Integrating Physical and Cognitive Interaction Capabilities in a Robot-Aided Rehabilitation Platform

Christian Tamantini ⬤, *Member, IEEE*, Francesco Scotto di Luzio ⬤, *Member, IEEE*, Claudiu Daniel Hromei ⬤, Lorenzo Cristofori, Danilo Croce ⬤, Marco Cammisa, Arcangela Cristofaro, Maria Vittoria Marabello, Roberto Basili ⬤, and Loredana Zollo ⬤

Abstract—The communication channels between physiotherapists and patients are many and varied. Rehabilitation robots are able to deliver intensive treatments and improve the patient's quality of life. However, rehabilitation robots in the literature do not integrate physical manipulation with natural verbal communication yet. This article proposes an innovative integrated system for motor rehabilitation based on the combination of physical and cognitive components to emulate the natural interaction between physiotherapists and patients. The proposed approach was validated in a laboratory setting with 20 healthy subjects. The cognitive system's ability to interact linguistically as well as the participants' kinematic performance and the emotional impact generated by two different robotic systems were assessed. The former integrates advanced linguistic capabilities and the latter lacks any verbal communication. The results showed that the presence of linguistic interaction promotes the quality of interaction, leading to improvements both in the execution of movements and in emotional terms.

Index Terms—Adaptive human–robot interaction, natural language understanding (NLU), robot-aided rehabilitation, upper limb, user state estimation.

I. INTRODUCTION

ROBOTICS is gaining more and more attention in the medical field [1]. Physical rehabilitation can be performed by means of robots to guide and monitor patients' movements in executing specific motor tasks [2]. Robots allow providing precise and controlled movements, objectively measuring patients' performance, and promoting active involvement, demonstrating promising rehabilitation outcomes for neurological and/or musculoskeletal disorders [3], [4]. The use of robotic platforms produces physical load relief for therapists [5]. While the robot actively assists patients during exercises, therapists can focus on other clinical activities, reducing the risk of musculoskeletal overload and improving their working conditions.

Indeed, human physiotherapists play a paramount role in accompanying patient during the rehabilitation process exploiting different communication modalities, involving physical contact and empathetic relationship achieved through verbal communication [6]. Robotic systems replicating the clinician's role can improve the physiotherapists' working conditions and promote patients' motor recovery.

This study aims at developing a robot-aided rehabilitation system integrating physical and cognitive interaction capabilities, as observed in the conventional rehabilitation treatment, and experimentally testing its performance by means of purposely developed metrics. A multimodal monitoring system is employed to gather data of different natures and the estimation of the complex patient's state is exploited to manage the human–robot interaction throughout the entire rehabilitation session. Furthermore, a complex cognitive component dynamically manages language on the basis of the user's state has been introduced. By considering the patient's performance and emotional state, the platform aims at enhancing the overall effectiveness and engagement of the rehabilitation process.

To validate the proposed approach, a quantitative analysis to compare the effectiveness of the proposed method, i.e., a robotic platform for physical therapy incorporating cognitive interaction capabilities, against a traditional robot-aided rehabilitation, i.e., without any verbal communication, is carried out in this work. The effects of both types of robotic interventions on various outcome measures, such as movement performance, emotional impact, and subjective perception were quantified.

The rest of this article is organized as follows. Section II presents the current state-of-the-art on the platforms for robot-aided rehabilitation. Section III provides a detailed description of the proposed approach experimentally validated in Section IV. Section V presents and analyzes the obtained results. Finally, Section VI concludes this article.

Christian Tamantini, Francesco Scotto di Luzio, and Loredana Zollo are with the Unit of Advanced Robotics and Human-Centred Technologies, Università Campus Bio-Medico di Roma, 00128 Rome, Italy (e-mail: c.tamantini@unicampus.it; f.scottodiluzio@unicampus.it; l.zollo@unicampus.it).

Claudiu Daniel Hromei, Lorenzo Cristofori, Danilo Croce, and Roberto Basili are with the Department of Enterprise Engineering, University of Rome "Tor Vergata," 00133 Roma, Italy (e-mail: hromei@ing.uniroma2.it; lorenzocristofori3 @gmail.com; croce@info.uniroma2.it; basili@info.uniroma2.it).

Marco Cammisa, Arcangela Cristofaro, and Maria Vittoria Marabello are with the Innovation Lab, Exprivia S.p.A, 70056 Molfetta, Italy (e-mail: marco.cammisa@exprivia.com; arcangela.cristofaro@exprivia.com; Vittoria.Marabello@exprivia.com).

II. RELATED WORKS

This section provides an overview of the current systems used in robot-aided rehabilitation to provide physical therapy. Specifically, they can be classified as robots for physical interaction, conversational agents, and social robots.

End-effector robots and exoskeletons have been largely adopted for motor rehabilitation of the upper and lower limbs of patients suffering from neurological and/or musculoskeletal disorders from the early stage of their recovery [7]. These robots are preprogrammed to guide the patient's affected limb in performing motor exercises by establishing a close physical interaction. The adoption of such robots may enhance the clinical outcomes of the conventional therapy thanks to their capabilities of objectively measuring the patient's motor performance [8], [9] and actively involving the patients by exploiting virtual reality games and/or biofeedback [10], [11]. More advanced systems can tune the control parameters based on the patient's performance and psychophysiological (PP) state [12]. Anyhow, these systems are not able to establish an empathetic relationship with the patient as they have no natural verbal communication capabilities. Furthermore, some robotic systems have been designed to offer physical assistance while patients engage in cognitive games [13]. In these scenarios, patients are engaged in cognitive training exercises while simultaneously receiving physical support from the robot [14]. However, it is important to emphasize that these systems do not establish a true cognitive interaction with the patient, similar to what occurs in natural language communication.

Recent advancements in digital systems have paved the way for the development of conversational interfaces, revolutionizing the natural interaction between humans and machines [15]. These interfaces have been employed in digital applications to engage patients through verbal interactions, offering both physical and cognitive exercises [16]. However, the lack of real-time monitoring of patient performance during the session limits their effectiveness in motor treatment. Conversely, integrating spoken language features onto physical platforms can augment the therapeutic impact, promoting improved outcomes for patients [17]. Social robots are demonstrated to be capable of establishing an empathetic relationship with the patient [18]. These robots are programmed to explain motor tasks by using video demonstrations or physical showcases and are able to evaluate performance during therapy [19], [20]. Additionally, such devices offer verbal feedback about task execution, utilizing a predetermined array of responses [21], and can adjust the complexity of the assigned tasks on patient needs [22]. However, these systems fall short in replicating the complete interaction methods employed by human physiotherapists. These social robots lack physical interaction capabilities due to hardware constraints that hinder their ability to adequately guide the weight of the limb needing treatment. In order to enable meaningful conversations with patients, it is essential for robots to possess the capability to understand and interpret natural language. In recent research, transformer-based models such as bidirectional encoder representations from transformers (BERT) [23] and generative pre-trained transformer (GPT) [24] have emerged as powerful tools for addressing natural language tasks, including

the interpretation and answer generation. However, these models have notable limitations, including an excessively large number of parameters, strict memory requirements, significant computational delays when implemented on standard hardware, and the need for dedicated resources such as GPUs that are unfeasible for a rehabilitation robot. Given that dialogue is the primary objective, the prolonged computational times will negatively impact the perceived interaction quality introducing a latency in the response.

From the literature analysis, it emerges that successful rehabilitation robots should encompass both physical and cognitive interaction modalities. Physical interaction guarantees accurate guidance in motor tasks, emulating the hands-on methodology of therapists. Cognitive interaction, achieved through verbal communication, offers immediate feedback, as in [25], encouragement, providing a customized element, and amplifying patient involvement. The fusion of physical and cognitive interaction yields a distinctive approach that mirrors the relationship between a human therapist and a patient. This study introduces a robotic physiotherapist combining these interactions to not only assist in motor recovery but also establish an empathetic relationship with the patient.

III. PROPOSED APPROACH

The proposed robot-aided rehabilitation platform aims at providing motor rehabilitation resembling the same communication channels of the patient–therapist interaction observed during conventional rehabilitation. This section presents the model adopted to match the patient–physiotherapist interaction onto a robotic platform and proposes the integration process between the physical and cognitive components.

