


RESEARCH ARTICLE

# Design and showcase of a stairs-based testbed for the benchmark of exoskeleton devices: The STEPbySTEP project

Marco Caimmi<sup>1</sup>, Nicole Maugliani<sup>1</sup>, Matteo Malosio<sup>1</sup>, Francesco Airoidi<sup>1</sup>, Tito Dinon<sup>1</sup>, Diego Borro<sup>2,3</sup>, Martxel Eizaguirre<sup>2</sup>, Iñaki Díaz<sup>2</sup>, Sergio Ausejo<sup>2</sup>, Gabriele Puzzo<sup>4</sup>, Federico Fraboni<sup>4</sup>, Luca Pietrantoni<sup>4</sup>, Marco Maccarini<sup>5</sup>, Asad Ali Shahid<sup>5</sup> and Loris Roveda<sup>5,6</sup> 

<sup>1</sup>Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing (STIIMA) of Italian National Research Council (CNR), Milan, Italy

<sup>2</sup>CEIT-Basque Research and Technology Alliance (BRTA) and Tecnun (University of Navarra), Donostia-San Sebastian, Spain

<sup>3</sup>Institute of Data Science and Artificial Intelligence (DATAI), University of Navarra, Pamplona, Spain

<sup>4</sup>Department of Psychology, University of Bologna, Bologna, Italy

<sup>5</sup>Department of Innovative Technologies, University of Applied Science and Arts of Southern Switzerland (SUPSI), Istituto Dalle Molle di studi sull'intelligenza artificiale (IDSIA), Lugano, Switzerland

<sup>6</sup>Politecnico di Milano, Dipartimento di Meccanica, Milano, Italy

**Corresponding author:** Loris Roveda; Email: [loris.roveda@polimi.it](mailto:loris.roveda@polimi.it)

**Received:** 13 December 2024; **Revised:** 17 February 2025; **Accepted:** 09 March 2025

**Keywords:** exoskeleton; benchmark; evaluation metrics; evaluation protocols

## Abstract

Wearable exoskeletons hold the potential to provide valuable physical assistance across a range of tasks, with applications steadily expanding across different scenarios. However, the lack of universally accepted testbeds and standardized protocols limits the systematic benchmarking of these devices. In response, the STEPbySTEP project, funded within the Eurobench framework, proposes a modular, sensorized, reconfigurable staircase testbed designed as a novel evaluation approach within the first European benchmarking infrastructure for robotics. This testbed, to be incorporated into the Eurobench testing facility, focuses on stairs as common yet challenging obstacles in daily life that provide a unique benchmark for exoskeleton assessment.

The primary aim of STEPbySTEP is to propose a modular framework – including a specialized staircase design, tentative metrics, and testing protocols – to aid in evaluating and comparing exoskeleton performance. Here, we present the testbed and protocols developed and validated in preliminary trials using three exoskeletons: two lower-limb exoskeletons (LLEs) and one back-support exoskeleton. The results offer initial insights into the adaptability of the staircase testbed across devices, showcasing example metrics and protocols that underscore its benchmarking potential.

## 1. Introduction and paper contribution

Exoskeleton systems are increasingly adopted in fields ranging from industrial applications to medical rehabilitation, where they offer valuable support and enhancement for human physical capabilities (Mauri et al., 2019; Dalla Gasperina et al., 2021). However, despite their growing use, the absence of standardized procedures for evaluating and benchmarking exoskeleton performance persists. Benchmarking plays a crucial role in establishing standardized metrics for assessing device efficacy, usability, and safety across

real-world conditions. Although the importance of such practices is widely recognized, benchmarking within wearable robotics remains underdeveloped (Torricelli et al., 2019; De Bock et al., 2022).

While recent studies have addressed various benchmarking methodologies (De Bock et al., 2022), there remains a critical need for instrumented testbeds capable of replicating real-world conditions and providing objective performance assessments across diverse exoskeleton designs. In response, the H2020 EUROBENCH project (Torricelli et al., 2019) aims to bridge this gap by offering standardized testing environments. Details on the project testbeds and infrastructure are available on its webpage <https://github.com/eurobench>.

As part of the EUROBENCH initiative, the STEPbySTEP subproject (Maugliani et al., 2020) introduces a modular, sensor-equipped, and adaptable staircase testbed. This system is intended to establish a standardized validation process for assessing exoskeleton capabilities, performance, and user experience, specifically during stair ascent and descent, common complex challenges in daily life. Stair climbing scenarios are essential in exoskeleton evaluation due to the high joint torques and power requirements they involve (Farris et al., 2012; Tricomi et al., 2023). Most studies to date have focused on non-standard, small-scale stair tests, often neglecting key factors such as stair inclination and step height, which significantly impact exoskeleton performance (Dollar and Herr, 2008; Benson et al., 2016).

The need for a versatile, adjustable staircase testbed, where inclination and step height can be modified independently of each other, has become increasingly apparent. Such a testbed would enable comprehensive analysis of exoskeleton adaptability across different real-world scenarios, providing a foundation for precise benchmarking protocols (Zhang et al., 2023). STEPbySTEP addresses this need by delivering a highly configurable testbed with integrated measurement tools, tentative metrics, and protocols for thorough performance evaluation and comparison of exoskeleton systems.

Emerging frameworks like the Exoworkathlon have underscored the importance of benchmarking within real-world environments, especially for industrial exoskeletons (Schmidt et al., 2017). These frameworks emphasize that benchmarking must extend beyond technical performance to include factors such as cognitive workload, usability, and user perception of comfort – crucial elements for understanding the overall impact of exoskeletons in applied contexts (Tricomi et al., 2023). Including human factors metrics, such as cognitive workload and user discomfort, allows for a holistic approach to exoskeleton evaluation, complementing technical data collected through systems like motion capture, force plates, and EMG (Lafranchi et al., 2011; Awad et al., 2017).

This paper presents a comprehensive benchmarking scheme specifically tailored to the stair climbing scenario. The scheme includes the modular staircase testbed along with associated metrics and evaluation protocols, forming a developing framework for systematic exoskeleton performance assessment in stair-related tasks. The configurability of the testbed – accommodating variable inclinations, step heights, and sensorized support systems – facilitates detailed data collection that is essential for benchmarking (De Bock et al., 2022).

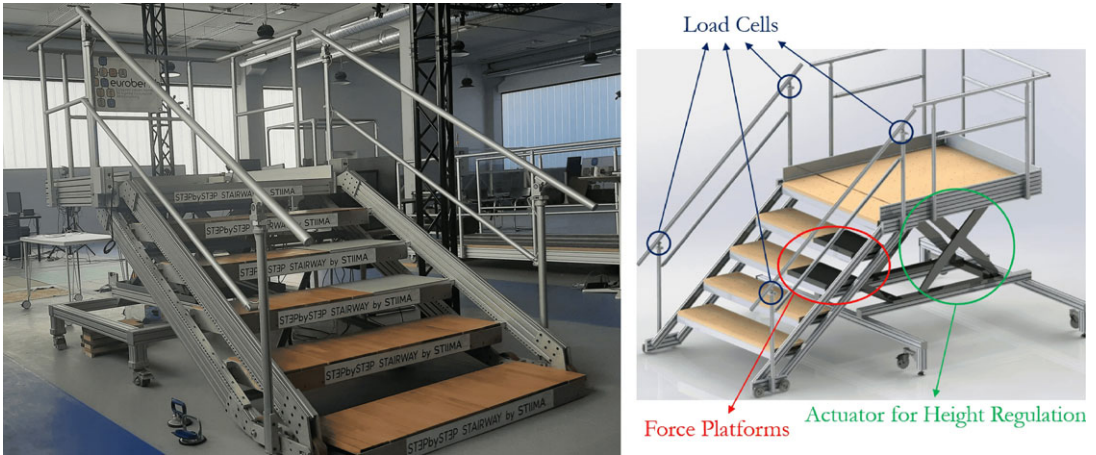
In the following sections, we provide an in-depth description of the designed and developed testbed and demonstrate its efficacy through case studies with two lower-limb exoskeletons (LLEs) and one back-support exoskeleton. These studies showcase the benchmarking capabilities of the proposed testbed for evaluating exoskeleton solutions within the stair climbing domain.

## 2. Staircase

### 2.1. Design requirements

The design of the testbed (Figure 1) had to comply with the set of technical and functional requirements listed hereafter, which were partly defined by the Eurobench project promoters and partly derived from the authors' analysis:

- compliant with technical and ergonomics norms of stair design;
- a minimum of five steps to allow the acquisition of two complete stance phases for each limb;



**Figure 1.** The STEPbySTEP staircase prototype.

- variable tread and riser lengths, by changing the staircase inclination to test the exoskeletons in different climbing conditions;
- two people, for example, the patient and the therapist, must be allowed to climb the staircase side by side, leading to a minimum width of the staircase of 1500 mm;
- a stair landing at the top of the staircase to allow the exoskeleton users and, if necessary, their companions to turn around and face the descent;
- motorized system to change the tread and riser lengths to facilitate and speed up the setup operations;
- a set of sensors to allow the measurement of kinematic and dynamic data while using the exoskeleton;
- a set of software tools, algorithms, and metrics to quantitatively analyze the user's motor capabilities and to provide a comprehensive platform to assess the exoskeleton performances.

They served as a guide for the mechanical design (Section 2.2), the selection and integration of sensors (Section 2.3), metrics definition (Section 3), and protocol design (Section 4).

