Evaluating Robot Posture Control and Balance by Comparison to Human Subjects using Human Likeness Measures

Lippi Vittorio[a], Christoph Maurer[b] and Thomas Mergner[c]

University Clinic of Freiburg, Neurology, Freiburg, Germany

Keywords: Humanoids, Benchmarking, Human Likeness, Posture Control, Balance.

Abstract: Posture control and balance are basic requirements for a humanoid robot performing motor tasks like walking and interacting with the environment. For this reason, posture control is one of the elements taken into account when evaluating the performance of humanoids. In this work, we describe and analyze a performance indicator based on the comparison between the body sway of a robot standing on a moving surface and the one of healthy subjects performing the same experiment. This approach is here oriented to the evaluation of human likeness. The measure is tested with three human-inspired humanoid posture control systems, the *independent channel* (IC), the *disturbance identification and compensation* (DEC), and the *eigenmovement* (EM) control. The potential and the limitations connected with such human-inspired humanoid control mechanisms are then discussed.

1 INTRODUCTION

The benchmarking of humanoid performance is gaining interest in the research community (Conti et al., 2018; Torricelli et al., 2014; Torricelli et al., 2020; Lippi et al., 2020c; Lippi et al., 2019). The performance of a humanoid is a complex subject that covers several issues, e.g. sensor fusion, cognitive and motor aspects, mechanics, and energy efficiency. In particular, a recent European project, EUROBENCH, is proposing to implement standard and repeatable experimental procedures to evaluate and compare the performance of different robots. This work describes one specific posture control performance indicator implemented within the project, evaluating human-likeness based on the similarity between responses to external disturbances , more are described in (Lippi et al., 2019) and (Lippi et al., 2020c). Balance has been taken into account because falling is one of the typical reasons of failure for humanoids (Atkeson et al., 2015; Atkeson et al., 2018; Guizzo and Ackerman, 2015). In particular, the aspect of human likeliness is studied in this work. A formal and unanimous definition of human likeliness is still missing, although the concept is relevant both for robotics and neuroscience (Torricelli et al., 2014). Human-like behavior is expected to
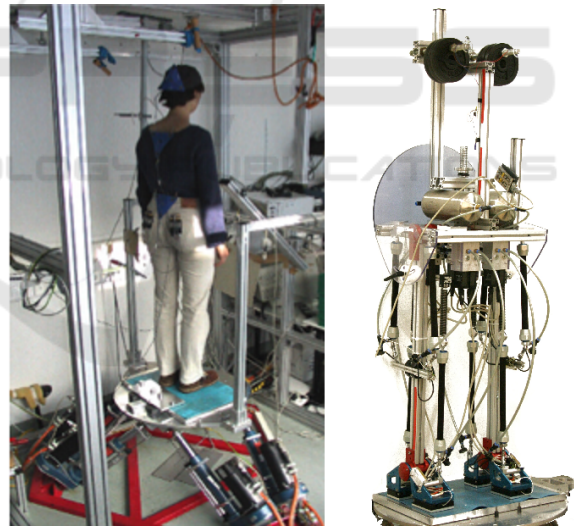

Figure 1: Posture control experiment with a moving support surface, a 6-Dof Stuart Platform. On the left, a human subject with active markers visible on the back (blue triangular plates) is shown. On the right, a robot standing on the same platform.

entail some desirable features such as mechanical compliance (that is important for safety) and low energy consumption. Both of these features are reasonably associated with the low feedback loop gain imposed on biological systems in face of the presence of neural delay. In robotics the delay associated with the trans-