A. Demonstrator–Observer–Helper (D-O-H) Approach

The relationship between physiotherapist and patient is very dynamic and complex to shape. Nevertheless, a model of patient–therapist interaction has been proposed in [26]. The two actors, i.e., the patient and the physiotherapist, play a certain role, in a predefined set, and the transition from one state to another occurs whenever a cue or stimulus generates an event. In particular, the simplified model of the rehabilitation session is composed of three states. In the first state, at the beginning of the treatment, the physiotherapist explains to the patient which motor exercise is to be performed. The clinician may physically perform the task or explain it verbally. In this state, therefore, the physiotherapist plays the role of the *demonstrator*, while the patient is an *observer*. Subsequently, the physiotherapist asks the patient to start performing the exercise independently, and his/her role then becomes monitoring the correct execution of the exercise. The patient is now the *performer* and the physiotherapist is the *observer*. While performing the exercise, there may be conditions in which the patient needs help, e.g., if they do not perform the exercise correctly or if they feel pain. These stimuli cause the transition into the last state: the physiotherapist is a *helper*, while the patient is a *performer with assistance*. The roles assumed by the physiotherapist in the proposed model, called D–O–H (from the *Demonstrator–Observer–Helper* robot
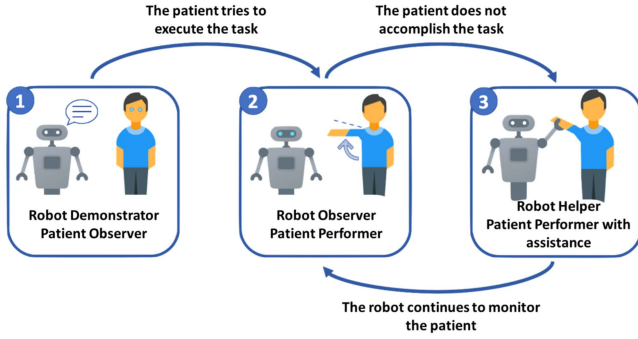
Fig. 1.   Implementation of the D–O–H model with a robotic physiotherapist.

roles) in the following, can therefore be retargeted on a robotic platform. Fig. 1 presents the implementation of the D–O–H model with a robotic physiotherapist.

B. Physical–Cognitive Architecture

The proposed approach aims at replicating the natural interactions that take place in the relationship between human physiotherapists and patients. To this purpose, Fig. 2 shows the architecture proposed to integrate physical and cognitive interaction modalities onto a robotic platform.

The physical interaction module (PHYI) refers to direct physical contact and communication between humans and robots. It involves the interaction between the robotic system and a human user, in which the robot is able to sense, interpret, and respond to human actions and intentions. The PHYI module includes the kinematic monitoring of the patient, the interaction control implemented to manage the physical interaction between the two actors, and the robot itself. Indeed, the robotic platform has to be capable of estimating the patient's kinematic performance and retrieving the position of anatomical landmarks where the robot has to physically intervene whenever he/she needs assistance. The interaction control allows the robotic arm to move until it contacts the user and assists him in performing the desired motor task. Additionally, the integration of haptic feedback would allow tactile and sensory communication between the robot and the patient, further enriching the overall interaction experience.

The cognitive interaction module (COGI) involves exchanging information about the cognitive sphere. It allows the robot to perceive and interpret the user's emotional status, goals, and intentions, and respond accordingly. As an example, multimodal responses could be exploited, such as verbal communication and facial expressions, to interpret the patient's emotional state and enable the robot to establish effective communication with the user. Specifically, to deal with linguistic tasks, the human–robot COGI module includes a Dialogue Manager (DM) implementing natural language processing to enable the robot to understand and verbally respond to human speech and actions. Both the PHYI and COGI modules exchange data with the Therapy Manager whose task is to manage the actions of the robot to ensure that the rehabilitation session is correctly carried out by verifying the completion of the prescribed steps. The Therapy Manager is detailed in Section III-B3. Furthermore,

other functionalities that have to be included in a robotic platform for robot-aided rehabilitation are a storage module and a therapist interface. Storage is necessary to collect the records of all patients and their medical information, such as the list of prescribed rehabilitation exercises, the affected limb, and their pathologies. Indeed, at the beginning of the session, the physiotherapists have the possibility to start the rehabilitation session of a specific patient by means of a dedicated interface. The patient information is retrieved from the database, and the Therapy Manager is able to accordingly set the therapy flow. Furthermore, during the rehabilitation session, all information flows, raw data, interpreted patient status, and dialogues are stored for future processing. The following explains in detail how the modules for physical and cognitive interaction are implemented by the robotic physiotherapist presented in this article.

1) Physical Interaction (PHYI): The main components of the PHYI module are the kinematic monitoring of the patient, the interaction control, and the robot. Skeleton-based approaches involving vision systems provide a valuable solution to retrieving the patient anatomical landmarks position that can be exploited to both assess the motor performance of the user and serve as input for the robot interaction control. Indeed, skeleton-based approaches are capable of estimating the upper limbs joints coordinates reported in Fig. 3(a).

Given the 3-D coordinates of the anatomical landmarks, the shoulder joints' angles can be geometrically retrieved. For the sake of brevity, the adopted procedure followed to extract the shoulder angles of the right shoulder only is reported. The same methodology can be applied also to the left side. First, the unit vector outgoing from the right shoulder and the unit vector outgoing from the user trunk are computed as

$$\vec{v} = \frac{\vec{p}_{RS} - \vec{p}_{LS}}{\|\vec{p}_{RS} - \vec{p}_{LS}\|} \qquad \vec{w} = \frac{\vec{v} \times \vec{p}_T}{\|\vec{v} \times \vec{p}_T\|} \tag{1}$$

where $\vec{p}_{RS}$ and $\vec{p}_{LS}$ are the positions of the right and left shoulder, respectively, $\vec{p}_T$ is the position of the user trunk, and the operator $\times$ represents the cross product between vectors. Finally, the unit vector $\vec{u}$ pointing downward can be retrieved as $\vec{u} = \vec{w} \times \vec{v}$. Given $\vec{v}$, $\vec{w}$, and $\vec{u}$, reported in Fig. 3(b), it is possible to project the arm vector ($\vec{p}_A = \vec{p}_{RE} - \vec{p}_{RS}$, where $\vec{p}_{RE}$ is the elbow position) into the sagittal and coronal planes to compute the shoulder flexion/extension (θ_{FE}) and abduction/adduction (θ_{AA}), respectively. In detail, the shoulder angles are computed as

$$\begin{cases} \theta_{FE} = a\cos\left(\frac{\vec{u} \cdot (\vec{p}_A - (\vec{p}_A \cdot \vec{v})\vec{v})}{\|\vec{u} \cdot (\vec{p}_A - (\vec{p}_A \cdot \vec{v})\vec{v})\|}\right) \\ \theta_{AA} = a\cos\left(\frac{\vec{u} \cdot (\vec{p}_A - (\vec{p}_A \cdot \vec{w})\vec{w})}{\|\vec{u} \cdot (\vec{p}_A - (\vec{p}_A \cdot \vec{w})\vec{w})\|}\right) \end{cases}. \tag{2}$$

The patient's kinematic performance is monitored during the rehabilitation session by assessing the target error (TE) with respect to an assigned target configuration, the completion time (CT) as the time required to complete the current repetition, and the motion smoothness in terms of mean arrest period ratio (MAPR). In detail, the TE is computed as

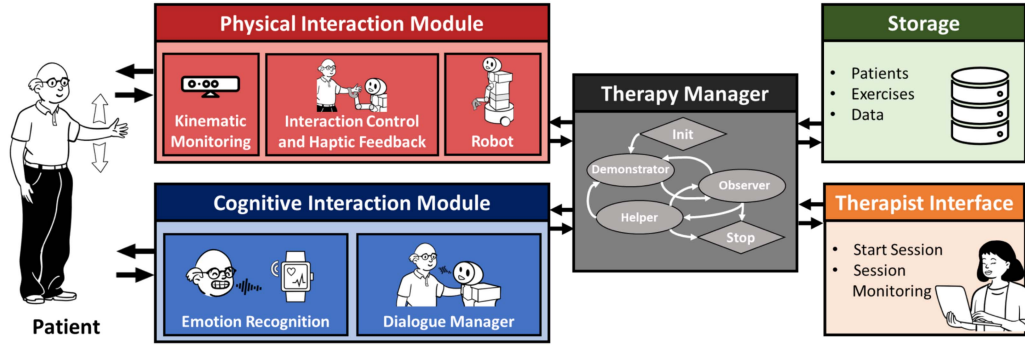$$\text{TE} = \theta(t_f) - \theta_T \tag{3}$$

