

The current document represents the accepted version of the following article:

Torricelli, D.; Gonzalez-Vargas, J.; Veneman, J.; Mombaur, K.; Tsagarakis, N.; del-Ama, A.; Gil-Agudo, A.; Moreno, J.; Pons, J., "Benchmarking Bipedal Locomotion: A Unified Scheme for Humanoids, Wearable Robots, and Humans," in *Robotics & Automation Magazine, IEEE*, vol.22, no.3, pp.103-115, Sept. 2015. DOI: 10.1109/MRA.2015.2448278

The published version of the article can be downloaded at:

<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7254255&isnumber=7254280>

### IEEE Copyright notice

**“©2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”**

# Benchmarking bipedal locomotion: A unified scheme for humanoids, wearable robots and humans

Diego Torricelli<sup>1</sup>, Jose Gonzalez<sup>1</sup>, Jan Veneman<sup>2</sup>, Katja Mombaur<sup>3</sup>, Nikos Tsagarakis<sup>4</sup>, Antonio J. del-Ama<sup>5</sup>, Ángel Gil-Agudo<sup>5</sup>, Juan C. Moreno<sup>1</sup>, and Jose L. Pons<sup>1</sup>

<sup>1</sup> Spanish National Research Council (CSIC), Cajal Institute, Neural Rehabilitation Group, Madrid, Spain

<sup>2</sup> Tecnalia Research and Innovation, Donostia-San Sebastian, Spain

<sup>3</sup> University of Heidelberg, Germany

<sup>4</sup> IIT, Istituto Italiano di Tecnologia, Department of Advanced Robotics, Italy

<sup>5</sup> National Hospital for Spinal Cord Injury (SESCAM), Biomechanics Unit, Spain

## Introduction

In the field of robotics, there is a growing awareness of the importance of benchmarking [1], [2]. Benchmarking not only allows assessing and comparing the performance of different technologies, but also defines and supports the standardization and regulation processes during their introduction to the market. Its importance has been recently emphasized by the adoption of the *technology readiness levels* (TRLs) in the Horizon 2020 ICT by the EU as an important guideline to assess when a technology can shift from one TRL to the other.

In the field of humanoid robots the main obstacle in identifying common benchmarks is that different methods and metrics are typically employed for specific robotic systems and functional scenarios. Benchmarking of humanoid locomotion is typically approached on a competition-based level, and is mostly focused on global functional goals (e.g. playing soccer, obstacle avoiding, stair climbing [1], [3]). In the field of wearable robots, performance is usually reported in terms of the effects on the user's motor function. New standards are highly expected, especially now that these products are appearing on the market. Yet, there are no accepted schemes for comparing wearable robots performance on a vast scale. The only initiative in this direction is represented by the upcoming CYBATHLON competition [4]. In the clinical and biomechanics field, many metrics and clinical scales have been defined and are regularly used to assess locomotion functions [5]. Most of these scales are based on observation by skilled personnel or defined on very general level, measuring variables like average speed or timed up-and-go. With the increasing application of sensorized and robotic technology in clinics, the expectation for new quantitative and reliable metrics is rapidly growing.

The objective of this article is to define the basis of a benchmarking scheme for the assessment of bipedal locomotion, which could be applied and shared across different research communities. Our approach does not aim to compare systems on a global level to see which one is better, but to assess the several aspects of multi-faceted performance, allowing a truthful comparison of each feature independently. We envision a scenario in which using this scheme will encourage the collaboration between different research groups towards the consolidation of standardized benchmarks and experimental procedures, and promote its use as a complementary tool to competition-based approaches. The scheme presented in this paper is the result of the joint efforts of five European projects, i.e. H2R<sup>1</sup>, BALANCE<sup>2</sup>, Koroibot<sup>3</sup>, Walkman<sup>4</sup> and Biomot<sup>5</sup>. We think that the proposed scheme can be taken as a starting point for a global iterative process that could lead to an international consensus, based on its practical use across different laboratories.

---

<sup>1</sup> [www.h2rproject.eu](http://www.h2rproject.eu)

<sup>2</sup> [www.balance-fp7.eu](http://www.balance-fp7.eu)

<sup>3</sup> [www.koroibot.eu](http://www.koroibot.eu)

<sup>4</sup> [www.walk-man.eu](http://www.walk-man.eu)

<sup>5</sup> [www.biomotproject.eu](http://www.biomotproject.eu)

