2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Assessment and Benchmarking of XoNLI: a Natural Language Processing Interface for Industrial Exoskeletons

Olmo A. Moreno F.[1], Raajshekhar Parameswari[1], Christian Di Natali[1], Darwin G. Caldwell[1], and Jesus Ortiz[1]

Abstract— **Industrial exoskeletons are a potential solution for reducing work-related musculoskeletal disorders during carrying or lifting tasks. Having sensors, electrical/pneumatic actuators, and control systems, active exoskeletons present a more versatile control system because it is possible to select different assistive strategies based on the performed task. From this perspective, human-machine interaction is required to safely open basic exoskeleton domains to the user and provide an adaptable setup system. This article presents the assessment and benchmarking of the novel XoLab Natural Language Interface, a voice user interface for interaction and configuration of industrial active exoskeletons. The evaluation of the novel interface was performed by 17 participants who completed the setup and operational activities while wearing the XoTrunk exoskeleton. The benchmark consisted of a comparison of the presented device with previous adaptable interfaces for the exoskeleton: the user command interface and the monitor system interface. The results showed that although the novel interface demonstrated a considerable lag in the time response, it was more attractive, stimulating and novel than the standard one. However, the standard interface obtained favourable results over the user command interface and the voice interface perspicuity and efficiency.**

I. INTRODUCTION

Despite the latest industrial technological advantages, workers in factories are still exposed to work-related musculoskeletal disorders (WMSDs) when performing manual material handling tasks (MMH) [1]. A potential solution to reduce the risk of injures and WMSD is the use of exoskeletons in the working place [2]. An exoskeleton is described as an electromechanical system worn in parallel to the human body [3]. It is capable of augmenting human ability in different body parts such as the knees, hips, lower back, elbows, and shoulders. By its application, exoskeletons can be categorized as medical, military, commercial, and industrial units [4]. Also, exoskeletons can be marked by the actuation type as passive or active exoskeletons [5]; being the active ones designed with sensors, electrical motors or pneumatic actuators, while the passive ones are composed of elastic bands and spring-type mechanisms. As mentioned in [6], active exoskeletons have intrinsic potential for researchers and future market interest. Active exoskeletons can be modulated by different control strategies and can be precisely adapted to diverse tasks and users. In the aim of proper force modulation, active exoskeletons need to open certain domains to the user (controller gains, parameters and user information) and be simple modified and suited to the task/user application.

In this paper, we introduce the XoLab Natural Language Interface (XoNLI), a voice user interface (VUI) designed to interact with industrial active exoskeletons. The goal is to verbally modify and adjust some domains of the exoskeleton XoTrunk [7] using the VUI: XoNLI. A VUI is a nonvisual user interaction interface (No-UI) , a part of the vocal interface technologies [8]. These systems rely on the use of natural language processing (NLP), as described in [9], it is an intersection of artificial intelligence and linguistics. To create an efficient NLP system, it is necessary to employ and index large volumes of text information retrieval. Different studies have been documented using NLP adapted to devices for medical [9], [10] and industrial [11], [12] purposes, the goal is to provide users a more natural interaction with speech capabilities. An important topic about the use of this technology is: how much more intuitive can a system be? As reported in [13], VUI's differ from typical graphical user interfaces (GUI) in the interaction type; users pass from commanding systems with their hands to having more natural verbal interactions. However, these interactions have their limits; as presented in [14], VUI's are far from having natural conversations, since the interaction mechanic consists of trial/error or question/answer conversations. There are several challenges regarding audio and speech to be considered during VUI design, however, the most important one is how to guide the user to a certain piece of information? [15].

To address this issue, human-machine interaction with a VUI should rely on user expectations and expertise. As described in [16], users have experience in human to human interaction, in this scenario, new users have expectations about the same interaction with the VUI on a conversational speech style known to humans. Therefore, a set of guidelines are required to maintain that conversational interaction; Nowacki has listed these in [17] as: i) guidance, ii) feedback, iii) error management, iv) user control, v) dialogue management, vi) system capabilities, vii) user diversity, and viii) user experience.

In this case study, we conduct usability benchmarking for the industrial exoskeleton XoTrunk using three different setup systems: a) the novel XoNLI-VUI, b) the User Command Interface (UCI), and c) the Monitor System Interface (MSI). We want to assess the quality attributes of the VUI as a new technology and compare the usability between systems previously integrated into the industrial exoskeleton XoTrunk. The heuristics-based [18] in this case study

---

[1] O. A. Moreno F., R. Parameswari, C. D. Natali, D. G. Caldwell, and J. Ortiz are with Advanced Robotics Department, Istituto Italiano di Tecnologia, Via Morego 30, 16163, Genova, Italy {`olmo.moreno, raajshekhar.parameswari, christian.dinatali, darwin.caldwell, jesus.ortiz`}`@iit.it`

979-8-3503-8457-4/24/$31.00 ©2024 IEEE

about the VUI are oriented to: i) help and information, ii) suitable feedback, and iii) natural dialogue structure. For the benchmark study, we decided to limit the functions to assess according to the current actions of the MSI interface (described in Sec. II-A.3). This is the standard framework to adapt and configure the industrial exoskeleton XoTrunk.

