Assessing the Children's Receptivity to the Robot MARKO

Jovica Tasevski, Milan Gnjatović, Branislav Borovac

Faculty of Technical Sciences, University of Novi Sad
Trg Dositeja Obradovića 6, 21101 Novi Sad, Serbia
e-mail: tasevski@uns.ac.rs, milangnjatovic@uns.ac.rs, borovac@uns.ac.rs

Abstract: This paper presents an experimental assessment of the children's receptivity to the human-like conversational robot MARKO. It reports on a production of a corpus that is composed of recordings of interaction between children, with cerebral palsy and similar movement disorders, and MARKO, in realistic therapeutic settings. Twenty-nine children participated in this study: 17 of them were recruited from among patients with cerebral palsy and similar movement disorders, and 12 healthy. Approximately 222 minutes of session time was recorded. All dialogues were transcribed, and nonverbal acts were annotated. A control group of 15 children (14 with cerebral palsy, one with spina bifida) was also included. The evaluation of the corpus showed that the positive effects go beyond social triggering – the children not only positively responded to MARKO, but also experienced increased motivation and engagement in therapy.

Keywords: child-robot interaction; robot-assisted therapy; robot MARKO; cerebral palsy, cognitive infocommunications

1   Introduction and Related Work

Although robot-assisted therapy for children with developmental disorders has drawn significant research attention, most research in this field is actually focused on children with autism [6, 7, 28, 34, 36]. It seems widely accepted that robots may induce positive social behavior in children with autistic spectrum disorders. However, these positive effects cannot be ad hoc generalized to children with other developmental disorders. In this paper, we report on an aspect of our research on robot-assisted therapy for children with cerebral palsy and similar movement disorders. The main features of cerebral palsy are abnormal gross and fine motor functioning and organization. They are often accompanied by other neurodevelopmental disorders or impairments, such as disturbances of sensation, perception, cognition, communication, etc. [29, pp. 8-10]. This target group of children has attracted less research attention in the field of robot-assisted therapy

(cf. [4,5,22]), despite the fact that cerebral palsy is considered to be the most common cause of serious physical disability in childhood [26, p. 3].

As a part of our previous work, we designed and developed the conversational human-like robot MARKO [11] (cf. Fig. 1) as an assistive tool for treatment of children with developmental disorders. The reported study particularly considers the possibility of applying MARKO for treatment of children with cerebral palsy and similar movement disorders. Its primary goal is to produce a corpus comprising recordings of interaction between children from the target group and the robot MARKO in realistic therapeutic settings, and to assess the children's receptivity to the robot, i.e., experimentally validate whether they positively respond to MARKO and engage in interaction with it.

The robot MARKO can perform selected therapy-relevant gross motor exercises, generate basic emotional facial expressions, and autonomously engage in natural language dialogue with the therapist [10-13, 25]. However, due to the sensitive nature of the research, in this study MARKO is strictly controlled by a trained human operator, and children activities during the experiments were monitored by a therapist.

2   The Experiment

The corpus of child-robot interaction was produced in the kinesitherapeutic room at the Clinic of Paediatric Rehabilitation in Novi Sad, Serbia.

2.1   Subjects

Twenty-nine children (13 female, 16 male, with an average age of 9.1, and a standard deviation of 3.54) participated in this study. Twelve children were healthy (7 female, 5 male, average age 6.75, st. dev. 2.45), and seventeen were recruited from among patients with cerebral palsy and similar movement disorders, admitted to the Clinic of Paediatric Rehabilitation in Novi Sad (6 female, 11 male, average age 10.76, st. dev. 3.27). The parents of all children were informed about the study and gave written permission for their children to participate. In addition, the children were also informed about the experimental settings – in an appropriate manner and to the extent to which they were capable of understanding – and each child above age five gave assent. The basic information on the children that participated in the study is given in Tables 1 and 2.

Table 1

Subjects recruited from among patients. All subjects from this group can comprehend speech, except subject s_{23} who can understand only simple verbal communications. Subjects s_{11}, s_{12}, s_{14}, s_{22}, s_{23} and s_{28} had encountered the robot MARKO previous to the experiment.

Subject ID	Age	Sex	Height [cm]	Weight [kg]	Diagnosis	Mobility
s_3	15	f	159	46	Morbus Alexander, disturbed walking pattern with increased muscle tension	can stand, can walk
s_4	12	m	167	69	spinal column deformity, scoliosis	can stand, can walk
s_5	12	m	165	41	spinal column deformity, scoliosis	can stand, can walk
s_6	13	f	153	55	paralysis cerebralis infantilis, paralysed right arm, partially paralyzed right leg	can stand, can walk
s_7	8	m	135	38	birth trauma nerve lesion in left arm, brachial plexus lesion	can stand, can walk, but uses arms for support
s_8	10	m	148	35	poor posture, scoliosis	can stand, can walk
s_{10}	11	m	142	41	birth trauma nerve lesion in left arm	can stand, can walk
s_{11}	6	f	120	33	paralysis cerebralis infantilis	can stand, can walk
s_{12}	9	f	129	27	paralysis cerebralis infantilis, vision problems	sitting, can stand, can walk a little with assistance
s_{13}	9	m	127	32	hemiparesis, left sided weakness (brain hemorrhage)	can stand, can walk
s_{14}	5	m	124	34	paralysis cerebralis infantilis	sitting, cannot stand or walk
s_{16}	13	m	173	75	car accident, left femur fracture, right clavicle fracture, comotio cerebralis	sitting, cannot stand or walk
s_{17}	13	m	160	51	paralysis cerebralis infantilis	sitting, can stand and walk
s_{22}	8	m	104	16	paralysis cerebralis infantilis, quadriparesis, difficulty speaking, epilepsy	sitting with assistance, stands and walks only with assistance
s_{23}	8	m	145	30	paralysis cerebralis infantilis, difficulty with speaking, difficulty with attention, sensorimotor integration disorder	sitting, cannot stand or walk
s_{28}	14	f	153	38	birth trauma nerve lesion in right arm, brachial plexus lesion	can stand, can walk
s_{29}	17	f	172	58	car accident polytrauma, basilar skull fracture, pelvic and pubic fractures, slow thought process	sitting, cannot stand or walk