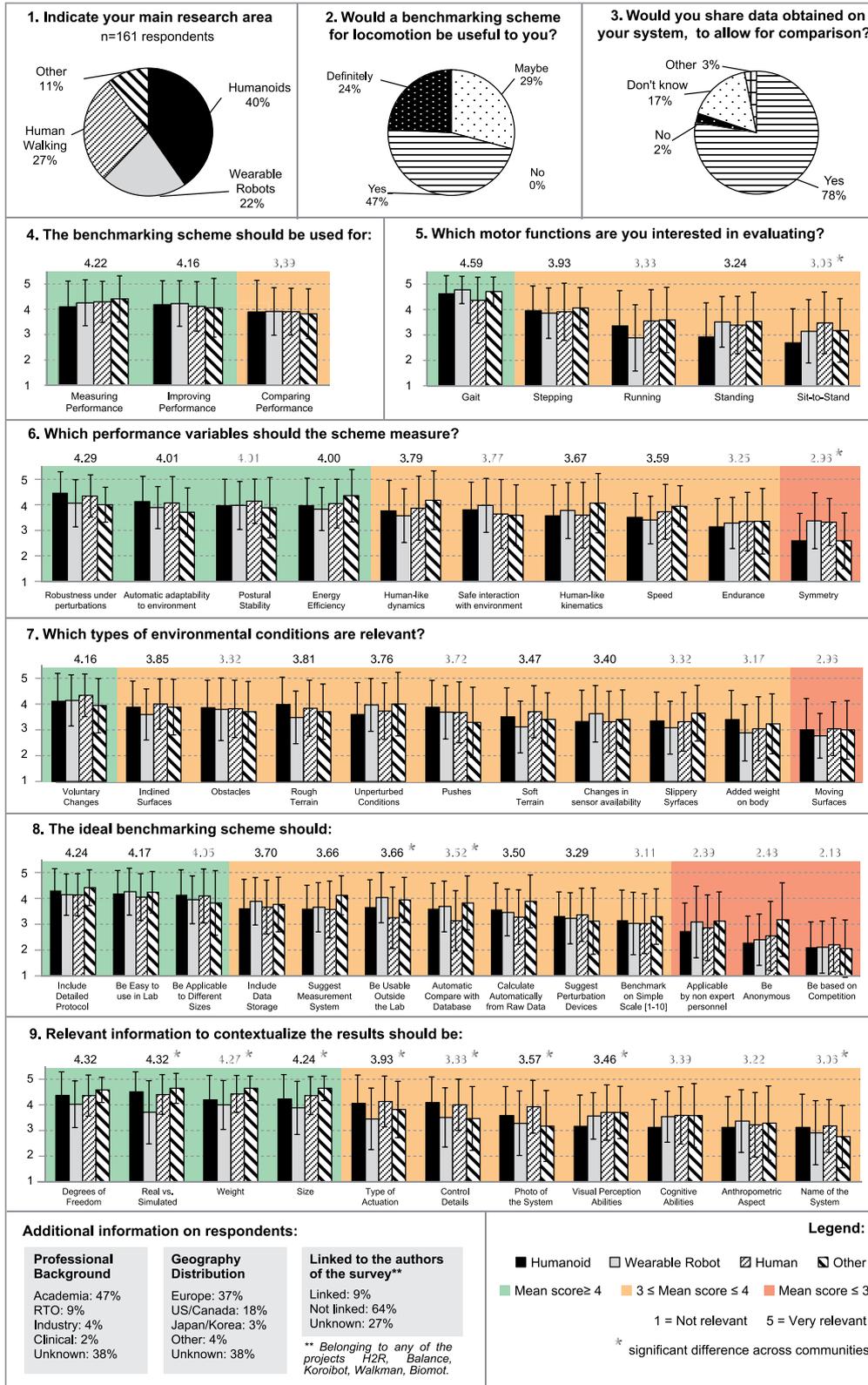
# 1 Design approach

## 1.1 Analysis of the needs: the web-based survey

A benchmark can be considered successful if and only if it is widely accepted by the community at which it is targeted. In order to reach this goal, a number of key principles for a successful benchmarking scheme have been identified [6]:

1. The benchmarks must be well defined, i.e. they really must serve their purpose. As a consequence, the *purpose* should be clear.
2. Benchmarks should be rigorously focused on limited, particular *sub-domains*.
3. It is more likely that a benchmark is successful within a scientific (sub) *community* if it arises from that community itself.

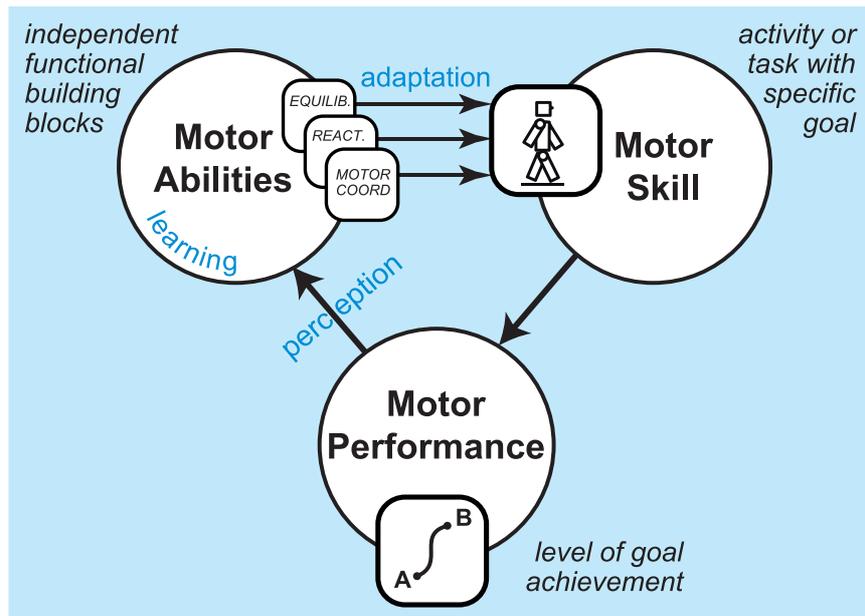
Our design process started with a web-based survey, in order to identify the needs of the different users to which the scheme is addressed. The research communities considered were humanoid robotics, wearable robotics, and human biomechanics. The last has been included because of the increasing need to merge insights from biomechanics and human motor control in robotic research. The survey (see Figure 1) comprised 9 questions, which overall address the first two aforementioned design principles. The first three questions aimed to collect general information about the respondents, such as their background and their overall interest in using a benchmarking scheme and in sharing the data obtained by its use. The last six questions focused on the contents of the ideal benchmarking scheme, in terms of: general purpose, motor function addressed, performance variables to be measured, conditions to be included, technical properties of the scheme, and information needed to contextualize results. In these questions, the user was asked to give a score from 1 to 5 to each of the predefined options. Results are represented in Figure 1 in terms of mean values and standard deviations, and divided by the background of the respondents. Statistical analysis of similarity across communities has been performed by one-way ANOVA (level of significance  $p=0.05$ ). Figure 1 also provides a rough classification of results in three classes, according to mean scores across all communities: items of high relevance (mean score over 4.00, highlighted in green), items of medium relevance (mean score over 3.00, in orange) and items of low relevance (mean score lower than 3.00, in red). An asterisk indicates which items presented a significant different response among the communities (excluding the “other” group).