[a] https://orcid.org/0000-0001-5520-8974
[b] https://orcid.org/0000-0001-9050-279X
[c] https://orcid.org/0000-0001-7231-164X

port of signals is negligible, but the robustness in face of delays may be important in limiting the required computational resources needed to implement the real-time control (Ott et al., 2016). Although being often considered important, human-likeness at the state of the art is defined in terms of perception from the point of view of human observers (von Zitzewitz et al., 2013; Oztop et al., 2005; Abel et al., 2020) or identified with the presence of a specific feature, e.g. reproducing human trajectories (Kim et al., 2009) or exhibiting compliance (Colasanto et al., 2015). The measure proposed here is based on body kinematics to make the evaluation of human likeness repeatable and objective instead of designed *ad hoc* and subjective. Specifically, the evaluation is based on a data-set of results from human experiments. The experimental data consist of the body sway of human body segments induced by an external stimulus, here specifically the tilt of the support surface in the body sagittal plane. The measured body sway is characterized as a frequency response function, FRF, i.e. an empirical transfer function between the PRTS stimulus and the response computed on specific frequency points. The data-set has been produced with healthy subjects. The comparison aims to assess the similarity in the balancing behavior between the robot and an average healthy subject. Such measure does not make explicit assumptions about specific advantages of the human behavior, in contrast to the performance indexes proposed in other works that identify some goal like minimum torque or energy consumption in robots (Kashiri et al., 2016), or a specific problem in human subjects (Singh, 2016). The implications of the proposed measure and the limitations connected with human likeliness will be discussed in the conclusions.

Comparing body sway profiles FRFs has been chosen as a base for the benchmarking because such analysis has been studied in several publications with human subjects. This provides a reference for comparison and tools for analysis. There are several reasons why the body sway induced by support surface tilt has been repeatedly used to study human posture control. As a repeatable stimulus it can be used to formally characterize the behavior in terms of responses to a well defined input. Furthermore, the tilting support surface requires the balancing mechanism to integrate vestibular input and proprioception (and vision, with eyes open), and hence it is well apt to study human sensor fusion.

The Sample Data-set here considered consists of data from human posture control experiments. A group of 38 young healthy subjects serves as a reference for the human-likeness criterion. The raw data include information like age, body weight, height, and the
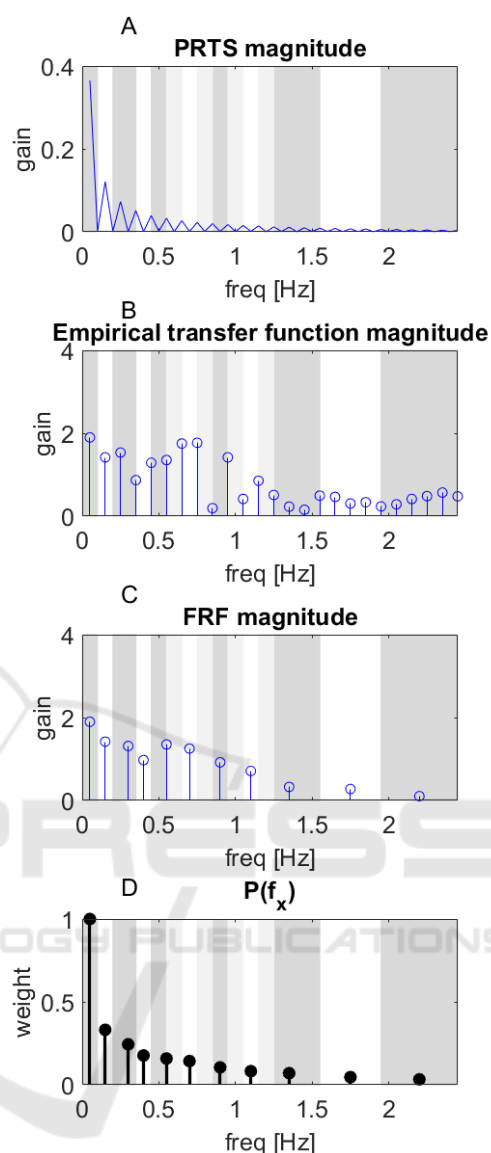

Figure 2: PRTS spectrum and averaging process. The magnitude of the discrete Fourier transform of the PRTS is shown in **A**. An empirical transfer function from a trial is shown in **B**, and the associated 11-component FRF are shown in **C**. Note that, as the FRF is averaged in the complex domain the magnitude of the average shown in the plot is not the average of the magnitudes. The bands on the background show the frequency ranges over which the spectrum is averaged: white and dark grey represent ranges associated with groups of frequencies. The sets of frequencies are overlapping, with light gray bands belonging to both the two contiguous groups, and a sample on the transition between two bands belongs to both the groups. The weights in **D** represents the values of $P(f_x)$ in eq. 4.