Table 2

Healthy subjects. Only subject s_9 had encountered the robot MARKO previous to the experiment.

Subject ID	s_1	s_2	s_9	s_{15}	s_{18}	s_{19}	s_{20}	s_{21}	s_{24}	s_{25}	s_{26}	s_{27}
Age	7	6	10	10	9	5	3	5	3	9	7	7
Sex	f	m	M	m	m	f	f	f	f	m	f	f
Height [cm]	120	110	145	125	125	110	90	95	104	110	105	104
Weight [kg]	24	19	30	34	32	25	17	20	20	25	20	19

2.2   Experimental Settings

In the experimental settings, a child is sitting or standing in front of the robot MARKO, at a distance of approximately one meter from each other. A small and lightweight toy (i.e., a plush giraffe) was placed just beside the robot, and visible to the child. The operator that controls the robotic system is sitting at a distance of approximately one meter from the robot, and 2.5 meters from the child. The parent that is accompanying the child and the therapist that monitors the child activities are behind the child.

All experimental sessions were captured by two digital video cameras placed on the stands, recording slightly obliquely towards the child and the robot, respectively. Both cameras were angled to capture both the child and the robot. However, one of them was primarily recording the child, at a distance of approximately 3 meters from the child. The other camera was primarily recording the robot, at a distance of approximately 5.5 meters from the robot. Sample images captured by these cameras are displayed in Fig. 1.

Figure 1

Experimental settings: Captured images from cameras at the moment when the child places the toy on the robot's wooden horse

2.3   Session Structure and Dialogue Management

For each child, a separate experimental session was conducted. The interaction was primarily evolving between the child and the robot, but other human participants were also allowed to interfere when appropriate or necessary (e.g.,

when explicitly prompted by the child, or to encourage the child, etc.). The language used in the study was Serbian.

Each session consisted of three parts. The first part was devoted to engaging the child in interaction. E.g., the robot introduces itself to the child, asks the child what their name is, and how old they are. It generates happy facial expression and says that it is happy to meet the child. Finally, it proposes to tell the child a story.

In the second part, the robot confronts the child with a very simple discourse and tries to induce the child to perform nonverbal actions. E.g., the robot generates sad facial expression and says that it is sad because it has lost his favorite toy – a yellow giraffe. Then, it asks the child whether she or he could help find the toy. The robot asks for help gradually. First, it says that maybe its friend could help, and if the child does not respond, the robot asks her or him directly for help. When the child points to the toy or place the toy on the robot's wooden horse (depending on the mobility of the child), the robot commends the child, generates happy facial expression, and says that it is not sad anymore.

In the third part, the robot performs selected therapy-relevant nonverbal acts (e.g., raising its arm, pointing to its head or stomach, looking leftwards and rightwards, etc.) and asks the child to repeat them. In all three parts, the robot addresses the child by name, and encourages and commends the child to perform its instructions. When necessary, the operator and the parent also encourage or additionally instruct the child.

However, in the scope of the exchange between the child and the robot, we were particularly interested in the children's responses to the robot's instructions. To examine this aspect of interaction, we mapped the specific types of speech roles that the children and the robot could adopt during the interaction onto more fundamental type of speech roles [17, pp. 106-111]:

- *command*, i.e., the robot demands from the child to perform a nonverbal act,

- *question*, i.e., the robot demands from the child to provide a verbal response,

- *statement*, i.e., the robot provide a verbal information,

- *offer*, i.e., the robot is performing a nonverbal act.

The child may respond in the following ways:

- *correct response*, i.e., the child performs the demanded nonverbal act, or provides the demanded information,

- *partial response*, i.e., the child understands the robot's command or question, but can only partially perform the demanded nonverbal act or provide the demanded information.

- *incorrect response*, i.e., the child misunderstands the robot's command or question, and provides an incorrect response,

- *no response*, i.e., the child does not respond because she or he does not understand the demand, does not want to respond, etc.

The operator was instructed to follow a preset dialogue strategy during the interaction. The robot works through a given sequence of therapeutic commands and questions. It instructs commands (*R-comm1*) or poses questions (*R-ques1*) on a one-by-one basis. If the child responds correctly (*C-corr*) or partially (*C-part*), the robot commends the child (*R-stat*), with some accompanying nonverbal action (*R-off*) when appropriate. Then, the robot proceeds with the next command of question. Otherwise, if the child does not respond (*C-nor*), the robot repeats the command (*R-comm2*) or the question (*R-ques2*). And, if the child responds incorrectly (*C-inc*), the robots reformulates the command (*R-comm3*) or the question (*R-ques3*), simplifying the formulation of its demand.