**Figure 1.** Results of the web-based survey. The 161 respondents to the questionnaire are international experts that were contacted through communication means of the following workshops, networks or forums: the involved research projects listed on page 1, euRobotics or Euron mailing list, Biomch-L mailing list, robotics\_worldwide mailing list, WeRob2014 - The 2014 International Workshop on Wearable Robotics (werob2014.org), European Network on Robotics for NeuroRehabilitation (EC COST Action TD1006) (www.rehabilitationrobotics.eu), Dynamic Walking (dynamicwalking.org/), RehabRobotics mailing list (associated with ICORR - http://www.rehabrobotics.org/).

## 1.2 Existing taxonomies for skills and abilities

Defining a common nomenclature is a basic purpose for a successful taxonomy. This is particularly true in our case, because the target is multidisciplinary and different terms like skill, function, ability, task, activity, action, performance, can have different meanings. Inspired by the approach of Magill [7], we will make use of three terms: *skill*, *ability* and *performance* (see Figure 2). We define *skill* as “a task or activity with a specified goal”. For instance, walking is a motor skill whose goal is to move from point A to point B. *Ability* can be defined as the “independent functional blocks needed to achieve a skill”. Usually several abilities – motor and/or cognitive - are needed for the achievement of one skill. *Performance* is the third relevant component of this scheme, defined as the level of achievement of the goal.



**Figure 2.** The basic components of our benchmarking taxonomy: motor skills, motor abilities and motor performance. These components have important interdependencies. In order to achieve a desired motor performance (e.g. moving from A to B), different motor abilities (e.g. coordination, equilibrium, reaction time) should be combined together, resulting in a functional motor skill (e.g. walking movements). These three motor aspects, whose quantitative measurement is the main objective of the proposed scheme, are associated to three corresponding internal processes (indicated in blue): i) the perception of the sensory feedback resulting from the actual performance, ii) the learning of new control strategies, and iii) the adaptation of motor abilities necessary to generate an improved motor skill. The analysis of these internal processes, very specific to each community, goes behind the scope of the proposed scheme.

Performance is a common aspect in clinical and robotic scenarios. Performance measures usually consist of discrete scales based on time, distance, or percentage of goal achievement, and can be obtained experimentally with no particular difficulty. Measures for skills and abilities are more difficult to obtain, because they rely on generic concepts (e.g. “walking”, “standing”) and depend on continuous variables such as kinematics, kinetics, muscular activity, which can hardly be translated into absolute metrics. For these reasons, appropriate classification methods can provide a useful basis for the organization of these concepts. If we look at a humanoid robot, or at a human in combination with a wearable device, as a sort of “impaired” version of the human machine, the potential benefit of using clinical-based taxonomies becomes apparent, since the process of (re)-learning is common to both rehabilitation and machine learning scenarios. Gentile [8] and Fleishman et al. [9] proposed successful taxonomies for motor skills and motor abilities, which are commonly used in physical therapy and psychology.

Gentile’s taxonomy (see Figure 3) classifies motor skills according to two general dimensions:

- 1) The *environment*, represented by the elements in contact with the person during the execution of the skill, which can be classified according to two intrinsic characteristics: i) its absolute motion

and ii) the presence of intertrial variability, which indicates whether the environmental condition changes from two consecutive trials.

- 2) The *function* of the motor skill is classified according to: i) the orientation of the body, which can be “maintained” (e.g. in standing) or “transported” (e.g. in walking), and ii) the presence of object manipulation during the execution of the task.

The resulting combination of these characteristics is normally represented in a bi-dimensional table (see Figure 3), organized in terms of increasing complexity, from top left to bottom right positions. A typical rationale during a motor learning procedure is to begin with stationary environment and no intertrial variability (e.g. repetitive trials of single movement), towards a complete moving environment with intertrial variability (e.g. real-life and out-of-the-lab conditions). Similarly, but from the perspective of the function, skills that require static body posture are simpler than those requiring body transport.

				FUNCTION			
				BODY POSTURE		BODY TRANSPORT	
				MANIPULATION		MANIPULATION	
				NO	YES	NO	YES
ENVIRONMENT	STATIONARY	INTERTR VARIAB	NO	<i>Most simple</i>			
		YES					
	IN MOTION	INTERTR VARIAB	NO				
		YES				<i>Most complex</i>	

**Figure 3.** The Gentile’s taxonomy classifies motor skills according to two main dimensions, environment and function, and four intrinsic characteristics, i.e. environment motion, intertrial variability, body motion, and manipulation. The table allows for 16 possible categories, ordered in a simple-to-complex progression, from top-left to bottom-right. Adapted from [8].

