Preliminary Development of Test Methods to Evaluate Lower Body Wearable Robots for Human Performance Augmentation

Brian Carlson, Adam Norton, and Holly Yanco New England Robotics Validation and Experimentation (NERVE) Center University of Massachusetts Lowell Lowell, Massachusetts 01854 brian_carlson@student.uml.edu, anorton@cs.uml.edu, holly@cs.uml.edu

Abstract-Wearable robotics are prevalent in the medical domain for prosthetic and rehabilitation uses, and those for performance augmentation of able bodied people for industrial and military domains are also on the rise. Some common metrics exist for evaluating these systems, such as metabolic cost, but they are incomplete with regards to the many other characteristics to be compared between systems. To this end, we are developing holistic test methods, specifically those for lower body wearable robots focused on performance augmentation. We discuss the test methods' structure, considerations, and development. Prototypes of the test methods have been exercised with a user wearing a B-Temia Dermoskeleton system. Our initial development has led to a baseline set of basic and applied tasks that can be evaluated comparatively between performing the task without and with the system, measuring simple task-based metrics based on time, repetitions, loading capacity, and range of motion. Future work includes exercising the test methods with more wearable robotic systems and formulating a working task group to assist in driving development.

I. INTRODUCTION

Wearable robots and exoskeletons are a growing technology that has seen rapid growth in recent years. The most common use of the modern day wearable robot is in medical rehabilitation, with exoskeletons such as the ReWalk 6.0 [1] and the Ekso GT [2], helping users regain limb control and movement while recovering from an injury. Exoskeleton systems are also used in the commercial space with exoskeletons such as the Ekso Works [2], demonstrating use in construction and industrial environments by providing passive augmentation to an able bodied user. For military use, exoskeletons from the Warrior Web Program [3] and the B-Temia Dermoskeleton [4] (see Figure 1), are providing active human augmentation to able bodied users, to extend and improve task performance.

From initial discussions stemming from a roundtable workshop held in December 2014 [5], it has become clear that a standardized way to evaluate these systems is needed. The use of common test methods will enable developers to test their exoskeletons using the same methodology that could later be used to guide procurements and meet safety specifications. Each of the three domains described have different goals in terms of determining the value of a system. There are also different designs for systems that are made to augment the upper and lower body. We are currently in the early stages of investigating how to develop task-based test methods, specifically those for lower body wearable robots to be used



Fig. 1. The B-temia Dermoskeleton [4] worn on the lower body with the battery pack in a backpack while performing a munitions loading task.

in the military domain, which also has bearings in industrial domains.

II. RELATED WORK

Many existing publications on wearable robotics are focused on the development of a particular exoskeleton and testing the design by measuring metabolic cost, commonly by performing simple motions or walking on a treadmill. The use of a treadmill has the subject remain in a fixed physical location, allowing for the use of biometrics for evaluation purposes, such as oxygen consumption (VO2), heart rate (HR), and electromyography (EMG).

Herr *et al.* [6] developed an exoskeleton that provided torque and positive power to the ankle during walking. During

testing, subjects were allowed to warm up while wearing the device in an assisted mode; they then performed three walking trials on a treadmill while wearing a weighted vest. The first trial was a control with no device, the second was with the device on assisted mode, and the third trial was another control with no device. Finally, the subjects stood to provide a resting metabolic state. The subjects wore a pulmonary gas exchange measurement instrument during all trials, which was used to measure metabolic cost.

A quasi-passive exoskeleton by Walsh *et al.* [7] was focused on load carriage during walking. Testing of this device was performed over a three day span to avoid fatigue, with an initial orientation session to ensure the subject was acclimated to the device. Data collection was performed on an indoor track, where the subject walked at a set pace for 10 minutes per trial under various conditions: loaded backpack weighing 36 kg (80 lb) with no exoskeleton, the quasi-passive exoskeleton with no backpack, a zero-impedance exoskeleton with no backpack, and a loaded backpack with the quasi-passive exoskeleton. Payload distribution to ground was measured by the strain gauges within the exoskeleton.

Walsh *et al.* [8] also developed a lightweight soft exoskeleton for gait assistance of able bodied healthy adults. Subjects performed numerous walking trials on a flat surface for a set distance. These trials were performed without the system, while wearing the system powered off, and while wearing system powered on. During testing, kinematic and metabolic data were collected, utilizing a motion capture system and a pulmonary gas exchange measurement instrument. It was found that the device had to be adjusted over several trials in order to find the correct actuation of the joints to provide proper assistance.