A. Motivation

Our exoskeletons were designed and developed in the Wearable Robots, Exoskeletons and Exosuits Laboratory (XoLab) at Istituto Italiano di Tecnologia (IIT) in partnership with the Italian Workers' Compensation Authority (INAIL). XoLab has achieved results for back-support assistance exoskeletons such as XoTrunk [19] and upper-limbs exoskeleton: Shoulder-sideWINDER [20]. Currently, giving access to the user into some domains of an exoskeleton is a matter of importance. However, the user will command a mechanical exosuit, so it is primordial to allow this access safely and limit some actions. Under laboratory conditions, tests and evaluation of an exoskeleton occur under the supervision of an exoskeleton manager; usually a scientific member of the group, who supervises the experimental session. In the past, the exoskeleton controller was adjusted by a command line interface (CLI) operated by the exoskeleton manager. In 2018, the group proposed the idea of developing novel tools and applications to improve the control, adjustment, and configuration of active exoskeletons beyond the CLI without the supervision of an exoskeleton manager, thus opening the idea of direct user interaction with the exoskeleton. In our previous work [21], we presented the UCI and MSI as a solution to overcome the challenge of secure user access to the exoskeletons. These solutions were intended to make the interaction between the user and exoskeleton closer, and today the MSI is the standard tool to adapt and modify the industrial exoskeleton XoTrunk. With these results, we feel motivated to integrate novel applications such as the VUI into our exoskeletons. The academic contribution of this work relies on the integration and benchmarking of emerging technologies for adaptable setup systems in the field of industrial exoskeletons, thus reducing the gap of user interaction with wearable supportive devices. We believe that it is important to simplify the tasks and improve the comfort and performance of users by testing disruptive technologies.

B. State of the art

We have revised the following literature closer to an NLP-exoskeleton interaction. The next research material contains VUI frameworks interacting with medical and industrial exoskeletons.

In [22] and [23] is presented "Bridge", an assistive upper-limb exoskeleton controlled by multimodal interfaces for severely impaired patients. The VUI is based on the Google Cloud Speech-to-text API, where the user provides standard instructions in Italian (no activation word, nor conversational algorithm is implemented) to control the motion of the end-effector of the exoskeleton.

(a) (b) (c) (d)

Fig. 1: XoLab adaptable setup systems for the industrial active exoskeletons. (a) XoLab Natural Language Interface (XoNLI). (b) User Command Interface (UCI). (c) Monitor System Interface (MSI). (d) Back-support exoskeleton XoTrunk.

Then, in the work of Kim et al. [24] a voice-activated exo-suit for upper-limb assistance is studied. The VUI they implemented in the industrial exoskeleton was integrated using a commercial off-the-shelf device. This system manages seven different words in English without using an activation command, these words directly affect the exoskeleton actuators to achieve desired arms positions.

"Remotion" is a lower-limb medical exoskeleton presented in [25], [26]. It uses a VUI based on a smartphone with a dedicated application for voice recognition. The system contains an activation word and eight-word commands in the Russian language.

Finally, Wang et al. in [27] presented another smartphone app-based VUI for a hand-restoring exoskeleton. The interface does not present activation words or natural conversation capabilities. Instead, the interface function relies on five-word commands in the English language to operate the exoskeleton.

II. METHODS AND MATERIALS

A. System description

1) XoLab natural language interface: XoNLI is a novel human-machine interaction system that configures industrial active exoskeletons. The system is composed of the portable wearable device XoNLI-VUI and the NLP server, as depicted in Fig. 2. The purpose of the VUI is to be an interactive element for the user. It consists of a touch sensor, a microphone, and a loudspeaker controlled by an embedded system; each time the user presses the button and speaks, the VUI records the audio file, sends it to the server, and plays the response back. The NLP server contains three main libraries: a) Automatic Speech Recognition (ASR), b) Natural Language Understanding (NLU), and c) Text-to-Speech (TTS). The ASR system [28] converts the audio to text and sends the output into the NLU library to decode the former voice message. From a list of keywords, the NLU detects the common words contained in the previous audio file. If any of these matches the list, it sends a predefined text message as an answer to the TTS system. Finally, the TTS system generates an audio output from the NLU message and sends it to the XoNLI-VUI to be reproduced [29], and a command to the exoskeleton. The NLU was designed with

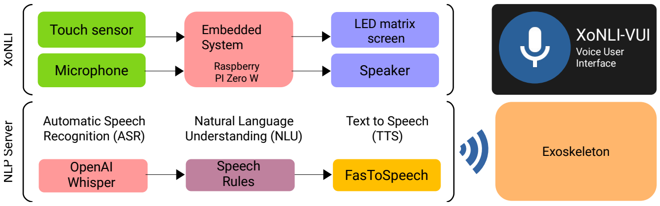

Fig. 2: XoLab Natural Language Interface (XoNLI) and NLP Server block diagram. The NLP server is wirelessly connected to the exoskeleton.

the speech rules of a natural conversation flow to achieve the interaction level with the functions from the MSI interface described in Sec. II-A.3. When a keyword match occurs in the NLU, the server sends the matched command to the exoskeleton to perform the action. The proposed dialogue model in this work follows the scheme of a natural language interface (NLI) given by [30], they presented a dialogue model $f^{dialogue}(P, u \leq_t, U) \rightarrow (s_1, s_2 ..., s_{|U|})$, where P is a given persona, $u \leq_t$ are the previous utterances (keywords to identify by the NLU rules system), U is the candidate of next utterances, and $s_1, s_2 ..., s_{|U|}$ the output scores dialogues from the model. In Sec. IV-B.5 is presented an example of the NLI dialogue model.

2) User command interface: This wearable gadget attached to the exoskeleton is an adaptable setup system designed as a unique hardware and software for the field of industrial exoskeletons. The UCI affords the exoskeleton's configuration with the principles of security and interaction using a button-based control and a digital screen. Besides the basic functions from the MSI (see Sec. II-A.3), the UCI provides secure fingerprint authentication, user database, and working profile sessions [31].

3) Monitor system interface: This is a visual framework running on a computer (see Fig. 1(c)) to setup and adjust the operational parameters of the back-support exoskeleton XoTrunk [21]. The interface was created because of the lack of user interaction with the exoskeleton; currently, this interface is used to perform four standard actions: a) calibration (R^{bn}), b) activation, c) introduction of user information (weight: M_{ub} and height: L_{ub}), and d) modification of the exoskeleton's force assistance (k_{acc}).