Motor abilities underlining motor skills directly influence the performance of their execution. Fleishman [9] proposed a list of 54 independent motor, cognitive and visual abilities at the basis of a wide variety of skills, from locomotion to complex manipulation. We identified a subset of significant motor abilities of the Fleishman’s list related to lower limb motion. They are inter-limb coordination, static and dynamic strength, limb flexibility, gross body equilibrium, reaction time, speed of limbs, and control precision.

## 2 The proposed benchmarking scheme

The proposed benchmarking scheme is composed of three sections.

- *Motor skills classification*: reports the most relevant motor skills related to locomotion and standing, classified according to the Gentile’s taxonomy.
- *Benchmarking methods*: includes the benchmarks that can be used to quantitatively assess the specific abilities behind motor skills, according to an extended version of the Fleishman’s taxonomy.
- *Experimental protocols*: we propose a template of ‘work sheet’ that can be used by researchers to design their own experimental protocols within our framework.

Internal properties, such as cognitive abilities (e.g. perception, learning, planning, prediction, adaptation, see Figure 2, blue items) or internal dynamics (e.g. inter segmental forces) are not considered in this scheme, because the goal of the scheme is to describe the different facets of resulting performance and not to quantify the possible causes.

## 2.1 Scheme for motor skills classification

Figure 4 shows our proposed taxonomy for motor skills, based on the Gentile's approach. Under the perspective of "function", we included the "body posture" conditions, corresponding to postural skills, and the "body transport" conditions, corresponding to locomotion skills. According to the "environment" characteristic, tasks are further classified depending on "motion of the environment" and "intertrial variability". For sake of clarity, we point out that the concept of intertrial variability can be assimilated to the concept of "unexpected disturbance", of more common use in the robotic field.

Motor skills related to *body posture* are depicted in the first column of Figure 4, and briefly defined below:

- Static horizontal surface: maintaining an upright posture.
- Static inclined surface: similar to the previous case, but using an inclined surface.
- Different static surfaces: automatically adapting to different and unknown inclinations.
- Continuous surface tilts: maintaining equilibrium on a support surface whose angular orientation is varied cyclically (e.g. sinusoidal), with constant amplitude and frequency.
- Continuous surface translations: similar to the previous case, but with translational displacements instead of angular displacements.
- Constant weight bearing: maintaining equilibrium after applying an external (and known by the system) weight to the body.
- Sudden surface tilts: equilibrating on a support surface whose angular orientation follows an unpredictable and variable pattern over time.
- Sudden surface translations: equivalent to the previous case, but in the horizontal direction.
- Body sway referenced platform (BSRP): equilibrating on a platform that is tilted so that the relative angle between foot and body is maintained constant, eliminating ankle proprioceptive information [10].
- Pushes: maintaining equilibrium after external pushes with short duration.

As for *body transport* (second column in Figure 4), different conditions have been identified:

- Horizontal ground at constant speed: maintaining a steady state walking over a static and horizontal ground, in absence of any external disturbance.
- Sloped ground: equivalent to the previous condition, but on a fixed slope.
- Variable slopes: maintaining a steady walking over various and unknown inclinations.
- Stairs: climbing stairs of constant and known dimensions.
- Irregular terrain: maintaining a steady walking over different kinds of irregularities in the ground, including obstacles, uneven surfaces and gaps.
- Slippery surface: maintaining equilibrium on a surface with unknown friction.
- Treadmill at constant speed: maintaining a steady state walking while the ground moves continuously at a constant speed.
- Soft terrain with constant compliance: maintaining a steady walking over a terrain made of a soft material with known compliance (e.g. sand, foam). The material cannot be changed throughout the trial.
- Weight bearing: walking on a static horizontal ground at constant speed with an additional weight placed on the back of the robot.
- Pushes: maintaining a steady walk after being pushed in different phases of the gait cycle. The pushes can have short or prolonged duration.
- Treadmill at variable speed: walking on a treadmill with variable velocity. Starting and stopping are also included as boundary conditions.
- Seesaw: steadily walking on an unstable surface free to pivot around an horizontal axis perpendicular to the direction of walking.
- Soft ground with variable compliance: steadily walking over a terrain made of different soft materials, whose location and compliance are not known a priori.

		FUNCTION					
		BODY POSTURE		BODY TRANSPORT			
ENVIRONMENT	STATIONARY	INTERTRIAL VARIABILITY	Static horizontal surface		Horizontal ground at constant speed		
			Static inclined surface		Sloped ground		
				Stairs			
		DIFFERENT STATIC SURFACES	Different static surfaces		Variable slopes		
					Irregular terrain		
					Slippery surface		
	MOTION	INTERTRIAL VARIABILITY	Continuous surface tilts		Treadmill at constant speed		
				Continuous surface translations		Soft terrain with constant compliance	
		IN MOTION	BEARING CONSTANT WEIGHT	Bearing constant weight		Bearing constant weight	
				Pushes		Pushes	
			Sudden or pseudorandom surface tilts		Treadmill at variable speed (including start-stop)		
				Sudden surface translations		Seesaw	
			Body sway referenced platform (BSRP)		Soft ground with variable compliance		

NOTE: All the disturbances may be applied in sagittal and frontal directions.

Figure 4. Motor skills considered in the benchmarking scheme. Since this scheme is limited to bipedal locomotion skills, the “manipulation” category originally included in the Gentile’s taxonomy (see Figure 3) has been omitted. The concept of intertrial variability is analogous to the concept of unexpected disturbances.