Shamaei *et al.* [9] also developed a quasi-passive exoskeleton, focusing on the knee and motor adaptation. The device was tested on healthy, able-bodied soldiers by first allowing them to become familiar with the device through practice on flat ground and a treadmill. Practice and test sessions were separated by a day or two to prevent fatigue. On the first day of orientation, test conditions were kept in the same order among all subjects. On the second day of orientation, test conditions were randomized among subjects. For data collection, test conditions were the same randomized order of the second orientation day. A motion capture camera system was used during test sessions record lower extremity motion.

Gams *et al.* [11] observed the effects of an exoskeleton for performing a simple stable motion squat. Subjects performed a squatting task without wearing the exoskeleton, then while wearing it under different control modes, resting approximately 1.5 hours between sessions. They were allowed to warm up while wearing the device before undergoing testing. A railing was present in case of loss of balance. Oxygen consumption, heart rate, and blood oxygen saturation were measured.

Sawicki *et al.* [12] employed the use of a powered ankle exoskeleton on healthy, able bodied users that performed walking at a comfortable stride length on a treadmill. Stride length and treadmill speed were varied across eight trials per user. EMG, metabolic cost, and joint kinematics were measured.

The DARPA Warrior Web [3] program investigates real



Fig. 2. The Marine Corps Load Effects Assessment Program (MC-LEAP) test course. Image from [10].

world tasks and requirements for soldiers for which researchers design exoskeletons. Initial testing was carried out by recording high precision biometric data [13]. Testing was then performed on a two mile cross country walk that traversed many terrain types, while continuing to collect biometric data such as HR and VO2. This a good example of an extension from lab testing into fielding in the real world.

Many suites of standard test methods exist in the robotics domain, such as those for emergency response robots [14] and autonomous industrial vehicles [15]. The test methods in those domains use directly observable performance within physical apparatuses that represent abstract versions of tasks and relevant environmental conditions, exercising different robot capabilities such as mobility, manipulation, and navigation, in a holistic fashion. The Marine Corps Load Effects Assessment Program (MC-LEAP) [10] functions similarly to these kinds of test methods, in that a series of physical tasks are performed as part of an obstacle course to measure the impact of worn or carried items on a soldier. A similar structure with respect to test method design can be employed for wearable robots.

III. APPROACH

Our approach is to build on a combination of the common testing elements prevalent in wearable robotics research and existing standard test method committees focused on evaluating robots (as suggested in [16]). The key characteristics from those described in section II are:

- Allow for orientation and practice sessions for the subject to become familiar with the device. These sessions occur on days prior to testing.
- Test the subject over numerous conditions with respect to the exoskeleton (such as its assistance level, donned on or off, etc.), in a randomized order.
- Collect objective data such as observable task completion, biometric data, and motion nuances.
- Collect subjective data through surveys.
- Analyze this data and compare it objectively across multiple users.

- Design test apparatuses whose physical features dictate performance and provide discernible scoring measurements.
- Record task-based metrics for holistic systems testing, abstracted to be broadly applicable to many applications.
- Fabricate test methods using readily available and inexpensive building materials.

When performing a task while wearing an exoskeleton for performance augmentation, the system reacts to the user's movements and is intended to impact his or her capability in a positive manner. The user must be able to perform the tasks in some capacity without the system, as his or her capability without and with the system should be compared. If the system is actively powered, performance of the latter can be further split into wearing the system when it is powered on or off.

Given that the system is designed to improve the performance of an able bodied person, such as elongating performance time or increasing payload capacity, tasks should elicit this extra work for the system to assist with. However, some baseline testing is needed to determine if the system restricts any basic movements of the wearer. This aspect introduces several areas where the impact of the system can be evaluated in a task-based manner, such as constriction of range of motion, allowing tasks to be performed for longer or at higher intensity (e.g., carrying more weight than usual), etc.