4) XoTrunk: The active back-support exoskeleton XoTrunk (see Fig. 1(d)) was designed to support MMH activities. The exoskeleton structure is composed of a rigid frame attached like a backpack to the user's body. It has three passive joints connected from the hips to the tights, and the assistance is powered by two DC brushless motors. As a result, force (up to $30Nm$) is applied in the sagittal plane between the torso and the tights [32]. The XoTrunk control strategy is described in [33], it uses an accelerometer raw signal from the inertia measurement unit (IMU) attached at the sternum of the user's trunk. The accelerometer measures the specific force f^b (the coordinate of frame b where the body is moving), this force is defined as:

$$f^b = R^{bn}(a^n - g^n) \tag{1}$$

where a^n and g^n represent the linear acceleration and gravity vector of the IMU sensor in the navigation frame n. An inverse rotation matrix R^{bn} is constructed when the exoskeleton is calibrated (to achieve it, the user must stand still), thus, the gravity direction in the sagittal plane x is registered. Finally, the acceleration control strategy τ_{acc} is described in the signal:

$$\tau_{acc} = K_{acc}(R^{nb} f^b)_x M_{ub} L_{ub} \tag{2}$$

in this equation, the accelerometer gain K_{acc} is an adimensional range between 0.0 and 1.0, M_{ub} is the user's upper body mass, and L_{ub} is the user distance between the hip and the center of mass. This strategy ensures the force assistance according to the user's statics and dynamics because it uses inclination-based and acceleration-based components.

B. Assessment metrics

The standardized assessment metrics used in this study were selected from the user-centered evaluation for wearable robotics devices (WRD), a user research method platform provided by the Interactive Usability Toolbox (IUT) [34].

NASA Task Load Index (NASA-TLX). The tool is designed to obtain workload estimates from one or more activities. It consists of six subscales that represent independent clusters of variables: mental, physical, and temporal demands, frustration, effort, and performance [35].

Net Promoter Score (NPS). It is a metric of a single question that measures how likely users are to recommend a product to others. The respond scale ranges from 0 to 10, and users are grouped into three categories: promoters, passively satisfied, and detractors [36].

Single Usability Metric (SUM). It is a usability tool that encapsulates most of information in four common metrics: task completion rates, average number of errors, average time on task, and post-task satisfaction [37].

Speech User Interface Service Quality (SUISQ). This is a standardized instrument for evaluating the usability of interactive VUI. It contains 25 items on a seven-point Likert scale. The questionnaire assesses variables such as the quality of the delivered service, in addition to the quality of speech input and output [38].

System Usability Scale (SUS). This measure of usability is a survey that consists of ten questions, each with a five-point Likert scale [39].

User Experience Questionnaire (UEQ). It consists of 26 items that cover six factors: attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty [40].

III. EXPERIMENTAL EVALUATION

A. Participants

Seventeen healthy subjects (weight: 70.17 ± 11.49 kg, height: 1.74 ± 0.07 m) with no history of low-back pain participated in the experiment. Among the participants, 8 wore glasses when watching a screen, and none of them presented with hearing difficulties. The experiment was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of Liguria (protocol no.: CER Liguria 001/2019).

B. Experiment design

The experimental design is shown in Fig. 3, which consists of the assessment of three different adaptable setup systems for the back-support exoskeleton XoTrunk: a) the standard MSI, b) the wearable UCI, and c) the novel XoNLI. Each participant was asked to perform four activities for every adaptable setup system previously mentioned while wearing the exoskeleton. The activities are: (i) calibrate the exoskeleton, (ii) activate the exoskeleton, (iii) introduce user weight and height, and (iv) perform the standard lifting. Activities "i, ii and iii" belong to the exoskeleton preparation sequence, and activity "iv" is the post-exoskeleton initialization, when the user usually starts performing a lifting task. After completing each activity (in every adaptable setup platform), the participant answered the following surveys: (1) CoP, (2) SUM, and (3) task completion time (TCT). At the end of the four activities in every adaptable setup platform (post experiment evaluation), the participant was requested to answer the following surveys: (4) NASA-TLX, (5) NPS, (6) SUS, and (7) UEQ. Only after completing all the activities using the XoNLI platform did the participant answer (8) SUISQ. The preparation sequence for the exoskeleton (activities "i, ii, and iii") is performed once the user interacts with one of the setup platforms. Then, activity "iv" (standard lifting task) is performed three times and on each trial the user modifies the exoskeleton force assistance gain (k_{acc}). When the user finishes the entire experiment, it will have completed all the activities ("i - iv") three times, each for every adaptable setup system. In Fig. 4 is depicted activity "iv" (standard lifting task, presented in [32] and [41]), as seen on the figure, this activity is a lifting sequence of a $10kg$ load that the user lifts from a fixed height ($50cm$) to the chest area, and the load is placed down to the base. The standard lifting task (iv-a) was performed using $k_{acc} = 0.0$ (transparency mode), then (iv-b) $k_{acc} = 0.1$ and (iv-c) $k_{acc} = 0.2$; after the second trial, the participant must notify if there was a change of perception (CoP) when modifying the force assistance (k_{acc}). Each participant was requested to say "I need your help" at the beginning of activity "i" with the XoNLI platform as a brief introduction to the system; then, they were requested to follow the system's instructions.

IV. RESULTS AND DISCUSSION

The average participant time for the experiment was $56m$, the participants evaluated the platforms in the following order: a) MSI, b) UCI, and C) XoNLI. The results are presented in two levels: A) post activity results and B) post experiment results. All statistical studies were executed with 17 participants ($n = 17$). Results IV-A.1, IV-A.2, IV-B.1, IV-B.2, IV-B.3, and IV-B.4 a are categorized as subjective results, while results IV-A.3 and IV-B.5 are objective results.