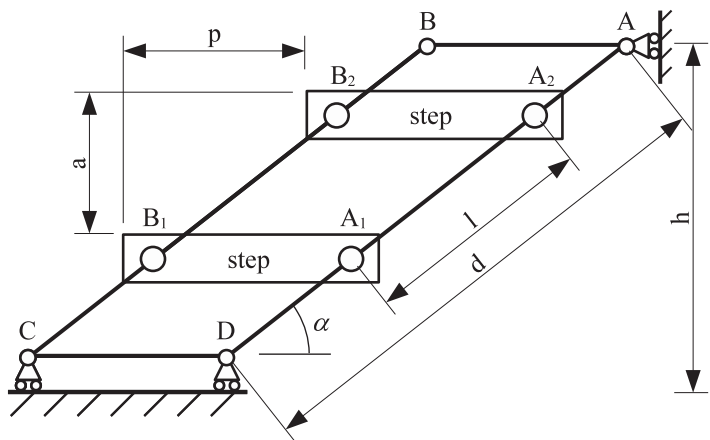
It has to be specified that there is no specific target user group w.r.t. the testbed. Therefore, the proposed design (including testing protocols and evaluation metrics) can be applied to any participant.

## 2.2. Mechanical design and assembly

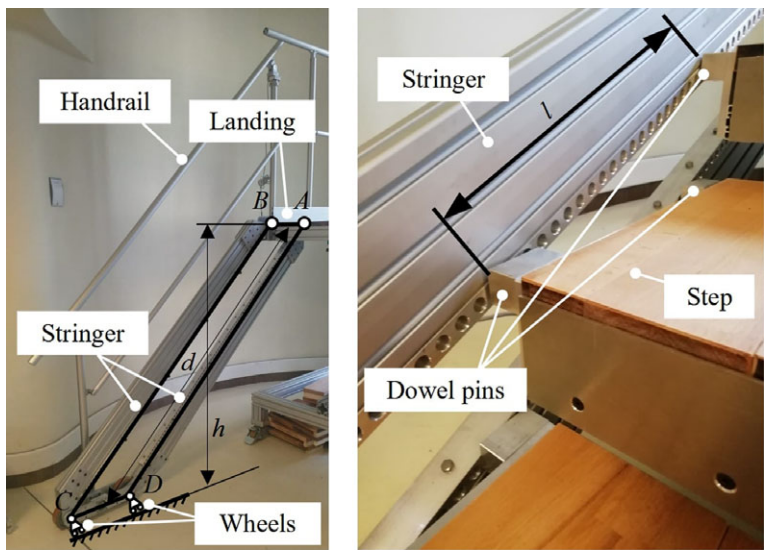
The STEPbySTEP testbed is a reconfigurable five-step staircase with appropriate adjustments to regulate the relative position among the steps and, consequently, the inclination of the staircase.

By referring to the lateral view of the adjustment mechanism depicted in Figure 2, the parallel-linkage four-bar linkage  $ABCD$  represents the structure of each stair stringer. Let us denote the vertical distance between  $\overline{AB}$  and  $\overline{CD}$  by  $h$ , and the total length of the stringer by  $d$ . The inclination of the staircase is  $\alpha = \arcsin(h/d)$ . Each step is constrained on the stringer by two hinges  $A_i$  and  $B_i$ . Their positions along the stringer are identical to guarantee that each step is parallel to the ground. The distance between two consecutive steps along the stringer is denoted by  $l$ . Consequently, the tread length is  $p = l \cos \alpha$ , and the riser length is  $a = l \sin \alpha$ . Therefore, the tread and the riser lengths can be adjusted by modifying  $h$  and  $l$ .

The realized adjusting mechanisms are depicted in Figure 3. The height of the staircase  $h$  is adjusted by a Pantograph lift table (Ameise®HIW 2.0, by Jungheinrich), denoted by *actuator for height regulation* in Figure 1. Moreover, the distance between two consecutive steps  $d$  is adjusted by



**Figure 2.** Kinematics of the staircase mechanism.



**Figure 3.** Adjustments realized in the prototype.

modifying the position of two Dowel pins, hinged to each side of a step, along a series of holes present in each stringer.

In order to represent the adjustment capabilities of the staircase, let us refer to the international norm ISO14122-3. It imposes the following conditions for realizing ergonomic staircases in civil works:

$$b_l \leq v \leq b_u, \tag{2.1}$$

with

- $v = p + 2a$
- $b_l = 600\text{mm}$  and  $b_u = 660\text{mm}$  denote the boundary limits of  $v$ .

Let us define the discrepancy of a generic geometrical configuration w.r.t. the acceptable range imposed by norms (2.1) by:

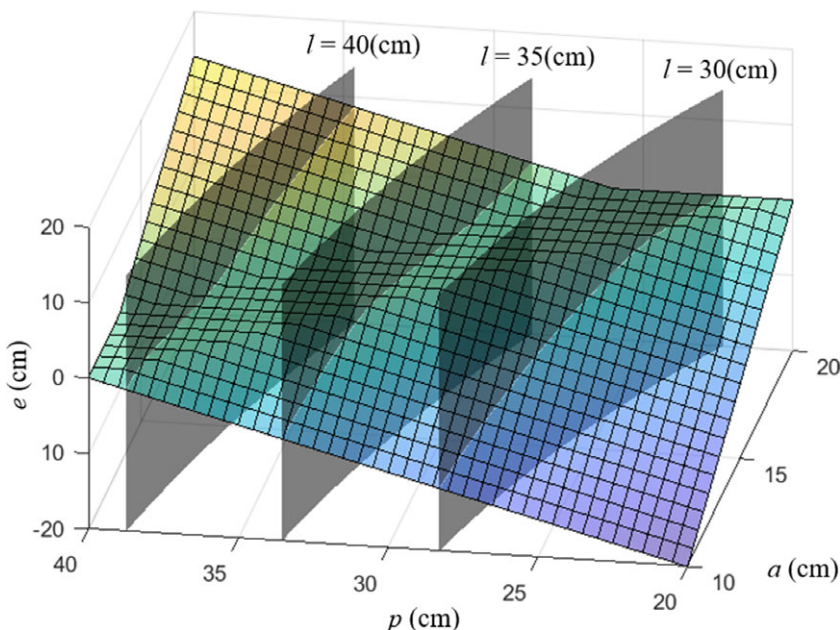
$$e = \begin{cases} v - b_l, & \text{if } v < b_l \\ 0, & \text{if } b_l \leq v \leq b_u \\ v - b_u, & \text{if } v > b_u. \end{cases} \quad (2.2)$$

The colored surface shown in Figure 4 shows the influence of geometrical parameters  $p$  and  $a$  on  $e$ . The values  $p$  and  $a$  for which  $e$  is equal to 0 comply with the norm ISO14122-3 (2.1). The gray surfaces represent the relationship between  $p$  and  $a$ , given a distance  $l$  between two consecutive steps along the stringer, modifying the inclination  $\alpha$  of the staircase. For clarity of representation, only surfaces with  $l = 30, 35$ , and  $40$  cm are represented. Intermediate configurations can be obtained by appropriately adjusting Dowel pins shown in Figure 3. Once  $l$  is set, the values of  $p$  and  $a$  that comply with the norm are given by the intersection line between the colored surface and the corresponding gray surface where  $e = 0$ .

### 2.3. Sensors selection

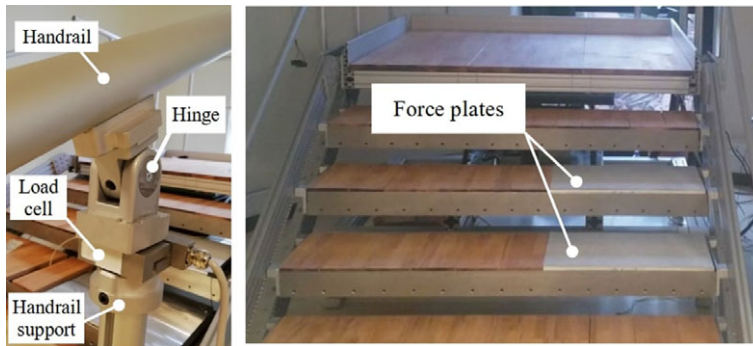
The testbed is provided with integrated sensors and acquisition systems to measure kinematic and dynamic data. The handrails of the testbed are sensorized with four load cells (Forsentek, F3G-500 N), two for each handrail, connected through the PhidgetBridge 4-Input modules bridge interfaces (Phidgets Inc., Canada), to measure the forces applied by the user wearing the exoskeleton (e.g., a patient using one crutch and the sustain of a handrail) during the ascent or descent of the stairs. Two force plates (P-6000, BTS Bioengineering, Italy) can be properly inserted into any of the steps to measure the ground reaction forces, thanks to the movable wooden tiles that allow for flexible positioning. This setup also permits the insertion of two force plates in the same step, which is useful, for example, when studying both the foot placement and the use of a crutch simultaneously. Load cells and force plates are installed as depicted in Figure 1 and in Figure 5.

A motion caption system (Perception Neuron Pro, Noitom Ltd., US) is part of the testbed to measure the user kinematics, available in the installation premises.



**Figure 4.** Discrepancy of staircase configurations concerning limits imposed by the International 14122–3 norm (1).





**Figure 5.** Load cells and force plates installed on the staircase.

### 3. Metrics

#### 3.1. Physical interaction metrics

One way to assess the performance of an exoskeleton is by investigating the user's motor capability in performing a task while wearing the exoskeleton. This can be done by analyzing physical measurements, such as the muscular activation pattern and interaction forces. This section discusses the developed metrics calculated from the Surface ElectroMyoGraphic (SEMG) signals and interaction forces measured with two force plates installed in the staircase steps and four load cells integrated into the handrails. Using the IMUs to calculate the duration of the gait phases, that is, stance (ST), swing (SW), double support (DS), and gait cycle (GC), the metrics are calculated for each phase.

