Towards a Conversational Robot-Assisted Therapy for Children: Challenges and Directions

Branislav Borovac, Vlado Delić, Milan Gnjatović, Branko Karan, Milutin Nikolić, Mirko Raković, Jovica Tasevski, Dragiša Mišković

University of Novi Sad
Faculty of Technical Sciences
Serbia
Overview of the presentation

- Some basic facts about Project.
- Cerebral palsy and its consequences.
- Conventional therapies.
- What robot can help.
- How it should look like and what it should be able to do.
- Conversation and cognition – our approach
- Conclusion
Basic facts about project

Participants:

- Medical doctors
- Psychologists
- Artists
- Engineers

Project is financed (from January 1, 2011) by the Ministry of education and science of the Republic of Serbia, under contract III 44008.
Cerebral palsy - basic facts

Cerebral palsy cause physical disability in human development chiefly in the various areas of body movement.

Treatment is usually symptomatic and focuses on helping the person to develop as many motor skills as possible or to learn how to compensate for the lack of them.
Cerebral palsy - basic facts

The brain, up to about age of 8, is not concrete in its development. It has the ability to reorganize and reroute signal paths that may have been affected by the initial trauma. The earlier it has help in doing this the more successful it will be.

In general, the earlier treatment begins and it is more intensive the better chance children have of overcoming developmental disabilities or learning new ways to accomplish the tasks that challenge them.
Cerebral palsy – examples of disorders

Gait disorder:

Video clip has been removed
Cerebral palsy – examples of disorders

Gait disorder:

Video clip has been removed
Cerebral palsy – examples of disorders

Gait disorder:

Video clip has been removed
Cerebral palsy – examples of disorders

Notion of himself and spatial relationships:

Video clip has been removed
Cerebral palsy – examples of disorders

Gross motor function and spatial relationships:

Video clip has been removed
Cerebral palsy – examples of disorders

Fine motor function:

Video clip has been removed
Robot as Assistive Technology

Robot should motivate children to work harder and longer in 3 therapeutic areas:

- Habilitation of Gross motor function and acquiring spatial relationship.
- Habilitation of Fine motor function.
- Speech therapy.
Robot as Assistive Technology

What characteristics robot must have?

- Robot must have pleasant appearance to be attractive to children,
- Robot must be able to develop affective relationship with children, robot must be able to communicate with them:
  - verbal communication (natural language),
  - non-verbal communication (facial expressions, gestures, ...)
- Robot must be able to demonstrate selected exercises,
Robot as Assistive Technology

Robot appearance
Robot as Assistive Technology
Robot as Assistive Technology

Robot appearance
Robot as Assistive Technology
Robot as Assistive Technology

Head of assistive robot Marko
Robot as Assistive Technology

Animation of some gross motor function exercises performed by robot
Robot as Assistive Technology

Animation of some gross motor function exercises performed by robot.
Robot as Assistive Technology

Animation of some fine motor function exercises performed by robot.
Robot as Assistive Technology

HUMAN

Verbal context
(natural language dialogue)

COGNITIVE TECHNICAL SYSTEM

Instruction

Visual information

Manipulation

Spatial context

ROBOT
Robot as Assistive Technology

- Mobilni (mobile)
- Antropomorfni (anthropomorphic)
- Robot (robot of)
- Kognitivnih (cognitive)
- Osobina (features)
Robot as Assistive Technology
To achieve all these goals robot must be able to simulate cognitive capabilities, including the language faculty as one of the most natural ways of communication.
Communication with the Robots

A widely accepted assumption underlying research in the field of robot-assisted therapy is that

- clinical benefits can be achieved through three-party interaction between the patient, the therapist, and the robotic system in the role of a facilitator.

E.g.:

- Recent work suggests that robots may induce positive social behaviors (e.g., imitation, eye gaze, and joint attention) in children with autism that are not observed while they are interacting with their peers, caregivers, and therapists.
Natural Language Dialogue

Relatively little attention has been devoted to robots’ capacity to autonomously engage in a natural language dialogue in the context of robot-assisted therapy.

• Why is language important for therapy for children?
The Importance of Language

The language faculty is a fundamental feature of humans that has an important role both in social phenomena and in the construction of an individual’s identity.

It is reasonable to expect that robots’ capacity to engage in a natural language dialogue may contribute to a great extent to establishing long-term social relations to robots.

This is particularly significant for therapy for children with developmental disorders, where one of the most crucial factors that determine the efficacy of the treatment is the child’s motivation to undergo a long-term therapy.
Functional Requirements

The functional requirements address both the analytical and generative aspects of the robot’s dialogue behavior.

• **The analytical aspect** relates to the research issue of capturing the meaning of spontaneously produced therapists’ linguistic inputs.