height of the different markers. In the experiments, the subjects were presented with a stimulus consisting of a tilt of the support surface in the body sagittal

plane, while the recorded output was the body sway. The typical set-up is shown in Fig. 1, the tracking was performed using active markers (Optotrak 3020; Waterloo, ON, Canada), attached to subjects' hips and shoulders and to the platform. A PC with custom-made programs was used to generate the support surface tilt in the body-sagittal plane. The marker positions were recorded at 100 Hz using software written in LabView (National Instruments; Austin, TX, United States). The profile used for the stimulus is a pseudorandom ternary signal, PRTS (Peterka, 2002). The peak-to-peak amplitude was set to 1°. The amplitude is rather small compared to what a healthy subject can withstand. Usually in similar studies the tilt may go up to 8 degrees and more (Peterka, 2002; Asslander et al., 2015). This was motivated by the aim to provide a data-set that could be compared safely with elderly subjects and patients and, in the specific case of robotic benchmarking, can be used to characterize the behavior of the robot without the risk of making the robot fall (and potentially break).

The Performance Indicator is here a measure of similarity with human behavior. The body sway profiles, i.e. the angular sway of the body COM with respect to the ankle joint, is used to characterize and compare the responses. The comparison is defined in terms of the norm of the difference between frequency response functions (FRFs). The PRTS power-spectrum has a profile with peaks at f_{peak} separated by ranges of frequencies with no power (Jilk et al., 2014; Peterka, 2002; Lippi et al., 2020b). The response is considered for a specific set of frequencies where the PRTS spectrum has peaks: $\mathbf{f_{peak}}$=[0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95, 1.05, 1.15, 1.25, 1.35, 1.45, 1.55, 1.65, 1.75, 1.85, 1.95, 2.05, 2.15, 2.25, 2.35, 2.45] Hz. Such a discrete spectrum is then transformed into a vector of 11 components by averaging the FRF over neighboring frequencies as illustrated in Fig. 2:

$$f_{x(k)} = \frac{\sum_{i \in B_k} f_{peak(i)}}{N_k} \qquad (1)$$

where k and i are the indexes of the components of the frequency vectors, B_k is a set of N_k frequencies averaged to obtain the kth sample. The B_k are shown in Fig. 2 as white and gray bands (notice that the bands are overlapping). Similarly the Fourier transform of the PRTS $P(f_x)$ and the Fourier transform of the responses are averaged over the bands B_k before computing the FRFs. The final representation of the FRF is a function of the 11 frequencies f_x =[0.1, 0.3, 0.6, 0.8, 1.1, 1.4, 1.8, 2.2, 2.7, 3.5, 4.4] [1]. The choice of the frequen-

cies in B_k and their overlapping follows the method described in (Peterka, 2002) but here is adapted to a vector of 11 frequencies. The rationale behind such a choice was to get a representation with the frequencies equally spaced on logarithmic scale, which is often used in posturography papers to present the FRF, with the overlapping providing a smoothing effect. In detail, the FRF is computed from the 11 components of the Fourier transform of the input U, and the output Y as

$$FRF = \frac{G_{UY}}{G_U} \qquad (2)$$

where $G_{UY} = U^* \odot Y$ and $G_U = U^* \odot U$ are empirical estimations of the cross power spectrum and the input power spectrum ("$\odot$" is the Hadamard, element-wise product). The peaks of the PRTS power-spectrum have larger values at lower frequencies (Jilk et al., 2014). This implies a better signal-to-noise ratio for the first components. A weighting vector $\mathbf{w}$ based on $P(f_x)$ then defined in a similar way to eq. 1, but considering the power

$$w_k = \sqrt{\sum_{i \in B_k} ||P(f_{peak(i)})||^2} \qquad (3)$$

The distance between two FRFs is defined and the norm of the difference weighted by the precision matrix, i.e. the inverse of the covariance matrix Σ, computed for the data-set of normal subjects. Before doing this the FRF is expanded into a vector with the real and imaginary components as separated elements, i.e. 22 components. This together with the foretold weighting leads to the definition of the norm:

$$D = \sqrt{S\Delta^T \Sigma^{-1} \Delta S} \qquad (4)$$

where $S = diag([\mathbf{w}, \mathbf{w}])$ is the diagonal matrix representing the re-weighting due to the power-spectrum, doubled to cover the 22 elements, and Δ is the difference between the two FRFs expanded to 22 components. This approach does not require model identification because it is performed on the basis of the data. The comparison can be performed between the tested robot and the average of the groups of humans (healthy or with special deficient conditions) or between two single samples to quantify how much two robots differ from each other. The score of human-likeness is obtained comparing the sample with the mean of the human sample set μ so that $\Delta = FRF - \mu$. The parameters μ and Σ defining the score are shown in Fig. 3.

---

[1] In a previous description of the system in (Lippi et al., 2020c) the frequency vector had 16 components as proposed

in other works using PRTS, e.g. (Peterka, 2002; Goodworth and Peterka, 2018; Lippi et al., 2020b; Lippi et al., 2020a). In this work, we decided for a shorter version of the signal which is considered easier for human subjects (sometimes with elderly patients) and convenient for robotics experimenters. The discrete Fourier transform of the signal was consequently shorter, as well as the resulting FRFs.
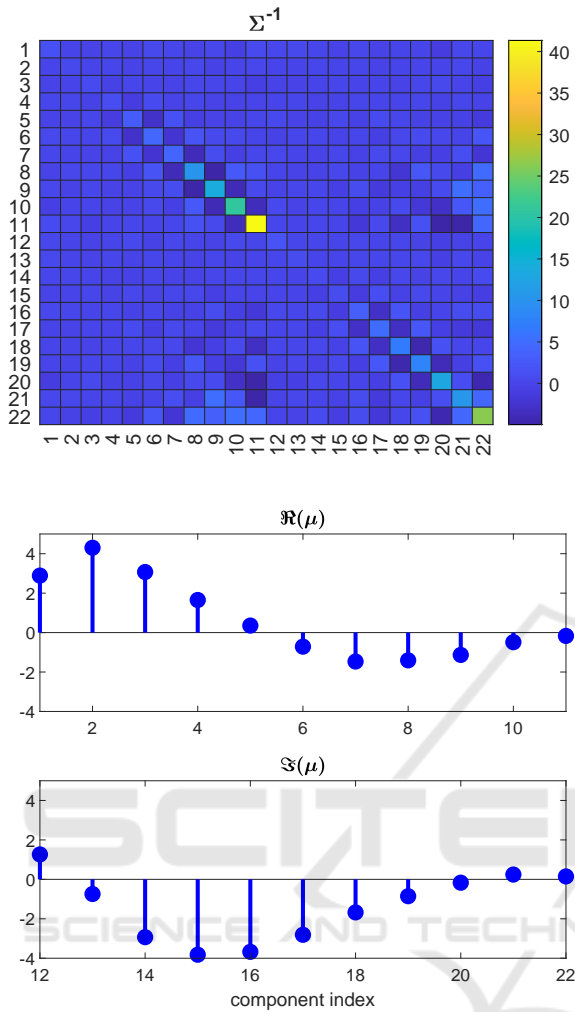
Figure 3: Parameters Σ (top) and μ (below) defining the score. Notice how the values on the diagonal corresponding to the highest frequencies are associated with high values (less variance in the dataset). This would make the metrics particularly sensitive to accidental changes in such components (e.g. due to noise). This disadvantage is compensated by the weighting profiles in Fig. 2 D, which associate almost zero weight to high-frequency components.

Three Control Systems are tested: the *independent channel model* (IC), the *disturbance estimation and compensation* (DEC), and the *eigenmovements* concept (EM). Next we provide a brief description of the concepts.

1) The IC model is a simplified descriptive linear model for human upright posture control. The model consists of a single inverted pendulum controlled by a feedback mechanism with a PD controller and a time delay. The sensory integration mechanism consists of a weighted sum of the sensory contributions. The weights are constrained to unity (Peterka, 2002). The three channels are: angle between body and foot perceived through proprioception, body-in-space orientation perceived by the vestibular system, and body-in-space orientation perceived by vision. The time delay represents all the delays in the control loop. In order to simulate a robot control experiment similar to the ones performed with the robot Posturob II in (Pasma et al., 2018) the system has been simulated in "closed-eyes" conditions. The coefficient W_{vis} is thus set to zero.