When performing a test method that mixes modalities or requires a quick succession of perceived mode changes, an exoskeleton's ability to transition between activities and poses can be evaluated. An actively powered exoskeleton essentially predicts the movements of its wearer by reacting to his or her movements. Some operate in disparate modes (generally aided by additional sensors on multiple limbs) for walking, climbing, etc., but can streamline the transition between them. The transition can also elicit instabilities when performing tasks that involve similar bodily movements, but use a different overall sequence of actions. Test methods in this domain should be agnostic to the software solution used by a particular system, and should simply allow for that solution to be exhibited.

The use of stationary performance spaces like treadmills can introduce high testing costs and can limit the ability to test combinations or transitions between tasks beyond that of walking/running speed [17]. Additionally, some wearable robots rely on physical movement of the user through space, rather than just local changes to the user's joints, to function properly. As such, our approach uses larger environments through which the user can traverse.

The test methods are designed from a high level to allow for biometric sensors to be added as part of a protocol if desired. Biometrics can be limited for our purposes due to high costs, lack of mobility, and potential obstructions caused by the wearable robot being tested. We have seen some success with wireless biometric sensors, but they can still be prone to detachment during task execution and the addition of sensors worn under a system could cause abrasions to the wearer.

There are also long-term effects that a system can have on a user, such as decreased capability when not wearing the



Fig. 3. Photo of a test method apparatus and props for munitions loading, 30 cm hurdle, uneven terrain, and slalom.

system. The system may also influence the movement of the user in a negative way that could cause injuries over time. In the future, we hope to be able to extrapolate data from these tests to predict effects of extended use, but for now is beyond the scope of our work.

Beyond objective measures, there is a need for qualitative assessment through surveys for characteristics such as comfort, effort required to use, donning and doffing complexity, etc. Survey techniques like NASA TLX have been used to allow subjects to self-assess mental demand, effort, and frustration when wearing an exoskeleton [18]. This aspect of testing will need to be further investigated before it is broadly implemented.

IV. TEST METHOD DEVELOPMENT

Following the approach previously described, a set of taskbased test methods are currently in development. The tasks are aimed at exhibiting basic capabilities to augment the performance of a human wearing a lower body exoskeleton. At a high level, the tasks are simple and goal oriented. At a low level, they require the user to enter many complex poses, forcing exoskeletons into modes of operation that can exhibit positive and negative impacts. The test methods are designed to work for multiple domains (e.g., stairs, walking), as well as for tasks in specific domain (e.g., munitions storage, underpass), allowing for an inheritance of basic tasks into specialized domains.

The test methods have been divided into four categories, each of which is not exclusive from one another, but have been used for our internal testing to specifically exercise the category each test method resides in. All test methods share elements of each category, which is to be expected with holistic systems testing. The categories are:

- Range of motion: exercising potential restrictions in range of motion due to how the system is worn and/or the mechanical influences it has on the body.
- Task performance, endurance: repetitive tasks that can be performed for longer or shorter periods of time due to wearing the system.

Category	Test method name	Task description	Body part(s) exercised
Range of motion			
	Stairs	Walking up and down 35 degree stairs	Lower body (focus on knees)
	Hurdle, walk over	Walk over 30 cm (1 ft) hurdle; left leg over first,	Hips, knees
		right leg over first	
	Hurdle, climb over	Walk over 61 cm (2 ft) hurdle; left leg step on with	Hips, knees
		right arm support, right leg step on with left leg	
		support	
	Jump	Jumping as high as possible in a stationary	Ankles, knees
		position	
	Leg lifts	Lifting foot off ground as high as possible, knee bent	Hips, knees
		or out straight; left leg, right leg	
	Forward split	Split legs forward as far as possible; left leg forward,	Lower body
		right leg forward	
	Lateral split	Split legs laterally as far as possible	Lower body
Task performance, endurance			
	Walk	Walking for X distance	Full body (focus on lower body)
	Run	Jogging/running for X distance	Full body (focus on lower body)
	Lunges	Lunging X number of times	Hips, quads, knees
	Squats	Squatting X number of times	Hips, quads, knees
Task performance, loading			
	Munitions storage	Arm-carried load moved between positions	Full body (focus on knees, lower back)
		without traversal	
	Lifting and carrying	Arm-carried load moved between positions after	Full body (focus on knees, lower back)
		traversing through apparatus	
Transitions between poses/modalities			
	Suicides	Running to other end of apparatus and touching	Lower body
		the ground	
	Underpass	Transitioning from standing to prone to standing	Full body (focus on knees, hips)
	Slaloms	Alternating between lateral movements while	Full body (focus on knees, ankles)
		avoiding obstacles	
	Pitched footfalls	Stepping on alternating pitched surfaces to direct	Ankles, knees
		ankle pitching forward and backward	
	Dynamic or	Walking on sand or gravel	Lower body
	compressible terrain	training on said of graver	Lower body
	Inclined plane	Uphill to downhill traverse	Ankles, knees

TABLE I. CATEGORIZATION OF EXAMPLE TEST METHODS FOR LOWER BODY WEARABLE ROBOTS.

