HeAL9000: an Intelligent Rehabilitation Robot

Lorenzo Cristofori[1], Claudiu D. Hromei[1], Francesco Scotto di Luzio[2], Christian Tamantini[2], Francesca Cordella[2], Danilo Croce[1], Loredana Zollo[2] and Roberto Basili[1]

[1]*Department of Enterprise Engineering University of Rome, "Tor Vergata", Rome, Italy*

[2]*Research Unit of Advanced Robotics and Human-Centred Technologies, Universitá Campus Bio-Medico di Roma, Rome, Italy*

Abstract

AI applications to health related processes include the adoption of robotic platforms for rehabilitation, aiming at the delivery of highly intensive, repeatable and accurate motion therapies and able to constantly monitor the patient and provide the suitable assistance levels. However, a comprehensive approach to robot-aided rehabilitation requires also a social level of interaction with the patient that implies cognitive modeling and linguistic communication. It is worth noticing that robotic platforms providing both physical and cognitive support to patients have not been proposed so far. In the HeAL9000 project an intelligent robot for rehabilitation of patients affected by musculoskeletal disorders is proposed with cognitive and linguistic abilities. It relies strongly on machine learning technologies whose aim is to support cost effective engineering of the platform as well as evolving capabilities across time. In the paper early experimental evidence is acquired through quantitative evaluation.

Keywords

AI for Rehabilitation, Rehabilitation Robotics, adaptive Human Robot Interaction, Natural Language Understanding, Dialogue

1. Introduction

The adoption of robotic devices in rehabilitation is widely increased in recent years since robots can deliver highly intensive, repeatable and accurate motion therapy for neurological and musculoskeletal disorders [1]. The goal of rehabilitation is the functional recovery of the body area of interest, the restoration of the functional range of motion and the recovery of muscle strength[2]. Robot-aided motor therapy has further advantages, such as the possibility of objectifying patient performance, modulating appropriately the level of assistance and providing feedback during treatment [3]. Several studies have been conducted to evaluate the effectiveness of robot-aided rehabilitation. The first clinical study was carried out in 1997 using the commercial robot MIT-MANUS on post-stroke patients [4]. In more recent studies, robotic assistance has been dynamically changed based on the subject's needs, evaluated by bio-mechanical and psycho-physiological monitoring systems [5]. These platforms establish a purely physical interaction with the patient, but they generally do not include a cognitive interaction, although it can be a fundamental tool for the user motivation and engagement [6]. However, robotic platforms capable of providing both physical and cognitive support to the

AIxIA 2021 SMARTERCARE Workshop, November 29, 2021, Milan, IT

patient with neurological or musculoskeletal disorders did not emerge from the analysis of the state-of-the-art. A platform capable of providing an adequate level of assistance, tailored to the individual patient's needs, favoring her participation may deliver a more effective training session.

The project HeAL9000 ("Healthcare Agents and Learning robots") is funded by Regione Lazio and aims at designing, developing and validating in the operative scenario a smart robotic platform to deliver rehabilitation to a patient affected by musculoskeletal disorders. The platform should be able to promote and facilitate the patient's motor recovery through a human-robot interaction exploiting communication channels typically used in therapist-patient interaction, i.e. verbal, physical, cognitive. HeAL9000 is a Service Oriented architecture that integrates a Robotic Platform and cognitive services whose aim is to control the verbal and non-verbal interaction between the robot and the patient. The cognitive components implement a Dialogue system that interprets the utterances of the patient, the non-verbal stimuli (e.g., the physiological input from dedicated devices, such as the heart rate or patient temperature) and plans the interaction. We modeled the interaction in the overall rehab session, by i) demonstrating the exercises to be performed, ii) observing and evaluating the patient during its practice and eventually correcting him with verbal signals, iii) actively supporting him with the robotic arm as a therapist would do. The entire interaction is also expected to account for emotional information, implicitly shown by the patient through Computer Vision (here devoted to Face Emotion Recognition [7]) and language processing according to the automatic analysis of the spoken utterance to extract Emotional stimuli both from the patient tones and the sentence contents. This allows to improve a more natural interaction with the robot and improve engagement. This paper is thus focused on presenting the overall architecture that is under development, with particular emphasis on the specific modules devoted to the Dialogue Management and the Natural Language Understanding. A dedicated experimental evaluation of these specific modules shows that they can be adopted for a more robust and effective HRI.

In the rest of the paper, Section 2 provides an overview to the HeAL9000 architecture, while Section 3 reports the experimental evaluation of some of the AI components and Section 4 draws the conclusions.