SEMG is a method to record the electrical muscle activity during contraction; the stronger the muscle contraction, the higher the electrical activity. Two metrics based on SEMG were used: an index to quantify the level of activity of a muscle, the amount of EMG of muscle  $i$  ( $AEMG^i$ ), and an index quantifying the level of co-contractions between two muscles  $i$  and  $j$  ( $ccEMG^{ij}$ ). These metrics have been previously used in studies related to exoskeletons and gait analysis (Roveda et al., 2019). The novelty in our approach lies in applying these metrics to stair climbing, specifically to assess the effect of exoskeleton assistance on the neuromuscular activation pattern across different gait phases.

The significance of  $AEMG^i$  and  $ccEMG^{ij}$  on gait functionality is supported by previous research, which has demonstrated that increased co-contraction may be associated with joint stiffness and altered gait patterns, particularly in populations with gait impairments (Damiano et al., 2000; Latash, 2018). Additionally, reduced or optimized muscle activation patterns have been shown to correlate with more efficient gait mechanics, potentially reducing the metabolic cost of walking (Collins et al., 2015; Galle et al., 2017). Thus, these metrics indicate the exoskeleton's ability to assist in achieving a more efficient and stable gait pattern during stair climbing.

Regarding the ground and handrail interaction forces, several metrics were defined for both feet: i) the maximum vertical and sagittal ground reaction force at the beginning stance ( $GFz_1$  and  $GFy_1$ ) and ii) at push-off ( $GFz_2$  and  $GFy_2$ ), iii) the maximum vertical and sagittal handrail reaction forces during double support ( $HFz$  and  $HFy$ ), and iv) the maximum lateral handrail reaction force during double support ( $HFx$ ). These metrics are used to assess the subject's capability to: i) accept the load, ii) push off before the swing phase, iii) maintain equilibrium without exerting excessive forces on the handrails in the vertical and sagittal directions, and iv) maintain equilibrium without exerting excessive lateral forces on the handrails.

The correlation between these ground reaction force (GRF) metrics and specific gait functionalities has been well-documented in the literature. For example, maximum vertical GRF during the stance phase is indicative of the body's ability to support itself and absorb impact, which is often compromised in individuals with gait deficits (Winter, 1991; Vaughan et al., 1999). Similarly, the push-off phase is crucial for forward propulsion, and deficiencies in this phase are linked to reduced gait speed and efficiency, often observed in elderly or neurologically impaired populations (Schenck and Kesar, 2017; Awad et al., 2020;

Alingh et al., 2021). The handrail forces, on the other hand, can indicate a reliance on external support, which may be necessary for maintaining balance and stability during stair ascent and descent, particularly in individuals with compromised motor control (Novak and Brouwer, 2014; Stessman et al., 2017).

The combined analysis of these metrics provides a comprehensive understanding of the physical interactions between the user and the exoskeleton, offering insights into the device's ability to support or enhance the user's natural movement patterns during stair navigation. This multifaceted approach ensures that the evaluation captures both the mechanical and neuromuscular aspects of gait, allowing for a more nuanced assessment of exoskeleton performance.

### 3.2. Temporal phases

In stair negotiation, defining sub-phases beyond the basic stance, swing, and double-support phases is essential for a deeper understanding of biomechanics. These sub-phases help capture the finer details of movement dynamics during each part of stair ascent and descent. By segmenting the movement into more specific intervals, such as weight acceptance, push-off, or controlled lowering, it becomes possible to analyze the biomechanical contributions during each part of the stair negotiation process. This level of detail is particularly important when studying complex systems like exoskeletons, which aim to replicate or assist human movement. A more detailed sub-phase framework allows for a nuanced evaluation of performance and movement assistance provided by these systems.

In this paper, we adopt the phases and sub-phases definition proposed by Harper et al. (2018)), as illustrated in Figure 6. The detailed breakdown of phases provides a structured foundation for analyzing exoskeleton performance in stair negotiation tasks.

The six phases (and their associated events) during ascent are as follows:

- **Stance phase:**
  - **Weight acceptance (WA).** From ipsilateral foot strike to contralateral toe-off.
  - **Pull-up (PU).** From the beginning of single-leg support to the mid-swing of the contralateral leg.
  - **Forward continuance (FCN).** From mid-swing of the contralateral leg to the contralateral foot strike.

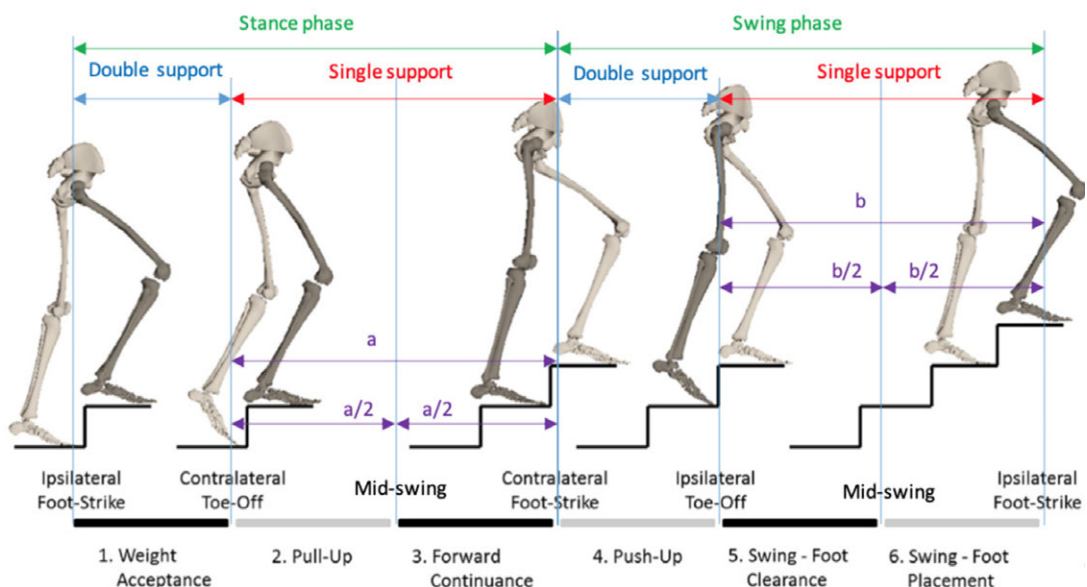


Figure 6. Figure adapted from Harper et al. (2018)).

- **Swing phase:**
  - **Push-up (PU).** From contralateral foot strike to ipsilateral toe-off.
  - **Swing foot clearance (SFC).** From ipsilateral toe-off to mid-swing of the ipsilateral leg.
  - **Swing foot placement (SFP).** From mid-swing of the ipsilateral leg to ipsilateral foot strike.

The five phases during the descent are as follows:

- **Stance phase:**
  - **Weight acceptance (WA).** From right foot contact to left toe-off.
  - **Forward continuance (FCN).** From left toe-off (single right-leg support) to mid-swing of the left leg.
  - **Controlled lowering (CL).** From mid-swing of the left leg until right toe-off.
- **Swing phase:**
  - **Leg pull-through (LP).** From right toe-off to mid-swing of the right leg.
  - **Foot placement (FP).** From mid-swing of the right leg to right foot contact.

Based on this phase definition, 11 temporal metrics can be established (see [Table 1](#)). This protocol aims to achieve two primary objectives: first, to compute the temporal metrics (refer to [Section 4.2](#)); and second, to provide the raw kinematic data to the end user, enabling its utilization as deemed appropriate. For each sub-phase, for example, the average angles for the different joints can be calculated, as well as the angle at the end of the sub-phase.

### 3.3. Human factors metrics

Several literature contributions highlight the importance of adopting a user-centered perspective on exoskeleton design (Bush and ten Hompel, 2017). For instance, among the most critical factors for successful gerontechnology acceptance and uptake, we can find usability (Shore et al., 2019), the possibility to contemporarily perform motor and cognitive tasks (O’Sullivan et al., 2015), and comfort (Wolff et al., 2014). Despite this, human factors and human–technology interaction aspects, such as usability and acceptance, are often overlooked when exoskeletons are designed, and there is currently no framework in the literature that allows for an evaluation of the usability and the impact on users’ perceptions and behaviors of exoskeletons about the task of climbing and descending stairs.

The cognitive workload is an important factor to consider when evaluating the performance of exoskeletons, as it can affect the user’s ability to use the device effectively and safely. The dual-task paradigm is commonly used for assessing cognitive workload, as it involves the simultaneous performance of two tasks that differ in their cognitive demands (Shaw et al., 2018). Usability is another key factor. The System Usability Scale (SUS) is a widely used tool for measuring usability. It assesses the user’s overall satisfaction, and it is widely adopted for exoskeleton usability assessment as well (Orekhov et al., 2021). Perceived musculoskeletal pressure and discomfort are, as well, crucial factors to consider

*Table 1. Temporal metrics*

Ascending	Descending
ascending_total_time	descending_total_time
ascending_weight_acceptance_time	descending_weight_acceptance_time
ascending_pull_up_time	descending_forward_continuance_time
ascending_forward_continuance_time	descending_controlled_lowering_time
ascending_push_up_time	descending_swing_leg_pull_through_time
ascending_swing_foot_clearance_time	descending_swing_foot_placement_time
ascending_swing_foot_placement_time	



when evaluating exoskeletons, as they can affect the user's ability to use the device comfortably and effectively over an extended period. The local perceived pressure method is a tool that can be used to assess the level of pressure or discomfort experienced by the user when using the exoskeleton (Kermavnar et al., 2021). Technology acceptance is another crucial factor to consider when evaluating exoskeletons. van der Laan Acceptance Scale is a tool that can be used to assess the user's acceptance of the technology. The scale measures the user's willingness to use the technology and the level of confidence they have in its ability to perform the intended tasks (Van Der Laan et al., 1997). The following cognitive metrics are thus selected, encompassing in particular: cognitive workload as measured by an adapted version of the dual-task paradigm (Plummer-D'Amato et al., 2008); usability as measured by the System Usability Scale (Lewis and Sauro, 2017); perceived musculoskeletal pressure/discomfort as measured by the local perceived pressure method (Hamberg-van Reenen et al., 2008), and technology acceptance as assessed by van der Laan Acceptance Scale (Van Der Laan et al., 1997).

