *Current Robotics Reports*, pages 1–16, 2020.
- [19] Caitlyn Clabaugh, Shomik Jain, Balasubramanian Thiagarajan, Zhonghao Shi, Leena Mathur, K Mahajan, Ragusa Gisele, and Matarić Maja J. Attentiveness of children with diverse needs to a socially assistive robot in the home. In *2018 International Symposium on Experimental Robotics*. University of Southern California Buenos Aires, 2018.
- [20] Caitlyn Clabaugh, Kartik Mahajan, Shomik Jain, Roxanna Pakkar, David Becerra, Zhonghao Shi, Eric Deng, Rhianna Lee, Gisele Ragusa, and Maja Matarić. Long-term personalization of an in-home socially assistive robot for children with autism spectrum disorders. *Frontiers in Robotics and AI*, page 110, 2019.
- [21] Mark B Colton, Daniel J Ricks, Michael A Goodrich, Behzad Dariush, Kikuo Fujimura, and Martin Fujiki. Toward therapist-in-the-loop assistive robotics for children with autism and specific language impairment. *autism*, 24:25, 2009.

- [22] Daniela Conti, Santo Di Nuovo, Serafino Buono, and Alessandro Di Nuovo. Robots in education and care of children with developmental disabilities: a study on acceptance by experienced and future professionals. *International Journal of Social Robotics*, 9(1): 51–62, 2017.
- [23] Daniela Conti, Allegra Cattani, Santo Di Nuovo, and Alessandro Di Nuovo. Are future psychologists willing to accept and use a humanoid robot in their practice? italian and english students' perspective. *Frontiers in psychology*, 10:2138, 2019.
- [24] Andreia P Costa, Georges Steffgen, Francisco Rodríguez Lera, Aida Nazarihorram, and Pouyan Ziafati. Socially assistive robots for teaching emotional abilities to children with autism spectrum disorder. In *3rd Workshop on Child-Robot Interaction at HRI*, 2017.
- [25] Cristina A Costescu, Bram Vanderborght, and Daniel O David. Reversal learning task in children with autism spectrum disorder: a robot-based approach. *Journal of autism and developmental disorders*, 45(11): 3715–3725, 2015.
- [26] Michael J Crawley. *The R book*. John Wiley & Sons, 2012.
- [27] Louise Demers, Rhoda Weiss-Lambrou, and Bernadette Ska. The quebec user evaluation of satisfaction with assistive technology (quest 2.0): an overview and recent progress. *Technology and Disability*, 14(3): 101–105, 2002.
- [28] Eric Deng, Bilge Mutlu, Maja J Mataric, et al. Embodiment in socially interactive robots. *Foundations and Trends in Robotics*, 7(4):251–356, 2019.
- [29] Laurie A Dickstein-Fischer, Darlene E Crone-Todd, Ian M Chapman, Ayesha T Fathima, and Gregory S Fischer. Socially assistive robots: current status and future prospects for autism interventions. *Innovation and Entrepreneurship in Health*, 5:15–25, 2018.
- [30] Stephanie Durrleman, H el ene Delage, Philippe Pr evost, and Laurice Tuller. The comprehension of passives in autism spectrum disorder. *Glossa*, 2(1):88, 2017.
- [31] Ver onica Egido-Garc a, David Est eviz, Ana Corrales-Paredes, Mar a-Jos e Terr on-L opez, and Paloma-Julia Velasco-Quintana. Integration of a social robot in a pedagogical and logopedic intervention with children: A case study. *Sensors*, 20(22):6483, 2020.
- [32] S. Short Elaine and Matari c Maja J. Understanding interaction dynamics in socially assistive robotics with children with asd. *International Meeting for Autism Research (IMFAR), Salt Lake City, Utah*, 2015.
- [33] David Est eviz, Mar a-Jos e Terr on-L opez, Paloma J Velasco-Quintana, Rosa-Mar a Rodr iguez-Jim enez, and Valle  lvarez-Manzano. A case study of a robot-assisted speech therapy for children with language disorders. *Sustainability*, 13(5):2771, 2021.