Table 3
Set Φ : the introduced general classification of dialogue acts in child-robot interaction

	Act-ID	Meaning
Robot	*R-comm1*	instructing a command for the first time
	R-comm2	repeating a command on which the child did not respond
	R-comm3	reformulating a command on which the child provided an incorrect response
	R-comm4	reformulating a command on which the child has previously provided a correct or partial response
	R-ques1	posing a question for the first time
	R-ques2	repeating a question on which the child did not respond
	R-ques3	reformulating a question on which the child provided an incorrect response
	R-ques4	reformulating a question on which the child has previously provided a correct or partial response
	R-stat	making a statement
	R-off	performing a nonverbal action
Child	*C-corr*	correct response
	C-part	partial response
	C-inc	incorrect response
	C-nor	no response

According to this dialogue strategy, the robot may any time repeat or reformulate a command (*R-comm4*) or a question (*R-ques4*) on which the child has previously provided a correct or a partial response, in order to additionally stimulate the child. However, the multiple successive repetition of a demand to which the child cannot provide a correct or a partial response may induce negative emotional states in the child. To prevent this, if the child responds incorrectly or does not respond, the robot repeats or reformulates the current command (*R-comm2 or R-comm3*) or question (*R-ques2 or R-ques3*) only once. The set Φ of all dialogue acts in this general classification is given in Table 3.

2.4   Corpus

In the reported experiment, 36 sessions were recorded, with a total duration of approximately 222 minutes. The basic information on the number and duration of the sessions is given in Table 4. All dialogues were transcribed, and nonverbal acts were annotated. The verbal dialogue act statistics are given in Tables 5 and 6. The nonverbal act statistics are given in Tables 7 and 8.

Table 4

Basic information on the number and duration of the sessions

	Healthy	Patients	Total
Number of sessions	12	24	36
Total duration (approx.)	69 min	153 min	222 min
Average duration	5 min 46 sec	6 min 9 sec	6 min 2 sec
Standard deviation	56 sec	2 min 16 sec	1min 56sec

Table 5

Verbal dialogue act statistics for the robot, operator and parents

		Interaction with healthy children	Interaction with patients	Total
Robot	Number of verbal dialogue acts	599	1172	1771
	Average number of words per act	8.69	8.66	8.67
	Standard deviation	5.43	5.5	5.48
Oper.	Number of verbal dialogue acts	24	65	89
	Average number of words per act	3	2.97	2.98
	Standard deviation	1.98	2.59	2.43
Parent	Number of verbal dialogue acts	15	73	88
	Average number of words per act	2.4	3.58	3.38
	Standard deviation	1.55	1.98	1.96

Table 6

Verbal dialogue act statistics for the children

	Healthy children	Patients	Total
Number of verbal dialogue acts	240	659	899
Average number of words per act	1.54	2.11	1.96
Standard deviation	1.52	2.23	2.08

Table 7
Children's nonverbal acts

Nonverbal act	Number of occurrences		Total
	Healthy	Patients	
looking at the operator	30	17	47
pointing to the toy or giving the toy	14	44	58
raising arm	38	88	126
pointing to her/his head	25	65	90
pointing to her/his stomach	13	55	68
looking leftwards	7	12	19
looking rightwards	4	12	16
nodding her/his head to express approval	4	12	16
shaking her/his head to express disapproval	22	30	52
searching for the toy	2	6	8
shrugging shoulders	0	2	2
applauding	0	4	4
pointing to herself/himself	0	1	1
using fingers to display number	0	2	2
looking at the parent	0	5	5
waving	5	3	8
Total	164	358	522

Table 8
Robot's nonverbal acts

Nonverbal act	Number of occurrences		Total
	Interaction with healthy children	Interaction with patients	
awakening (opening eyes)	0	1	1
pointing to its head	34	66	100
generating happy facial expression	16	3	19
looking upwards	0	6	6
pointing to its stomach	23	48	71
raising arm	47	93	140
generating sad facial expression	13	20	33
looking leftwards	9	20	29
looking rightwards	3	11	14
looking at its wrist-watch	0	3	3
Total	145	271	416

3   Evaluation

The corpus is evaluated with respect to the robot's dialogue behavior, the children's verbal production, and the children's motivation to undergo the therapy.

3.1   Evaluating the Robot's Dialogue Behavior

The first aspect of the evaluation relates to the robot's dialogue behavior, i.e., it is aimed at showing that (i) the operator consistently followed the preset dialogue strategy (as introduced in Section 2.3) across the experimental sessions, and that (ii) the applied strategy did not restrict the expressive conduct of the children. In order to evaluate this, we introduce an approach to profiling child-robot dialogues based on dialogue acts n-grams.

The point of departure for our approach to dialogue profiling is that a dialogue structure is not given beforehand, but rather evolves as the dialogue proceeds [16, 31]. This is also true in the case when the underlying dialogue domain is rather simple and a priori given (e.g., a specific therapeutic session), and one dialogue participant (e.g., a therapist) has an elaborated plan for the dialogue. Even then, the specific intentions, foci of attention, and linguistic constructions of the other dialogue participant (e.g., a child) significantly influence the dialogue structure. Therefore, in our approach we do not attempt to determine the child-robot dialogue structure or infer its constitutive rules. We rather try to profile a dialogue unfolding between two parties, one of which follows a preset dialogue strategy.