2. Integrating robot-aided rehabilitation and language learning

The role of the therapist is paramount in motor rehabilitation. The healthcare operator has to establish a physical and cognitive relationship with the patient and his intervention cannot ignore the patient's clinical and emotional state. The HeAL9000 project aims at replicating the interaction between therapist and patient established during a conventional rehabilitation session, to improve the effectiveness of the rehabilitation treatment. Such a platform could represent a disruptive source of technological innovation in modern robot-aided rehabilitation and it can be an enormous step forward in terms of human-robot interaction, robot autonomy, safety for the patient, reliability of the robot and effectiveness of the rehabilitation treatment. To do this, HeAL9000 robotic platform will (i) consist of a service robot capable of replicating the behavior of the human therapist thanks to the use of Machine Learning and Learning by Demonstration techniques; (ii) have a highly adaptive behavior concerning the characteris-

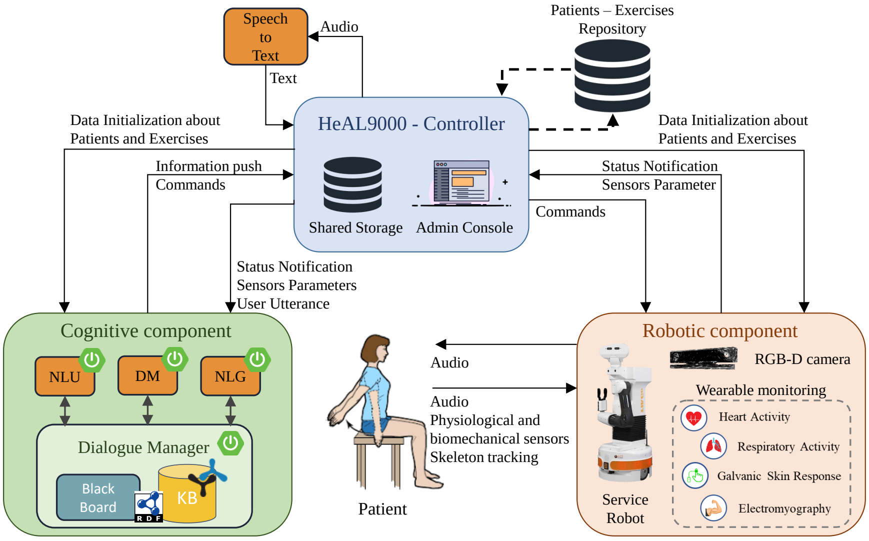

Figure 1: HeAL9000 Architecture

tics of the patient and the context thanks to the multi-modal monitoring of the patient; (iii) establish physical and cognitive interaction with the patient, similar to the one observed in the combination therapist-patient. It has the twofold purpose of motivating the patient to actively participate in the treatment and to strongly personalize the rehabilitation session according to the physical and cognitive state of the patient.

Figure 1 summarizes the overall architecture. The patient physically and verbally interacts with the Robotic component that is also devoted to measuring her physiological and biomechanical information, while visually tracking the patient's movements and emotional status. This body of information is provided to the so-called cognitive component, which processes such input, plans the interaction (through dedicated sub-modules presented hereafter) and provides instruction to the robot, both in terms of actions to be performed or utterances to be pronounced. Both components and their communication are orchestrated by the HeAL9000 Controller, which also stores the shared information. Moreover, the controller provides the Monitoring and Administration console (also for security purposes) and interfaces with external repositories (with clinical information about patients and exercises) or external modules such as Speech to Text modules. The Robotic Components are implemented on the adopted TIAGo robot[1] based on the ROS (Robot Operating System) middleware, while the other components are implemented as Service Oriented Architectures hosted in a dedicated Cloud.

2.1. Machine Learning for Patient-therapist interaction

The natural interaction between the patient and a therapeutic robot is crucially dependent on the ability to integrate learning at the physical level (as adaptive control mechanisms and data-

[1] https://robots.ieee.org/robots/tiago/

driven machine vision are involved) and at a cognitive level, related to the ability to recognize people, profile them and support linguistic communication with them.

Cognitive aspects. Starting from the analysis of the conventional rehabilitation sessions, it is possible to distinguish the roles played by the patient and the therapist. They assume dynamic behaviors based on circumstances and stimuli that the two exchange reciprocally. At the beginning of the rehabilitation session, the clinician carries out the demonstration: the therapist explains to the patient the task to be performed, not only verbally but also with the help of his/her own body. In this context, the patient is a listener: he/she does not perform any movement. However, the patient may asks for clarifications about the activity to be performed. At the end of the demonstration, the therapist starts the second phase of the treatment: the observation. At this time, the patient plays the role of the main actor as he/she is asked to perform the proposed exercise independently. In turn, the therapist monitors the subject and encourages and/or warns him/her to carry out the assigned motor task in the best way. Whenever the therapist decides to intervene to correct any patient error and/or the patient complains of pain or fatigue, the role of the therapist turns into the helper one. In this context, the real physical interaction between the two actors begins and takes place. As soon as the exercise is completed, the cycle can start over and iteratively continue until the rehabilitation session is completed [8]. In order to develop an effective robot-aided system for rehabilitation, it is necessary to implement such roles onto a robotic platform, able to handle both physical and cognitive interactions, as shown in Figure 2.