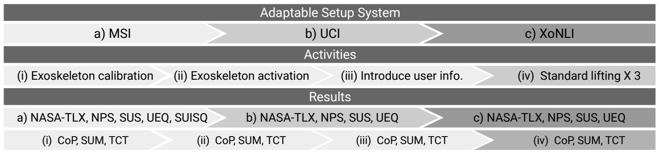

Fig. 3: Experiment design structure, three adaptable setup systems to assess: a) MSI, b) UCI, and c) XoNLI; four activities to perform for each setup system, activities "i - iii" are for the exoskeleton preparation, and activity "iv" is task performance; lastly, two levels of results: a) post-activity results, and b) post-experiment results.

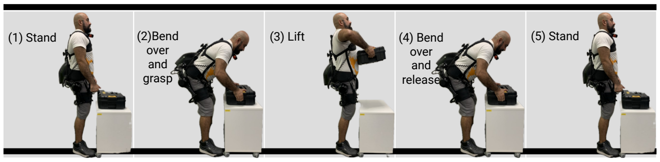

Fig. 4: Standard lifting sequence.

A. Post activity

1) Change in Perception Survey: Participants reported change of perception when performing the lifting task (see Fig. 3(b)) with every one of the setup platforms. Activity "iv-a" was carried out with a $k_{acc} = 0.0$, which indicates transparency mode. Force assistance was modified to $k_{acc} = 0.1$ and $k_{acc} = 0.2$ for the second and third trials (activities "iv-b" and "iv-c"), respectively. Table I shows that all participants using the UCI platform perceived a change in force assistance from transparency mode ($k_{acc} = 0.0$) to the following configuration ($k_{acc} = 0.1$). All participants in all platforms perceived a change in force assistance for $k_{acc} = 0.2$.

TABLE I: Change in Perception Survey

Activity	**Force assistance gain**	**MSI**	**UCI**	**XoNLI**
(iv-b)	$k_{acc} = 0.1$	76.47%	100.0%	88.24%
(iv-c)	$k_{acc} = 0.2$	100.0%	100.0%	100.0%

2) Single Usability Metric: Fig. 5 shows the results of the SUM evaluation for three different attributes: a) task ease, b) satisfaction, and c) time on task. These attributes were assessed for each activity using three different setup platforms. The UCI platform registered the lowest task ease score in activity "ii" with mean 3.94 ± 1.29. Activity "iv" had the highest task ease score with mean 4.88 ± 0.33. The lowest satisfaction score was in activities "i" with mean 4.0 ± 1.17 using the UCI system. In contrast, the highest satisfaction score was in activities "i" and "iv" with mean 4.88 ± 1.11 using the XoNLI system. Finally, the highest score in time satisfaction was recorded in activity 'iv-c" with mean 4.87 ± 0.56 using the MSI system, however, the lowest satisfaction time score was recorded on the XoNLI platform with mean 3.94 ± 1.34. A Friedman test was conducted using the three setup platforms for each activity. The results showed that there were statistically significant differences on the attribute

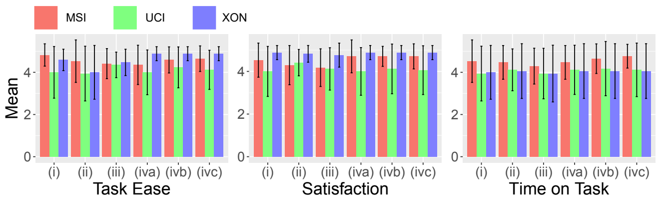

Fig. 5: SUM benchmarking post activity. Three attributes evaluated during six activities using the three different platforms.

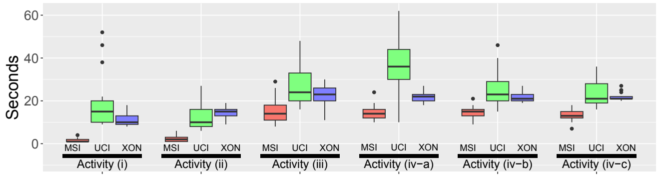

Fig. 6: Task completion time scores of each activity using the three different setup platforms.

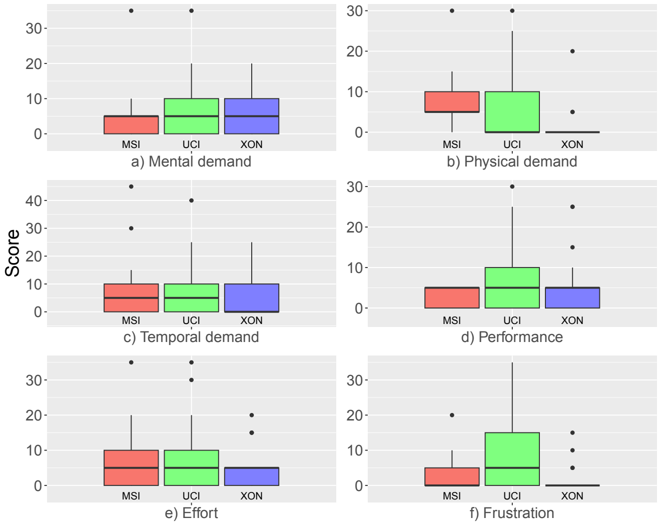

Fig. 7: NASA TLX Scores. Comparison of the three setup platforms evaluated in six different attributes.

task ease in activity “iv” with values [$6.7 < X^2(2) < 9.2$], [$7.8e^{-3} < p < 3.4e^{-2}$]; for the attribute satisfaction, there were significant differences on activities “i”, “iii” and “iv” with values [$5.1 < X^2(2) < 12.2$], [$2.2e^{-3} < p < 4.9e^{-2}$]; time on task attribute did not present any statistical difference. On a Kruskal-Wallis test ($n = 17$) , the pairs MSI-XoNLI and UCI-XoNLI presented a significant difference in the attribute task ease during activity “i” with values [$4.1 < X^2(1) < 11.9$], [$5.5e^{-4} < p < 4.2e^{-2}$]. For the attribute satisfaction, the pairs MSI-UCI, MSI-XoNLI, and UCI-XoNLI reported significant differences on activities “i”, “iii” and “iv” with values [$4.8 < X^2(1) < 10.1$], [$1.4e^{-3} < p < 2.7e^{-2}$].