Fig. 2. Architecture of the robotic physiotherapist proposed in this article.
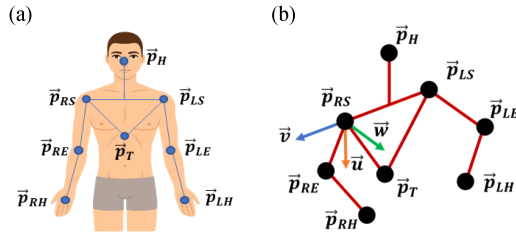

Fig. 3. (a) Anatomical landmarks reconstructed by means of a skeleton-based approach. (b) Stick model of the tracked skeleton along with the unit vector used to extract the shoulder angles.

where θ_T is the angular target to be reached and $\theta(t_f)$ is the shoulder angle executed by the participant at the end of the performed movement ($t = t_f$), during a single repetition of the task. The TE is expressed in [degrees]. The little TE evidences the capability to accurately reach the assigned target. CT is measured in [seconds], from the patient movement initiation until the reaching of the target. As already stated, the MAPR was used as a measure of motion smoothness [27], [28]. In particular, it exploits indirectly the velocity of the patients to determine whether the patient is in motion or not at a certain timestamp. In particular, the MAPR was computed as

$$\text{MAPR} = \frac{T_{\text{vel}}}{T_{\text{tot}}} \qquad (4)$$

where T_{vel} is the amount of time in which the velocity of the patient arm exceeds 10% of the maximum velocity recorded and T_{tot} is the current movement total duration. Patients affected by motor disabilities obtain lower MAPR values since they tend to perform fragmented movements [29].

Robot control has to be capable of managing physical interaction. Compliance and/or impedance control can be implemented to achieve indirect force control in unstructured environments via motion control. In this way, the robotic arm is capable of assisting the patient's arm in executing a desired motor exercise while managing the exchanged forces. The robot that should be used to replicate a human physiotherapist has to be as anthropomorphic as possible. In fact, it was demonstrated that the social acceptance of a robot in clinical settings is improved as its aspect resembles the human one [30]. At the same time, the robot should be capable of establishing physical interaction with the patient by guaranteeing a payload to the end effector compatible with the application scenario.

2) Cognitive Interaction (COGI): The COGI module has the role of managing the dialogue, exploiting the recognition of the patient status. The status of a person can be estimated in a multimodal manner. In particular, the analysis of facial expressions, patient sentences, and physiological parameters are ways to perform emotion recognition. For this purpose, the robot involved in the platform should be provided with an RGB camera to look at the patient's face during the rehabilitation treatment. The phrases pronounced by the patients are transcribed to extract the sentiment. At the same time, physiological processes such as cardiorespiratory ones and the galvanic skin response (GSR) should be measured by means of wearable instrumentation. Methods capable of combining such heterogenous signals are required to perform both emotion recognition and facial expression detection [31].

In addition, the robotic therapist is stimulated by multimodal information, such as visual input, kinematic performance, and emotional information. In synthesis, the cognitive component of a robotic therapist should at least be capable of the following:

1) having a strategy for verbal communication, for the physical demonstration of the target exercises, and for the *on-the-fly* monitoring of the patient execution with patient helping initiatives when necessary;
2) gathering and reacting to the rich and heterogeneous stimuli generated from the environment, taking forms of verbal stimuli, sensor signals, images, or known objects in the environment;
3) using the overall information for modeling the patient's emotional state in an accurate manner;
4) dynamically planning possible departures from the standard execution of therapy, and proactively acting toward the patient through empathetic stimuli and/or verbal corrections.

The DM, purposely developed for rehabilitation, should be capable of managing a verbal interaction that is not only task-oriented [32] but is also *grounded*. Grounded means that the interaction, and consequently, the interpretation, strictly depends on the environment in which the dialogue takes place. Indeed, it is possible to refer to objects, e.g., weights or sticks involved in the therapy, as they are located in the environment where the interaction takes place [as in [33], where spatial references to
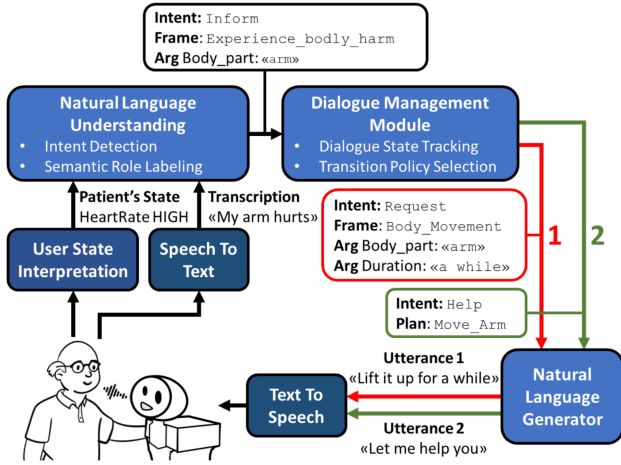
Fig. 4.   Dialogue workflow.

real objects are used to improve semantic role labeling (SRL)].
An example of the dialogue workflow is depicted in Fig. 4.

The neural architecture discussed in [34] is applied for *Speech to Text* transcription when the patient provides a verbal input. Consider a patient statement such as "*My arm hurts*" that indicates he is having trouble performing the given exercise. The *natural language understanding* (NLU) module processes the transcription by exploiting the inductive method outlined in [33]. In this case, semantic interpretation in terms of frames semantics [35], which models input sentences into meaning representation graphs, is integrated with the recognition of user intent. The interpretation as an *SRL* is decomposed into the traditional cascade of tasks: frame prediction (FP), in order to extract the evoked frames; argument identification to identify the properties mentioned in the transcription; and argument classification to classify the arguments identified in the previous step. The tasks are here resolved by a cascade of kernel-based classifiers, acting over the words of the transcription for the SRL cascade, and a single kernel-based classifier acting over the whole transcription for the intent recognition. Notably, the lightweight and high-speed nature of the kernel-based classifiers enable seamless integration directly onto the robot's on-board system if needed, differently from state-of-the-art models such as transformer-based architectures [24], recurrent neural networks (RNNs) [36], or convolutional neural networks (CNNs) [37], which require dedicated GPU resources. The NLU module enriches the transcripted sentence with behavioral signals, making thus the interpretation grounded and sensible to the state of the patient. They are processed through a dedicated biometric processor and generate a structured object summarizing behavioral signals: sensor numerical scores are mapped into predefined categorical labels. As an example, the patient's *heart rate* is processed into labels such as {LOW, NORMAL, HIGH}. The resulting interpretation is given to the *dialogue management module*, which uses a semantic graph to identify user states relying on an independent *dialogue state tracking* component. It also plans robot responses to the input, updates current states when necessary, and finally, compiles the desired linguistic output and/or triggers some actions. "*Lift it up for a while*" is

an example of a sentence constructed from the output frame by natural language generation and fed to the robot's text-to-speech module. In workflow labeled as (1) in Fig. 4, the request to the patient defines the *move* (i.e., the BODY_MOVEMENT frame) of his *arm* (as the argument BODY_PART) for a specified DURATION, i.e., a *while*. The natural language generator (NLG) module is a template rule-based approach [38] that combines the information (e.g., the previous frames and arguments) from the DM with the predefined sentences in order to answer back to the patient. The main advantage is the combination of ontological knowledge for the therapy session with predefined answers. Notice that state-of-the-art architectures, such as BART or GPT, could be implied, however, these models suffer from hallucinations. In the medical scenario, is crucial not to provide misleading or invented information to the patient, but rather answers that are grounded in the environment and sensible to the patient state. The cognitive architecture combines inductive modules, such as the language understanding one, with knowledge-based components. This heavily depends on pragmatic resources specific to a given domain (like the state tracking needed for the therapy) as well as on the involved medical knowledge bases. Notice that the DM aims at managing the patient–robot interaction in a critical scenario. It is thus recommendable to use a system that is maximally flexible and portable to new scenarios and largely verifiable in terms of the generated utterances of the dialogue as well as safe robotic actions. 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 verbal one, as in the following example:

Robotic Therapist: *How are you feeling?*
Patient: *I'm doing just fine!*
 sensor: **Patient HeartRate High**
 sensor: **Patient Sad**
Robotic Therapist: *Please take a break,*
 sit down and breathe deeply.