#### 4. Protocol design

Protocols exploiting the metrics, proposed in this section, are intended to form the basis for future ad hoc protocols to investigate specific features of the exoskeletons.

##### 4.1. Physical interaction protocol

It is based on the interaction metrics; the rectus femoris (RF) and the biceps femoris (BF) muscles were selected as showcases as highly involved in the task. More muscles should be studied to have a picture of the neuromuscular activation pattern underneath the movement.

The setup is as follows:

- four-EMG channels (four probes of the Delsys system) to assess bilaterally the activity of the selected muscles;
- two IMUs (two probes of the Delsys system) are placed at the shanks to identify the gait phases (Salarian et al., 2004);
- two force plates were inserted in the second and third steps to evaluate a complete gait cycle; and
- two load cells integrated into one of the handrails to measure the interaction forces.

Data acquisition protocol:

1. User preparation: donning of the exoskeleton and EMG electrode placement.
2. Resting and standing acquisition: resting acquisition with the subject sitting on a chair.
3. Place the subject at a one-step distance from the first step of the stairway.
4. Start data acquisition: launch the data collection software.
5. Ascending stairs: at the go command, the subject makes a left footstep and climbs the stairs, beginning with the right foot, till reaching the upper platform (usually, subjects use the support of a crutch and the handrail).
6. Stop the acquisition when the user stands with two feet on the upper platform.
7. Turning: the subject turns and prepares to descend the stairs.
8. Descending stairs.
9. Trials repetitions: steps 3–7 are repeated (four other trials).
10. Change the staircase inclination: the operator changes the staircase inclination using the platform elevator.
11. Second set of trials: steps 3–7 are repeated (five trials).

The execution of the protocol, donning the exoskeleton excluded, lasts 30–40 min. Metrics are calculated offline with the software provided to Eurobench.

#### 4.2. Temporal phases protocol

The protocol aims to define the phases of stair negotiation during both ascent and descent, facilitating a detailed understanding of human movements in these contexts. To achieve this, a subset of sensors from a motion capture (mocap) system is utilized, specifically tailored for the knee joint angle calculation and phase definition.

The setup consists of the Perception Neuron Pro system (Perception Neuron Pro, Miami, US), which uses inertial measurement units (IMUs) for capturing kinematic data.

The protocol is defined as follows:

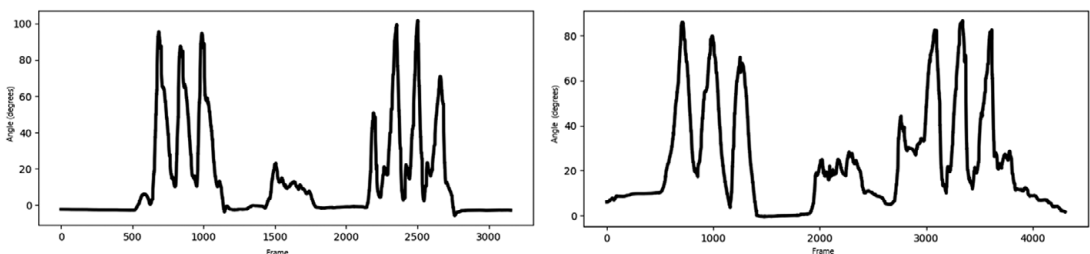
1. Place the mocap system onto the subject's body (focusing on the lower limbs). The placement is optimized based on the geometric constraints of each exoskeleton and the recommended locations provided by the capture system manufacturer, allowing for flexibility in sensor positioning.
2. Perform mocap calibration to ensure accurate data capture.
3. Position the subject at the starting point, which is one step away from the stairway.
4. Launch the data collection software to begin recording.
5. Conduct the experiment, where the subject ascends and descends the stairs using a step-over-step approach.
6. Stop the data collection once the subject returns to the initial position.
7. Repeat the procedure five times to ensure reliability and consistency in the data.

The total of five runs takes approximately 22 min.

For the gait events and phase identification, only knee joint angle data have been used (Figure 7). Looking at the figure, the differences between exo and no-exo are particularly noticeable in the three peaks during the ascending phases. These peaks indicate a slightly more rigid and less fluid movement in the exo case, suggesting that the knee does not rotate as much as it does without the exoskeleton.

The algorithm used to identify gait events and phases from knee angle data is based on the K-nearest-neighbors (KNN) machine learning (ML) algorithm, with results shown in Figure 10). KNN is a supervised classification method that identifies patterns in data by determining the category of a data point based on its proximity to other, already categorized points. In the context of gait phase identification, the KNN algorithm classifies IMU data into different gait phases by comparing the current data to a previously labeled training dataset.

Two separate ML models were trained: one for stair ascent and another for descent. The training dataset consists of 34 trials from five participants (three females and two males) without exoskeletons. The participants' mean age was 31.4 years (SD: 7.27), their mean weight was 75.6 kg (SD: 4.39), and their mean height was 1.71 m (SD: .09). The first three participants completed six trials each, while the last two completed eight trials. The trained algorithms were subsequently tested using an exoskeleton, and the results are presented and discussed in Section 5.



**Figure 7.** Knee joint angle of a healthy subject during stair ascent and descent, with (left) and without (right) a lower limb exoskeleton (LLE).

This protocol not only enables the calculation of phases and sub-phases but also allows for the assessment of joint angles using IMUs (Recinos et al., 2020). Additionally, it is compatible with other signal acquisition systems, such as EMG and GRF. This combination offers a more comprehensive analysis of biomechanical performance, extending beyond traditional phases like stance, swing, and double support. In this paper, we focus specifically on the definition of temporal phases.

#### 4.3. Human factors protocol

Human factors metrics are assessed through an adapted version of the dual-task paradigm – applied to measure cognitive workload – and through various questionnaires and scales measuring usability, acceptance, and discomfort – administered at the end of the task. Regarding the cognitive workload, this protocol emulates cognitively demanding situations, as participants are asked to simultaneously perform the STEPbySTEP testbed task (going up and down the stairs continuously) and a sound recognition task. The latter consists of presenting a sequence of sound stimuli (phonemes) to the participants, who have to say either “no” (each time the phoneme is different from the previously listened one) or “yes” (each time the phoneme is the same as the previous one). To discern differences in cognitive performance due to the exoskeleton design characteristics, every participant must perform the task under two conditions: the control condition (without wearing the exoskeleton) and the experimental condition (while wearing the exoskeleton). The control condition is necessary since the dual-task paradigm allows the cognitive workload to be assessed by calculating the difference in the number of errors and latency of responses between the single-task baseline (only the sound recognition task) and the dual-task condition (when adding exoskeleton-assisted movement). Therefore, analyzing differences in the cognitive workload levels between the two conditions allows us to discern the cognitive load generated by wearing the exoskeleton alone. Higher latency and higher number of errors correspond to higher levels of cognitive workload (Koch et al., 2018).

In other words, our protocol envisages conducting a cognitively demanding task (the sound recognition task) in two different conditions: without wearing the exoskeleton and with the exoskeleton on. Since the sound recognition task does not change between the two conditions, and the only feature that changes between the two conditions is the presence of the exoskeleton, we can infer that changes in cognitive load are caused by the exoskeleton design. To minimize bias, participants were trained to familiarize themselves with the use of the exoskeleton. The STEPbySTEP testbed task (going up and down the stairs continuously) is performed in both conditions to ensure that participants are actively interacting with the exoskeleton. Therefore, analyzing differences in the cognitive workload levels between the two conditions allows us to discern the cognitive load generated by wearing the exoskeleton alone. Higher latency and higher numbers of errors correspond to higher levels of cognitive workload (Koch et al., 2018).

W.r.t. the human factors questionnaires, usability is measured through the 10-item System Usability Scale (Lewis and Sauro, 2017), a 5-point Likert scale going from 1 = strongly disagree to 5 = strongly agree. Example item: “I think that I would like to use this exoskeleton frequently.”

Acceptance is measured using the 9-item van der Laan Acceptance Scale (Van Der Laan et al., 1997). Participants are asked to rate the exoskeleton on a 5-point semantic differential scale with two opposite adjectives (i.e., useless–useful). The score ranges from −2 (negative opposite) to +2 (positive opposite). Four items related to the usefulness sub-dimension (i.e., “I found the XSPINE prototype to be ... Effective – Superfluous”), while five items to the satisfaction sub-dimension (i.e., “I found the XSPINE prototype to be ... Pleasant – Unpleasant”).

Discomfort is measured through a modified version of the 12-item Local Perceived Discomfort Scale (Hamberg-van Reenen et al., 2008) that includes lower limbs. Participants are asked to rate to which extent the exoskeleton exerted pressure on different body parts on an 11-point scale (0 = no pressure at all to 10 = maximal pressure).

In both conditions, an experimenter registers the participant’s responses manually. Data are collected through the “OpenSesame” software, which records the audio and the participants’ number of errors and

response times. Moreover, participants are asked to fill out usability, acceptance, and discomfort scales through an online questionnaire on “Qualtrics” at the end of the task.

The summary of the protocol is detailed below:

1. A detailed explanation of the experiment and the two phases of the task;
2. Earphones wearing and volume calibration;
3. Running “OpenSesame” software;
4. Place the subject at the starting point;
5. Trial run: perform the dual task;
6. Addressing participant’s eventual doubts and questions, eventual volume re-calibration, and “OpenSesame” re-running;
7. Experimental run: perform the dual task;
8. Donning the exoskeleton;
9. Repeat the procedure described from step 3 to step 8 (with the participant wearing the exoskeleton);
10. Doffing the exoskeleton;
11. Saving audio recordings and data collected through “OpenSesame” (i.e., response time in milliseconds and error number);
12. Ask participants to fill out the Local Perceived Discomfort Scale (Hamberg-van Reenen et al., 2008) to assess discomfort;
13. Administration of the System Usability Scale (Lewis and Sauro, 2017) and van der Laan Acceptance Scale (Van Der Laan et al., 1997) to assess usability and acceptance.