3) Task Completion Time: Results in activity “i” using the MSI platform reported the shortest in time (mean 1.5 ±0.89s), and activity “iv-a” using the UCI registered the longest time (mean 35.62 ± 15.77s). A Friedman test was conducted for all activities (“i-iv”) using the three different setup platforms. Among the six activities, there was a statistically significant difference ([$18.2 < X^2(2) < 28.7$], [$5.7e^{-7} < p < 81.06e^{-4}$]). Consequently, a Kruskal-Wallis test was conducted ($n = 17$) for the six activities pairing the setup platforms as: a) MSI-UCI, b) MSI-XoNLI, and c) UCI-XoNLI. This resulted in a statistically significant difference between the MSI-UCI platforms in the six activities with values [$12.03 < X^2(2) < 25.94$] and [$3.5e^{-7} < p < 5.2e^{-4}$]. In addition, a significant difference was found between MSI-XoNLI for all activities with values [$11.2 < X^2(2) < 25.9$] and [$3.4e^{-7} < p < 8.02e^{-4}$]. Finally, in the pair UCI-XoNLI it was reported a significant difference in activities “i” and “iv-a” with values [$6.6 < X^2(2) < 14.6$] and [$1.2e^{-4} < p < 9.8e^{-3}$]. Fig. 6 shows the TCT scores for the six activities using the three setup platforms.

B. Post experiment

1) NASA Task Load Index Metric: The results of the NASA-TLX survey are shown in Fig. 7, the three setup platforms (MSI, UCI and XoNLI) are compared for each survey attribute. The MSI platform registered the lowest mean scores on attributes “a” and “d”, and the XoNLI platform obtained the lowest mean scores on the attributes “b, c, e and f”. A Friedman test was conducted on 17 participants to examine statistically significant differences in each attribute using the three different platforms. Results showed that there are statistically significant differences in attribute “b” with values $X^2(2) = 11.375$ and $p = 0.003388$. A Kruskal-Wallis test was performed pairing the setup platforms and there was a significant difference between MSI-XoNLI with values $X^2(2) = 10.594$, and $p = 0.001135$.

2) NPS and SUS Metrics: Table II shows the NPS and SUS scores for the three setup platforms. According to the NPS score, the XoNLI interface is the highest ranked and the MSI and UCI interfaces have an equal score. The lowest SUS individual score registered belongs to the UCI interface. The MSI platform was the highest ranked in the SUS combined mean score.

TABLE II: Net Promoter and System Usability Scale Scores

Platform	**NPS score**	**SUS low**	**SUS high**	**SUS mean**
MSI	52.94	75.0	100.0	90.88
UCI	52.94	37.5	100.0	82.35
XoNLI	76.47	52.5	100.0	89.35

3) User Experience Questionnaire Benchmark: The UEQ is a direct measurement for user experience and usability. This 26-item metric is designed to produce six attributes (shown in column UEQ of Table III), according to [42], a Cronbach’s Alpha coefficient (CA) ≥ 0.7 indicates consistency in the scale of the items. The score ranking (Mn) for the attributes varies from -3 to 3, these are depicted in Fig. 8. Finally, each attribute is labelled with the corresponding benchmark interval indicator (L) as presented in [43], the attribute can described each platform in terms of quality to be excellent (E), good (G), above average (A), bellow

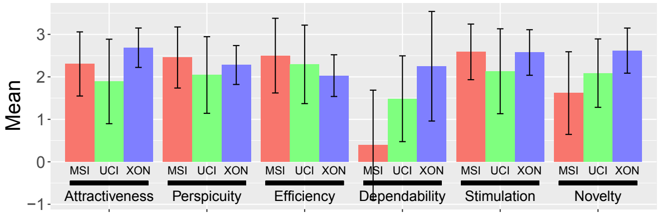

Fig. 8: UEQ benchmarking results of the three setup platforms for six user-experience and usability attributes.

average (B) and bad (D). In this benchmark, the attribute "Novelty" of the MSI platform was ranked with the lowest interval indicator. A $t-test$ was performed for all attributes paired as a) MSI/UCI, b) MSI/XON, and c)UCI/XON. The only attributes with statistical significance were "Novelty" in MSI/UCI with $p \geq 0.01$, in MSI/XON "Stimulation" with $p \geq 0.0008$, "Novelty" with $p \geq 0.0001$, and in UCI/XON "Attractive" with $p \geq 0.0054$ and "Stimulation" with $p \geq 0.03$.

TABLE III: User Experience Questionnaire Benchmark

UEQ	**MSI**				**UCI**				**XoNLI**			
	CA	Mn	SD	L	CA	Mn	SD	L	CA	Mn	SD	L
Att.	0.88	2.30	0.75	E	0.93	1.89	0.99	E	0.92	2.68	0.46	E
Per.	0.78	2.58	0.65	E	0.76	2.13	1.00	E	0.85	2.57	0.53	E
Eff.	0.77	2.50	0.87	E	0.80	2.29	0.92	E	0.52	2.02	0.49	E
Dep.	0.73	2.45	0.71	E	0.58	2.04	0.90	E	0.28	2.27	0.45	E
Stm.	0.85	1.61	0.97	G	0.81	2.08	0.80	E	0.91	2.61	0.53	E
Nov.	0.82	0.39	1.29	B	0.76	1.48	1.01	G	0.93	2.25	1.01	E

4) Speech User Interface Service Quality: Table IV shows the results of the SUISQ-R questionnaire, which was administered only after completing all activities using the XoNLI platform. The questionnaire contains four factors: a) user goal orientation (UGO), b) customer service behavior (CSB), c) speech characteristics (SC), and verbosity (V).