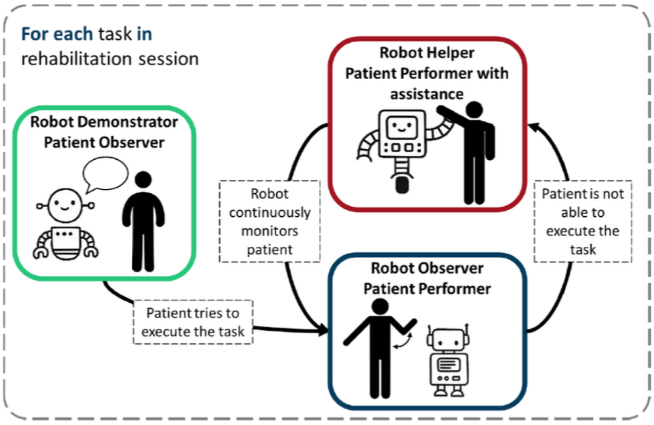
For each task in rehabilitation session
Robot Demonstrator Patient Observer
Robot Helper Patient Performer with assistance
Robot continuously monitors patient
Patient is not able to execute the task
Robot Observer Patient Performer
Patient tries to execute the task

Figure 2: Proof-of-concept of patient-robot interaction.

Physical interaction. The robotic platform will be able to play one role among demonstrator, observer and helper, as reported in Figure 2. In the first scenario, the robot will demonstrate the motor task to be performed by the patient. When the patient will try to execute the task, the robot will constantly evaluate the motor performance of the patient exploiting RGB-D cameras and the skeleton tracking algorithm. In this way, the robot will be aware of patient errors, pain and/or risk conditions to assist him/her when needed. In the helper role, the robot will guide the limb of the patient to correctly execute the assigned task. To do this, it is essential to model the physical interactions of the traditional rehabilitation treatment to tune the optimal behavior

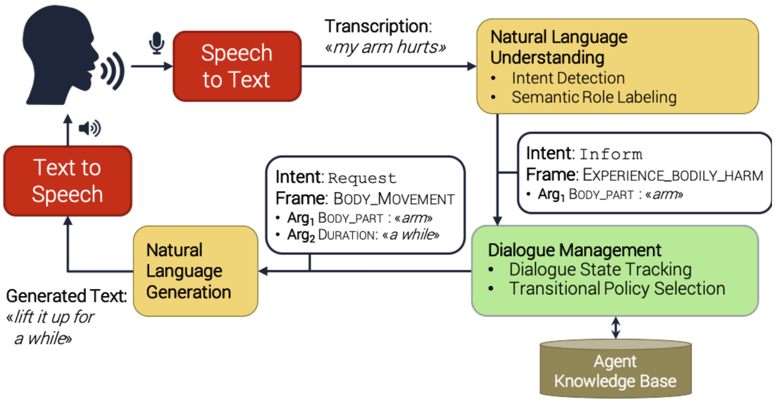
Speech to Text
Transcription: «my arm hurts»
Natural Language Understanding
• Intent Detection
• Semantic Role Labeling
Intent: Inform
Frame: EXPERIENCE_BODILY_HARM
• Arg$_1$ BODY_PART : «arm»
Dialogue Management
• Dialogue State Tracking
• Transitional Policy Selection
Agent Knowledge Base
Intent: Request
Frame: BODY_MOVEMENT
• Arg$_1$ BODY_PART : «arm»
• Arg$_2$ DURATION: «a while»
Natural Language Generation
Text to Speech
Generated Text: «lift it up for a while»

Figure 3: The workflow behind the cognitive interaction

of the robotic platform. Learning by demonstration approaches based on Dynamic Movement Primitives [9] will allow encoding the therapists-patient physical interaction to re-target the recorded motions onto the robotic platform.

2.2. The role of Dialogue

The cognitive dimension of the therapist-patient interaction is supported by managing natural dialogue able to integrate control aspects (e.g. the visual detection of critical phenomena for the patient, such as wrong positions or expressions of pain) into natural managed by a dedicated set of modules devoted to the acquisition of input from the environment (from verbal input to physical stimuli acquired through the dedicated sensors), to the tracking of the whole dialogue and the planning of individual reactions. The workflow is depicted in Figure 3.