The total time to perform the human factors protocol (for one subject) is around 20 min.

## 5. Results

### 5.1. Methods

All experiments in the EUROBENCH project, including those conducted in the STEPbySTEP project and reported in this paper, were approved by the Ethics Committee of the Consejo Superior de Investigaciones Científicas (CSIC).

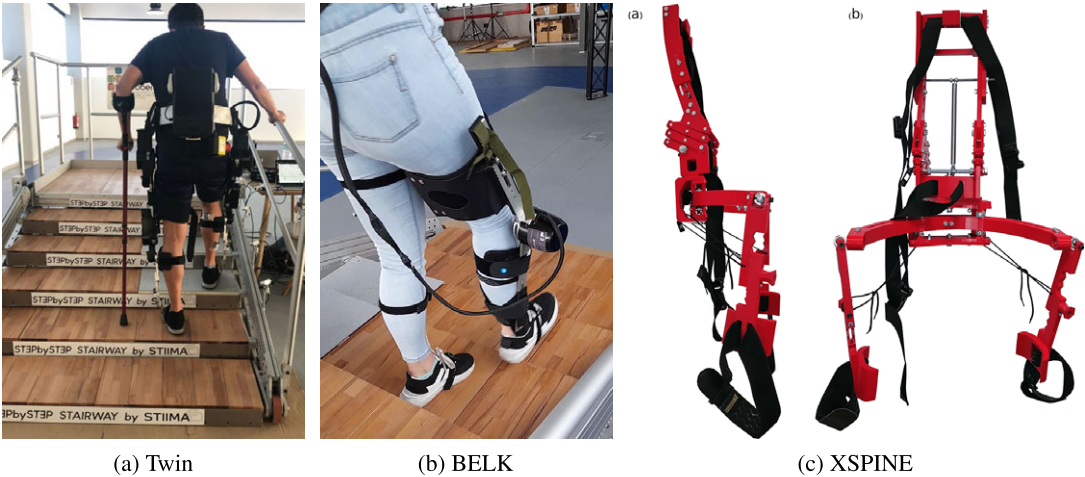
All methods were carried out in accordance with relevant guidelines and regulations, with informed consent obtained from all subjects.

### 5.2. Experimental tests

To evaluate the effectiveness of the proposed protocols, three exoskeletons from different manufacturers were tested: two lower-limb exoskeletons (LLEs) and one back-support exoskeleton (Figure 8). Each exoskeleton is described in the relevant subsection, along with a summary of protocol testing and example results, followed by a brief discussion. While these results are not intended to provide a comprehensive assessment of exoskeleton performance, they serve as a valuable demonstration of the versatility of the testbed and the potential of the proposed metrics. These initial findings illustrate the broad applicability of the approach, highlighting its capacity to effectively evaluate a variety of devices.

#### 5.2.1. The TWIN: physical interaction protocol

The *physical interaction protocol* was tested using the TWIN, an LLE developed by the Italian Institute of Technology (IIT) with support from the Istituto Nazionale per la Assicurazione contro gli Infortuni sul Lavoro (INAIL). TWIN is a novel, modular LLE designed for personal use by individuals with spinal cord injuries. A detailed description of the system and its functionalities is available in Laffranchi et al. (2021). Testing was conducted on a healthy subject under two conditions: stair pitches with step heights of 11 and 17 cm.



**Figure 8.** (a) A user, wearing TWIN, ascending the STEPbySTEP staircase in the 11 cm step-height configuration; a frame of the right double-support phase. (b) BELK exoskeleton while descending the STEPbySTEP staircase. (c) XSPINE back-support exoskeleton.

**Table 2.** Gait phases in seconds and gait cycle percentage, relating to a step height of 11 (Cond 1) and 17 cm (Cond 2). St = Stance; SW = Swing; DS = Double Support; GC = Gait Cycle

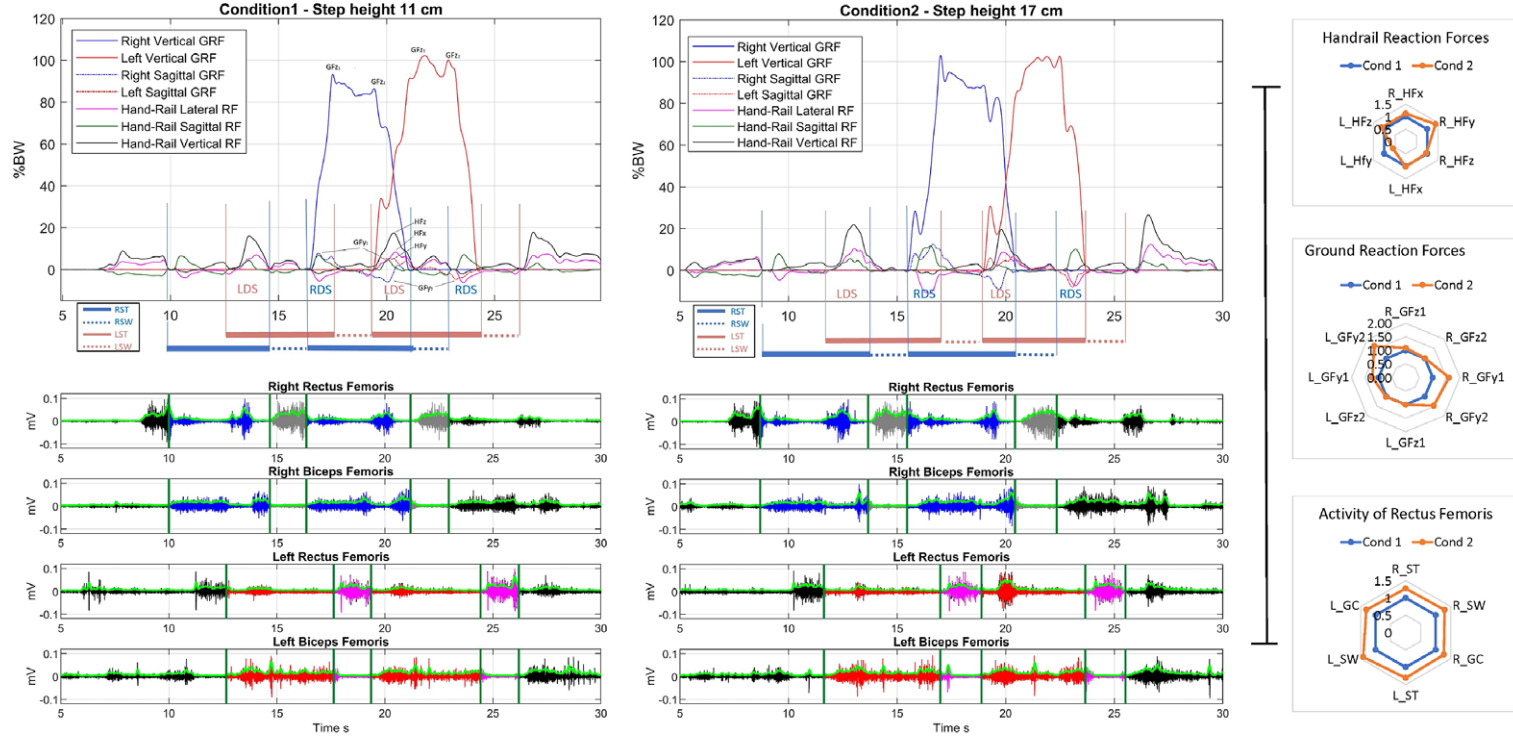
	Left foot				Right foot			
	St	Sw	DS	GC	St	Sw	DS	GC
Cond 1	5 s	1.7 s	1.9 s	6.8 s	4.8 s	1.7 s	1.3 s	6.5 s
Cond 2	5.1 s	1.9 s	1.8 s	6.9 s	5 s	1.9 s	1.5 s	6.8 s
	<b>St</b>	<b>Sw</b>	<b>DS</b>	<b>GC</b>	<b>St</b>	<b>Sw</b>	<b>DS</b>	<b>GC</b>
Cond 1	74%	26%	28%	100%	73%	27%	20%	100%
Cond 2	73%	27%	25%	100%	73%	27%	22%	100%

Example results from two trials in these conditions are presented in both numerical and graphical formats, followed by a brief discussion. Table 2 shows the gait phase durations for each condition. Interestingly, the relative and absolute temporal parameters remain stable despite the increased step height, indicating consistent support from the exoskeleton. Compared to regular walking, the stance phase is slightly prolonged (73% vs. 60% in walking (Alamdari and Krovi, 2017), while the swing phase is reduced. The gait cycle (GC) duration is approximately doubled relative to normal walking (around 7 vs. 3 s (Alamdari and Krovi, 2017), although left and right gait remains symmetrical.

Figure 9 displays the ground and handrail reaction forces and the EMG activity of the left and right rectus femoris (RF) and biceps femoris (BF) for condition 1 (left image) and condition 2 (central image). The forces are expressed as percentages of the combined weight of the subject and exoskeleton.

In both conditions 1 and 2, the left and right muscle activation patterns appear physiological (lower graphs), indicating that the subject is comfortable using the exoskeleton without requiring compensatory movements. All reaction forces (Figure 9, left panels, upper figures) exhibit repeatable patterns across all gait phases. Differences in the handrail force component are observed in the double-support (DS) phases. Specifically, HFz is 10 times larger during the left DS phase (i.e., when the left foot precedes the right) than during the right DS phase. Additionally, the maximum vertical component of the ground reaction force (GRF) does not reach 100% of the total weight, as the subject leans on the handrail during the left DS phase but leans on a cane during the right DS phase (note that cane forces were not measured in this experiment). All the above considerations apply to both conditions.