## 2.2 Scheme for benchmarks

Figure 5 shows a schematic overview of the most relevant motor abilities related to bipedal functions. We have classified them in two main categories, *performance* and *human likeness*, in which performance is related to the accomplishment of the goal of a motor skill, and human likeness represents in what manner the task is executed, which can or cannot be correlated with the level of accomplishment of the goal. Each ability is associated to one or more benchmarks, which allows the quantitative measurement of the corresponding ability. To allow for truthful application across a wide variety of bipedal systems, all benchmarks should be made independent from weight and size.

### 2.2.1 Benchmarks of performance

In our view two features can describe performance: *stability* and *efficiency*. We define stability as the ability of maintaining the equilibrium during the execution of a motor skill. Loss of equilibrium can be easily detected by the occurrence of a falling event. To assess stability within a single trial, we identified two benchmarks: *time until falling* and the *cycles until falling*. The time until falling should be used in all static postural conditions (e.g. quiet standing on a static surface), because the detection of a cycle cannot be easily determined. The number of cycles (e.g. walking stride cycle, or tilting platform cycle) is more suitable during dynamic conditions, or when robots with different sizes are considered, because of the influence of speed and size on time. In order to measure stability across different trials, the *success rate* should be also measured. Another benchmark of stability is the ability of maintaining the CoM above the polygon of support, reflecting what Fleishman referred to as *gross body equilibrium*. The ability can be measured analytically by the Energy Stability Margin (ESM) [11], or by identifying the maximum accepted disturbance, in terms of amplitude and frequency.

Measuring energy efficiency of robots and humans can be done by the specific cost of transport ( $c_t$ ) [12], [13], defined as the ratio of the energy consumed and the weight times the distance travelled. In robotics, to isolate the effectiveness of the mechanical design and controller from the efficiency of the actuators, the specific *energetic* cost of transport ( $c_{et}$ ), comprises the total energy consumed, and specific *mechanical* cost of transport ( $c_{mt}$ ), which only considers the positive mechanical work of the actuation system, have been introduced. A further way of assessing the energetic aspects of locomotion has been recently introduced with the concept of “passivity”, defined as the ability of optimizing the use of gravity and inertia to move the body forward. The resulting Passivity Gait Measure (PGM) [14], appears to be a potential benchmark because of its practical use in robotic and human scenarios. Another aspect of efficiency is the ability of reacting promptly to an external command or perturbation, usually referred to as *reaction time*.

### 2.2.2 Benchmarks of human likeness

Human likeness is a term widely used in humanoid robot community to define the similarity with human behavior. The concept of *healthy behavior* is used instead in the fields of wearable robotics and human biomechanics. In our scheme, we propose to maintain the term “human likeness”, due to its conciseness. In order to translate this concept into a number of abilities and related benchmark, we divided human-likeness into two categories, *kinematics* and *dynamics* (see Figure 5).

Under the kinematics category we included all the abilities that can be analyzed by observing only the motion of the body. We have identified three further sub-categories:

- i) *Whole body motion* can be generally described by the motion of the CoM, and compared with humans through correlation techniques, such as dynamic time warping [15]. Recently, other techniques for global movement assessment have been introduced, such as the gait harmony [16]. In the specific case of posture, the human-like whole-body sway is commonly considered [17], [18]. Global motion can be also assessed by visual inspection from human observers [19].
- ii) *Individual joint motion* can be easily measured and compared with healthy humans [15]. Foot motion is also a crucial aspect in walking. In the ideal benchmarking scheme, the assessment of

basic wheel-like mechanisms of the foot - namely heel, ankle, and forefoot rockers – should be included [20], [21].

- iii) *Coordination*, which includes *inter-limb* coordination, such as symmetry [22] and trunk/arm motions used for regulating body momentum [23]; and *intra-limb* coordination, i.e. ankle-knee-foot synergies [24].

		ABILITIES		BENCHMARKS		
		NAME	DESCRIPTION	BENCHMARK	APPLICABILITY	
					Posture	Transport
P E R F O R M A N C E	S T A B I L I T Y	Intra-trial stability	Ability to maintain the equilibrium within a sing trial	Time until falling	X	
				Cycles until falling	X	X
		Inter-trial stability	Ability to maintain the equilibrium across different trials	Success rate across N different trials	X	X
	Gross body equilibrium	Ability to maintain equilibrium over the base of support	Energy stability margin (ESM)	X		
			Max accepted disturbance amplitude	X	X	
			Max accepted disturbance frequency	X	X	
	E F F I C I E N C Y	Global energy consumption	Ability to transport body with low energetic costs	Specific energetic cost of transport $C_{et}$		X
				Specific mechanical cost of transport $C_{mt}$		X
		Passivity	Ability to minimize joint torques during walking	Passive Gait Measure		X
		Reaction time	Ability to promptly react to disturbance or external command	Time from input and initiation of motor action	X	X
H U M A N L I K E N E S S	K I N E M A T I C S	Gross body motion	Motion of the whole body expressed by global variables	CoM trajectory (correlation, dynamic time warping)	X	X
				Gait harmony		X
				Body sway (Frequency Response Function)	X	
				Natural looking motion	X	X
	Individual joint motion	Motion of the single joints or limbs taken separately	Joint trajectory (correlation, dynamic time warping)	X	X	
			Knee, ankle forefoot rocker		X	
	Inter-limb coordination	Ability to coordinate between different body parts	Symmetry (Ratio Index)	X	X	
			Trunk/arm motion	X	X	
	Intra-limb coordination	Ability to move multiple joints of the same limb coordinately	Kinematic synergies	X	X	
	D Y N A M I C S	Gross body kinetics	Forces exerted between the whole body and the environment	Ground reaction forces (correlation, dynamic time warping)	X	X
		Single joint kinetics	Force exerted among limbs	Joint torques (correlation, dynamic time warping)	X	X
		Dynamic similarity	Ability of having leg pattern dynamically similar to most legged animals.	Froude number (Dimensionless gait velocity)		X
		Dynamicity	Ability to use falling state for body progression	Dynamic Gait Measure		X
External compliance		Ability to respond resiliently to external disturbances	Impulse Response Function (IRF)	X	X	
Internal compliance		Ability to store and release energy	Active/net joint torque	X	X	