2) The DEC model was proposed as a model of human posture control in steady state (Mergner et al., 2003). It was implemented into a humanoid control system (Lippi and Mergner, 2017; Lippi et al., 2016) and tested on robotic platforms (Hettich et al., 2014; Lippi, 2018; Ott et al., 2016; Zebenay et al., 2015). The DEC exploits sensor-fusion-derived internal reconstructions of the external disturbances having impact on body posture and balance. The reconstructed disturbances are compensated using a servo controller. The model is nonlinear because it includes sensory thresholds and the disturbance estimators are non-linear functions. The control loop includes a lumped delay (one for each joint). Its formulation leads to a modular control system with one control module for each degree of freedom as shown in Fig. 4 (C).

3) The EM (Alexandrov et al., 2005) as implemented in (Alexandrov et al., 2017). This control system of bio-mechanical inspiration is based on the diagonalization of the inertial matrix B_0 describing the linearized body dynamics, i.e.

$$B_0\ddot{\mathbf{q}} - G_0\mathbf{q} = \tau_{con} \tag{5}$$

where $\mathbf{q}$ is the vector of joint angles, G_0 is the gravity matrix, and τ_{con} the control torques applied to the joints. EMs have independent dynamics, i.e. PD controller is implemented for each EM. In the original formulation the body kinematics is expressed in the joint angles q_1 and q_2 of Fig. 4 E. The control of balance in space is here implemented by integrating a vestibular system (technically an IMU, *inertial measurement unit*) and defining q_1 as leg-in-space sway angle (i.e. the sum of the hip angle and vestibular trunk in space orientation). When simulating the human system, there are different neural delays for the ankle and trunk control loops. The additional delays in Fig. 4 F are added to make the delays the same for both the joints, whic here is assumed as a condition for decoupling the dynamics of the EM.