As in [33], sensor data are combined with linguistic information by the NLU module. The recognition of complex events (e.g., the one in the aforementioned example) is achieved by training on data combining linguistic and sensory features. No matter what stage of the physiotherapy dialogue is in progress, the dialogue needs to be revised and a break must be proposed to the patient. Moreover, after the break, the dialogue state must be resumed to invite the patient to restart his exercise.

3) Therapy Manager: The main functionalities of the Therapy Manager contribute to a service-oriented architecture, possibly hosted in a dedicated Cloud. Its role is to orchestrate all the interaction modules until the completion of the physiotherapy session. The involved behaviors of the platform correspond to actions (or plans), each organized through an *ad hoc* graph. Graphs are added incrementally to a stack of plans depending on the environmental conditions.

As the PHYI and COGI modules collect stimuli from the contexts (e.g., verbal stimuli, images, information from sensors,..), their combinations thus trigger the graph activation required at each time.

Examples of graphs are: the WELCOME graph that governs the starting phase of a session; the DEMONSTRATOR for the
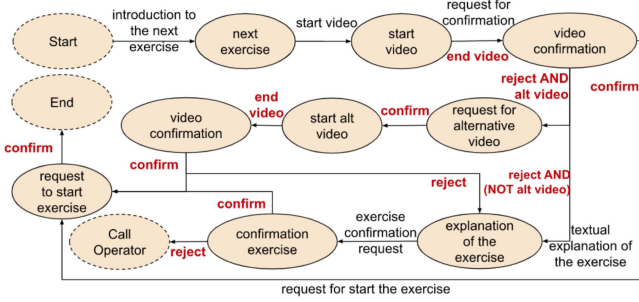
Fig. 5. DEMONSTRATOR graph. On the transitions, we report: in black, the robotic actions (utterances, e.g., *introduction to the next exercise*, or commands, e.g., *start video*); and in red, conditions satisfying the transition (user input, e.g., *confirm*, or synchronization messages, e.g., *end video*).

Fig. 6. Experimental setup used to test the proposed rehabilitation robotic platform.

visual or verbal explanations of an exercise to be performed (see Fig. 5); the OBSERVER that supports the monitoring of the patient execution, and when necessary, triggers the corrections and motivations actions needed; the HELPER, where the robotic platform physically aids the patient to correctly perform repetitions; WAITING to handle pauses requested by the patient or planned between two series of the exercise; DEFINITION REQUEST to handle patient questions about the meaning of medical terms, e.g., *flexion-extension* or *abduction*; THERAPY STATUS REQUEST for questions about the remaining repetitions or exercises; EXECUTION COMPLIANCE CHECK to deal with questions about the quality of an *on going* execution; and CLARIFICATION to trigger system's requests about additional information due to possible robot's misunderstandings.

The division into various graphs allows the stack manager to combine, sort, remove, or update them, without losing track of the dialogue status. Graphs are thus linked to each other, to solve subtasks and progress within the physiotherapy. For example, the completion of a DEMONSTRATOR graph pushes the stacking of the OBSERVER graph whenever the patient is ready and the exercise has been demonstrated as well as the patient has understood it. Similarly, the OBSERVER graph triggers the stacking of the HELPER graph if and only if the patient has committed to 2 or more significant errors or the patient asks explicitly for help.

Notice that the graph stacking approach provides a general and modular framework for dialogue management: possibly needed novel graphs can be added simply to the framework, without changing existing ones and they are ready for being used by the Stack Manager. Fig. 5 depicts the DEMONSTRATOR graph, which has the task of showing the exercise to be performed through a video and assessing the patient's understanding. Starting from the *Start* state (top left-hand corner), the video is introduced and shown (*start video* state), the patient is requested to confirm the overall understanding, and against a positive response, the D–O–H role is changed to Observer (*End* state): the corresponding OBSERVER graph is stacked accordingly. In the case of a negative response, an alternative video may be available, and it will be shown, otherwise, the robot explains verbally the exercise, and a new confirmation is requested. In the case of further rejection, the platform interrupts the session and calls a human operator (*call operator* state).
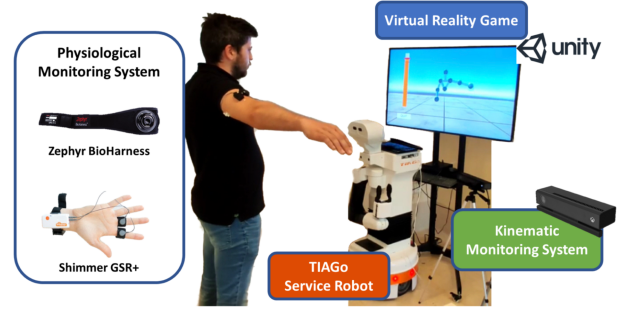
As an example, we report here some fragments of a real dialogue with a subject during a rehabilitation session, to give an idea about the linguistic and unexpressed elements involved in the dialogue, where R stands for *Robotic Platform* and P stands for *Patient*. Usually, it starts with the confirmation of the patient's data (R: *"What's your name?"* - P: *"I am... "*) and a simple chatty question (R: *"How are you today?"* - P: *"Not so well, thanks"* - R: *"Sorry to hear that, let's concentrate on the therapy."*).

Notice that the system answers using an empathetic utterance in reaction to the negative emotion expressed by the patient. Usually, the patient may be requested to explicitly confirm he/she is ready and able to use his/her limb to perform the exercises. The robotic platform (as a *demonstrator*) also initializes the next exercises (e.g., R: *"The next exercise is called Flexion-Extension of the shoulder"*) and projects a video to show to the patient the movements to be executed.

During execution, the platform monitors the quality of each repetition, in terms of kinematic indicators that show the movement and the relative errors with respect to the target. Errors made available to the cognitive component may trigger system prompts to correct the patient (e.g., R: *"Remember that you should always keep the same speed"*). Moreover, the patient may ask for information about the therapy (e.g., P: *"How many repetitions are left?"* with R: *"You need to perform N more repetitions"*). The patient may also be misunderstood (as in P: *"Can you reproduce TZXXXZYT's last song?"*) with the system asking for clarifications (e.g., R: *"Sorry, I didn't understand. Can you please repeat?"*). Finally, the robot can actively help the patient to execute a repetition upon request, or motivate him when he is performing well. In the end, it greets the patient and concludes the session.

IV. EXPERIMENTAL SETUP AND PROTOCOL

The validation of the proposed robotic physiotherapist was carried out by designing an experiment involving healthy participants. The experimental setup used in the validation process is reported in Fig. 6. The PHYI integrates the Kinect One RGB-D camera to reconstruct the participant's kinematics of upper limbs exploiting the NITE library to retrieve the 3-D position of the head, trunk, shoulders, elbows, and hands [39]. Starting from the monitored joint position, the shoulder angles are computed

by means of (1) and (2), and the number of repetitions for each exercise is counted/tracked.

The robot integrated into the PHYI module relies on the service robot TIAGo [40] (PAL Robotics S.L., Spain). The robot has an anthropomorphic arm with 7 DoFs, a lifting torso, and a wheeled mobile base. The movements of the robotic arm are controlled with compliance control in the operational space under robot operating system (ROS) at a frequency of 100 Hz. Moreover, TIAGo also includes a microphone and speakers to manage audio input and output and one head with a pan and tilt degrees of freedom mounting an RGB-D camera providing RGB and depth images of 640×480 resolution at a frame rate of 30 Hz. A screen displaying a Virtual Reality game implemented in Unity provides visual feedback about the stick model of the arm by exploiting the computed joint angles and shows demonstration videos recorded by physiotherapists. The COGI module estimates the emotional status of the participants from the FE using a convolutional neural network trained on FER-2013 dataset images to recognize three discrete emotions, i.e., "Happy," "Neutral," and "Sad" [41]. The images retrieved from the RGB camera integrated into the TIAGo head are exploited.