Figure 5. Motor abilities and related benchmarks classified in two categories: *performance* and *human likeness*. The *performance* category includes all those abilities related to stability (ability of maintaining equilibrium) and efficiency. The *human likeness* category includes all those abilities related to typical human behavior, under the perspective of kinematics

and dynamics. For each ability, a specific benchmark has been identified. The last column specifies in what classes of motor skills (i.e. the “function” category of Figure 4) the corresponding benchmark is applicable.

The category of dynamics includes all the abilities that are correlated with forces behind movements. The ground reaction forces are mostly used as a descriptor of the global kinetics of the body. Beyond the direct measurement of forces, some other interesting features related to dynamics can be considered and assessed. One of them is the *dynamic similarity*, introduced by Alexander et al. [25], which is defined with the following six criteria: (i) geometric similarity, (ii) equal phase relationships, (iii) equal duty factors, (iv) equal relative stride lengths, (v) equal relative ground reaction forces, and (vi) equal relative mechanical power outputs. They verified experimentally that different-sized animals meet these six criteria when they move with the same Froude number. Therefore the Froude number can be taken as a compact way to describe dynamic similarity between a robot and human, irrespective to size [26]. Mummolo et al. [14] recently proposed an indicator of “dynamicity”, i.e. the dynamic gait measure (DGM), defined as the ability of a legged system to maintain dynamic stability while statically unbalanced, therefore useful to distinguish between ZMP-based control approaches vs. natural dynamics systems (e.g. passive walkers). A further relevant characteristic in biological locomotion is the compliance, defined as the reciprocal of stiffness [27]. Compliance can be assessed by measurement and derivation of the displacement-force relationship in consequence to an external stimulus, usually referred to as impulse response function (IRF) [18]. At the joint level it can be measured through the derivation of angle-torque relationship.

### 2.3 Scheme for experimental protocols

Establishing unified experimental protocols is one of the major challenges of the proposed scheme. To facilitate this process, we have developed a template for proposal, called “work sheet” (see Figure 6). This sheet has the main purpose of encouraging researchers and external collaborators to provide practical proposals for simplified experimental scenarios, which can be shared with other researchers and tested in different laboratories. Through an iterative approach, these experimental methods will be then refined and eventually get to feasible and agreed protocols. The work sheet is composed of two main sections (see Figure 6). The first column of the sheet should be used to define the specific motor skill, both qualitative, by classifying the skill according to the taxonomy, and quantitative, by specifying a set of parameters that characterize the type of disturbance, e.g. the location, direction, magnitude and frequency of disturbance, or the duration and number of trials. The second column of the work sheet should be used to allow the replication of the experimental protocol in different laboratories. Four kinds of information should be included: the experimental procedure, the applicable benchmarks, the variables to be measured, and way of representing the results, namely numerical, graphical, or single-scale.



### 3 Discussion

The need for quantitative metrics of bipedal motor skills is becoming more and more relevant in humanoids, wearable robotics and human biomechanics research. The web-based survey showed that, despite this heterogeneous scenario, the different communities share similar needs, with some minor differences. In humanoid research it appears to be especially relevant to benchmark the performance under different perturbed conditions. In the wearable robotics community, there is a general interest on natural motion and postural stability. In human biomechanics, and particularly in the clinical fields, benchmarking should be focused on the detection of specific abnormal patterns with higher precision and reliability with respect to the current clinical scales. It is important to consider that the survey's respondents were mostly from the humanoid fields (40%), which could have biased the results towards this community. Nevertheless, we consider that the results obtained through the survey are a good starting step towards a unified framework. In general, respondents did not support "anonymous" and "competition based" approaches. This does not deny the importance of robotic competitions, but states that competitions are not being perceived as proper performance benchmarks. In this respect, our scheme can represent a complementary tool that can help researchers to find causal relationships between the performance during competition and the performance in each of the sub-functions identified in our scheme. This may provide additional clues to improve the technology, which is the (common) ultimate goal of all benchmarking efforts.

We have included in our benchmarking scheme most of the conditions and features that received higher scores in the survey. A preliminary version of this scheme has been also discussed with the humanoid and exoskeleton communities in two recent international workshops<sup>6</sup>. One aspect that was discussed extensively was the importance of a common terminology. This discussion led us to put more emphasis on the definitions of specific terms, such as motor skills, abilities and performance, and resulted in the proposal of the different taxonomies.