- Task performance, loading: lifting, carrying, or dragging weight that is enabled or inhibited due to wearing the system.
- Transitions between poses/modalities: tasks that involve dynamic maneuvers between perceived system modes, such as transitioning between standing to prone, incline to decline, etc.

Eighteen test methods are listed in Table I, tagging each as one of the four categories and the body parts that each exercises. Fourteen of these test methods have been exercised by a human user with and without wearing an exoskeleton. The function of the test methods is not dissimilar to elements of the benchmarking scheme presented in Torricelli *et al.* [19], although disturbances like pushes and the use of treadmills are not considered.

Stills from recorded video of performance in some of the test methods can be seen in Figure 4. All testing has been performed with the addition of a weighted vest of 18 kg (40 lb) to introduce additional burden to the user, simulating the load carried by that of a soldier.

A. Apparatus

The test method apparatuses are environments and props that feature elements for the user to interact with. Physical structures that require certain body pose maneuvering (e.g., stepping up onto a platform, crouching underneath a bar) are used to ensure proper body motions are being engaged. Lines and markings on the apparatuses are used to convey traversal paths and footstep placements, and provide visual indicators as to whether or not a proper movement is being performed. These elements are used to dictate the user how to perform a test, and as directly observable actions that can be used to evaluate performance.

For our development, a simple test bed measuring $2.4 \times 7.3 \text{ m}$ (8 x 24 ft) made up of wood panels covered in foam padding was fabricated. Test props and markings have been fabricated using wood panels, 2x4s, PVC pipes, and duct tape (see Figure 3). Many of the existing apparatuses and props from the ASTM E54.08 [14] test methods have been leveraged, such as the 15 degree ramps (from E2826) for uneven terrain, 35 degree stairs (from E2804), and inclined plane (from E2803).

The test apparatus can be marked by simple means (e.g., duct tape) to indicate user movement, such as a line to follow, or noting areas that should be crossed to constitute a task. Some elements are not solidly fixed, such as the PVC pipes used for the slalom test. These are simply slid into a PVC cap on the ground such that if they are bumped by the user (indicating a misestimation in movement) they will fall over, indicating a fault condition.



Fig. 4. Photos of the B-Temia system being worn by a user performing in six draft test methods. Clockwise from top left: leg lifts, lateral split, lunges, lifting and carrying, slaloms, underpass.

B. Procedure

Before any testing can occur, multiple training sessions must be held. Wearable robots and exoskeletons are not designed for immediate use and require practice to acclimate to the manner in which it changes the user's movement. However, a user's ability to learn how to use the system is also a very pertinent data point when it comes to evaluation. While a definitive training protocol has not yet been developed, our suggestion is to use the test method apparatuses and props to practice maneuvers.

It is also important that any variable settings on the device, both hardware (e.g., strap tightness and length) and software (e.g., maximum and minimum torque), be set appropriately. Each system has a different set of parameters, so the only specification provided is that the user is comfortable and that all manufacturer guidelines are followed.

For any test methods in this domain, the tasks should be performed in three states: without wearing the system, wearing the system powered off, and wearing the system powered on. Practice of a task must be allowed prior to testing to ensure that the user knows the proper way to perform a task. For instance, if a test method involving lunges or squats is not performed properly it could risk injuring the user or skewing the performance data. See Figure 5 for an example of different ways to perform the hurdle, climb over test. Both are viable methods to perform within the test method.

The tests can be run multiple ways, depending on the intended result. For instance, tests for range of motion can be simply administered by having the user place his or her feet at specific positions (e.g., steps of stairs, on top of or over a hurdle). Variables like the total number of repetitions, amount of time to repeat a task, etc., are not explicitly specified. At this point in development, there is no single correct value for these variables, so they are left to the test administrator based on what aspect of a system is being tested. Regardless of the value of each of these variables, they should be held constant within the same test method for the three performance states, allowing for proper comparative analysis. Range of motion tests should be performed before anything task-based, as limitations in range of motion could have significant effects on other testing.