- [34] World Health Organization et al. *International Classification of Diseases, 11th Revision (ICD-11)*. Number 2018. Retrieved from <http://www.who.int/classifications/icd/en> (2018), 2018.
- [35] Special Eurobarometer. Attitudes towards the impact of digitisation and automation on daily life, 2017.
- [36] Nikolaos Fachantidis, Christine K Syriopoulou-Delli, and Maria Zygopoulou. The effectiveness of socially assistive robotics in children with autism spectrum disorder. *International Journal of Developmental Disabilities*, 66(2):113–121, 2020.
- [37] David Feil-Seifer and Maja J Mataric. Defining socially assistive robotics. In *9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005.*, pages 465–468. IEEE, 2005.
- [38] Andy Field. *Discovering statistics using IBM SPSS statistics*. sage, 2013.

- [39] Samantha Finkelstein, Amy Ogan, Caroline Vaughn, and Justine Cassell. Alex: A virtual peer that identifies student dialect. In *Proc. Workshop on Culturally-aware Technology Enhanced Learning in conjunction with EC-TEL*, 2013.
- [40] Lisa Furlong, Meg Morris, Tanya Serry, and Shane Erickson. Mobile apps for treatment of speech disorders in children: An evidence-based analysis of quality and efficacy. *PLoS One*, 13(8):e0201513, 2018.
- [41] Natasa Georgiou and George Spanoudis. Developmental language disorder and autism: Commonalities and differences on language. *Brain Sciences*, 11(5):589, 2021.
- [42] Maria Teresa Guasti. *Language acquisition: The growth of grammar*. MIT press, 2017.
- [43] Maria Teresa Guasti, Silvia Palma, Elisabetta Genovese, Paolo Stagi, Gabriella Saladini, and Fabrizio Arosio. The production of direct object clitics in pre-school- and primary school-aged children with specific language impairments. *Clinical linguistics & phonetics*, 30(9):663–678, 2016.
- [44] Marcel Heerink, Ben Kröse, Vanessa Evers, and Bob Wielinga. Assessing acceptance of assistive social agent technology by older adults: the almere model. *International journal of social robotics*, 2(4):361–375, 2010.
- [45] Rebecca Isbell, Joseph Sobol, Liane Lindauer, and April Lowrance. The effects of storytelling and story reading on the oral language complexity and story comprehension of young children. *Early childhood education journal*, 32(3):157–163, 2004.
- [46] Shomik Jain, Balasubramanian Thiagarajan, Zhonghao Shi, Caitlyn Clabaugh, and Maja J Matarić. Modeling engagement in long-term, in-home socially assistive robot interventions for children with autism spectrum disorders. *Science Robotics*, 5(39):eaaz3791, 2020.
- [47] Cory D Kidd and Cynthia Breazeal. Robots at home: Understanding long-term human-robot interaction. In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3230–3235. IEEE, 2008.
- [48] Richard A Krueger. *Focus groups: A practical guide for applied research*. Sage publications, 2014.
- [49] Hawon Lee and Eunja Hyun. The intelligent robot contents for children with speech-language disorder. *Journal of Educational Technology & Society*, 18(3):100–113, 2015.
- [50] Laurence B Leonard, Anita M-Y Wong, Patricia Deevy, Stephanie F Stokes, and Paul Fletcher. The production of passives by children with specific language impairment: Acquiring english or cantonese. *Applied Psycholinguistics*, 27(2):267–299, 2006.
- [51] Rosemarijn Looije, Mark A Neerincx, Johanna K Peters, and Olivier A Blanson Henkemans. Integrating robot support functions into varied activities at returning hospital visits. *International Journal of Social Robotics*, 8(4):483–497, 2016.
- [52] Tom Loucas, Tony Charman, Andrew Pickles, Emily Simonoff, Susie Chandler, David Meldrum, and Gillian Baird. Autistic symptomatology and language ability in autism spectrum disorder and specific language impairment. *Journal of Child Psychology and Psychiatry*, 49(11):1184–1192, 2008.
- [53] Ester Martinez-Martin, Felix Escalona, and Miguel Cazorla. Socially assistive robots for older adults and people with autism: An overview. *Electronics*, 9(2):367, 2020.