Finally, the physiological monitoring system measures the heart activity, respiration, and GSR of the participants. Both the electrical and respiratory activities of the enrolled participants are monitored by using the BioHarness 3.0 chest belt, developed by Zephyr Technology [42]. Starting from the raw signals, i.e., the electrocardiogram and the breathing waveform, collected at 250 and 20 Hz, respectively, and heart and respiration rates are returned by the device at 1 Hz. GSR is measured by using two electrodes of the Shimmer 3 GSR+ Unit placed on the index and middle fingers of the nondominant hand. The raw GSR is returned by the sensor at a sampling rate of 50 Hz. A fuzzy logic model is implemented to estimate the PP state of the user among "Excited," "Relaxed," "Bored," and "Stressed" from the cardiorespiratory activity and GSR [43].

The entire architecture is grounded on service-oriented architectures hosted in a dedicated cloud. All the collected data are synchronized and stored for postprocessing.

The dialogue-based platform is evaluated with respect to the patient's ability to self-correct and through quantitative performance indicators as compared to a traditional robot-aided rehabilitation robot (making no use of dialogue). The objective of the traditional robot-aided rehabilitation system, namely the simple user interface (SUI), is to show the patient the trajectory describing the exercise execution: this is rendered through virtual reality, supporting the comprehension of the error and self-corrections. On the other hand, the advance cognitive component (ACC) is the robot-aided rehabilitation system including all the functional components shown in Fig. 2 to provide verbal feedback guiding the patient whenever they commit errors or execute movements with a lower smoothness. The ACC provides suggestions and instructions to the participants to enhance their movement quality, actively helping them to correct their movement, and motivating them.

Twenty healthy participants were enrolled in this study (mean age 29.5 ± 5.1, 14 males and 6 females, right-handed). The
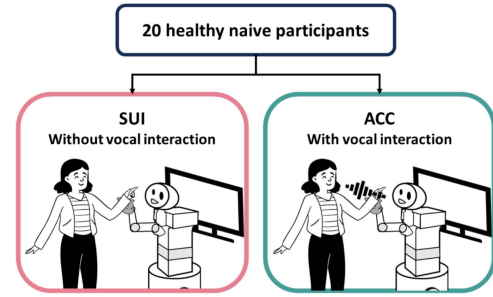

Fig. 7. Experimental protocol of the current study. Twenty healthy naive participants were enrolled and randomized into the SUI and ACC groups implementing the traditional and proposed robot-aided rehabilitation paradigms, respectively.

research was carried out with the approval of the Ethical Committee (Ethical Approval N. 03/19 PAR ComEt CBM) and in compliance with the principles of the Declaration of Helsinki. Prior to the study, all patients were fully informed about the objectives of the research and provided written consent. The volunteers were randomized into two groups, i.e., SUI and ACC, to verbally interact with the robotic system. A schematic representation of the experimental protocol is depicted in Fig. 7. All the participants in this study were naive, meaning they did not have any prior experience with rehabilitation robots or physical therapy. This deliberate choice allowed for a controlled analysis of the impact of the robot interaction, both with and without linguistic capabilities. By including participants without prior knowledge or familiarity with physiotherapy, the study aimed to assess the effectiveness and potential benefits of the robot's interaction in an unbiased manner, free from any preconceived notions or expectations. All the participants were asked to perform the following exercises 15 times.

1) Shoulder abduction/adduction: The participants were asked to perform abduction/adduction movements of their dominant upper limb to reach $\theta_T = 90°$.
2) Shoulder flexion/extension (sFE): The participants were asked to perform a flexion/extension movement of their dominant upper limb to reach $\theta_T = 90°$.
3) Functional task: The participants were asked to perform a reaching-moving-homing task inside the virtual reality game. In particular, starting from an initial testing position, the user was asked to reach a red sphere inside the 3-D virtual environment. Once the user touches the ball, another target point appears. After the second movement, the red target is moved toward the resting position. Since the task is accomplished as the participant touches the target, only the CT and MAPR of the movement were computed to determine the kinematic performance.

During the experimental sessions, participants had the freedom to move autonomously and engage in verbal interaction with the robotic system, without continuous physical guidance as typically observed in conventional rehabilitation treatments involving human therapists. In instances where participants encountered difficulties to reach the assigned target or to smoothly perform movements, the robot provided verbal interaction. However, if errors persisted over time or if participants explicitly

requested assistance, the robot assumed the role of a helper and physically interacted with the subject, providing assistance.

In the helper role, the robot acted as an end-effector machine connected to the distal part of the limb being treated (see Fig. 1). It extended its arm forward, prompting the participants to place their arm on the end effector, and subsequently, guided their limb to achieve the desired target configuration. This interaction was facilitated by the robot compliance control in the operational space, expressed as follows:

$$\tau_{\mathbf{c}} = \mathbf{g}(\mathbf{q}) + \mathbf{J}_A^T \mathbf{K}_P \tilde{\mathbf{x}} + \mathbf{J}_A^T \mathbf{K}_D \dot{\tilde{\mathbf{x}}} \tag{5}$$

where $\tau_{\mathbf{c}}$ is the joint torques, $\mathbf{g}(\mathbf{q})$ represents the gravitational torque compensation depending on the current joint configuration $\mathbf{q}$, $\mathbf{J}_A^T$ is the transpose of the analytic Jacobian, and $\mathbf{K}_P$ and $\mathbf{K}_D$ are diagonal gain matrices for proportional and derivative action, respectively. $\dot{\tilde{\mathbf{x}}}$ are the pose errors (computed as $\tilde{\mathbf{x}} = (\mathbf{x}_d - \mathbf{x}_a)$, given $\mathbf{x}_d$ and $\mathbf{x}_a$ the desired and actual poses, respectively) and its derivative. The proportional and derivative gains were set at 1000 N/m and 0.8 Ns/m, respectively. This approach allows the robot to act as an operative machine that grasps and holds the patient's limb at its end effector. This design enables easy extension of the approach to include the lower limbs, broadening the scope of rehabilitation possibilities.

A. Performance Metrics

During the experiments, metrics were calculated to assess the goodness of the NLU and the effect of the robotic system on the user in terms of kinematic performance and emotional impact.

1) Linguistical Assessment: Linguistically, the system is evaluated in terms of the accuracy in understanding the user input, that is the percentage of predicates (i.e., frames and frame elements [44]) correctly recognized at each turn. In addition, the ability to identify the patient's intent, i.e., its goal, is also assessed, to distinguish informational inputs (INFORM) from requests (REQUEST). Every utterance by the patient is thus interpreted by the robot, by performing SRL, i.e., FP, argument identification, and classification (AIC) [33]. Each input is evaluated in terms of the following.

1) Precision (P): The ratio of correctly predicted positive observations to the total predicted positive observations.
2) Recall (R): The ratio of correctly predicted positive observations to all observations in the actual class.
3) F-Measure (F1): The harmonic mean of the P and R. Therefore, this score takes both false positives and false negatives into account providing a highly informative metric.

2) Participant Kinematic Assessment: The kinematic performance assessed during the experiments, i.e., the TE, CT, and MAPR, are compared in the two experimental conditions. Moreover, the derivative of the target error (DTE) is computed as

$$\text{DTE} = \frac{1}{N} \sum_{i=2}^{N} \Big(\text{TE}(i) - \text{TE}(i-1) \Big) \tag{6}$$

where N represents the number of repetitions of a motor exercise and $\text{TE}(i)$ is the target error committed at the ith repetition highlights the effect of the interaction overtime on the kinematic

performance. The DTE is expressed in [deg/rep]. A positive DTE means that the users increase their TE over time. On the other hand, a negative DTE highlights a decrease in errors, repetition by repetition.

3) Participant Emotional Assessment: The emotional impact of the robotic physiotherapist on the users estimated by the COGI, i.e., the FE and the PP state, are collected during the experiments. In particular, the percentage of time spent in one of the possible emotional states with respect to the total duration of the therapy is assessed. In addition, subjective perception of the interaction is quantified by means of the following questionnaires, administered to the participants.