**Figure 9.** Ground and handrail reaction forces and the EMG activity of the rectus and biceps femoris (left panels). All signals are synchronized, and the gait phases are shown for clarity. In the right panel: an example of comparison between data of condition 2 (step height = 17 cm) and condition 1 (step height = 11 cm), which is taken equal to 1 as reference.

The physical interaction metrics (Section 3.1) for condition 2 are plotted in Figure 9 (right panel) against those for condition 1, used here as reference values. This data presentation format emphasizes differences between the two conditions, with the following observations. The vertical and lateral components of the handrail forces are similar across the two conditions (see Figure 9, right panel, upper image). In contrast, the sagittal component HFy (green line in Figure 9, left panels) in condition 2 is larger during the right DS phase and smaller during the left DS phase compared to condition 1 (see Figure 9, right panel, upper image). This suggests that with higher steps, the user exerts more pull on the handrail when the right foot precedes the left, while less pull is needed when the left foot precedes the right, as seen with lower steps. Correspondingly, left and right GFy1 and GFy2 (Figure 9, right panel, central image) have higher values in condition 2 than in condition 1, indicating that braking and pushing forces increase with step height. Finally, RF EMG activity also rises with step height (Figure 9, right panel, lower image), which aligns with the increased joint torques and power requirements associated with ascending higher steps (Dollar and Herr, 2008).

### 5.2.2. BELK – temporal phases protocol

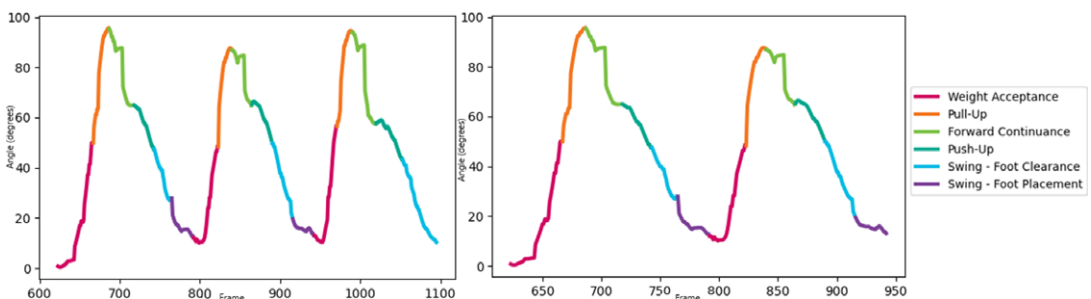
The *temporal phases protocol* was tested using the BELK, an LLE developed by Gogoa (<https://www.gogoa.eu/belk>). BELK is an active knee exoskeleton (Figure 8b) designed to support the rehabilitation of patients recovering from surgical procedures or managing neurological conditions. The system allows customization of the angular movement range and walking speed, and it can adjust the level of assistance by generating a supportive force field tailored to the patient's individual needs and recovery progress.

The objective of this test was to evaluate the effectiveness of the ML algorithm in identifying temporal phases. The models were trained on a dataset of healthy subjects ascending and descending stairs without exoskeletons. The goal was to assess the performance of the model when applied to subjects using an exoskeleton, even though it was initially trained on data from individuals without one. To demonstrate the validity of the algorithm, the identified sub-phases are highlighted in the knee joint angle graph during the ascending phase (Figure 10). Notably, the sub-phase identification using KNN observed in trials with the exoskeleton (Figure 10, right) shows a clear definition when compared to the manually labeled ground truth of a subject ascending stairs without an exoskeleton (Figure 10 left).

Please insert figure 10 here

Table 3 shows the ascending computed temporal metrics for Figure 10. The figure does not present descending data but the ML algorithm works perfectly well for descending scenarios as well.

As seen in Figure 10, the ML algorithm correctly identifies the gait phases. Failures in identification may occur if the exoskeleton cannot adequately replicate typical knee motion. When analyzing the walking pattern of a healthy individual without an exoskeleton ascending stairs, certain patterns emerge in the peaks of the knee rotation signal, which correlate with knee–ankle movements essential for stair ascent and descent. If a user with an exoskeleton that has incorrect ascending or descending movement patterns performs the test, these patterns may change, as either the knee or the ankle may not rotate properly.



**Figure 10.** Ascending mocap data example. Ground truth sub-phases manually labeled for a healthy user without an exoskeleton (left) compared to ML algorithm predictions with an exoskeleton (right).

**Table 3.** Temporal ascending metrics for Figure 10. Gray cells correspond to the % with respect to the gait cycle. Stance is the sum of the stance sub-phases, and swing is the sum of the swing sub-phases

Sub-phase	Time (s) and % of the total									
	Ground truth labeled						ML prediction			
	1° Peak		2° Peak		3° Peak		1° Peak		2° Peak	
	Time	%	Time	%	Time	%	Time	%	Time	%
Stance – weight acceptance	.367	26.1	.258	20.5	.233	–	.372	26.3	.259	20.04
Stance – pull-up	.167	11.9	.142	11.3	.158	–	.171	12.1	.145	11.4
Stance – forward continuance	.258	18.3	.208	16.5	.250	–	.246	17.4	.202	15.9
Stance – push-up	.208	14.8	.208	16.5	.275	–	.215	15.2	.215	16.9
Swing – foot clearance	.183	13.0	.217	17.2	.358	–	.183	12.9	.215	16.9
Swing – foot placement	.225	16.0	.225	17.9	–	–	.227	16.1	.233	18.4
Total stance	1.000	71.0	.816	64.9	–	–	1.004	71.0	.821	64.7
Total swing	.408	29.0	.442	35.1	–	–	.410	29.0	.448	35.3
Gait cycle	1.408	100	1.258	100	–	–	1.414	100	1.269	100

This can be observed visually as a “rigid” movement of the knee and ankle, indicating limited rotation. These results could help exoskeleton designers consider necessary mechanical and software modifications to better mimic natural joint motions during stair ascent and descent.

Using this method, particularly alongside other biomechanical measures such as kinematics and dynamics, could lead to more detailed analysis, potentially supporting more precise benchmarking of exoskeleton effectiveness across various movement phases. The algorithm shows effectiveness in detecting temporal phases, especially when knee motion during exoskeleton use closely mirrors physiological patterns. However, some limitations may emerge if movement patterns deviate significantly from the expected norm. For example, variations could arise if the ankle movement is restricted or actively controlled, potentially impacting algorithm performance. In such cases, retraining the model for specific exoskeleton configurations could be beneficial. Alternatively, if the system allows adjustments to assistance settings, these results could guide the optimization of those settings.

This approach may also find relevance beyond rehabilitation. Accurately identifying sub-phases could contribute to advancements in assistive technologies for individuals with chronic conditions. Additionally, applying machine learning algorithms trained on healthy individuals to those with varying movement patterns suggests a degree of adaptability, which could improve exoskeleton effectiveness across diverse user populations.

5.2.3. XSPINE – the human factors protocol

A back-support exoskeleton, XSPINE, was tested, with a complete description available in Roveda et al. (2020, 2022)).

The human factors metrics and protocol were applied to evaluate the exoskeleton. The dataset included six participants (three men and three women, aged between 25 and 35) who voluntarily participated in the study and completed the previously described human factors protocol. Table 4 provides descriptive statistics for the scales used with participants. Results indicate that participants rated the XSPINE exoskeleton positively in terms of usability ( $4.0 \pm .7$ ), with favorable assessments of its usefulness ( $4 \pm .4$ ) and overall satisfaction ( $.9 \pm .8$ ).

In addition, the exoskeleton was reported to exert very low to low pressure across various body areas. The shoulders ( $3.4 \pm 1.8$ ), the right lower back and hip ( $2.8 \pm 2.0$ ), and the left lower back and hip ( $2.7 \pm 2.0$ ) were identified as the most uncomfortable areas.

Regarding cognitive workload, the variables used to operationalize cognitive performance are the average response time and the number of errors in both conditions (without and with the exoskeleton). While we acknowledge that our statistical power may be limited, we ran a paired-sample t-test on IBM SPSS 22 to analyze the average score difference of each participant in the two conditions – since the dual-

**Table 4.** Descriptive statistics for usability, acceptance, and local perceived discomfort scales

Human factors scale	M	SD	Max	Min
Usability	4.0	.7	5	1
Acceptance				
Usefulness	.4	.4	2	−2
Satisfaction	.9	.8	2	−2
Local perceived discomfort scale				
Shoulders	3.4	1.8	6	0
Right upper back	1.5	2.0	6	0
Right lower back and hip	2.8	2.0	7	0
Right upper leg	1.2	1.3	4	0
Left upper leg	1.1	1.4	4	0
Left lower back and hip	2.7	2.0	7	0
Left upper back	1.7	2.1	6	0
Right lower arm	0	0	0	0
Right upper arm	1.4	2.3	6	0
Chest	1.2	1.7	4	0
Left upper arm	1.4	2.2	6	0
Left lower arm	0	0	0	0

task paradigm is a within-subject experimental design. Table 5 displays the descriptives and results of the paired-sample t-test. The results show no significant performance difference between the two conditions for both response time and the number of errors. In particular, the response time discrepancy between the no-exoskeleton ( $1128 \pm 203$  ms) and exoskeleton one ( $1072 \pm 161$  ms) conditions displays no statistically significant difference ( $t = 2.3$ ;  $p > .05$ ). The same applies to the difference in the number of errors between the no-exoskeleton condition ( $.3 \pm .5$ ) and with exoskeleton 2 ( $1.7 \pm 1.6$ ), which is not statistically relevant ( $t = -2.0$ ;  $p > .05$ ). Therefore, these preliminary findings point to a possible trend for future research, so that wearing the exoskeleton does not seem to influence users' cognitive workload and performance for both exoskeletons tested in this study. This is particularly relevant for the tasks envisaged by the STEPbySTEP testbed described previously (e.g., climbing and descending stairs).