A relevant issue arisen from the discussions is related to the benchmarking of control algorithms and other internal cognitive processes. On the one hand, we found this topic of extreme importance, being the basis of the resulting performance measured by our scheme. On the other hand, we observed that the internal processing strategies have great variability across the different communities to which this scheme is addressed. Therefore, we believe that this topic should be considered and discussed within each community independently. This process will eventually result in benchmarks that can be either added to this unified scheme or included as community-specific add-ons.

One problem when defining similarity between different systems is that the dynamic and kinematic properties, including elementary properties such as weight, size, mass distribution or number of degrees of freedoms (DOFs), but also the corresponding kinematic and dynamic constraints, have to be taken into account. As for weight and size, some of the proposed benchmarks already take into account these differences in their scores (e.g. Froude number, cost of transport). Other methods did not explicitly include scaling laws (e.g. GRF, reaction time, joint torques), therefore requiring further discussion within the community in order to establish clear rules for scaling. As for the differences in the DOFs, in the cluster of the European projects authoring this paper, some groups are currently investigating this issue, and work on how to best compare similarity in the common degrees of freedom while taking into account the effect of the non-common ones.

Certainly, some of the proposed benchmarks might not be considered effective in specific scenarios or systems. However, the proposed benchmarking scheme should not be used as a whole. Researchers are encouraged to choose only those features that are in line with their objectives. At the same time, this scheme is conceived as a flexible platform, open to new contributions and extensions resulting from international discussions. For instance, motor skills copying with voluntary transitions, such as changes in walking speeds, transitions from standing to walking, or turning, are still not present and will be considered for their inclusion. As for the experimental procedure, it appears necessary to ensure replicability of the benchmarking protocols. At the same time, the scheme should leave certain degree of freedom in the

---

<sup>6</sup> Workshop on "Benchmarking bipedal locomotion", IEEE-RAS International Conference on Humanoid Robots, Nov 18, 2014, Madrid. [http://orb.iwr.uni-heidelberg.de/koroibot/?page\\_id=492](http://orb.iwr.uni-heidelberg.de/koroibot/?page_id=492), and European Robotics Forum 2015 ([www.erf2015.eu](http://www.erf2015.eu)), Session on "Replicable robotics research and benchmarking".

application of the experimental procedures, due to the wide range of systems and laboratory conditions across the different communities. The “work sheet” has been conceived for this purpose. The major goal of the work sheet is to standardize the design process of a benchmarking protocol, therefore maximizing its potential use across different scenarios and end users. To this aim, special efforts should be done to translate the method currently used in human motion analysis (based on GRF, COM, or COP measurements) into minimal experimental setups, which allow at the same time fast, versatile and sufficiently accurate results across platform with different hardware

## 4 Conclusions and roadmap

In this paper we have set the foundations of a general structure for benchmarking bipedal motor skills. The originality of our approach is threefold. First, the proposed scheme is *comprehensive*, i.e. it arranges the great majority of bipedal motor functions into a meaningful taxonomic structure, using the classification scheme based on motor skills, abilities and performance. This global way of classifying motor functions, inherited from the field of rehabilitation and psychology, has not been proposed or applied in the robotics scenario previously. Second, it is *function-based*, i.e. it analyses specific sub-functions of the global motor behavior instead of evaluating the general accomplishment of a goal. This approach is innovative, because, if applied in combination with existing goal-based benchmarking analysis (e.g. DARPA), can provide clues on the causal relationships between the sensorimotor mechanisms and resulting behavior. Third, the scheme is *collaborative*, i.e. it requires the participation of the community in proposing and refining new protocols and benchmarks, e.g. by means of the work sheet tool provided in this paper. In order to encourage this collaborative process, we recently created a website ([www.benchmarkinglocomotion.org](http://www.benchmarkinglocomotion.org)), which will allow researchers to participate actively in the definition and improvement of the scheme. Similarly, we created a mailing list ([https://listas.csic.es/wws/info/benchmarking\\_list](https://listas.csic.es/wws/info/benchmarking_list)), which is currently used to disseminate related events and topics.

In the roadmap towards an interdisciplinary and international consensus, we have identified some crucial steps. The first step should be to identify and test the experimental protocols on different bipedal systems, in order to verify to what extent two different systems/laboratory can share the same procedures. In this respect, a crucial factor will be the involvement of the robotic platforms currently available in the literature, and preferably those already participating in other benchmarking initiatives, to start defining standard procedures and calculating representative scores. The second step would be the refinement of the benchmarking scheme in order to formalize additional/specific goals for each (sub) community. In particular we envision the development of community-specific schemes, such as those related to benchmarking of cognitive and algorithmic processes, currently not included in this scheme. The third step should be directed to discuss with current standardization work groups on robotic technology<sup>7</sup> the appropriate strategies to translate the proposed benchmarks and metrics into future standards. This last step will be essential for an appropriate market introduction of the new robotic technologies.

## Acknowledgments

This research has been supported by the following FP7 European projects: H2R (grant No. 600698), BALANCE (grant No. 601003), Koroibot (grant No. 611909), Walk-man (grant No. 611832), and Biomot (grant No. 611695), and by the EC COST TD1006 - European Network on Robotics for NeuroRehabilitation. We would like to especially thank all the participants in the web-based survey and the attendees at the past workshops for their participation in the discussion.

---

<sup>7</sup> IEC SC62A & ISO TC184/SC2/JWG 9 "Medical electrical equipment and systems using robotic technology"; ISO TC184/SC2/WG1 "Vocabulary and coordinate systems ", WG7 "Personal care robot safety" and WG8, "Service robots".