At the surface level (and only at this level) a dialogue may be considered as a sequence of dialogue acts. Let $\Phi=\{d_1,d_2,...,d_n\}$ be a set of possible dialogue acts that may occur in a given dialogue domain. Then, a dialogue instance can be represented as a sequence $D_i=d_{i1},d_{i2},...,d_{ik}$, where $(\forall 1 \leq j \leq k)\,(d_{ij} \in \Phi)$. In our approach, we represent a dialogue as if it were a *bag of dialogue act trigrams*, i.e., a set of dialogue act trigrams that occur in a given dialogue instance. For example, a dialogue act sequence $d_1 d_2 d_1 d_2 d_1 d_2 d_1 d_2 d_1$ is represented as a set of distinct trigrams $\{d_1 d_2 d_1, d_2 d_1 d_2\}$. This set is unordered because the position and frequency of the contained trigrams in a given dialogue are ignored.

In Section 2.3 (cf. Table 3), we have already defined a general set of possible dialogue acts that are relevant for the given domain. In addition, the choice of the n-gram size (i.e., the order of the language model) was also not arbitrary, but in accordance with the introduced dialogue strategy. We recall that if the child responds incorrectly or does not respond at all to the robot's command or question, the robot repeats or reformulates the current instruction only once. Due to this constraint, we opted for trigrams when deciding on the n-gram size. In other words, dialogue act trigrams provide enough context to capture the class of the children's responses.

At the implementation level, a dialogue profile may be represented as a binary vector of size $\left|\Phi^3\right|$, where Φ is set of dialogue acts that may occur in a given dialogue domain. Each element of this vector is bijectively assigned to a trigram from Φ^3, and represents the weight of the trigram: 1 if the assigned trigram occurs in a given dialogue, and 0 otherwise. This conceptualization allows for applying distance metrics for binary data.

To evaluate the similarity between two dialogue profiles D_1 and D_2, we apply two similarity measures [8, pp. 299, 304]:

- the Rand similarity on $\{0,1\}^n$ (i.e., the Sokal-Michener's simple matching), to evaluate *the similarity of dialogue strategies applied by the robot*:

$$R(D_1, D_2) = 1 - \frac{\left|D_1 \Delta D_2\right|}{n},$$

- the Jaccard similarity of community on $\{0,1\}^n$ (also called Tanimoto similarity), to evaluate *the similarity of the children's interaction styles*, under the assumption that the robot applies the same dialogue strategy in both profiles:

$$J(D_1, D_2) = \frac{\left|D_1 \cap D_2\right|}{\left|D_1 \cup D_2\right|} = 1 - \frac{\left|D_1 \Delta D_2\right|}{\left|D_1 \cup D_2\right|},$$

where n is the number of possible trigrams, and $D_1 \Delta D_2$ represents the symmetric difference of sets D_1 and D_2, i.e., $(D_1 \setminus D_2) \cup (D_2 \setminus D_1)$. Table 3 describes fourteen dialogue acts that may occur in the observed dialogue domain, ten of which are the robot's dialogue acts, and the rest are the children's dialogue acts. In principle, the parameter n is equal to $\left|\Phi^3\right| = 14^3 = 2744$, but the adopted dialogue act classification (i.e., we also annotate the interaction situation when the child does not respond) allows only trigrams of the following forms: robot-child-robot, robot-robot-child, child-robot-robot, child-robot-child. This restricts the number of possible dialogue act trigrams to: $n = 3 \cdot (10 \cdot 4 \cdot 10) + 4 \cdot 10 \cdot 4 = 1360$.

In both formulas, the similarity is defined as one minus the corresponding normalized Hamming distance between two profiles, where the distance is conceptualized as proportionate to the symmetric difference of sets D_1 and D_2. However, the above formulas differ in their denominators. When the operator consistently follows the dialogue strategy, its dialogue acts are determined by the immediately preceding child's response and the preset sequence of therapeutic commands and questions. In other words, the number of dialogue act trigrams that are expected to occur in this case is significantly smaller than n. Therefore, to

evaluate the similarity of dialogue strategies applied by the robot across the experimental sessions, the denominator in the formula for the Rand similarity is set to n, i.e., the number of trigrams contained in the symmetric difference of sets D_1 and D_2 is normalized with respect to the number of all possible trigrams that may occur in the observed domain.

In contrast to this, the denominator in the formula for the Jaccard similarity is $\left| D_1 \cup D_2 \right|$, i.e., the number of trigrams contained in the symmetric difference of sets D_1 and D_2 is normalized with respect to the number of different trigrams that actually occur in the compared profiles. Thus, if the robot applies the same dialogue strategies in both dialogues, this measure evaluates the similarity of the children's interaction style.

We annotated the corpus with respect to the general categorization of dialogue acts given in Table 3, and created a profile for each of 36 sessions. The average number of trigrams occurring in the profiles is 23.26, with the standard deviation of 7.98. The minimum number of trigrams in a profile is 6, while the maximum number of trigrams in a profile is 44. The number of all different dialogue act trigrams that occur in the profiles is 127, which is significantly smaller than the number of possible trigrams n.

Using the Rand and Jaccard similarity measures, we compared each of $\binom{36}{2} = 630$ possible pairs of dialogue profiles contained in the corpus. We recall that the values of the Rand and Jaccard similarity coefficients range from 0 to 1. The closer the value is to 1, the more similar are the profiles. The average Rand similarity between dialogue profiles is 0.9855, with the standard deviation of 0.0057. The Rand similarity of the least similar dialogue profiles is 0.9654, while the Rand similarity of the most similar dialogue profiles is 0.9978. The high values of the Rand similarity imply that the operator has consistently applied the introduced dialogue strategy.