### 5.3. Discussion

Evaluating exoskeleton performance across varied physical tasks is essential for understanding their capabilities and refining the evaluation protocols that enable such testing. Stair climbing is of particular interest due to its prevalence in daily life and the unique biomechanical challenges it presents. While research on stair ascent and descent has historically been limited, the development of exoskeletons capable of addressing these complex movements highlights the need for structured evaluation protocols as exoskeleton technology advances.

In this work, we introduced a reconfigurable, sensorized staircase testbed designed for evaluating and benchmarking exoskeleton functionality during stair negotiation. While the testbed itself provides a comprehensive and adaptable platform for assessing exoskeletons, our focus here was on testing the

**Table 5.** Descriptive and paired-sample t-test results for response time and error number

Variable	M	SD	t	p value
Pair 1				
1. Response time (no exoskeleton)	1128.1	203.3		
2. Response time (with exoskeleton)	1072.0	161.4		
1–2	56.1	59.4	2.3	.07
Pair 2				
3. Errors number (no exoskeleton)	.3	.5		
4. Errors number (with exoskeleton)	1.7	1.6		
3–4	−1.3	1.6	−2.0	.10

protocols and demonstrating their potential. The preliminary results show that the protocols can capture relevant biomechanical and cognitive metrics. These initial measurements are tentative and serve as a showcase for the platform's ability to support diverse exoskeleton types rather than offering definitive benchmarks.

We applied the protocols to two lower-limb exoskeletons (LLEs) with distinct characteristics. TWIN, a modular exoskeleton developed by the Italian Institute of Technology, demonstrated reliable gait phase consistency, but its slower-than-natural gait – particularly in its extended stance phase (7 s compared to the typical 3 s for unaided walking) – illustrated the testbed ability to capture unique device characteristics. In contrast, BELK, an active knee exoskeleton by Gogoa, achieved faster, more dynamic movements. These differences highlight the versatility of the testbed in accommodating various device configurations and performance levels.

The *temporal phases protocol* was tested with BELK using an ML algorithm trained on healthy subject data. Results demonstrated that the algorithm could identify gait sub-phases even when using the exoskeleton, suggesting that the protocol could work across different user profiles and devices. However, further adaptations – such as retraining the algorithm with data from exoskeleton users – could improve accuracy when movement patterns deviate from physiological norms, such as limited ankle mobility due to the device.

Additionally, we tested cognitive workload metrics to assess human–exoskeleton interactions. Preliminary tests using the XSPINE back-support exoskeleton showed no significant difference in cognitive performance between exoskeleton and no-exoskeleton conditions, illustrating the testbed potential to integrate cognitive metrics into exoskeleton evaluation. While these results do not provide conclusive trends, they emphasize the importance of considering cognitive aspects in future research on exoskeleton performance, especially for tasks like stair navigation where both physical and cognitive demands are critical.

The STEPbySTEP testbed, along with its protocols and metrics, is now available at the EUROBENCH facility in Brunete, Spain (<https://github.com/eurobench>). This standardized environment offers a valuable platform for exoskeleton testing and benchmarking, supporting the field as new devices for stair navigation continue to emerge. Currently, there are protocols and metrics for validating exoskeletons, but, to the best of our knowledge, there is a lack of standardized benchmarking frameworks, specifically for these devices. The Eurobench initiative was created to address this gap by developing an infrastructure dedicated to the benchmarking of wearable robotics, ensuring a more consistent and comparable assessment across devices and real-world contexts. The protocols and metrics presented in our work align with Eurobench guidelines and requirements, ensuring compatibility with this benchmarking framework. Such a testbed, together with the other testbeds within the EUROBENCH project, can enhance exoskeleton development. In particular, from the industrial and commercialization point of view, it will be possible to perform standardized and structured evaluations of the performance of an exoskeleton product, being market ready.

While this study highlights the potential of the STEPbySTEP testbed and its associated protocols to capture meaningful metrics for stair climbing scenarios, certain limitations warrant further investigation. Two main aspects deserve particular consideration.

First, the protocols and metrics were tested using exoskeletons worn only by healthy subjects. While we expect kinematics and dynamics to differ when used by individuals with neurological impairments, the testbed is fully capable of collecting and processing relevant data. The results obtained will naturally vary depending on the user group, but this variation is precisely what the platform is designed to capture and analyze.

Second, the number of exoskeletons used was limited to two lower limbs and one back-support device. However, this does not constrain the testbed applicability. The modular and customizable design of the staircase allows the evaluation of a broad range of exoskeletons, including those developed for non-rehabilitative applications. By adjusting the step configuration, the platform can be adapted for testing high-performance exoskeletons or even legged robots (e.g., humanoids) used in military, industrial, or



space environments. The testbed supports step heights ranging from 11 cm to over 1 m, enabling assessments across different operational contexts.

One final limitation of this study is the lack of a direct comparison between two exoskeletons using the same protocol, making it impossible to determine any particular strengths or weaknesses of each device. However, the primary objective of this study was to assess the scale along with the protocols, verifying their applicability and sensitivity. Nevertheless, we are confident that we have developed a benchmarking framework capable of highlighting both the strengths and weaknesses of different exoskeletons. Supporting this claim, the protocols and metrics have proven to be sensitive in detecting differences between subjects using the same exoskeleton, as well as intra-subject differences when applying the scale in two different setups.

In summary, despite these limitations, the STEPbySTEP testbed and its associated protocols demonstrate the potential to capture important metrics in realistic stair climbing scenarios. The observed gait patterns, usability metrics, and biomechanical data provide insights into how exoskeletons might perform in real-world applications, such as rehabilitation clinics and industrial settings. To bridge the gap between controlled testing and real-world application, the staircase testbed can be configured to replicate specific real-life scenarios. This adaptability allows for early identification of potential challenges and facilitates the development of targeted solutions, improving exoskeleton integration into diverse environments. These protocols, metrics, and the adaptability of the testbed together provide a foundation for future research, enabling the development of standardized benchmarks, supporting exoskeleton improvement, and contributing to the advancement of both rehabilitative and assistive exoskeleton technologies.

## 6. Conclusions

This paper presented STEPbySTEP, a reconfigurable, sensorized staircase testbed created to support standardized exoskeleton evaluation within Eurobench. Through trials with three exoskeletons, we showcased the adaptability of our protocols and tentative metrics, demonstrating the platform's potential for versatile benchmarking across various devices. While preliminary, these findings underscore the testbed promise as a flexible evaluation tool. Future work will focus on refining protocols and expanding metrics to support broader benchmarking applications, advancing the development of standardized approaches for exoskeleton performance assessment.

**Data availability statement.** Data will be made available upon request. Please contact [loris.roveda@supsi.ch](mailto:loris.roveda@supsi.ch).

**Acknowledgments.** The authors want to thank Roberto Bozzi and Joao Carlos Dalberto (CNR-STIIMA) for their support in designing and building the testbed; Indya Ceroni, Stefano Maludrottu, Angelo Musumarra, Silvia Scarpetta, Marianna Semprini, Christian Vassallo, and Gaia Zinni (Rehab Technologies Lab, IIT) for the management and control of the TWIN exoskeleton; Maurizio Ferrarin, Giorgia Fusaroli, Johanna Jonsdottir, and Tiziana Lencioni (LAMoBIR and LaRiCE at IRCCS Fondazione Don Carlo Gnocchi) for their expertise and management of the acquisitions.

**Authors contribution.** The authors contributed to the paper as follows: Conceptualization, M.C., M.M., D.B., F.F., and L.R.; Realization of the testbed, M.C., N.M., M.M., F.A., T.D., D.B., M.E., I.D., S.A., and L.R.; Software, M.C., N.M., D.B., G.P., F.F., M.M., A.A.S., and L.R.; Supervision, M.C., M.M., D.B., L.P., and L.R.; Paper writing and editing, M.C., M.M., D.B., G.P., F.F., and L.R.

**Funding statement.** This paper has received funding from the European Union's Horizon 2020 research and innovation program, via an Open Call issued and executed under Project EUROBENCH (grant agreement no. 779963) – STEPbySTEP project, XSPINE project, and REMOTE\_XSPINE project.

This paper has received funding from the EIT Manufacturing – SUPERHUMAN project.

**Competing interests.** The authors declare no competing interests exist.

**Ethical standard.** Ethical approval for this study was provided by the CSIC Ethics Committee (approval number 091/2021).