• **The generative aspect** relates to the research questions of adaptive dialogue management and generation of appropriate dialogue acts (and non-verbal acts).
The Analytical Aspect

One of the most fundamental research questions in the field of human–machine interaction is:

• How to enable dialogue systems to capture the meaning of **spontaneously produced** linguistic inputs **without** explicit syntactic expectations?

The principle behind this requirement:

• The users should not be forced to intentionally adapt their dialogue acts to a preset grammar. Instead, they should be allowed, as far as possible, to express themselves **naturally**.
Challenge

Spontaneously produced utterances are often ungrammatical:
  • elliptical, minor, context-dependent, etc.

A dialogue system should be able to cope with such dialogue phenomena (without a preset grammar).
To Address this Research Question ...

We introduce:

- the **Focus Tree** model of attentional information in task-oriented human-machine interaction.

This model is representational:

- it is non-statistical, computationally and analytically tractable, and with explanatory power.

This model is cognitively-inspired:

- it integrates insights from behavioral and neuroimaging studies on working memory operations and language-impaired patients (i.e., Broca’s aphasics).
Why Working Memory?

There is a broad consensus among researchers that working memory plays an essential role in human cognition.

It is considered to be:

• a temporary storage system of limited capacity
• that refers to the ability to retain some information active for further use,
• and to do so in a flexible way allowing information to be prioritized, added or removed.

Since the human language faculty is a cognitive capacity, it implies that working memory also constitutes an important aspect of natural language processing.
Conceptualization of Working Memory

The current understanding of working memory:

• **Working memory is a functional state** (not a separate system) that allows a direct access to activated part of long-term memory (cf. Cowan, Oberauer and Lange).

Or, in other words:

• It is a **limited** resource that is distributed flexibly among all items to be maintained in long-term memory (Ma et al.)
**An Insight into Working Memory**

The activated part of long-term memory: holds representations that are activated above baseline (black nodes).

The region of direct access: holds a limited number of activated representations available for ongoing cognitive processes.

The focus of attention: a single representation is selected to be in the focus of attention.
What is Missing in this Conceptualization?

However, it should be noted that

- this cognitive model neither specifies the topology of the network of the available memory representations,
- nor explains on which principles these representations are related.

In addition,
- the processing aspect of working memory is not specified.
Therefore ...

We introduce a technical and detailed approach to constructing the focus tree model of attentional information in human-machine interaction.

The model addressed both the research questions of:

- storage, and
- processing

of attentional information.
Structure and storage of Attentional Information

Focus tree is a hierarchical structure:

- human cognition is believed to have generally hierarchical properties,
- a hierarchical structure reflects the principle that available entities in long-term memory are distinguished with respect to their access status.
Comparison

Comparison between
(a) the concentric model of working memory, and
(b) the focus tree model of attentional state.
Comparison (2)

The focus tree represents the activated part of the long-term memory (black nodes).

At any moment, the current focus of attention is placed on exactly one node in the focus tree.

The region of direct access is a subtree in the focus tree determined by the current focus of attention as its root node.
Example 1: Spatial context

In a dedicated therapeutic exercise:

- The spatial context shared between the child and the therapist is a set of objects (i.e., cuboids, cylinders, and triangular-based prisms), randomly positioned on a table, that differ in size of the base, height, and color.
Example 1: Spatial context

For the purpose of illustration, let us adopt a simplified spatial context:

• It contains only two figures – a square and a triangle.
• Each figure can be moved in two directions (i.e., left/right and up/down) or rotated (clockwise/counterclockwise)
Example I: Simplified Focus Tree

Objects:
- ▲
- ■

Actions:
- translate
- rotate
- translate
- rotate

Directions:
- Up
- Left
- Down
- Right
- Up
- Left
- Down
- Right
- Up
- Left
- Down
- Right
Recall …

We want to enable dialogue systems to capture the meaning of spontaneously produced linguistic inputs without explicit syntactic expectations?

E.g., the system should be able to process spontaneously produced utterances:

• *Triangle to right.* (ambiguous: Move to right, or rotate to right?)

• *Rotate it rightward.* (context-dependent: The square or the triangle?)

• *Do it!* (ellipsys-substitution: what to do?)

• etc.

How does the focus tree model help?
Some (Fundamental) Questions

To what extent is syntax distinguished from semantics?

Is it possible, and in which cases, to recover semantic information without a complete syntactic analysis of the sentence?

If we do not take syntactic information into account, what kind of information is needed to recover semantic information?

The answers to these questions derive from separate but interrelated insights from neurolinguistics.
A Neurocognitive Insight

Language is a modularly organized neurological entity, in which syntax is anatomically distinguished from semantics and the lexicon.