In contrast to this, the average Jaccard similarity between dialogue profiles is 0.4197, with the standard deviation of 0.0640. The Jaccard similarity of the least similar dialogue profiles is 0.1, while the Jaccard similarity of the most similar dialogue profiles is 0.8571. This wide range of values of the Jaccard similarity coefficients implies that, although the operator consistently applied the dialogue strategy, the children's interaction styles vary across the experimental sessions. In other words, the robot's dialogue strategy did not restrict the expressive conduct of the children.

3.2   Evaluating the Children's Verbal Production

The evaluation of the verbal production was primarily focused to patients. To perform the evaluation, we included a control group of children, i.e., we used a corpus of recordings that the therapists from the Clinic of Paediatric Rehabilitation in Novi Sad selected as typical examples of therapeutic exercises (i.e., without the robot) for children with cerebral palsy. Fifteen children (6 female, 9 male, with an average age of 6.8, and a standard deviation of 3.19) participated in these exercises. All of them were recruited from among patients admitted to the Clinic of Paediatric Rehabilitation in Novi Sad. The basic information on the children from the control group is given in Table 9. The control corpus contains 20 recordings, with a total duration of approximately 10 minutes and 20 seconds. The average duration of a recording is 31 seconds, with the standard deviation of 17 seconds. All recordings were transcribed and annotated. The verbal dialogue act statistics is given in Tables 10.

Table 9
The control group of patients

Subject ID	Sex	Age	Diagnosis	Mobility
o_1	m	6	paralysis cerebralis infantilis	can stand, can walk
o_2	f	6	paralysis cerebralis infantilis	can stand, can walk
o_3	m	7	paralysis cerebralis infantilis	can stand, can walk
o_4	f	6	paralysis cerebralis infantilis	can stand, can walk
o_5	m	10	paralysis cerebralis infantilis	can stand, can walk
o_6	f	12	paralysis cerebralis infantilis	can stand, can walk
o_7	f	12	spina bifida	can stand, can walk
o_8	m	8	paralysis cerebralis infantilis	can stand, can walk
o_9	f	5	paralysis cerebralis infantilis	can stand, can walk
o_{10}	m	10	paralysis cerebralis infantilis	can stand, can walk
o_{11}	f	7	paralysis cerebralis infantilis	can stand, can walk
o_{12}	m	3	paralysis cerebralis infantilis	can stand, can walk
o_{13}	m	2	paralysis cerebralis infantilis	can stand, can walk
o_{14}	m	2	paralysis cerebralis infantilis	can stand, can walk
o_{15}	m	6	paralysis cerebralis infantilis	can stand, can walk

We compared the verbal production of the experimental group of patients that were interacting with the robot, described in Table 1, and the control group of patients that were not confronted with the robot during exercises, described in Table 9.

It can be observed that the children in the experimental group were more engaged in the interaction. The average number of words per dialogue act for children in the experimental group is 1.52 times greater than the average number of words per dialogue act for children in the control patient group (i.e., 2.11/1.39, cf. Tables 6

and 10). In addition, the children in the experimental group produced in average 2.97 (i.e., 659/222, cf. Table 6) dialogue acts per minute, while the children in the control group produced 1.74 (i.e. 18/10.33, cf. Table 10) dialogue acts per minute. These differences become even more important if we keep in mind that the children from the target group (i.e., cerebral palsy and similar movement disorders) often suffer from impairments in communication, and thus are not verbose.

Table 10

Verbal dialogue act statistics for the children in the control group, therapist and parents

Children (Patients)	Number of verbal dialogue acts	18
	Average number of words per act	1.39
	Standard deviation	0.70
Therapist	Number of verbal dialogue acts	17
	Average number of words per act	3
	Standard deviation	2.81
Parent	Number of verbal dialogue acts	31
	Average number of words per act	3.06
	Standard deviation	2.21

However, what is also important is that the level of parent engagement in the interaction is significantly smaller for the experimental group. In the experimental group, the number of the children's verbal dialogue acts is one order of magnitude greater than the numbers of parents' and the operator's dialogue acts. In the control group, the number of parent's verbal dialogue acts is 1.72 times greater than the number of the children's verbal dialogue acts. This indicates that the children in the experimental group needed significantly less parental support in order to engage in the interaction.

3.3 Evaluating the Children's Motivation

A preliminary qualitative insight into the corpus showed that the children reacted positively to the robot, and that the robot was a strong motivational factor. The increased motivation in the children was observed by their respective long-term therapists. As a small illustration, we give several examples. In the points given below, the term "normal behavior" refers to the behavior of the children under normal therapeutic conditions (i.e., without the robotic system),

- Child s_4 normally avoids using his right arm in therapeutic exercises, but during interaction with MARKO, he used his right arm to point to his head.

- During the experimental session, child s_7 soon got tired and the intensity and amplitude of his movements decreased with time. Finally, the child

gave up on performing the robot's instruction. However, when MARKO insisted, the child performed the given instruction.

- Child s_{11} was aware where the plush giraffe is, but pretended that she was still searching for it, in order to prolong the interaction.

- Child s_{12} was affectively attached to MARKO, asked it whether she walks properly, and insisted that MARKO should confirm their friendship. In this child, the robot has induced increased motivation to undergo the therapy.