When a verbal input is provided by the patient, the neural architecture discussed in [10] is applied for *Speech to Text* transcription. Let us consider a patient during the exercise who says "*My arm hurts*" to express some difficulties due to the requested movements. Content is processed by the *Natural Language Understanding* module that implements the inductive method described in [11]: here the semantic interpretation in terms of Frames Semantics [12] that models input sentences into meaning representation graphs is also coupled with the recognition of the user intent. The semantic graph is then provided to the *Dialogue Management Module* used to recognize user states on *Dialogue State Tracking*, to plan the robot reactions to the input, to update current states, accordingly, and finally to compile the requested linguistics output. In the workflow, the resulting semantic frame is Body_movement that ask the patient to move its arm (the Body_part as argument of the input frame) for a given Duration, i.e. *a while*. The output frame is compiled by the *Natural Language Generation* into a sentence like "*Lift it up for a while*" used to feed the robot text-to-speech module. The cognitive architecture in HeAL9000 integrates inductive modules such as the language understanding one with knowledge-based components, strongly dependent on domain-specific pragmatic (e.g dialogue state tracking) resources as well as medical knowledge bases.

The tasks of Dialogue Manager (DM) and Dialogue State Tracking (DST) in recent years are often addressed with the use of end-to-end methods for example using transformers to encode the input, like user sentences, and generate the output, the response of the system. This emerges also from the different dialogue state tracking challenges [13] such as the last of these, the DST9 [2] in which most of the systems use the technologies mentioned above. In HeAL9000 the Dialogue Manager has the ambition to controlling the interaction between a robot and a patient in a critical scenario. Based on this consideration we decide to adopt a system that was i) as controllable as possible in terms of dialogues produced and actions performed by the robot; ii) flexible and adaptable to new scenarios. The DM was modeled as a set of State Machines, each of which performs a specific activity within the flow of the physical therapy session, i.e. in the initial phase a state to welcome the patient or during the execution of the exercise a state to stop the patient when he/she makes mistakes. Each state represents a specific action (verbal or not) to be performed by the robot, each edge can be used for state change if and only if the conditions of the edge are satisfied. Examples of conditions can be a combination of particular user utterances, a set of information in the knowledge base, or different events such as sensor data or facial expressions of the user. Thus, the response of a patient who is performing a physical therapy exercise, to feedback requested by the system, can be processed by considering a wide set of signals, not only the vocal one, as in the following example:

HeAL9000: *are you okay?*

Patient: *yes, everything is fine!*
sensor: Patient HeartRate High
sensor: Patient BloodPressure High
sensor: Patient Sad

HeAL9000: *please take a break and breathe deeply.*

Information from the sensors is added to the linguistic features for the Natural Language Understanding module, similarly to [11]. The model is thus trained on data containing a mixture of linguistic and sensory features in order to be able to distinguish events as in the example above.

The Dialogue Management module is modeled as a set of non-deterministic finite state machines. The dialogue can be represented as a set of quintuple (I, S, s_0, T, F) where:

- I is the input user utterances, sensors or other signals and knowledge base information.
- S is the set of states of dialogue.
- s_0 is the initial states of each dialogue phase.
- T is the function that makes each state and input to correspond a sub-set of possible states: $T : S \times I \rightarrow U$, where $U \subset S$.
- F is the set of final states.

Figure 4 shows a simplified version for the demonstration phase in which the robot shows the video of the exercise (Start_video_exercise), confirms that the user has understood the exercise to be performed (Confirm_exercise) and asks the patient to start the activity (Req_activity_start). If the patient does not start the exercise, a second explanation of the

[2] https://dstc9.dstc.community/

exercise is provided (EXPLAIN_EXERCISE). Where the user fails to start the activity, the system provides a special end state and then calls a human operator (END_CALL_OPERATOR). In order to avoid not planned behaviors, there is always an error state (END_W FAILURE). When the system is not able to understand the intentions of the user, it enters in a special state (CLARIFICATION) with which the intentions of the user are clarified.

During the entire dialogue, the system acquires information and exploits the data in its possession to choose and generate the answers to be provided to the user. All data is stored in a knowledge base structured as an RDFS/OWL ontology. Ontologies are often used in the context of dialogue systems, such as in [14]. Within the knowledge base, we defined several concepts so that the system would be able to use patient information such as first name, last name, birth date. The information about the exercises such as the correct movements to perform, the number of repetitions, the series, the parts of the body involved in the exercise and other information useful to the system to help the patient to perform the exercise through the use of dialogue. This technological choice has two main advantages, i) the use of ontologies allows a formal and explicit description of the concepts of the domain of interest, and this allows the system to query structured data and ii) the Open World Assumption of OWL allows the system to progressively enrich its knowledge with information from the web.