1) Propensity-to-trust scale (PTT): This validated scale is used to assess the propensity level of a user in interacting with technology. In this context, the PTT scale can be used to assess the presence of any bias in the two groups of enrolled participants [45].
2) Self-assessment manikin (SAM): The SAM allows the participants to declare their valence (V) of the response (from positive to negative), perceived arousal (A) (from high to low levels), and perceptions of dominance (D) (from low to high levels), evoked by rehabilitation robot use [46].
3) NASA-TLX: Assess the perceived workload in interacting with the robot [47]. In particular, the participants were asked to rate from 0 to 100 their experience in terms of mental demand (MD), physical demand (PD), temporal demand (TD), performance (PER), effort (EF), and frustration (FR).
4) System usability scale (SUS): Evaluate the usability of the implemented systems. It is a ten-item questionnaire with five response options for respondents ranging from strongly agree = 0 to strongly disagree = 4. The participant's scores for each question are converted into a 0–100 scale by multiplying the cumulative results by 2.5.
5) Subjective assessment of speech system interfaces (SASSI): A sensitive measure of users' subjective experiences with speech recognition systems. It is a 34-item questionnaire, with responses ranging from strongly disagree = 1 to strongly agree = 5, to evaluate the system about: system response accuracy (SRA), in terms of how precise the understanding of the system was perceived by the patients; the likeability (L), in terms of entertainment and joy felt during the dialogue; the cognitive demand (CD) including stress and mental workload; the annoyance (A) meaning how repetitive or boring the interaction was; the habitability (H) relating to whether the user knows what to say and knows what the system is doing; and the speed (S) of reaction.

Since the SASSI is used to assess the users' experience in verbally interacting with the robot, it was administered only to the ACC participants.

B. Statistical Analysis

The impact of the two tested experimental conditions is evaluated by means of statistical analysis on the aforementioned performance indicators. The Wilcoxon rank-sum test is

Fig. 8. Robotic physiotherapist roles. (a) Demonstrator: The exercise is displayed on the screen. (b) Observer: The robot monitors the kinematic performance and the emotional status of the participant. (c) Helper: The robot physically interacts with the participant to provide aid in sFE.

TABLE I
INTERPRETATION RESULTS FOR THE NLU MODULE IN THE COGNITIVE COMPONENT, WHERE AIC IS ARGUMENT IDENTIFICATION AND CLASSIFICATION, ON REAL SPOKEN DIALOGUES

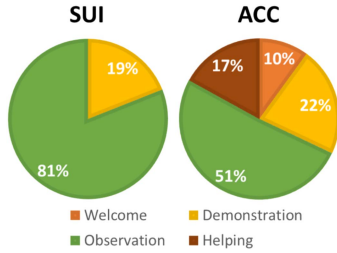| Task	Precision	Recall	F1-score
FP	92.55%	93.55%	93.05%
AIC-ExactMatch	89.03%	68.23%	77.25%
AIC-HeadMatch	91.11%	79.34%	84.81%

Fig. 9. Percentage of time the robotic physiotherapist assumed the various roles in the SUI and ACC conditions.

performed on the computed metrics. This test assesses whether a significant difference exists between the investigated conditions. In particular, the significance level was set for p-value ≤ 0.05.

V. RESULTS AND DISCUSSIONS

Frames of the experiments carried out in this work are reported in Fig. 8. The figure shows the robotic physiotherapist assuming all the implemented roles during the experiments. It is worth observing the screen displaying the demonstration video recorded with the aid of a human physiotherapist [demonstrator role, Fig. 8(a)], the system monitoring the participant autonomously executing the task [observer role, Fig. 8(b)], and the TIAGo assisting the user arm to reach the angular target during the sFE exercise [helper role, Fig. 8(c)]. As evident, when the robot played the helper role, it worked as an end-effector machine, grasping and holding the participants' distal part of the upper limb, i.e., their wrist.

Fig. 9 shows the percentage of time the robotic physiotherapist spent in the different implemented roles in the SUI and ACC conditions. It is evident that the robot performed different roles in the ACC condition. This is because the robot is provided with linguistic capabilities and it can manage cognitive roles such as the welcome one. Indeed, the robot introduces the patient to the treatment before starting with the exercises themselves. Moreover, the demonstration of the rehabilitation exercise takes more time in the ACC condition. The ACC participants asked for clarification and the robot verbally provided a detailed description of the motor task to perform. It is worth observing that the robot did not enter the helper role in the SUI condition. The enrolled healthy participants did not trigger the Therapy Manager to start the helper role in the SUI experiments. On the other hand, the ACC participants asked for aid to test the assisting capabilities of the PHYI module. The robotic physiotherapist

demonstrated the ability to manage all the implemented roles and successfully guide all the participants in completing the therapy. Table I shows the results of the interpretation for the NLU module in the cognitive component on real spoken dialogues. Data used to evaluate the system are dialogues collected during rehabilitation sessions carried out by the 20 healthy participants. It should be noted that verbal user utterances are first transcribed by the *speech to text (STT)* module, which may produce errors due to microphone failures or external noise. Ungrammatical sentences may be thus generated contributing to errors in the SRL module.

The performance measured in the FP task is very positive: the system recognizes 93.55% of the correct frames (R) and 92.55% of its output is correct (P). This means that the robot understands 93.05% (F1) of the information spoken by the user, whether asking about his/her exercise or asking it about technical terms.

The second SRL task is the identification and classification of predicate arguments (*AIC*), where the system is less accurate. This latter task is more complicated, as finer grain details are involved. In fact, within a sentence, such as *"Can I take a break to go to the bathroom,"* the system must correctly detect *"go to the bathroom"* as the provided explanation of the *break*. Similarly, if the patient wants to hear a particular song, *"Ludwig van Beethoven's fifth symphony"* is a multiword expression that must be identified and classified correctly for the model to increase P and R in the AIC task. On the contrary, when just one wrong word is missed (possibly due to some noise in the *STT* process), the system gets no performance reward. Notice that the rehabilitation session is not much affected by such errors as the system misunderstandings may always trigger requests for clarifications and ambiguities can be resolved in a few turns.

Fig. 10 reports the kinematic performance computed for the two groups. The TE resulted to be ≥ 0 in both the motor tasks and experimental conditions. This means that the participants exceeded the 90° angular target assigned. In addition, statistically significant differences are present for both tasks (p-value ≤ 0.05). The ACC participants were able to minimize the error repetition by repetition during the session. The DTE performance indicator highlights this effect. The robot verbally corrects any errors made in reaching the target by generating additional feedback to improve the motor performance. DTE assumes negative values whenever the volunteer exhibits an average reduction in error during the session. As evident, the SUI group shows a null mean DTE: there is no evidence of any improvement or deterioration in performance over time. On the other hand, the ACC group rapidly corrected their execution
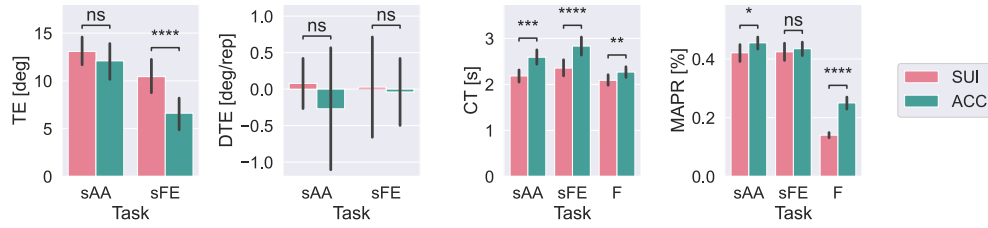
Fig. 10. Kinematic indicators computed during the experiments for the two experimental groups.
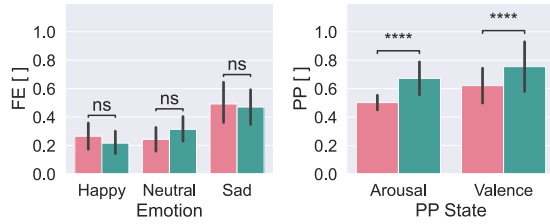

Fig. 11. Emotion indicators computed during the experiments for the two experimental groups.

during the first task. Moreover, they outperformed the SUI group also in the second exercise.

Speed and smoothness of movement also pointed out differences between the two groups. The participants who interacted with the cognitively complex platform took longer on average to perform the single repetition of both motor exercises. The robot informed the subjects in real time about the correct execution of the task regarding speed and fluidity of movement.