- [54] Maja J Matarić and Brian Scassellati. Socially assistive robotics. *Springer handbook of robotics*, pages 1973–1994, 2016.
- [55] Jauwairia Nasir, Barbara Bruno, Mohamed Chetouani, and Pierre Dillenbourg. What if social robots look for productive engagement? *International Journal of Social Robotics*, pages 1–17, 2021.

- [56] Anton J Nederhof. Methods of coping with social desirability bias: A review. *European journal of social psychology*, 15(3):263–280, 1985.
- [57] Corrado Pacelli, Tharushi Kinkini, Micol Spitale, Eleonora Beccaluva, Franca Garzotto, et al. “how would you communicate with a robot?”: People with neurodevelopmental disorder’s perspective. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 968–972. IEEE, 2022.
- [58] Philippe Prévost, Laurice Tuller, Racha Zebib, Marie Anne Barthez, Joëlle Malvy, and Frédérique Bonnet-Brilhault. Pragmatic versus structural difficulties in the production of pronominal clitics in french-speaking children with autism spectrum disorder. *Autism & Developmental Language Impairments*, 3: 2396941518799643, 2018.
- [59] Nick G Riches, Tom Loucas, Gillian Baird, Tony Charman, and Emily Simonoff. Sentence repetition in adolescents with specific language impairments and autism: An investigation of complex syntax. *International journal of language & communication disorders*, 45(1):47–60, 2010.
- [60] V Robles-Bykbaev, M Guamán-Heredia, Y Robles-Bykbaev, J Lojano-Redrován, F Pesántez-Avilés, D Quisi-Peralta, M López-Nores, and J Pazos-Arias. Onto-speltra: A robotic assistant based on ontologies and agglomerative clustering to support speech-language therapy for children with disabilities. In *Colombian Conference on Computing*, pages 343–357. Springer, 2017.
- [61] Vladimir Robles-Bykbaev, Mario Ochoa-Guaraca, Marco Carpio-Moreta, Daniel Pulla-Sánchez, Luis Serpa-Andrade, Martín López-Nores, and Jorge García-Duque. Robotic assistant for support in speech therapy for children with cerebral palsy. In *2016 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)*, pages 1–6. IEEE, 2016.
- [62] Vladimir E Robles-Bykbaev, Martin Lopez-Nores, Jose Juan Pazos-Arias, and Jorge Garcia-Duque. Ramses: a robotic assistant and a mobile support environment for speech and language therapy. In *Fifth International Conference on the Innovative Computing Technology (INTECH 2015)*, pages 1–4. IEEE, 2015.
- [63] Everett M Rogers, Arvind Singhal, and Margaret M Quinlan. *Diffusion of innovations*. Routledge, 2014.
- [64] Brian Scassellati, Henny Admoni, and Maja Matarić. Robots for use in autism research. *Annual review of biomedical engineering*, 14: 275–294, 2012.
- [65] Brian Scassellati, Laura Boccanfuso, Chien-Ming Huang, Marilena Mademtzi, Meiying Qin, Nicole Salomons, Pamela Ventola, and Frederick Shic. Improving social skills in children with asd using a long-term, in-home social robot. *Science Robotics*, 3(21), 2018.
- [66] Zhonghao Shi, Thomas R Groechel, Shomik Jain, Kourtney Chima, Ognjen Rudovic, and Maja J Matarić. Toward personalized affect-aware socially assistive robot tutors in long-term interventions for children with autism. *arXiv preprint arXiv:2101.10580*, 2021.
- [67] Jiro Shimaya, Yuichiro Yoshikawa, Yoshio Matsumoto, Hirokazu Kumazaki, Hiroshi Ishiguro, Masaru Mimura, and Masutomu Miyao. Advantages of indirect conversation via a desktop humanoid robot: Case study on daily life guidance for adolescents with autism spectrum disorders. In *2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN)*, pages 831–836. IEEE, 2016.
- [68] Elaine S Short, Eric C Deng, David J Feil-Seifer, and Maja J Mataric. Understanding agency in interactions between children with autism and socially assistive robots. 2017.