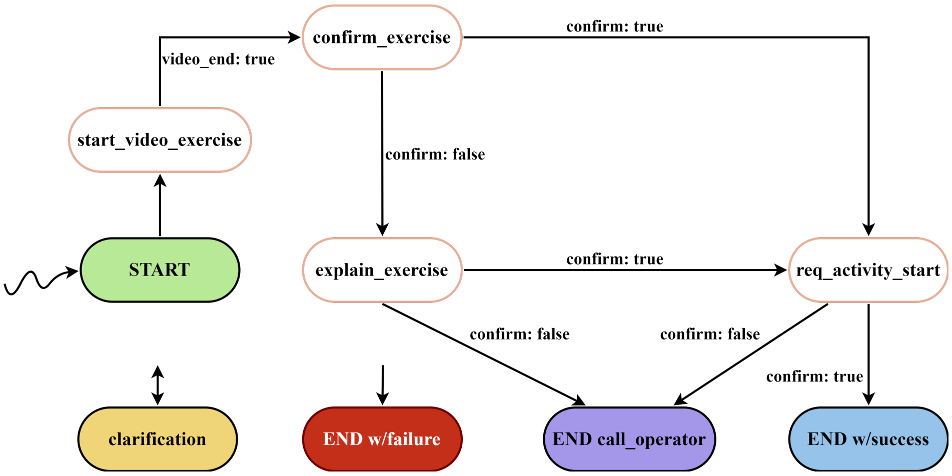

Figure 4: State Machine for Demonstration phase.

3. Evaluating learning robots in therapeutic scenarios

Natural Language Understanding as an emerging ability. Machine learning for natural language has been traditionally applied to induce cognitively plausible interpretation models, usually based on theories about the nature and semantics of human communication. Frame semantics [15] has played often the role of reference theory for semantic interpretation and representation. Frames [15] are cognitive devices able to represent and encode properties of eventualities, i.e. situations, world states and subject's personal state, and support interpretation and reasoning about a context and a domain. They play at the same time the role of formalism for the representation of the operational context of a robot and of a knowledge repository to express world knowledge.

Frames are thus useful during dialogue as they can express (and constraint) the intentions and content related to a patient utterance, as a guide of the interpretation process, as well as a storage device for maintaining the dialogue state. Frames have been often used for automatic Information Extraction through Machine Learning ([16]) whereas interpretation is seen as a structured sentence classification process. In line with this perspective, we propose an interpretation framework consisting of a cascade of classification steps aiming at recognizing purposes and content semantics in support of meaningful dialogue. Firstly, a sentence classification step (namely Intent Classification) is applied to input (i.e. the patient's) utterances and then sequence labeling is applied for automating Semantic Role Labeling (*SRL*) as in [17]. In *Intent Classification* each sentence is associated with the intent, consisting of a label such as REQUEST or INFORM): the objective is that we need to detect the patient's intent of a sentence during the dialogue to understand his goal or state, i.e. if he's *asking* for something or *providing information* to the system.

During *Semantic Role Labeling* words are associated with labels expressing the role they play in the semantic frame vehiculated by the sentence. Roles establish the differences between predicates f, the so-called Lexical Units denoting the frame, and arguments, that are the roles r_i, called Frame Elements of f, activated by the sentence. Multiple frames in a sentence are the norm and establish ways of detecting and storing contextual knowledge during the dialogue. For its complexity, Semantic Role Labeling is further divided into three subtasks: *Frame Prediction* predicts which Frames f is expressed by an input sentence and labels the words (lexical units) responsible for evoking f; *Boundary Detection* detects the starting and ending positions of individual arguments of each recognized frame f; *Argument Classification* assigns the semantic roles, i.e. the Frame elements associated with the predicted Frame f, to the arguments detected in the previous step. In the example of Figure 3, the sentence "*my arm hurts*" evokes the frame EXPERIENCE_BODILY_HARM through the lexical unit *hurt*. Then the notion of BODY_PART is expressed by the fragment "*my arm*" as the injured part of the EXPERIENCER, here implicitly referred to the speaker, i.e. the patient.

Semantic Role Labeling acts on the entire sentence and is modeled as a Markovian formulation of a structured SVM (SVM^{hmm} as in [18, 11]). The learning algorithm combines a local discriminative model, which estimates the individual observation probabilities of a sequence, with a global generative approach to retrieve the most likely tag sequence that better explains the semantics of the whole sequence. The labeling obtained by the SVM^{hmm} onto the example sentence "*my arm hurts*" is as follows:

$$[\texttt{Speaker}]_{\textsc{Experiencer}}\ [\textit{my arm}]_{\textsc{Body\_part}}\ [\textit{hurts}]_{LU:\textsc{Experience\_bodily\_harm}}$$

where the pseudo token [Speaker] is used to denote the implicit argument EXPERIENCER not related to any text portion.