A single performance of a test is also not fully indicative of the system's capabilities. Multiple tests over time must be performed with a variety of body types, heights, weights, etc. However, these aspects are beyond the scope of our development.

C. Metrics

For performing range of motion testing, the user can be asked to report if he or she feels limited when wearing the system. In addition, simple objective measures can be performed by using recorded video, keeping the video camera in a static position in between test method performance with and without the system being worn, then performing a sideby-side comparison of user movement. The top left and center images in Figure 4 show simple scales within the video frame that can be used for comparison.

If range of motion is determined to not be limited, directly observable performance metrics such as the total number of task repetitions can be recorded. These are noted by an administrator during test method performance. By performing the test method both with and without the system, a comparison between recorded metrics can be made in terms of relative increases or decreases.

Additionally, subjective user experiences such as comfort, cognitive load, and fatigue can be evaluated. At this point in development, a standard survey has not yet been generated.



Fig. 5. Two examples of a user performing in the "Hurdles, climb over" test method. Left: placing a footfall on top of the hurdle and balancing with a hand. Right: stepping completely over the hurdle.

D. Internal Validation

The test methods are still in development, so any recorded data up until this point is not properly representative of any one system's capabilities. As such, no internally recorded test data is presented in this paper.

To validate these test methods, a human user has performed them with a 18 kg (40 lb) weighted vest and a B-Temia Dermoskeleton system, first in a control condition with no system and then in an assisted condition with the system. These test sessions were recorded using a four camera setup, while heart rate was recorded using a mobile monitor. Reviewing this data has shown that the test methods can highlight potential exoskeleton instabilities and what tasks an exoskeleton can aid a user in performing.

The aspect of transitions between tasks was not heavily considered during the first round of test method designs, but proved to have significant impact on task performance, modifying our approach. A system like B-Temia's allows for custom profiles for each user that adjusts its autonomous behaviors and applied torque based on user preference. Performing many of the tasks in Table I allowed the wearer to properly tune these settings for optimal performance. This experience has influenced the idea of using the test methods as a training and practice tool when acclimating to a system.

V. CURRENT STATUS AND FUTURE WORK

The set of test methods discussed in this paper are in very early development. By internally working to validate the test methods with the B-Temia system, we have been able to identify other potential areas for testing. We will continue to exercise and develop test methods, eventually leading to a larger experiment involving human subjects; an institutional review board (IRB) protocol application is underway.

While continuing development, other wearable robot systems must be exercised in the test methods. The B-Temia system actuates the knee, while other lower body systems actuate ankles and hips. Based on the body part being exercised, some test methods will apply and some will not. For instance, the use of forward and backward pitched ramps in the "Pitched footfalls" test method could be rotated such that the ramps are pitched left or right for "Rolled footfalls." Such a test method would only be viable for an exoskeleton that actuates the ankle.

Additional effort is needed to validate the directly observable test performance through biometric data. We are investigating using a wireless heart rate monitor, which has proven to be useful for showing reductions in heart rate when comparing performance states within a test method. However, there were many instances where the monitor detached from the user due to obstructions from the exoskeleton, weighted vest, etc. A motion capture system may also be used to validate changes in gait when wearing an exoskeleton, but the use of such a system would not be intended as a standard practice in the final test protocol due to its high cost.

More input from industry, academia, and end users is needed to further development. As it stands, no standards or test methods committee exists that is focused on wearable robotics. This type of testing could fall into some existing ASTM standards committees, such as F45 [15] for industrial autonomous vehicles, but the many additional considerations to testing with a human wearing the system may warrant its own dedicated committee.

VI. ACKNOWLEDGMENTS

This research has been supported in part by the U.S. Army Natick Soldier Research, Development and Engineering Center (NSRDEC) and the Oak Ridge Institute for Science and Education (ORISE). Thank you to B-Temia for lending their system to be used as a test candidate to assist in developing and validating the test methods.