The parameters for the simulations are based on the experiments previously performed with the robot Posturob II using an IC, DEC, and EM approach as presented in (Pasma et al., 2018), (Hettich et al., 2014), and (Alexandrov et al., 2017), respectively. The experiments are simulated in MATLAB/Simulink®. The robot is simulated using the double inverted model
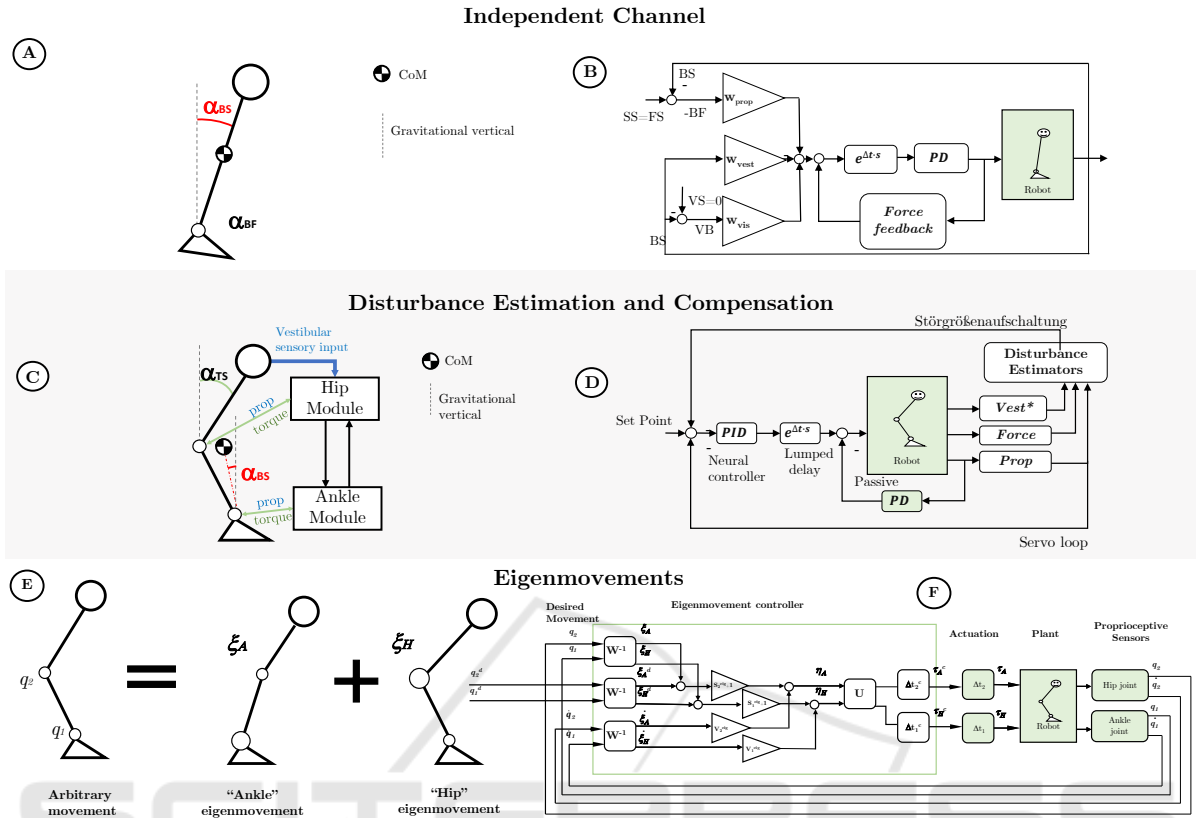
Figure 4: Overview of the tested three control systems. The notation is slightly different to match the one proposed by the respective authors in previous works. (A) The single inverted pendulum mode for the IC, characterized by the body orientation in space α_{BS} and the angle between body and foot α_{BF}. (B). The IC control scheme as presented in (Pasma et al., 2018). The three weights W_{prop}, W_{vest}, and W_{vis} are associated with proprioception (α_{BF}), vestibular signal (α_{BS}) and vision (α_{BS}) respectively. The lumped delay $e^{\Delta t \cdot s}$ represents all the motor and sensory delays in the loop. The positive forcre feedback includes a gain and a first-order low pass filter: $K_F/(T_f s + 1)$. It is included to explain the observed behavior at low frequencies (Peterka, 2003). (C) The DEC, implemented as double inverted pendulum (DIP) (Hettich et al., 2014; Hettich et al., 2013; Hettich et al., 2015), with two control modules: the "Hip Module" controlling the orientation of the trunk in space α_{TS}, and the "Ankle Module" controlling the orientation of the body CoM in space α_{BS}. (D) The scheme of a module of the DEC control. The system includes a passive control loop (PD) and a neural controller (PID) implementing the servo loop and the compensation of estimated disturbances as a *störgrößenaufschaltung*, i.e. feedforward compensation based on sensory input. There is a lumped delay $e^{\Delta t \cdot s}$ in each of the modules. The vestibular input *vest** for the ankle module uses a sensor fusion derived information (vestibular + proprioceptive) to reconstruct the orientation of legs in space. The vestibular input provides the orientation of the head in space (and hence of trunk in space) directly. (E) The decomposition of an arbitrary movement (q_1 = ankle angle, q_2 = hip angle) in two EMs. (F) The EM control scheme.

from (AlBakri, 2008). In order to implement the SIP control for the IC, the hip stiffness is set to a large value, $\approx 10^5$ Nm, leading to virtually no movement ($\leq 10^{-3}$ deg, as shown in Fig. 5). This is similar to what was implemented in (Pasma et al., 2018) blocking the hip joint of the robot with a screw. This choice of simulation parameters aims to reproduce the robot behavior. The parameters of the models can be optimized to reproduce the human FRF our in simulation as previously done in posture control studies, e.g. (Assländer et al., 2015). While this is useful to discuss the parameters themselves, e.g. to estimate the relative weight of different sensory channels in the