## References

- [1] S. Behnke, “Robot Competitions Ideal Benchmarks for Robotics Research,” in *IROS2006 - Workshop on Benchmarks in Robotics Research*, 2006.
- [2] A. P. del Pôbil, “Why do We Need Benchmarks in Robotics Research?,” in *IROS2006 - Workshop on Benchmarks in Robotics Research*, 2006.
- [3] “DARPA Robotics Challenge Trials (DRC).” [Online]. Available: <http://www.theroboticschallenge.org/>.
- [4] “CYBATHLON 2016: The Championship for Robot-Assisted Parathletes.” [Online]. Available: <http://www.cyathlon.ethz.ch>.
- [5] S. Tyson and L. Connell, “The psychometric properties and clinical utility of measures of walking and mobility in neurological conditions: a systematic review,” *Clin. Rehabil.*, vol. 23, pp. 1018–1033, 2009.
- [6] EURON, “Introduction: Benchmarks in Robotics Research,” *from : Survey and Inventory of Current Efforts in Comparative Robotics Research*. [Online]. Available: <http://www.robot.uji.es/benchmarks>.
- [7] R. A. Magill, *Motor learning and control: concepts and applications*. McGraw-Hill, 2007.
- [8] A. M. Gentile, “Skill acquisition: Action, movement, and neuromotor processes,” in *Movement Science: Foundations for Physical Therapy*, 2000, pp. 93–154.
- [9] E. A. Fleishman and M. K. Quaintance, *Taxonomies of human performance*. Orlando, FL: Academic Press.
- [10] A. Ishida, S. Imai, and Y. Fukuoka, “Analysis of the posture control system under fixed and sway-referenced support conditions,” *IEEE Trans. Biomed. Eng.*, vol. 44, pp. 331–336, 1997.
- [11] D. Messuri and C. Klein, “Automatic body regulation for maintaining stability of a legged vehicle during rough-terrain locomotion,” *IEEE J. Robot. Autom.*, vol. 1, 1985.
- [12] G. Gabrielli and T. von Kármán, “What price speed? Specific power required for propulsion of vehicles,” *Mech. Eng. ASME*, vol. 72, no. 10, pp. 775–781, 1950.
- [13] S. Collins, A. Ruina, R. Tedrake, and M. Wisse, “Efficient bipedal robots based on passive-dynamic walkers,” *Science*, vol. 307, pp. 1082–1085, 2005.
- [14] C. Mummolo and J. H. Kim, “Passive and dynamic gait measures for biped mechanism: formulation and simulation analysis,” *Robotica*, vol. 31, no. 04, pp. 555–572, Oct. 2012.
- [15] N. E. Helwig, S. Hong, E. T. Hsiao-Wecksler, and J. D. Polk, “Methods to temporally align gait cycle data,” *J. Biomech.*, vol. 44, no. 3, pp. 561–6, Feb. 2011.
- [16] M. Iosa, F. Paradisi, S. Brunelli, A. S. Delussu, R. Pellegrini, D. Zenardi, S. Paolucci, and M. Trallesi, “Assessment of gait stability, harmony, and symmetry in subjects with lower-limb amputation evaluated by trunk accelerations,” *J. Rehabil. Res. Development*, vol. 51, pp. 623–634, 2014.

- [17] T. Mergner, G. Schweigart, and L. Fennell, "Vestibular humanoid postural control," *J. Physiol. Paris*, vol. 103, pp. 178–194, 2009.
- [18] A. D. Goodworth and R. J. Peterka, "Contribution of sensorimotor integration to spinal stabilization in humans," *J. Neurophysiol.*, vol. 102, no. 1, pp. 496–512, Jul. 2009.
- [19] N. F. Troje, "Decomposing biological motion: a framework for analysis and synthesis of human gait patterns," *J. Vis.*, vol. 2, no. 5, pp. 371–87, Jan. 2002.
- [20] J. Perry, "Gait analysis: normal and pathological function," *SLACK Inc.*, 1992.
- [21] A. H. Hansen, D. S. Childress, and E. H. Knox, "Roll-over shapes of human locomotor systems: effects of walking speed," *Clin. Biomech. (Bristol, Avon)*, vol. 19, no. 4, pp. 407–14, May 2004.
- [22] H. Sadeghi, P. Allard, F. Prince, and H. Labelle, "Symmetry and limb dominance in able-bodied gait: a review," *Gait Posture*, vol. 12, no. 1, pp. 34–45, Sep. 2000.
- [23] M. Popovic, A. Hofmann, and H. Herr, "Angular momentum regulation during human walking: biomechanics and control," *IEEE Int. Conf. Robot. Autom. 2004. Proceedings. ICRA '04. 2004*, vol. 3, 2004.
- [24] F. Lacquaniti, Y. P. Ivanenko, and M. Zago, "Kinematic control of walking," *Archives Italiennes de Biologie*, vol. 140. pp. 263–272, 2002.
- [25] R. M. Alexander, "The Gaits of Bipedal and Quadrupedal Animals," *Int. J. Rob. Res.*, vol. 3, no. 2, pp. 49–59, Jun. 1984.
- [26] C. L. Vaughan and M. J. O'Malley, "Froude and the contribution of naval architecture to our understanding of bipedal locomotion," *Gait Posture*, vol. 21, no. 3, pp. 350–62, Apr. 2005.
- [27] A. J. Ijspeert, "Biorobotics: Using robots to emulate and investigate agile locomotion," *Science (80-. )*, vol. 346, no. 6206, pp. 196–203, Oct. 2014.