The input of the models is composed, besides the linguistic features in line with [11], also of other features such as the intent of the patient's sentence and the requested information (i.e. the frames and arguments that HeAL9000 expects). In addition, information from the sensors the patient wears, the tone and intensity of the voice volume and the emotion recognized through the Face Emotion Recognition Model are added. This composes a more complete picture of what the patient communicates to the robot, both verbally and non-verbally. Finally, each model adds the result of the previous models as a feature to the input. During the dialogue, the system needs to store and use some information about the interlocutor. When HeAL9000 is a *Demonstrator* (i.e. in the Demonstrator stage, Figure 3) it can be effective to use the patient's real name and always be aware of the body part involved during the rehabilitation session. During the *Observer* part of the session, the robotic platform should consider the age of the patient to better evaluate the movements (for example, an older patient may not be able to fully execute some exercises). Finally, in the *Helper* stage HeAL9000 needs to consider the body part involved, whether the patient is in pain and, eventually, the intensity to better help him complete the session.

The Frames are thus adopted to model the knowledge about the stages and knowledge incoming through sentences. We studied the involved frames over the current repository created by the Framenet[3] project [12]. In particular, we focused on medical and informational aspects according to the existing frames, an excerpt of which is provided below:

- BEING_NAMED: Concerns Entities (PATIENT) conventionally being referred to by particular names (NAME and SURNAME).
- MEDICAL_CONDITIONS: Medical conditions or diseases that a PATIENT suffers from. Contains the part or area of the body (BODY_PART) affected by the condition, the CAUSE of the condition, a prominent SYMPTOM and others.
- ACTIVITY_START: An Agent (renamed PATIENT for our case) initiates the beginning of an ongoing ACTIVITY in which he will be continuously involved. Used to model the exercise activity during the rehab session.
- ACTIVITY_RESUME: An Agent (PATIENT) resumes participation in an ACTIVITY.
- ACTIVITY_FINISH: An Agent (PATIENT) finishes an ACTIVITY, which can no longer logically continue.
- EXPERIENCE_BODILY_HARM: An EXPERIENCER is involved in a bodily injury to a BODY_PART, even though in some cases, no BODY_PART need be indicated.
- MEDICAL_INTERACTION_SCENARIO: A PATIENT interacts with one or more MEDICS, usually, the PATIENT has an AFFLICTION.
- LEVEL_OF_FORCE_EXERTION: An EXERTER, ACTION or FORCE is capable of exerting or does exert a physical force at a level specified by the target. The Frame could be used in the *Helper* phase to describe the force used by the robotic arm.
- INHIBIT_MOVEMENT: An AGENT (the physical robot in our case) restricts the movement of a THEME (may be the patient's arm) despite the THEME's desire, plan, or tendency towards motion; the AGENT may also use an INSTRUMENT (robotic arm).

Some simulated interactions between a patient and a robotic therapist were registered in a Wizard-of-Oz (WoZ) method and manually labeled to train the machine learning algorithm, whose evaluation is reported hereafter.

Evaluating Semantic Role Labeling. The dataset used to train the SVM models consists of about 2,000 sentences representing interactions between a patient and a therapist, equally

[3] https://framenet.icsi.berkeley.edu/fndrupal/frameIndex

distributed among the various stages of the dialogue and split into train and test sets with an 80/20 ratio. All steps are modeled as classification tasks. Intent classification corresponds to a multi-class classification where each sentence has to be assigned to one possible class (from 8 total classes) reflecting the user's intent. Frame Prediction corresponds to a multi-label classification task, where each sentence has to be assigned to zero, one or more classes reflecting the evoked linguistic predicates f (here 10 possible frames are considered). Boundary detection is modeled as a sequence labeling task where arguments are annotated according to the BIO notation[4]. Finally Argument classification is a multi-class classification task where each informative chunk has to be associated one of the 28 possible classes. The system was evaluated according to different metrics as the two tasks (i.e., Intent Classification and SRL) have different objectives and needs. Accuracy simply calculates the ratio between correct prediction and total predictions. Sentence Level Accuracy is similar to Accuracy, but for a sentence to be considered correct, it is necessary for all its word labels to be correctly predicted to make a perfect match. This is the case with the subtasks of Semantic Role Labeling. We have also reported Precision and Recall metrics to evaluate the performance of the models at the level of the entirety of the entities (Frames, Boundaries or Arguments) to be predicted. As far as the Frames Prediction model is concerned, it is necessary that all the words belonging to the Frame are correctly labeled in order for it to be considered correct.

Table 1
Results of Sentence level task and Semantic Role Labeling tasks.