- [69] Candace L Sidner, Cory D Kidd, Christopher Lee, and Neal Lesh. Where to look: a study of human-robot engagement. In *Proceedings of the 9th international conference*

- on *Intelligent user interfaces*, pages 78–84, 2004.
- [70] S Silleresi, L Tuller, H Delage, S Durrelaman, F Bonnet-Brilhault, J Malvy, and P Prévosti. Sentence repetition and language impairment in french-speaking children with asd. *On the acquisition of the syntax of romance*, pages 235–258, 2018.
- [71] David Silvera-Tawil and Christine Roberts-Yates. Socially-assistive robots to enhance learning for secondary students with intellectual disabilities and autism. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 838–843. IEEE, 2018.
- [72] Ramona E Simut, Johan Vanderfaellie, Andreea Peca, Greet Van de Perre, and Bram Vanderborght. Children with autism spectrum disorders make a fruit salad with probio, the social robot: an interaction study. *Journal of autism and developmental disorders*, 46(1):113–126, 2016.
- [73] Henrique Siqueira, Alexander Sutherland, Pablo Barros, Mattias Kerzel, Sven Magg, and Stefan Wermter. Disambiguating affective stimulus associations for robot perception and dialogue. In *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*, pages 1–9. IEEE, 2018.
- [74] Wing-Chee So, Chun-Ho Cheng, Wan-Yi Lam, Ying Huang, Ka-Ching Ng, Hiu-Ching Tung, and Wing Wong. A robot-based play-drama intervention may improve the joint attention and functional play behaviors of chinese-speaking preschoolers with autism spectrum disorder: A pilot study. *J. Autism Dev. Disord.*, 50(2):467–481, February 2020.
- [75] Micol Spitale, Silvia Silleresi, Giulia Cosentino, Francesca Panzeri, and Franca Garzotto. Whom would you like to talk with? exploring conversational agents for children’s linguistic assessment. In *Proceedings of the Interaction Design and Children Conference*, pages 262–272, 2020.
- [76] Micol Spitale, Chris Birmingham, R Michael Swan, and Maja J Matarić. Composing harmony: An open-source tool for human and robot modular open interaction. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3322–3329. IEEE, 2021.
- [77] Micol Spitale, Silvia Silleresi, Giulia Leonardi, Fabrizio Arosio, Beatrice Giustolisi, Maria Teresa Guasti, and Franca Garzotto. Design patterns of technology-based therapeutic activities for children with language impairments: A psycholinguistic-driven approach. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–7, 2021.
- [78] Micol Spitale, Sarah Okamoto, Mahima Gupta, Hao Xi, and Maja J Matarić. Socially assistive robots as storytellers that elicit empathy. *ACM Transactions on Human-Robot Interaction*, 2022.
- [79] Alireza Taheri, Ali Meghdari, Minoo Alemi, and Hamidreza Pouretamad. Teaching music to children with autism: a social robotics challenge. *Scientia Iranica*, 26 (Special Issue on: Socio-Cognitive Engineering):40–58, 2019.
- [80] Adriana Tapus, Mataric Maja, and Brian Scassellatti. The grand challenges in socially assistive robotics. 2007.
- [81] David Williams, Nicola Botting, and Jill Boucher. Language in autism and specific language impairment: Where are the links? *Psychological bulletin*, 134(6):944, 2008.
- [82] Kacie Wittke, Ann M Mastergeorge, Sally Ozonoff, Sally J Rogers, and Letitia R Naigles. Grammatical language impairment in autism spectrum disorder: Exploring language phenotypes beyond standardized testing. *Frontiers in psychology*, 8:532, 2017.
- [83] Laura Zampini, Paola Zanchi, Chiara Suttora, Maria Spinelli, Mirco Fasolo, and Nicoletta Salerno. Assessing sequential reasoning skills in typically developing children. *BPA-Applied Psychology Bulletin (Bollettino di*

Psicologia Applicata), 65(279), 2017.

- [84] Ran Zhao, Tanmay Sinha, Alan W Black, and Justine Cassell. Socially-aware virtual agents: Automatically assessing dyadic rapport from temporal patterns of behavior. In *International conference on intelligent virtual agents*, pages 218–233. Springer, 2016.