Fig. 11 presents the computed emotional impact of the robotic platform on the enrolled participants in terms of FE and PP state. No significant differences in FE were found between the two experimental conditions. The complex verbal interaction managed by the DM does not affect the FE of the participants. On the other hand, the physiological responses of the autonomic nervous system are able to capture certain differences induced by experimental conditions. As evident from Fig. 11, both A and V are significantly higher in the ACC condition. This means that the ACC generates physiological responses indicating greater engagement and the quality of the interaction is more positive. In particular, it was possible to appreciate physiological responses generated by verbal interaction. Whenever the robot asked a participant and/or gave feedback, a physiological response to the interaction itself was generated in the ACC participants. Table II reports the results of the self-reported questionnaires.

First of all, the results of the PTT questionnaire show that the two groups are comparable: no significant differences are evident in the results of the two groups. Volunteers belonging to both groups reported a high propensity for technology. This means that the evaluation of the SUI and ACC platforms is not affected by any bias between the two groups. On the other hand, it should be noticed as a limitation that the two groups are only composed of particularly tech-savvy users. Different results could be found by extending the study to subjects less likely to use technologically advanced devices, such as the target population of conventional motor rehabilitation.

TABLE II
RESULTS ABOUT THE SELF-REPORTED QUESTIONNAIRES

Questionnaire	ITEM	SUI	ACC	p-value
PTT	–	24.22 ± 4.08	23.36 ± 3.04	0.56
SAM	V	6.66 ± 1.41	7.90 ± 1.04	**0.05**
	A	4.77 ± 2.72	5.09 ± 2.51	0.93
	D	6.22 ± 2.16	6.45 ± 1.75	0.87
NASA-TLX	MD	3.11 ± 1.53	3.90 ± 2.46	0.64
	PD	4.88 ± 1.76	5.27 ± 3.03	0.64
	TD	4.00 ± 1.50	3.72 ± 1.90	0.52
	PER	8.22 ± 1.09	8.09 ± 1.92	0.75
	EF	3.88 ± 1.76	4.36 ± 2.15	0.64
	FR	2.88 ± 2.26	2.09 ± 0.70	0.77
SUS	–	75.00 ± 9.84	71.59 ± 6.35	0.17
SASSI	SRA	–	3.55 ± 0.90	–
	L	–	4.25 ± 0.77	–
	CD	–	3.58 ± 0.88	–
	A	–	2.86 ± 0.86	–
	H	–	2.59 ± 0.97	–
	S	–	3.14 ± 1.08	–

The emotional subjective evaluation of the platforms was carried out by the SAM administration. The participants of the ACC group rated their experience more positively than the SUI did. Indeed, the p-value was 0.05 in the V assessment. The subjective ratings of the V are in line with those computed from the PP estimation module. On average, all the volunteers declared that the interaction with the robot was exciting and they had control over the situation. These aspects are reflected in the A and D scores. From the point of view of workload perception, no significant differences are evidenced in the NASA-TLX questionnaire. In general, the interaction with the robot did not load the users in terms of MD, TD, and FR. Higher scores were obtained for PD and EF since the proposed session consisted purely of physical exercises. All participants reported a high level of success, as highlighted by the high PER scores. Moreover, the participants were asked to rate the usability of the two implemented experimental conditions. The SUS scores of 75.00 ± 9.84 and 71.59 ± 6.35 were obtained for SUI and ACC, respectively. Both the scores are above average since an SUS score ≤ 68 is considered below average [48].

Finally, the SASSI questionnaire, used to assess the spoken interaction with the robotic physiotherapist, has been compiled only by the group involved in the ACC scenario. Notice that measures such as *Accuracy* (SRA), L, H, and S are to be maximized, as higher values reflect better-perceived performance. On average, human subjects were assigned a good score, suggesting that the system interaction is accurate, enjoyable, and quite

fast. Other factors, e.g., CD and A, are meant to be minimized. Results, in this case, show that the system is still perceived as somehow repetitive ($A = 2.86 \pm 0.86$).

VI. CONCLUSION

This article presented the integration of physical and cognitive components on the same robotic platform for rehabilitation to build up a novel system, able to emulate the common interaction between patients and physiotherapists. In particular, the proposed approach envisages a vision of integrating systems of various kinds to improve the quality of human–machine interaction, and thus, guarantee a more effective motor rehabilitation treatment. The ability to physically interact to manually move the limb to be treated together with natural language skills along with the quantitative monitoring of multimodal parameters allows the development of a cognitively complex system capable of autonomously managing a rehabilitation session. The approach presented in this article was implemented on a service robot and an experiment was carried out to evaluate the performance of the linguistic component as well as the impact generated on the user in terms of kinematic performance and emotional status. The experiments carried out revealed how the presence of a module that handles natural verbal communication is able to improve the kinematic performance of the participants, who receive feedback during the execution of the proposed motor exercises. After a few repetitions, ACC participants reduced the target error to 5 deg. Furthermore, the verbal interaction allows an empathetic relationship to be established between man and machine, evidenced both by the emotional state analyzed by the cognitive component and the subjective ratings of the participants (the perceived V of the interaction was about 18% higher in the ACC condition).

This work demonstrated how the integration of different signals, such as facial expressions, biosensor data, and text can effectively lead to more natural human–robot interactions in the robot-aided rehabilitation context. The intention of the work was not to extensively model the patient's cognitive status but rather to incorporate these signals to improve the interaction, conversely from conventional robot-aided rehabilitation methods that do not involve verbal interaction.

As this study exclusively involved healthy participants to prove the feasibility and potential advantages of the proposed approach, future works will be devoted to implement the robotic physiotherapist into a clinical setting to quantitatively evaluate its rehabilitative effectiveness on patients affected by pathological conditions.

VI. ACKNOWLEDGMENT

REFERENCES

[1] H. M. Qassim and W. Wan Hasan, "A review on upper limb rehabilitation robots," *Appl. Sci.*, vol. 10, no. 19, 2020, Art. no. 6976.

[2] E. D. Oña, J. M. Garcia-Haro, A. Jardón, and C. Balaguer, "Robotics in health care: Perspectives of robot-aided interventions in clinical practice for rehabilitation of upper limbs," *Appl. Sci.*, vol. 9, no. 13, 2019, Art. no. 2586.

[3] H. Rodgers et al., "Robot assisted training for the upper limb after stroke (RATULS): A multicentre randomised controlled trial," *Lancet*, vol. 394, no. 10192, pp. 51–62, 2019.

[4] C. Tamantini et al., "Patient-tailored adaptive control for robot-aided orthopaedic rehabilitation," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2022, pp. 5434–5440.

[5] N. Karanikas and B. D. Jani, "Frequency of examination and perceived contribution of factors relating to work-related musculoskeletal disorders of physiotherapists," *Int. J. Occup. Saf. Ergonom.*, vol. 28, no. 2, pp. 1112–1129, 2022.

[6] A. Langer, R. Feingold-Polak, O. Mueller, P. Kellmeyer, and S. Levy-Tzedek, "Trust in socially assistive robots: Considerations for use in rehabilitation," *Neurosci. Biobehavioral Rev.*, vol. 104, pp. 231–239, 2019.

[7] J. Wu, H. Cheng, J. Zhang, S. Yang, and S. Cai, "Robot-assisted therapy for upper extremity motor impairment after stroke: A systematic review and meta-analysis," *Phys. Ther.*, vol. 101, no. 4, 2021, Art. no. pzab010.

[8] L. Zollo, F. J. Badesa, R. Morales, N. Garcia-Aracil, J. M. Sabater, C. Perez-Vidal, and E. Fernandez, "Multimodal interfaces to improve therapeutic outcomes in robot-assisted rehabilitation," *Springer Tracts Adv. Robot.*, vol. 94, no. 6, pp. 321–343, 2014.

[9] D. Formica, L. Zollo, and E. Guglielmelli, "Torque-dependent compliance control in the joint space for robot-mediated motor therapy," *J. Dyn. Syst. Meas. Control*, vol. 128, no. 1, pp. 152–158, 2006.

[10] D. C. Porras, P. Siemonsma, R. Inzelberg, G. Zeilig, and M. Plotnik, "Advantages of virtual reality in the rehabilitation of balance and gait: Systematic review," *Neurology*, vol. 90, no. 22, pp. 1017–1025, 2018.