- Child s_{14} is normally not motivated to undergo the therapy, while child s_{17} is only occasionally motivated. During interaction with MARKO, significantly increased motivation is observed in both children.

However, to evaluate the children's motivation to undergo the therapy in a more systematic manner, we engaged a group of five evaluators (healthy, native Serbian speakers; 2 female, ages 28, 30; and 3 male, age 27, 30, 32; one of them had educational background in psychology). They were allowed to see and hear the recordings from the corpus, and were asked to annotate, separately from each other, the children's emotional state with respect to their motivation to undergo the therapy. The set of annotation labels was predefined: *positive* (i.e., the child is motivated to undergo the therapy), *negative* (i.e., the child is not interested in the therapy or has an aversion to the therapy or the environment), and *neutral*.

Table 11
Evaluation results

	Positive	Negative	Neutral	No-majority-rating
Total agreement	49.3036%	1.1142%	1.3928%	48.1894%
Strong majority	66.8524%	3.3426%	6.1281%	23.6769%
Weak majority	77.3481%	4.1436%	16.2983%	2.2099%

One of the experimental sessions contained in the corpus was not suitable for evaluation (due to the very serious condition of the child, his parent was holding him throughout the session, occasionally blocking the camera from seeing the child). The rest 35 sessions were divided into slots of 30 seconds each. These slots represented evaluation units, i.e., the total number of evaluation units was 359. We used majority rating to attribute labels to the evaluation units. We differentiate among three types of majority rating (cf. [14]):

- weak majority agreement: at least three evaluators agreed

- strong majority agreement: at least four evaluators agreed,

- total agreement: all five evaluators agreed.

Table 11 shows the percentage of the evaluation units with majority rating. These results are in favor of the conclusion that the robot was a strong motivational

factor. For each type of majority ratings, the positively annotated evaluation units represent the most dominant class, while the class of negatively annotated units is significantly smaller.

4   Discussion

In this section, we discuss how this contribution fits into the field of cognitive infocommunications [1-3, 30]. One important aspect of this field is devoted to improving the well-being of people [35], including assistive technologies for people with non-standard cognitive characteristics [18]. Recent research in this subfield of cognitive infocommunications relates to socially assistive robots with interactive behavioral capability [24, 27], games for cognitive competence of children with learning difficulties [33], motion detection sensors-based exercise games for elderly people [19], and ambient assisted living services for elderly with cognitive impairments [21].

In line with the concept of the generation of cognitive entities (cf. [1, 2, pp. 20-1]), the research reported in this paper represents a specific combo of natural cognitive capability of human and the information and communication technologies, aimed at overcoming the problem of lack of collective intentional behavior in the conventional therapy for children with cerebral palsy.

4.1   Problem: The Lack of Collective Intentional Behavior

The conventional therapy for children with cerebral palsy (i.e., therapy without a robot) is fundamentally based on two-party interaction between the child and the therapist. One of the most important problems of the conventional therapy relates to poor motivation of the child to undergo therapeutic exercises. We recognize that the reason for poor motivation lies in the fact that the child and the therapist do not (and often cannot not, due to the health condition of the child) perform collective intentional behavior during therapy, but rather individual intentional behavior (the notions of collective and individual intentions in interaction are discussed in [32]). The intentions of the therapist are to motivate the child to perform therapy-relevant exercises that are targeted at the parts of the body which are affected by cerebral palsy. However, the child cannot recognize the therapist's intentions or the long-term goal of the therapy. From the child's point of view, the requested actions are often perceived as boring (e.g., if the child has to repeat them many times) or unpleasant (e.g., if the child has to move an affected limb), and lacking a clear or playful goal. Thus, children are not really motivated to participate in therapy.

4.2   Solution: Inter-Cognitive, Representation-Bridging Communication

In this paper, we proposed three-party interaction between the child, the therapist, and the robot MARKO, which is designed to overcome the problem of lack of collective intentional behavior. The visual appearance of the robot, its apparent physical autonomy, and its ability to engage in interaction convey an impression to the child that MARKO has its own intentionality. This impression is a crucial interaction catalyst enabling the child to emotionally attach to the robot and engage in interaction, as shown in Section 3. Following [15], we differentiate between goal intentions and implementation intentions in the observed therapeutic context. Goal intentions specify a desired end point of motivating the child to undergo long-term therapeutic exercises, whereas implementation intentions are subordinate to goal intentions and specify a particular plan of engaging the child in interaction in order to induce goal-directed responses (although the child does not necessarily recognize the overall goal). After establishing an affective relation of the child toward MARKO, the robot dialogue behavior is applied to translate the goal intentions of the therapist into the implementation intentions of the robot that are perceived by the child.

In terms of cognitive infocommunications, MARKO is a cognitive technical agent that mediates inter-cognitive, representation-bridging communication directed from the therapist to the child. The communication is inter-cognitive to the extent that information transfer occurs between two humans with different cognitive capabilities, as the motor disorders of cerebral palsy are often accompanied by disturbances of sensation, perception, cognition, communication, and behavior [29]. The communication is representation-bridging to the extent that different representations of intentions are used on the two ends of communication.

```
if (correct or partial response)
  commend the child and go to the next instruction;
else if (no response)
  repeat the instruction;
else if (incorrect response)
  reformulate the instruction;
else
  go to the next instruction;
```

Figure 2
Simplified dialogue strategy

From the therapist's point of view, the goal intentions are formally represented as a dialogue strategy conceptualized as a sequence of if-else statements. The conceptualization and implementation of the robot's dialogue strategies are discussed in [11] in more details. For the purpose of illustration, a simplified

version of the dialogue strategy introduced in Section 2.3 is given in Fig. 2. It is important to note that this dialogue strategy is defined in a general manner, i.e., independent of a particular therapeutic exercise. However, when it is applied in a given exercise, it generates exercise-dependent dialogue behavior of the robot that reflects implementation intentions in a form accessible to children from the target group (cf. Table 12).