TABLE IV: SUISQ-R Factor Scores for XoNLI Platform

Measure	**UGO**	**CSB**	**SC**	**V**	**Overal**
Alpha	0.857	0.884	0.807	0.829	NA
Mean	6.485	6.661	5.482	2.573	6.013
SD	0.234	0.112	1.004	0.502	0.650

5) Dialogue accuracy: A total of 262 dialogues (mean: 15.41 ± 39.8 dialogues) were introduced to the ARS in the NPL server using the VUI XoNLI. There were 51 mistakes (mean: 3.0 ± 1.96 dialogues) when executing a dialogue, the system accuracy was 80.5%; all the participants were not native English speakers. Table V shows the dialogues interaction according to the model presented in Sec. II-A.1, the dialogue interaction follows the heuristics-based presented in Sec. I. The intro message is: "Hello, I am XoTalk, I can assist you to configure the exoskeleton, activate it, introduce your weight and height, and modify the force assistance. What can I do for you?".

TABLE V: Dialogue

Act.	**Dial.** P	**Utt.** $u \leq_t$	**Utt.** U	**Dialogue output** $s_{\|U\|}$
Ini.	"I need help"	help	need can could	Intro
(i)	"Could you please calibrate the exoskeleton?"	calibrate	calibration can could start	"Please stand still, the calibration has been completed, I suggest you to introduce your weight / heigh"
(ii)	"My weight/ height is x kilograms / meters"	weight / height	set introduce	"Your weight / height has been set to x kilograms / meters, what else can I dot?"
(iii)	"Please activate the exoskeleton"	activate	activation enable start	"XoTrunk is now active, have a great session"
(iv)	"Set the force assistance to x"	assistance	set change modify switch	"The assistance has been changed to x"

V. CONCLUSIONS AND FUTURE WORK

Active industrial exoskeletons are more versatile in terms of control strategy because they can be modulated according to the task performed. Because active industrial exoskeletons require pre-performance preparation (calibration, user information and activation), a user-interaction system is needed to achieve this. In this scenario, the setup platforms play a critical role in the exoskeleton configuration. The experimental results presented in this study showed that participants were more satisfied using the novel VUI XoNLI than the standard and wearable interfaces. In addition, the participants rated a more favourable score in task ease for the XoNLI platform than the others. However, satisfaction at the time of the task was not the highest with the VUI. In fact, the activity time results showed that the fastest platform to achieve the activity goals was the standard MSI. The UCI interface registered the highest perceived workload for the activities in the experiment, and the XoNLI system the lowest. The user experience benchmarking results showed that the XoNLI interface registered the highest scores in attractiveness, stimulation, and novelty over the others; nevertheless, the standard MSI interface registered the highest efficiency score. The XoNLI platform is a robust voice-interaction system. Although the experiment was conducted in English all the participants were not native English speakers, and the accuracy of the dialogue interaction was above 80.5%. Finally, the results on CSB (highest score) from the SUISQ-R showed that participants were satisfied with the XoNLI interface interaction while performing their activities. In the future, we plan to correct the dialogue model and increment the number of utterances during the conversation; we also plan to reduce the time response for user-interaction after speaking a dialogue and integrate a language large model (LLM) to the ASR system.

ACKNOWLEDGMENT

This work was supported by the Italian Workers' Compensation Authority (INAIL) and Istituto Italiano di Tecnologia (IIT) within the project Sistemi Cibernetici Collaborativi (SCC).

REFERENCES

[1] M.-L. Lu, T. R. Waters, E. Krieg, and D. Werren, “Efficacy of the Revised NIOSH Lifting Equation to Predict Risk of Low-Back Pain Associated With Manual Lifting: A One-Year Prospective Study,” *Human factors*, vol. 56, p. 73, Feb. 2014. Publisher: NIH Public Access.

[2] T. Poliero, M. Iurato, M. Sposito, C. D. Natali, S. Toxiri, S. Anastasi, F. Draicchio, L. Monica, D. G. Caldwell, V. Sanguineti, and J. Ortiz, “A case study on occupational back-support exoskeletons versatility in lifting and carrying,” in *The 14th PErvasive Technologies Related to Assistive Environments Conference*, (Corfu Greece), pp. 210–217, ACM, June 2021.

[3] K. Anam and A. A. Al-Jumaily, “Active Exoskeleton Control Systems: State of the Art,” *Procedia Engineering*, vol. 41, pp. 988–994, Jan. 2012.

[4] R. A. R. C. Gopura and K. Kiguchi, “Mechanical designs of active upper-limb exoskeleton robots: State-of-the-art and design difficulties,” in *2009 IEEE International Conference on Rehabilitation Robotics*, pp. 178–187, June 2009.

[5] M. P. de Looze, T. Bosch, F. Krause, K. S. Stadler, and L. W. O’Sullivan, “Exoskeletons for industrial application and their potential effects on physical work load,” *Ergonomics*, vol. 59, pp. 671–681, May 2016.

[6] M. Lazzaroni, S. Toxiri, D. G. Caldwell, S. Anastasi, L. Monica, E. D. Momi, and J. Ortiz, “Acceleration-based Assistive Strategy to Control a Back-support Exoskeleton for Load Handling: Preliminary Evaluation,” in *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)*, pp. 625–630, June 2019. ISSN: 1945-7901.