control loop in different conditions, it is not obviously leading to a human-like behavior when applied on real robots. The responses are compared with eyes-closed human experiments and hence the visual signal is not considered: a trunk-in-space orientation is assumed to be available in the simulations with no specific modeling of the dynamics of the vestibular system (which is, in general, faster but less accurate than the visual signal). This is because, at the state of the art, humanoids are not (yet) using visual input for balance control. This may change in the future considering the current advancements in computer vision and their integration in humanoid robotics. Data-sets for the
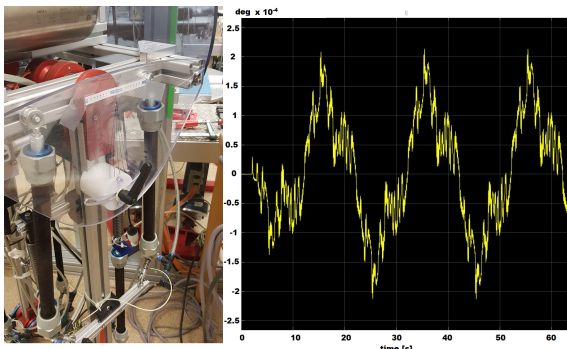
Figure 5: The hip joint of the robot Posturob II blocked by a screw (left) and the trunk to leg sway over time in simulation in degrees (right).

eyes-open condition are however available for future benchmarking.

2 RESULTS

The simulation with the three control systems produced the FRFs shown in Fig. 6 together with the average of the human FRFs in the reference data-set. The results obtained were:

control	score	CDF
IC	2.2034	75.2212 %
DEC	2.7318	87.6106 %
EM	3.9616	97.3451 %

The values in the third column give estimates of the cumulative distribution function (CDF) of the data-set at the given score. It is computed by counting the fraction of samples with a score smaller than the one produced in the specific simulation. The comparison in Fig. 6 shows that all three robotic controllers exhibit a smaller gain than the one observed on average in humans (especially the EM) and differences of phase. In Fig. 7 the result for the three controllers is compared with the distribution of the human data-set.

3 DISCUSSION, CONCLUSIONS AND FUTURE WORK

The Main Contribution of the Present Work is a set of data from human experiments that can be used as a benchmark for human-like posture control to be compared with robots, and the evaluation of a metric proposed for such comparison.

The Tests for the Bio-inspired Control Systems showed that the three robot control systems produced
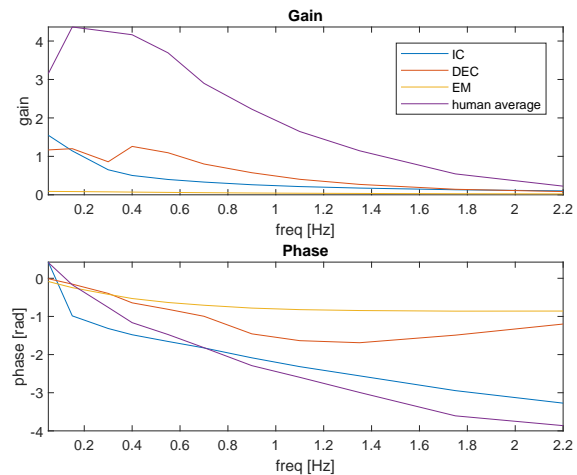

Figure 6: FRF obtained with the IC, DEC, and EM as well as the average FRF of the human data-set.

small gains compared to humans: according to the proposed measure they were "less human-like" than most of the samples in the data-set. Notice that, as explained in the methods section, the parameters of the models can be optimized to create a FRF that is more similar to the human average (as it is actually done in several posture control studies that include a control model). Here, the aim of the simulation was not to evaluate the three models in general (the conclusion here is not that the IC is more human-like from an absolute point of view) or to have a "fair competition" between the three control concepts: that would require to perform the test on the real robot or, at least, optimize the control parameters based on the controlled system. The main idea was to get real models as they were used in real robot experiments (i.e. with parameters calibrated for that specific scenario) to test the proposed metric of human likeness.