Table 12

Dialogue fragment between the child and the robot MARKO

Participant	Verbal act	Nonverbal act	Description
MARKO:	[*Name of the child*] yellow giraffe is my favorite toy. Do you know where my toy is?	-	Request
Child:	No.	Shrugs shoulders	Incorrect response
MARKO:	[*Name of the child*] did you maybe see my toy? It was somewhere here, but now I can't see it.	-	Reformulation
Child:	It is down there.	Points to the toy	Correct response
MARKO:	Great. It is just great.	-	Commending

Conclusions

The robot MARKO can autonomously engage in natural language interaction with users ([10-13, 25]). We believe that this technical ability is essential for establishing a long-term attachment of children to the robotic system, which in turn has an important role in facilitating human-machine coexistence and cognitive infocommunications in the robot-assisted therapeutic setting (cf. [2, 3, 20, 23, 30]). Corpora of children-robot interaction have an important role in this field because they are fundamental (if not the only) empirical foundation for validation of this requirement for therapeutic social robots.

In this paper, we reported on a production of a corpus that is comprised of recordings of interaction between children with cerebral palsy (and similar movement disorders) and the robot MARKO, in realistic therapeutic settings. In this study, due to the sensitive nature of the research, the robot MARKO was controlled by a human operator who consistently applied a preset therapeutic dialogue strategy. It was shown that, at the first (and second) encounter with the robot MARKO, the children positively responded to it. It is important to note that these positive effects go beyond social triggering. During the interaction, the MARKO was demanding from the children to perform selected therapy-relevant nonverbal acts, and the evaluation showed that the children experienced increased motivation and engagement in therapy.

Note: This paper is a significantly extended version of the paper [9]. Anonymized annotation data are available from the authors for research purposes on request.

Acknowledgement

References

[1]   P. Baranyi, A. Csapo: Revisiting the Concept of Generation CE – Generation of Cognitive Entities, 6th IEEE International Conference on Cognitive Infocommunications, Gyor, Hungary, pp. 583-586, 2015

[2]   P. Baranyi, A. Csapo, G. Sallai: Cognitive Infocommunications (CogInfoCom), Springer International Publishing, 2015

[3]   P. Baranyi, A. Csapo: Definition and Synergies of Cognitive Infocommunications, Acta Polytechnica Hungarica, 9(1) pp. 67-83, 2012

[4]   M. Belokopytov, M. Fridin: Motivation of Children with Cerebral Palsy during Motor Involvement by RAC-CP Fun, Proc. of the Workshop on Motivational Aspects of Robotics in Physical Therapy, IEEE/RSJ International Conference on Intelligent Robots and Systems, Vilamoura, Algarve, Portugal, 6 pages, no pagination, 2012

[5]   M. P. Blázquez: Clinical Application of Robotics in Children with Cerebral Palsy, In: J. L. Pons, D. Torricelli, M. Pajaro (eds), Converging Clinical and Engineering Research on Neurorehabilitation. Biosystems & Biorobotics 1, Springer Berlin Heidelberg, pp. 1097-1102, 2013

[6]   M. B. Colton, D. J. Ricks, M. A. Goodrich, B. Dariush, K. Fujimura, M. Fujiki: Toward therapist-in-the-loop assistive robotics for children with autism and specific language impairment, Proc. of the AISB 2009 Symposium on New Frontiers in Human-Robot Interaction. Edinburgh, Scotland, 5 pages, no pagination, 2009

[7]   K. Dautenhahn, C. L. Nehaniv, M. L. Walters, B. Robins, H. Kose-Bagci, N. A. Mirza, M. Blow: KASPAR – A Minimally Expressive Humanoid Robot for Human-Robot Interaction Research, Applied Bionics and Biomechanics 6(3):369-397, 2009

[8]   M. M. Deza, E. Deza: Encyclopedia of Distances, Springer-Verlag Berlin Heidelberg, 2009

[9]   M. Gnjatović, J. Tasevski, D. Mišković, S. Savić, B. Borovac, A. Mikov, R. Krasnik: Pilot Corpus of Child-Robot Interaction in Therapeutic Settings, 8th IEEE International Conference on Cognitive Infocommunications, Debrecen, Hungary, pp. 253-7, 2017

[10] M. Gnjatović: Changing Concepts of Machine Dialogue Management, 5th IEEE Conference on Cognitive Infocommunications, Vietri sul Mare, Italy, 6 pages, no pagination, 2014

[11]   M. Gnjatović: Therapist-Centered Design of a Robot's Dialogue Behavior, Cognitive Computation, 6(4):775-788, 2014

[12]   M. Gnjatović, V. Delić: Cognitively-inspired Representational Approach to Meaning in Machine Dialogue, Knowledge-Based Systems, 71:25-33, 2014

[13]   M. Gnjatović, M. Janev, V. Delić: Focus Tree: Modeling Attentional Information in Task-Oriented Human-Machine Interaction, Applied Intelligence, 37(3):305-320, 2012