[7] M. Sposito, V. Fanti, P. Sencandan, D. G. Caldwell, and C. Di Natali, “Measuring Anthopometric Fit for Exoskeletons: Methodologies and Preliminary Assessment,” in *2022 9th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*, pp. 01–08, Aug. 2022. ISSN: 2155-1782.

[8] D. Swoboda, J. Boasen, P.-M. Léger, R. Pourchon, and S. Sénécal, “Comparing the Effectiveness of Speech and Physiological Features in Explaining Emotional Responses during Voice User Interface Interactions,” *Applied Sciences*, vol. 12, p. 1269, Jan. 2022. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

[9] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, “Natural language processing: an introduction,” *Journal of the American Medical Informatics Association*, vol. 18, pp. 544–551, Sept. 2011.

[10] C. L. Petersen, R. Halter, D. Kotz, L. Loeb, S. Cook, D. Pidgeon, B. C. Christensen, and J. A. Batsis, “Using Natural Language Processing and Sentiment Analysis to Augment Traditional User-Centered Design: Development and Usability Study,” *JMIR mHealth and uHealth*, vol. 8, p. e16862, Aug. 2020.

[11] J. Norberto Pires, “Robot-by-voice: experiments on commanding an industrial robot using the human voice,” *Industrial Robot: An International Journal*, vol. 32, pp. 505–511, Jan. 2005. Publisher: Emerald Group Publishing Limited.

[12] S. Papanastasiou, N. Kousi, P. Karagiannis, C. Gkournelos, A. Papavasileiou, K. Dimoulas, K. Baris, S. Koukas, G. Michalos, and S. Makris, “Towards seamless human robot collaboration: integrating multimodal interaction,” *The International Journal of Advanced Manufacturing Technology*, vol. 105, pp. 3881–3897, Dec. 2019.

[13] S. L. Hura, “Usability Testing of Spoken Conversational Systems - JUX,” *JUX - The Journal of User Experience*, vol. 12, pp. 155–163, Aug. 2017.

[14] C. Murad and C. Munteanu, “”I don’t know what you’re talking about, HALexa”: the case for voice user interface guidelines,” in *Proceedings of the 1st International Conference on Conversational User Interfaces*, CUI ’19, (New York, NY, USA), pp. 1–3, Association for Computing Machinery, 2019.

[15] D. Schnelle and F. Lyardet, “Voice User Interface Design Patterns,” in *EuroPLoP’ 2006, Eleventh European Conference on Pattern Languages of Programs, Irsee, Germany, July 5-9, 2006* (U. Zdun and L. B. Hvatum, eds.), pp. 287–316, UVK - Universitaetsverlag Konstanz, 2006.

[16] C. Kamm, “User interfaces for voice applications.,” *Proceedings of the National Academy of Sciences*, vol. 92, pp. 10031–10037, Oct. 1995. Publisher: Proceedings of the National Academy of Sciences.

[17] C. Nowacki and A. Gordeeva, “Improving the Usability of Voice User Interfaces: A New Set of Ergonomic Criteria,” Apr. 2020.

[18] V. Farinazzo, M. Salvador, A. L. S. Kawamoto, and J. S. d. O. Neto, “An Empirical Approach for the Evaluation of Voice User Interfaces,” in *User Interfaces*, IntechOpen, May 2010.

[19] T. Poliero, M. Lazzaroni, S. Toxiri, C. D. Natali, D. G. Caldwell, and J. Ortiz, “Lifting and Carrying: Do We Need Back-Support Exoskeleton Versatility?,” in *Wearable Robotics: Challenges and Trends* (J. C. Moreno, J. Masood, U. Schneider, C. Maufroy, and J. L. Pons, eds.), Biosystems & Biorobotics, (Cham), pp. 499–503, Springer International Publishing, 2022.

[20] D. Park, S. Toxiri, G. Chini, C. D. Natali, D. G. Caldwell, and J. Ortiz, “Shoulder-Sidewinder (Shoulder-Side Wearable Industrial Ergonomic Robot): Design and Evaluation of Shoulder Wearable Robot With Mechanisms to Compensate for Joint Misalignment,” *IEEE Transactions on Robotics*, pp. 1–12, 2021.

[21] O. A. Moreno, F. Draicchio, L. Monica, S. Anastasi, D. G. Caldwell, and J. Ortiz, “Designing an Integrated Tool Set Framework for Industrial Exoskeletons,” in *Wearable Robotics: Challenges and Trends* (J. C. Moreno, J. Masood, U. Schneider, C. Maufroy, and J. L. Pons, eds.), Biosystems & Biorobotics, (Cham), pp. 583–588, Springer International Publishing, 2020.

[22] M. Gandolla, S. Dalla Gasperina, V. Longatelli, A. Manti, L. Aquilante, M. G. D’Angelo, E. Biffi, E. Diella, F. Molteni, M. Rossini, M. Gföhler, M. Puchinger, M. Bocciolone, F. Braghin, and A. Pedrocchi, “An assistive upper-limb exoskeleton controlled by multi-modal interfaces for severely impaired patients: development and experimental assessment,” *Robotics and Autonomous Systems*, vol. 143, p. 103822, Sept. 2021.

[23] S. Dalla Gasperina, M. Gandolla, A. Calcagno, A. Costa, L. Aquilante, M. Puchinger, M. Gfoehler, F. Braghin, and A. Pedrocchi, “Multi-Modal Human-Machine Control Interfaces of Upper Limb Motorized Exoskeletons for Severely Impaired Patients,” in *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*, pp. 491–496, Aug. 2018. ISSN: 2155-1782.