[11] A. A. Chellal et al., "Robot-assisted rehabilitation architecture supported by a distributed data acquisition system," *Sensors*, vol. 22, no. 23, 2022, Art. no. 9532.

[12] C. Tamantini et al., "Tailoring upper-limb robot-aided orthopedic rehabilitation on patients' psychophysiological state," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 3297–3306, 2023.

[13] F. Yuan, E. Klavon, Z. Liu, R. P. Lopez, and X. Zhao, "A systematic review of robotic rehabilitation for cognitive training," *Front. Robot. AI*, vol. 8, 2021, Art. no. 605715.

[14] I. Aprile et al., "Robotic rehabilitation: An opportunity to improve cognitive functions in subjects with stroke. An explorative study," *Front. Neurol.*, vol. 11, 2020, Art. no. 588285.

[15] T. Bickmore and T. Giorgino, "Health dialog systems for patients and consumers," *J. Biomed. Informat.*, vol. 39, no. 5, pp. 556–571, 2006.

[16] J. Hocking and A. Maeder, "Motivational embodied conversational agent for brain injury rehabilitation," in *Telehealth Innovations in Remote Healthcare Services Delivery*. Amsterdam, The Netherlands: IOS Press, 2021, pp. 37–46.

[17] A. Russo et al., "Dialogue systems and conversational agents for patients with dementia: The human-robot interaction," *Rejuvenation Res.*, vol. 22, no. 2, pp. 109–120, 2019.

[18] P. Kellmeyer, O. Mueller, R. Feingold-Polak, and S. Levy-Tzedek, "Social robots in rehabilitation: A question of trust," *Sci. Robot.*, vol. 3, no. 21, 2018, Art. no. eaat1587.

[19] S. M. Nguyen, P. Tanguy, and O. Remy-Neris, "Computational architecture of a robot coach for physical exercises in kinaesthetic rehabilitation," in *Proc. 25th IEEE Int. Symp. Robot Hum. Interactive Commun.*, 2016, pp. 1138–1143.

[20] A. Lotfi, C. Langensiepen, and S. W. Yahaya, "Socially assistive robotics: Robot exercise trainer for older adults," *Technologies*, vol. 6, no. 1, p. 32, 2018.

[21] M. H. Lee, D. P. Siewiorek, A. Smailagic, A. Bernardino, and S. B. Badia, "Towards personalized interaction and corrective feedback of a socially assistive robot for post-stroke rehabilitation therapy," in *Proc. 29th IEEE Int. Conf. Robot Hum. Interactive Commun.*, 2020, pp. 1366–1373.

[22] A. Martín, J. C. Pulido, J. C. González, Á. Garcia-Olaya, and C. Suárez, "A framework for user adaptation and profiling for social robotics in rehabilitation," *Sensors*, vol. 20, no. 17, 2020, Art. no. 4792.

[23] J. D. M.-W. C. Kenton and L. K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. naacL-HLT*, 2019, vol. 1, p. 2.

[24] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," 2018. [Online]. Available: https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf

[25] N. Maclean, P. Pound, C. Wolfe, and A. Rudd, "A critical review of the concept of patient motivation in the literature on physical rehabilitation," *Soc. Sci. Med.*, vol. 50, no. 4, pp. 495–506, 2000.

[26] M. J. Johnson, M. Mohan, and R. Mendonca, "Therapist-patient interactions in task-oriented stroke therapy can guide robot-patient interactions," *Int. J. Social Robot.*, vol. 14, no. 6, pp. 1527–1546, 2022.

[27] V. Di Lazzaro et al., "Combining robotic training and non-invasive brain stimulation in severe upper limb-impaired chronic stroke patients," *Front. Neurosci.*, vol. 10, 2016, Art. no. 88.

[28] D. Hwang, J.-H. Shin, and S. Kwon, "Kinematic assessment to measure change in impairment during active and active-assisted type of robotic rehabilitation for patients with stroke," *Sensors*, vol. 21, no. 21, 2021, Art. no. 7055.

[29] F. Scotto di Luzio et al., "Bio-cooperative approach for the human-in-the-loop control of an end-effector rehabilitation robot," *Front. Neurorobot.*, vol. 12, 2018, Art. no. 67.

[30] B. Liu, P. Markopoulos, and D. Tetteroo, "How anthropomorphism affects user acceptance of a robot trainer in physical rehabilitation," in *Proc. 12th Int. Conf. Health Inform.*, 2019, pp. 30–40.

[31] L. Cristofori et al., "HeAL9000: An intelligent rehabilitation robot," in *Proc. SMARTERCARE Workshop*, 2021, pp. 29–41.

[32] D. R. Traum and E. A. Hinkelman, "Conversation acts in task-oriented spoken dialogue," *Comput. Intell.*, vol. 8, no. 3, pp. 575–599, 1992.

[33] A. Vanzo, D. Croce, E. Bastianelli, R. Basili, and D. Nardi, "Grounded language interpretation of robotic commands through structured learning," *Artif. Intell.*, vol. 278, 2020, Art. no. 103181.

[34] A. Baevski, H. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0: A framework for self-supervised learning of speech representations," in *Proc. 34th Int. Conf. Neural Inf. Process. Syst.*, 2006, pp. 12449–12460.

[35] C. F. Baker, C. J. Fillmore, and J. B. Lowe, "The Berkeley framenet project," in *Proc. 36th Annu. Meeting Assoc. Comput. Linguistics, 17th Int. Conf. Comput. Linguistics*, 1998, pp. 86–90.

[36] D. W. Otter, J. R. Medina, and J. K. Kalita, "A survey of the usages of deep learning for natural language processing," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 604–624, Feb. 2021.

[37] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[38] S. W. McRoy, S. Channarukul, and S. S. Ali, "An augmented template-based approach to text realization," *Natural Lang. Eng.*, vol. 9, no. 4, pp. 381–420, 2003.

[39] R. A. Clark, B. F. Mentiplay, E. Hough, and Y. H. Pua, "Three-dimensional cameras and skeleton pose tracking for physical function assessment: A review of uses, validity, current developments and kinect alternatives," *Gait Posture*, vol. 68, pp. 193–200, 2019.

[40] J. Pages, L. Marchionni, and F. Ferro, "TIAGo: The modular robot that adapts to different research needs," in *Proc. Int. Workshop Robot Modularity, IROS*, 2016, vol. 290.

[41] P. Giannopoulos, I. Perikos, and I. Hatzilygeroudis, "Deep learning approaches for facial emotion recognition: A case study on FER-2013," in *Advances in Hybridization of Intelligent Methods*. Berlin, Germany: Springer, 2018, pp. 1–16.

[42] R. Cittadini, C. Tamantini, F. Scotto di Luzio, C. Lauretti, L. Zollo, and F. Cordella, "Affective state estimation based on Russell's model and physiological measurements," *Sci. Rep.*, vol. 13, no. 1, 2023, Art. no. 9786.

[43] F. J. Badesa et al., "Dynamic adaptive system for robot-assisted motion rehabilitation," *IEEE Syst. J.*, vol. 10, no. 3, pp. 984–991, Sep. 2016.

[44] C. Fillmore, "Frames and the semantics of understanding," *Quaderni di Semantica*, vol. 6, pp. 222–254, 1985.

[45] S. A. Jessup, T. R. Schneider, G. M. Alarcon, T. J. Ryan, and A. Capiola, "The measurement of the propensity to trust automation," in *Proc. Int. Conf. Human.- Comput. Interact.*, 2019, pp. 476–489.

[46] T.-M. Bynion and M. T. Feldner, *Self-Assessment Manikin*. Berlin, Germany : Springer, 2017, pp. 1–3.

[47] L. M. A. La Bara, L. Meloni, D. Giusino, and L. Pietrantoni, "Assessment methods of usability and cognitive workload of rehabilitative exoskeletons: A systematic review," *Appl. Sci.*, vol. 11, no. 15, 2021, Art. no. 7146.

[48] A. Bangor, P. Kortum, and J. Miller, "Determining what individual SUS scores mean: Adding an adjective rating scale," *J. Usability Stud.*, vol. 4, no. 3, pp. 114–123, 2009.