Task	Span Level		Sentence Level
	Precision	Recall	Accuracy
Frame Prediction	0.85	0.83	0.86
Boundary Detection	0.93	0.92	0.91
Argument Classification	0.99	0.98	0.96

In terms of Accuracy, 96% of the time the system correctly predicts the Intent of a user sentence. Table 1 then shows the results of the Semantic Role Labeling (SRL) tasks that more straightforward. Indeed, in the SRL pipeline, each model assumes that the predictions in the previous step are correct. As a consequence, Argument Classification is almost perfect, as it takes advantage of such a gold standard input, where informative chunks (i.e. arguments) are already perfectly matched.

Evaluating Dialogue. We demonstrate by simulation that the Dialogue Manager is robust to adversarial interactions with the system and that it tries to complete the conversation in a successful end state with as few turns as possible. We thus prepare an experimental setup consisting of a dialogue made up of 4 phases, in three of which the robot plays the 3 roles shown in Figure 2. The remaining phase (called *Information Gathering*) is used when the interaction starts to welcome the patient and collect her personal information. In the *Demonstration* phase, HeAL9000 shows the exercise to be performed. In *Observation* phase, the patient is observed performing the exercise and HeAL9000 responds to stimuli coming from sensors and user

[4] As an example: [*my*]$_{\_}$ [*left*]$_{\text{B}}$ [*arm*]$_{\text{O}}$ [*hurts*]$_{\_}$, where _ denotes a non informative chunk, B is the beginning of a chunk, I is used for the elements in the middle and O is for the last element of the informative chunk.

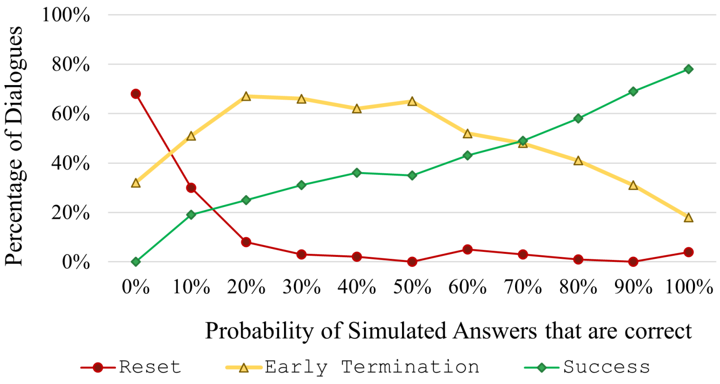

Figure 5: Dialogue success rates as the percentage of relevant responses increases

verbal input. In *Helping* phase, HeAL9000 physically helps the patient to execute the exercise. For each output of the dialog manager, during all phases, we created 3 categories of possible responses. A category of consistent answers to simulate a user collaborating with the system. The second category reflects answers that only partially help the system to continue the dialogue, by introducing not completely consistent answers, or requests for further explanations (thus increasing the length of dialogue). The last category reflects not consistent responses to simulate an adversarial user.

Simulated data are made of conversations where, in each turn, we select with probability p an answer uniformly at random among the consistent answers and with probability $1 - p$ select an answer among the other categories. We simulated 100 dialogues $\forall p \in \{0.0, 0.1, 0.2, \ldots, 1.0\}$ for a total number of 1100 dialogues. Dialogue ending states fall into three categories: dialogues that correctly terminate in the final state of the system (`Success` in Figure 5), with an average length of 51 dialogue turns; dialogues that correctly terminate in final states but anticipating the end of the therapy session, e.g., the patient feeling pain (`Early Termination` in Figure 5), with an average length of 44 dialogue turns; finally, dialogues that terminate in an error state not handled by the system or conversations that are too long (more than 100 dialogue turns) and therefore terminated early (`Reset` in Figure 5). Figure 5 shows the percentage of the three categories of dialog termination (`Success`, `Early Termination`, `Reset`) as the probability value p increases. Tests showed the system to be tolerant of misinterpretation of user sentences, even with low probability p values all dialogues, indicated with a reset termination, were terminated for reaching the maximum allowed length. The system never needed to make use of special termination states for unexpected errors such as the `Reset` state of the dialogue.

4. Conclusions

In this paper, a cognitively inspired robotic architecture for orthopedic rehabilitation is described. The system integrates motion control capabilities with dialogue and natural language understanding in order to harmonize and personalize the relationship with the patients. The approach discussed in the paper is strongly focused on the adaptive abilities supported by Machine

Learning algorithms. The results obtained in the acquisition of language processing abilities and in the dialogue control are more than encouraging, and pave the way to a robotic system able to support operational adoption of this technology, data acquisition and incremental improvement over time. This is a core property in the enabling of rapid and beneficial penetration of this technology in daily practices.

Acknowledgments

The research for this paper was partly funded in project ”Healthcare Agents and Learning robots – HeAL9000” by Regione Lazio (prot. A0320-2019-28108).