[24] Y. G. Kim, K. Little, B. Noronha, M. Xiloyannis, L. Masia, and D. Accoto, “A voice activated bi-articular exosuit for upper limb assistance during lifting tasks,” *Robotics and Computer-Integrated Manufacturing*, vol. 66, p. 101995, Dec. 2020.

[25] I. Kagirov, A. Karpov, I. Kipyatkova, K. Klyuzhev, A. Kudryavcev, I. Kudryavcev, and D. Ryumin, “Lower Limbs Exoskeleton Control System Based on Intelligent Human-Machine Interface,” in *Intelligent Distributed Computing XIII* (I. Kotenko, C. Badica, V. Desnitsky, D. El Baz, and M. Ivanovic, eds.), Studies in Computational Intelligence, (Cham), pp. 457–466, Springer International Publishing, 2020.

[26] I. Kagirov, A. Kapustin, I. Kipyatkova, K. Klyuzhev, A. Kudryavcev, I. Kudryavcev, Y. Loskutov, D. Ryumin, and A. Karpov, “Medical exoskeleton “Remotion” with an intelligent control system: Modeling, implementation, and testing,” *Simulation Modelling Practice and Theory*, vol. 107, p. 102200, Feb. 2021.

[27] X. Wang, P. Tran, S. M. Callahan, S. L. Wolf, and J. P. Desai, “Towards the development of a voice-controlled exoskeleton system for restoring hand function,” in *2019 International Symposium on Medical Robotics (ISMR)*, pp. 1–7, Apr. 2019.

[28] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, “Robust Speech Recognition via Large-Scale Weak Supervision,” Dec. 2022. arXiv:2212.04356 [cs, eess].

[29] C. Wang, Y. Tang, X. Ma, A. Wu, S. Popuri, D. Okhonko, and J. Pino, “fairseq S2T: Fast Speech-to-Text Modeling with fairseq,” June 2022. arXiv:2010.05171 [cs, eess].

[30] S. Welleck, J. Weston, A. Szlam, and K. Cho, “Dialogue Natural Language Inference,” Jan. 2019. arXiv:1811.00671 [cs].

[31] O. A. Moreno Franco, J. Ortiz, and D. G. Caldwell, “Evaluation of the User Command Interface, an Adaptable Setup System for Industrial Exoskeletons,” in *2022 9th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*, pp. 01–07, 2022.

[32] M. Lazzaroni, V. Fanti, M. Sposito, G. Chini, F. Draicchio, C. D. Natali, D. G. Caldwell, and J. Ortiz, “Improving the Efficacy of an Active Back-Support Exoskeleton for Manual Material Handling Using the Accelerometer Signal,” *IEEE Robotics and Automation Letters*, vol. 7, pp. 7716–7721, July 2022.

[33] M. Lazzaroni, A. Tabasi, S. Toxiri, D. G. Caldwell, E. D. Momi, W. v. Dijk, M. P. d. Looze, I. Kingma, J. H. v. Dieën, and J. Ortiz, “Evaluation of an acceleration-based assistive strategy to control a back-support exoskeleton for manual material handling,” *Wearable*

Technologies, vol. 1, p. e9, Jan. 2020. Publisher: Cambridge University Press.

[34] J. T. Meyer, N. Tanczak, C. M. Kanzler, C. Pelletier, R. Gassert, and O. Lambercy, "Design and validation of a novel online platform to support the usability evaluation of wearable robotic devices," *Wearable Technologies*, vol. 4, p. e3, 2023. Publisher: Cambridge University Press.

[35] S. G. Hart, "Nasa-Task Load Index (NASA-TLX); 20 Years Later," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, pp. 904–908, Oct. 2006. Publisher: SAGE Publications Inc.

[36] N. I. Fisher and R. E. Kordupleski, "Good and bad market research: A critical review of Net Promoter Score," *Applied Stochastic Models in Business and Industry*, vol. 35, no. 1, pp. 138–151, 2019. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/asmb.2417.

[37] J. Sauro and E. Kindlund, "Using a single usability metric (SUM) to compare the usability of competing products," in *Proceedings of the Human Computer Interaction International Conference (HCII)*, 2005.

[38] J. R. Lewis and M. L. Hardzinski, "Investigating the psychometric properties of the Speech User Interface Service Quality questionnaire," *International Journal of Speech Technology*, vol. 18, pp. 479–487, Sept. 2015.

[39] R. A. Grier, A. Bangor, P. Kortum, and S. C. Peres, "The System Usability Scale: Beyond Standard Usability Testing," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 57, pp. 187–191, Sept. 2013. Publisher: SAGE Publications Inc.

[40] B. Laugwitz, T. Held, and M. Schrepp, "Construction and Evaluation of a User Experience Questionnaire," in *HCI and Usability for Education and Work* (A. Holzinger, ed.), Lecture Notes in Computer Science, (Berlin, Heidelberg), pp. 63–76, Springer, 2008.

[41] T. Poliero, V. Fanti, M. Sposito, D. G. Caldwell, and C. D. Natali, "Active and Passive Back-Support Exoskeletons: A Comparison in Static and Dynamic Tasks," *IEEE Robotics and Automation Letters*, vol. 7, pp. 8463–8470, July 2022. Conference Name: IEEE Robotics and Automation Letters.

[42] M. Rauschenberger, M. Schrepp, M. Perez-Cota, S. Olschner, and J. Thomaschewski, "Efficient Measurement of the User Experience of Interactive Products. How to use the User Experience Questionnaire (UEQ).Example: Spanish Language Version," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 2, no. Special Issue on Artificial Intelligence and Social Application, pp. 39–45, 2013.

[43] M. Schrepp, J. Thomaschewski, and A. Hinderks, "Construction of a Benchmark for the User Experience Questionnaire (UEQ)," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. Regular Issue, pp. 40–44, 2017.