REFERENCES

- [1] "New ReWalk Personal 6.0 ReWalk More Than Walking," 2016, http://rewalk.com/rewalk-personal-3/. Accessed April 2016.
- [2] "Ekso Bionics An exoskeleton bionic suit or a wearable robot that helps people walk again," 2016, http://www.eksobionics.com/. Accessed April 2016.
- [3] C. Orlowki, "Warrior Web," 2014, http://www.darpa.mil/program/warrior-web. Accessed April 2016.
- [4] "B-Temia Human augmentation systems, restore and enhance mobility Homepage - B-Temia Human augmentation systems, restore and enhance mobility," 2016, http://www.b-temia.com/. Accessed March 2016.
- [5] "Human Augmentation and Army Vision 2025," 2014, http://nerve.uml.edu/human-augmentation/. Accessed April 2016.
- [6] L. M. Mooney, E. J. Rouse, and H. M. Herr, "Autonomous exoskeleton reduces metabolic cost of walking," in 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2014, pp. 3065–3068.
- [7] C. Walsh, K. Endo, and H. Herr, "A quasi-passive leg exoskeleton for load-carrying augmentation," 2007, International Journal of Humanoid Robotics Vol. 4, No. 3 (2007).
- [8] M. Wehner, B. Quinlivan, P. M. Aubin, E. Martinez-Villalpando, M. Baumann, L. Stirling, K. Holt, R. Wood, and C. Walsh, "A lightweight soft exosuit for gait assistance," in *Robotics and Automation* (*ICRA*), 2013 IEEE International Conference on. IEEE, 2013, pp. 3362–3369.
- [9] K. Shamaei, M. Cenciarini, A. A. Adams, K. N. Gregorczyk, J. M. Schiffman, and A. M. Dollar, "Design and evaluation of a quasipassive knee exoskeleton for investigation of motor adaptation in lower extremity joints," 2014, IEEE Transactions on Biomedical Engineering, Vol. 61, No. 6.
- [10] HumanSystems Incorporated, "Building an instrumented performance course to assess under load," 2016 http://www.humansys.com/portfolio-items/leap2/. Accessed April 2016.

- [11] A. Gams, T. Petric, T. Debevec, and J. Babic, "Effects of robotic knee exoskeleton on human energy expenditure," *IEEE Transactions* on Biomedical Engineering, vol. 60, no. 6, pp. 1636–1644, 2013.
- [12] G. S. Sawicki and D. P. Ferris, "Powered ankle exoskeletons reveal the metabolic cost of plantar flexor mechanical work during walking with longer steps at constant step frequency," *Journal of Experimental Biology*, vol. 212, no. 1, pp. 21–31, 2009.
- [13] D. McNally, "DARPA's Warrior Web project," 2014, http://www.army.mil/article/125315/DARPA_s_Warrior_Web_project_ may_provide_super_human_enhancements/. Accessed April 2016.
- [14] ASTM, "Committee E54 on Homeland Security Applications, Subcommittee E54.08 on Operational Equipment," 2015, http://www.astm.org/COMMIT/SUBCOMMIT/E5408.htm. Accessed March 2016.
- [15] ASTM, "Committee F45 on Driverless Automatic Guided Industrial Vehicles," 2015, http://www.astm.org/COMMITTEE/F45.htm. Accessed March 2016.
- [16] R. Bostelman, E. Messina, and S. Foufou, "Cross-Industry Standard Test Method Developments - from Manufacturing to Wearable Robots," in *Proceedings of the International Conference on Wearable Sensor and Robot*, 2015.
- [17] P. Malcolm, P. Fiers, V. Segers, I. Van Caekenberghe, M. Lenoir, and D. De Clercq, "Experimental study on the role of the ankle push off in the walk-to-run transition by means of a powered ankle-footexoskeleton," *Gait & Posture*, vol. 30, no. 3, pp. 322–327, 2009.
- [18] A. Schiele, "Ergonomics of exoskeletons: Subjective performance metrics," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009. IEEE, 2009, pp. 480–485.
- [19] D. Torricelli, J. Gonzalez-Vargas, J. F. Veneman, K. Mombaur, N. Tsagarakis, A. J. del Ama, A. Gil-Agudo, J. C. Moreno, and J. L. Pons, "Benchmarking bipedal locomotion: A unified scheme for humanoids, wearable robots, and humans," *IEEE Robotics & Automation Magazine*, vol. 22, no. 3, pp. 103–115, 2015.