References

[1] M. Babaiasl, S. H. Mahdioun, P. Jaryani, M. Yazdani, A review of technological and clinical aspects of robot-aided rehabilitation of upper-extremity after stroke, Disability and Rehabilitation: Assistive Technology 11 (2016) 263 – 280.

[2] P. Maciejasz, J. Eschweiler, K. Gerlach-Hahn, A. Jansen-Troy, S. Leonhardt, A survey on robotic devices for upper limb rehabilitation, Journal of neuroengineering and rehabilitation 11 (2014) 1–29.

[3] F. Scotto di Luzio, C. Lauretti, F. Cordella, F. Draicchio, L. Zollo, Visual vs vibrotactile feedback for posture assessment during upper-limb robot-aided rehabilitation, Applied ergonomics 82 (2020) 102950.

[4] M. L. Aisen, H. I. Krebs, N. Hogan, F. McDowell, B. T. Volpe, The effect of robot-assisted therapy and rehabilitative training on motor recovery following stroke, Archives of neurology 54 (1997) 443–446.

[5] C. Rodriguez-Guerrero, K. Knaepen, J. C. Fraile-Marinero, J. Perez-Turiel, V. Gonzalez-de Garibay, D. Lefeber, Improving challenge/skill ratio in a multimodal interface by simultaneously adapting game difficulty and haptic assistance through psychophysiological and performance feedback, Frontiers in Neuroscience 11 (2017) 242. doi:10.3389/fnins.2017.00242.

[6] A. Aly, A. Tapus, Towards an intelligent system for generating an adapted verbal and nonverbal combined behavior in human–robot interaction, Autonomous Robots 40 (2016) 193–209.

[7] S. Li, W. Deng, Deep facial expression recognition: A survey, IEEE Transactions on Affective Computing (2020) 1–1. doi:10.1109/TAFFC.2020.2981446.

[8] M. Johnson, M. Mohan, R. Mendonca, A stimulus-response model of therapist-patient interactions in task-oriented stroke therapy can guide robot-patient interactions, in: Proceedings of the Annual Rehabilitation Engineering and Assistive Technology Society of North America (RESNA) Conference, New Orleans, USA, 2017.

[9] S. Schaal, J. Peters, J. Nakanishi, A. Ijspeert, Control, planning, learning, and imitation with dynamic movement primitives, Workshop on Bilateral Paradigms on Humans and Humanoids, IEEE Int. Conf. on Intelligent Robots and Systems, Las Vegas, NV (2003).

[10] A. Baevski, H. Zhou, A. Mohamed, M. Auli, wav2vec 2.0: A framework for self-supervised

learning of speech representations, CoRR abs/2006.11477 (2020). URL: https://arxiv.org/abs/2006.11477. arXiv:2006.11477.

[11] A. Vanzo, D. Croce, E. Bastianelli, R. Basili, D. Nardi, Grounded language interpretation of robotic commands through structured learning, Artif. Intell. 278 (2020). doi:10.1016/j.artint.2019.103181.

[12] C. F. Baker, C. J. Fillmore, J. B. Lowe, The berkeley framenet project, in: Proceedings of COLING-ACL 1998, 1998. doi:10.3115/980845.980860.

[13] J. D. Williams, A. Raux, M. Henderson, The dialog state tracking challenge series: A review, Dialogue & Discourse 7 (2016) 4–33.

[14] D. Milward, M. Beveridge, Ontology-based dialogue systems, in: Proc. 3rd Workshop on Knowledge and reasoning in practical dialogue systems (IJCAI03), 2003, pp. 9–18.

[15] C. J. Fillmore, Frames and the semantics of understanding, Quaderni di Semantica 6 (1985) 222–254.

[16] D. Gildea, D. Jurafsky, Automatic Labeling of Semantic Roles, Computational Linguistics 28 (2002) 245–288. arXiv:https://direct.mit.edu/coli/article-pdf/28/3/245/1797857/089120102760275983.pdf.

[17] E. Bastianelli, G. Castellucci, D. Croce, R. Basili, D. Nardi, Effective and robust natural language understanding for human-robot interaction, in: ECAI 2014 - 21st European Conference on Artificial Intelligence, 18-22 August 2014, Prague, Czech Republic, volume 263 of *Frontiers in Artificial Intelligence and Applications*, IOS Press, 2014, pp. 57–62. URL: https://doi.org/10.3233/978-1-61499-419-0-57. doi:10.3233/978-1-61499-419-0-57.

[18] S. Filice, G. Castellucci, G. D. S. Martino, A. Moschitti, D. Croce, R. Basili, Kelp: a kernel-based learning platform, Journal of Machine Learning Research 18 (2018) 1–5. URL: http://jmlr.org/papers/v18/16-087.html.