## References

- Alamdari A and Krovi V** (2017) Chapter two - a review of computational musculoskeletal analysis of human lower extremities. In Ueda J and Kurita Y (eds), *Human Modelling for Bio-Inspired Robotics*: Academic Press, pp. 37–73. <https://doi.org/10.1016/B978-0-12-803137-7.00003-3>.
- Alingh JF, Groen BE, Kamphuis JF, Geurts ACH and Weerdesteyn V** (2021) Task-specific training for improving propulsion symmetry and gait speed in people in the chronic phase after stroke: A proof-of-concept study. *Journal of Neuroengineering and Rehabilitation* **18**(1), 69
- Awad LN, Bae JH, O'Donnell K, De Rossi SMM, Hendron K, Slood LH, Kudzia P, Allen S, Holt KG, Ellis TD and Walsh CJ** (2017) A soft robotic exosuit improves walking in patients after stroke. *Science Translational Medicine* **9**(400), eaai9084
- Awad LN, Lewek MD, Kesar TM, Franz JR and Bowden MG** (2020) These legs were made for propulsion: Advancing the diagnosis and treatment of post-stroke propulsion deficits. *Journal of Neuroengineering and Rehabilitation* **17**(1), 139
- Benson I, Hart K, Tussler D and van Middendorp JJ** (2016) Lower-limb exoskeletons for individuals with chronic spinal cord injury: Findings from a feasibility study. *Clinical Rehabilitation* **30**(1), 73–84
- Bush P and ten Hoppel S** (2017) An integrated craft and design approach for wearable orthoses. *Design for Health* **1**(1), 86–104
- Collins SH, Wiggin MB and Sawicki GS** (2015) Reducing the metabolic cost of walking with an untethered exoskeleton. *Nature* **522**(7555), 212–215
- Dalla Gasperina S, Roveda L, Pedrocchi A, Braghin F and Gandolla M** (2021) Review on patient-cooperative control strategies for upper-limb rehabilitation exoskeletons. *Frontiers in Robotics and AI* **8**
- Damiano DL, Martellotta TL, Sullivan DJ, Granata KP and Abel MF** (2000) Muscle force production and functional performance in spastic cerebral palsy: Relationship of cocontraction. *Archives of Physical Medicine and Rehabilitation* **81**(7), 895–900. <https://doi.org/10.1053/apmr.2000.5579>
- De Bock S, Ghillebert J, Govaerts R, Tassignon B, Rodríguez-Guerrero C, Crea S, Veneman J, Geeroms J, Meeusen R and De Pauw K** (2022) Benchmarking occupational exoskeletons: An evidence mapping systematic review. *Applied Ergonomics* **98**, 103582
- Dollar AM and Herr H** (2008) Lower extremity exoskeletons and active orthoses: Challenges and state-of-the-art. *IEEE Transactions on Robotics* **24**(1), 144–158
- Farris RJ, Quintero HA and Goldfarb M** (2012) Performance evaluation of a lower limb exoskeleton for stair ascent and descent with paraplegia. *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1908–1911
- Galle S, Malcolm P, Collins SH and De Clercq D** (2017) Reducing the metabolic cost of walking with an ankle exoskeleton: Interaction between actuation timing and power. *Journal of Neuroengineering and Rehabilitation* **14**(1), 35. <https://doi.org/10.1186/s12984-017-0235-0>
- Hamberg-van Reenen HH, van der Beek AJ, Blatter BM, van der Grinten MP and van Mechelen W** (2008) Does musculoskeletal discomfort at work predict future musculoskeletal pain? *Ergonomics* **51**(5), 637–648
- Harper NG, Wilken JM and Neptune RR** (2018) Muscle function and coordination of stair ascent. *Journal of Biomechanical Engineering* **140**(1), 10
- Kermavnar T, de Vries AW, de Looze MP and O'Sullivan LW** (2021) Effects of industrial back-support exoskeletons on body loading and user experience: An updated systematic review. *Ergonomics* **64**(6), 685–711
- Koch I, Poljac E, Müller H and Kiesel A** (2018) Cognitive structure, flexibility, and plasticity in human multitasking—An integrative review of dual-task and task-switching research. *Psychological Bulletin* **144**(6), 557
- Laffranchi M, D'Angella S, Vassallo C, Piezzo C, Canepa M, De Giuseppe S, Di Salvo M, Succi A, Cappa S, Cerruti G, Scarpetta S, Cavallaro L, Boccardo N, D'Angelo M, Marchese C, Saglia JA, Guanziroli E, Barresi G, Semprini M, et al.** (2021) User-Centered design and development of the modular TWIN lower limb exoskeleton. *Frontiers in Neurobotics* **15**, 709731
- Lafranchi M, Botsch M, Tonietti G and Bicchi A** (2011) Sizing actuators in wearable robotics: A data-driven model for the actuation of the lower limb in human locomotion. *2011 IEEE International Conference on Rehabilitation Robotics*, 1–8
- Latash ML** (2018) Muscle coactivation: Definitions, mechanisms, and functions. *Journal of Neurophysiology* **120**(1), 88–104. <https://doi.org/10.1152/jn.00084.2018>
- Lewis JR and Sauro J** (2017) Revisiting the factor structure of the system usability scale. *Journal of Usability Studies* **12**(4), 183–192
- Maugliani N, Caimmi M, Malosio M, Airoidi F, Borro D, Rosquete D, Sergio A, Giusino D, Fraboni F, Ranieri G, et al.** (2020) Lower-limbs exoskeletons benchmark exploiting a stairs-based testbed: The stepbystep project. *International Symposium on Wearable Robotics*, 603–608
- Mauri A, Lettori J, Fusi G, Fausti D, Mor M, Braghin F, Legnani G and Roveda L** (2019) Mechanical and control design of an industrial exoskeleton for advanced human empowering in heavy parts manipulation tasks. *Robotics* **8**(3), 65
- Novak AC and Brouwer B** (2014) Relationship between stair ambulation with and without a handrail and Centre of pressure velocities during stair ascent and descent. *Gait & Posture* **39**(1), 158–164. <https://doi.org/10.1016/j.gaitpost.2013.06.005>
- O'Sullivan L, Power V, Virk G, Masud N, Haider U, Christensen S, Bai S, Cuyppers L, D'Havé M and Vonck K** (2015) End user needs elicitation for a full-body exoskeleton to assist the elderly. *Procedia Manufacturing* **3**, 1403–1409
- Orekhov G, Fang Y, Cuddeback CF and Lerner ZF** (2021) Usability and performance validation of an ultra-lightweight and versatile untethered robotic ankle exoskeleton. *Journal of Neuroengineering and Rehabilitation* **18**(1), 1–16

- Plummer-D'Amato P, Altmann LJP, Saracino D, Fox E, Behrman AL and Marsiske M** (2008) Interactions between cognitive tasks and gait after stroke: A dual task study. *Gait and Posture* **27**(4), 683–688
- Recinos E, Abella J, Riyaz S and Demircan E** (2020) Real-time vertical ground reaction force estimation in a unified simulation framework using inertial measurement unit sensors. *Robotics*, **9**(4), 88. MDPI AG. ISSN:2218-6581. <http://doi.org/10.3390/robotics9040088>.
- Roveda L, Haghsheenas S, Caimmi M, Pedrocchi N and Tosatti L** (2019) Assisting operators in heavy industrial tasks: On the design of an optimized cooperative impedance fuzzy-controller with embedded safety rules. *Frontiers in Robotics and AI* **6**, 75
- Roveda L, Pesenti M, Rossi M, Covarrubias M, Pedrocchi A, Braghin F and Gandolla M** (2022) *User-centered back-support exoskeleton: Design and prototyping*. *CIRP Annals*.
- Roveda L, Savani L, Arlati S, Dinon T, Legnani G and Tosatti LM** (2020) Design methodology of an active back-support exoskeleton with adaptable backbone-based kinematics. *International Journal of Industrial Ergonomics* **79**, 102991
- Salarian A, Russmann H, Vingerhoets FJ, Dehollain C, Blanc Y, Burkhard PR and Aminian K** (2004) Gait assessment in Parkinson's disease: Toward an ambulatory system for long-term monitoring. *IEEE Transactions on Biomedical Engineering* **51**(8), 1434–1443
- Schenck C and Kesar TM** (2017) Effects of unilateral real-time biofeedback on propulsive forces during gait. *Journal of Neuroengineering and Rehabilitation* **14**(1)
- Schmidt B, Böhm V, Günther T, Kretschmer F, Waschek S, Baumann M, Wörteler M and Riedl M** (2017) Wearable technologies for factory workers: Exoskeletons, wearable sensors, and smart personal protective equipment in industry 4.0. *Wearable Technologies* **1**(2), 17–32
- Shaw EP, Rietschel JC, Hendershot BD, Pruziner AL, Miller MW, Hatfield BD and Gentili RJ** (2018) Measurement of attentional reserve and mental effort for cognitive workload assessment under various task demands during dual-task walking. *Biological Psychology* **134**, 39–51
- Shore L, Power V, Hartigan B, Schüle S, Graf E, de Eyto A and O'Sullivan L** (2019) Exoscore: A design tool to evaluate factors associated with technology acceptance of soft lower limb exosuits by older adults. *Human Factors*, 1–20
- Stessman J, Rottenberg Y and Jacobs JM** (2017) Climbing stairs, handrail use, and survival. *The Journal of Nutrition, Health & Aging* **21**(2), 195–201
- Torricelli D, Veneman J, Gonzalez-Vargas J, Mombaur K and Remy DC** (2019) Editorial: Assessing bipedal locomotion: Towards replicable benchmarks for robotic and robot-assisted locomotion. *Frontiers in Neurorobotics* **13**
- Tricomi E, Mossini M, Missiroli F, Lotti N, Zhang X, Xiloyannis M, Roveda L and Masia L** (2023) Environment-based assistance modulation for a hip exosuit via computer vision. *IEEE Robotics and Automation Letters* **8**(5), 2550–2557
- Van Der Laan JD, Heino A and De Waard D** (1997) A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies* **5**(1), 1–10
- Vaughan CL, Davis BL and O'Connor JC** (1999) *Dynamics of Human Gait*. Human Kinetics.
- Winter DA** (1991) *The Biomechanics and Motor Control of Human Gait: Normal, Elderly and Pathological*. University of Waterloo Press.
- Wolff J, Parker C, Borisoff J, Mortenson BW and Mattie J** (2014) A survey of stakeholder perspectives on exoskeleton technology. *Journal of Neuroengineering and Rehabilitation* **11**
- Zhang X, Chen X, Huo B, Liu C, Zhu X, Zu Y, Wang X, Chen X and Sun Q** (2023) An integrated evaluation approach of wearable lower limb exoskeletons for human performance augmentation. *Scientific Reports* **13**(1), 4251