[14]   M. Gnjatović, D. Rösner: Inducing Genuine Emotions in Simulated Speech-Based Human-Machine Interaction: The NIMITEK Corpus, IEEE Transactions on Affective Computing, 1(2) pp. 132-144, 2010

[15]   P. M. Gollwitzer: Implementation Intentions: Strong Effects of Simple Plans, The American Psychologist, 54(7), pp. 493-503, 1999

[16]   B. Grosz, C. Sidner: Attention, Intentions, and the Structure of Discourse, Computational Linguistics, 12(3), pp. 175-204, 1986

[17]   M. A. K. Halliday, C. M. I. M. Matthiessen: An Introduction to Functional Grammar, third edition, Hodder Arnold, 2004

[18]   L. Izsó: The Significance of Cognitive Infocommunications in Developing Assistive Technologies for People with Non-Standard Cognitive Characteristics: CogInfoCom for People with Nonstandard Cognitive Characteristics, 6th IEEE International Conference on Cognitive Infocommunications, Győr, Hungary, pp.77-82, 2015

[19]   N. Katajapuu, M. Luimula, A. Pyae, Y. L. Theng, T. P. Pham, J. Li, K. Sato: Benefits of Exergame Exercise on Physical Functioning of Elderly People, 8th IEEE International Conference on Cognitive Infocommunications, Debrecen, Hungary, pp. 85-90, 2017

[20]   J. Kinugawa, Y. Sugahara, K. Kosuge: Co-Worker Robot - "PaDY", Acta Polytechnica Hungarica, 13(1), pp. 209-221, 2016

[21]   A. Konstadinidou, N. Kaklanis, I. Paliokas, D. Tzovaras: A Unified Cloud-based Framework for AAL Services Provision to Elderly with Cognitive Impairments, 7th IEEE International Conference on Cognitive Infocommunications, Wroclaw, pp. 145-150, 2016

[22]   H. I. Krebs, B. Ladenheim, C. Hippolyte, L. Monterroso, J. Mast: Robot-assisted Task-Specific Training in Cerebral Palsy, Developmental Medicine & Child Neurology, 51(4):140-145, 2009

[23]   G. Kronreif: Mechatronic Assistance for Surgical Applications, Acta Polytechnica Hungarica, 13(1), pp. 31-42 2016

[24]   B. Lewandowska-Tomaszczyk, P. Wilson: Compassion, Empathy and Sympathy Expression Features in Affective Robotics, 7th IEEE International Conference on Cognitive Infocommunications, Wroclaw, pp. 65-70, 2016

[25] D. Mišković, M. Gnjatović, P. Štrbac, B. Trenkić, N. Jakovljević, V. Delić: Hybrid Methodological Approach to Context-Dependent Speech Recognition, International Journal of Advanced Robotic Systems, 14(1), 12 pages, 2017

[26] C. Morris: Definition and Classification of Cerebral Palsy: a Historical Perspective, Dev Med Child Neurol Suppl, 109:3-7, 2007

[27] D. C. Nguyen, G. Bailly, F. Elisei: Conducting Neuropsychological Tests with a Humanoid Robot: Design and Evaluation, 7th IEEE International Conference on Cognitive Infocommunications, Wroclaw, pp. 337-42, 2016

[28] D. J. Ricks, M. B. Colton: Trends and Considerations in Robot-assisted Autism Therapy, Proc. of the IEEE International Conference on Robotics and Automation (ICRA) pp. 4354-4359, Anchorage, AK 2010

[29] P. Rosenbaum, N. Paneth, A. Leviton, M. Goldstein, M. Bax, D. Damiano, B. Dan, B. Jacobsson: A Report: the Definition and Classification of Cerebral Palsy April 2006, Dev Med Child Neurol Suppl, 109:8-14, 2007

[30] Gy. Sallai: The Cradle of Cognitive Infocommunications, Acta Polytechnica Hungarica, 9(1) pp. 171-181, 2012

[31] J. Searle: Conversation, In J.L. Mey, H. Parret, J. Verschueren (eds) (On) Searle on conversation. John Benjamins, Philadelphia, pp. 7-29, 1992

[32] J. Searle: Collective Intentions and Actions, in P. R. Cohen, J. Morgan, M. Pollack (eds), Intentions in Communication, MIT Press, pp. 401-415, 1990

[33] C. Sik-Lanyi, V. Szucs, T. Guzsvinecz, S. Shirmohammadi, B. Abersek, K. Van Isacker, A. Lazarov: How to Develop Serious Games for Cognitive Competence of Children with Learning Difficulties, 8th IEEE International Conference on Cognitive Infocommunications, Debrecen, Hungary, pp. 321-6, 2017

[34] S. Thill, C. A. Pop, T. Belpaeme, T. Ziemke, B. Vanderborght: Robot-assisted Therapy for Autism Spectrum Disorders with (partially) Autonomous Control: Challenges and Outlook, Paladyn. Journal of Behavioral Robotics, 3(4):209-217, SP Versita 2012

[35] A. Vagner: Cognitive Infocommunication for Monitoring and Improving Well-being of People, 8th IEEE International Conference on Cognitive Infocommunications, Debrecen, Hungary, pp. 103-7, 2017

[36] B. Vanderborght, R. Simut, J. Saldien, C. Pop, A. S. Rusu, S. Pintea, D. Lefeber, D. O. David: Using the Social Robot Probo as a Social Story Telling Agent for Children with ASD, Interaction Studies, 13(3):348-372, 2012