At the level of linguistic behavior, this anatomical distinction can be observed in Broca’s aphasia.
Broca’s aphasia is an impairment of language ability caused by an injury to a part of the brain called Broca’s area, and its vicinity.

Broca’s aphasics suffer, inter alia, a highly restricted receptive disorder of syntax.
What They can(not) Comprehend?

Broca’s aphasics can comprehend basic phrase syntactic structures (the canonical clauses, e.g., Subject-Verb-Object clauses of English).

E.g.:

“A dog is chasing a brown cat”.

Broca’s aphasics have difficulties in understanding sentences with syntactic movement out of the object position.

E.g.:

“The cat that the dog is chasing is brown”.
**But ...**

They are able to use **semantic cues** (i.e., the lexicon and the general knowledge of the world) **to get around their deficit**.

E.g.:

They can correctly interpret the following sentence

**“The apple that the boy is eating is red.”**

although they cannot comprehend its underlying syntactic structure (they know from their general knowledge of the world that apples do not eat boys).
Processing the User‘s Commands

Understanding the cognitive mechanisms that underlie the behavior of Broca’s aphasics may be valuable for researchers that aim at enabling dialogue systems to process spontaneously produced user’s linguistic inputs with no explicit syntactic expectations.
Processing the User‘s Commands (2)

Mapping of the user's command onto the focus tree is context dependent,

- i.e., it is performed with respect to the position of the current focus of attention.

The module for natural language understanding does not take the syntactic information into account, but only keywords and phrases.

The main idea:

- Step A: *Updating the focus of attention* – Try to map the command onto the current region of direct access.
- Step B: *Mnemonic selection* – If it is not possible, transit the focus of attention to its parent. Go to Step A.
Example 1: Processing

$C_1$: triangle to right

$C_2$: rotate it rightward

Updating the focus of attention

Resolving ambiguity
Example 1: Processing

$C_1$: triangle to right

$C_2$: rotate it rightward

Mnemonic selection

Updating ...
Example 11: Conversation with a Robot
Robot as Assistive Technology
You have noticed ...

In the case when the user’s command was incomplete, the system asked for a clarification.

In general, the system applies **adaptive dialogue strategy**, i.e., it **dynamically adapts its dialogue behavior** according to the **current context of interaction**.

In our approach, the dialogue strategy is **not hard-coded**.

Instead, we enable the therapsit to **define** the dialogue strategy of the system.
The second assumption underlying research in the field of robot-assisted therapy:

- therapists should be directly involved in the development of a robotic system intended to be used as an assistive tool.

We introduce:

- a therapist-centered programming platform that enables the therapist to design and test a robot’s dialogue behavior.
Therapist-Centered Programming Platform

To the extent that the platform is therapist-centered,

• it is motivated by and tailored to real-life difficulties that therapists encounter while trying to specify a robot’s dialogue behavior.

To the extent that the introduced platform is domain-independent,

• it can be applied to different therapeutic settings.
Defining a Dialogue Strategy

- Multi-tab environment
- New context feature
- New context variable
- Definition of a dialogue strategy
- Feedback on syntax errors
Defining a Focus Tree
The introduced approach is not limited only to verbally uttered commands. It supports also non-verbal dialogue acts produced by the user, e.g.:

- pointing a finger (visual information),
- touching (haptic information),
- etc.

Such non-verbal acts may also change the attentional state. The transition of the focus of attention can be modeled by the focus tree.
Discussion

There is an intellectual divide in the field of human-machine dialogue:

• **Representational approach** relates to studying and modeling the mechanisms underlying the human language processing system (e.g., attention, memory, perception, etc.) in order to computationally simulate their features.

• **Statistical approach** is primarily intended to simulate language behavior without a deep understanding of the underlying phenomena. The principal approach is to use data derived from language corpora, and apply automated analysis methods to empirically derive rules and structures to manage machine dialogues.
The state-of-the-art

Statistical approaches have proved to be relatively successful when applied in certain aspects of machine learning such as automated speech recognition or machine translation.

However, the last two decades have shown that they are insufficient to address the more general research question of managing human-machine dialogues that require some sort of contextual analysis and interpretation.
A Way to Go

Why we need a new direction?

- It cannot be expected that language corpora, no matter how carefully produced, will contain surface manifestations of all relevant dialogue phenomena.

A possible solution:

- We need to involve linguistic structures that underlay the human language processing system and cannot be extracted from language corpora.

- The focus tree model is a step in this direction.
Generalizability

The level of detail contained in the specification of the focus tree model is sufficient for a computational implementation …

… while the level of abstraction is sufficient to enable generalization of the model over different interaction domains.

Various adaptations of this model were successfully applied in several prototypical dialogue systems with diverse domains of interaction.
Thank you very much for your attention