Human Like Behavior can represent a standard for humanoid posture control, and with the data-set presented in this work, we propose a way to quantify it. Of course, this does not exclude the possibility that a robot can benefit from super-human capabilities or peculiar behaviors that are optimized for its non-human hardware, as motors are obviously different from muscles and almost all humanoid robots have limb kinematics that is simpler than the corresponding human one. Comparing FRFs is not the only method available in the literature. Models can be tested versus human data with statistical tests such as analysis of variance, focusing on some specific kinematic variables, as presented, for example, in (Bayon et al., 2020) where the range of motions of the body links are considered. The direct comparison of FRFs has the main advantage of being based just on experimental data. In any case, mimicking human behavior may

be intrinsically valuable in human-robot interactions for making movements understandable and physical interaction natural (Villalobos and Zielińska, 2017). The presented comparison principle can be extended using several FRFs obtained with different modalities, e.g. support surface translation and tilt, and different amplitudes. Performing different tests with the same control system parameters can highlight some specificity of human behavior such as nonlinear responses and sensory re-weighting (the importance of specific sensory inputs can change depending on the conditions). From the practical point of view, humans can get an advantage from a nonlinear response in the sense that smaller disturbances can be compensated in a more "relaxed" way saving energy and effort of the nervous system. This means that the stimulus-to-sway gain may be relatively larger for smaller stimuli, and smaller for larger and more dangerous perturbations.

Human Likeness is a "human likelihood" once it is defined in terms of the distance between the tested sample and a distribution of given human behaviors. In particular the score in eq. 4 resembles a Mahalanobis distance $D = \sqrt{\Delta^T \Sigma^{-1} \Delta}$, but with the addition of weights. Assuming a joint normal distribution for the weighted FRFs, $\mathcal{N}(\mu, \Sigma)$, the Mahalanobis distance defines probability density, where a smaller distance associated with higher probability density [2]. It comes naturally to ask how the proposed metric is distributed over the data-set and how it relates to the likelihood of a sample. Fig 7 shows the distribution of the human likeness metric over the data-set of human experiments (cumulative distribution function and relationship between the metric and the Mahalanobis distance) together with the results on the tested control systems. The metric we used in this work opens the possibility to make geometrical considerations for the performance score: i.e. in terms of "How much two behaviors are different?" or, "Which is the smallest change in the FRF that would bring the behavior to a certain score?". This represents an advantage when one aims to define solutions with better scores.

Translation and Tilt of the body support surface can be used in combination. Exploiting the structure of the PRTS described in §1, two PRTS signals can be designed with no overlapping peaks $P(f_{Peaks})$ in the power-spectrum. These two signals can be used together as profiles for the two stimuli and hence two independent FRF can be produced in a trial. Preliminary studies suggest that, although human responses to perturbations are in general non-linear, they may

---

[2]Normality of variables describing posturography may be implicitly assumed when the effects on posture are tested through an analysis of variance, e.g. (Akçay et al., 2021; Jilk et al., 2014).
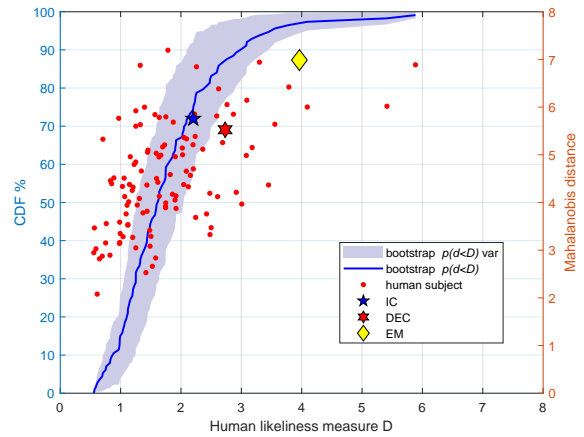

Figure 7: Mahalanobis distance for the samples in the dataset (red dots) as a function of the metric from eq. 4, and the bootstrap estimated cumulative distribution of the metric (blue line) and its variance (light blue bands). The human-likeness score is smaller than the Mahalanobis distance because if the coefficients in Fig. 2 D that are ≤ 1, and in general there is a spread of the Mahalanobis distances of samples associated with the same D because the weighting almost removes the variation due to high-frequency components. The three examples are shown for comparison.

exhibit a linear superposition effect between these two modalities.

Future Work will focus on tests with real humanoid robots as soon as possible. In this work, simulations were used in lieu of real experiments because of the limited laboratory access due to the pandemic. A performance indicator evaluating transient response will be developed using a raised cosine velocity profile instead of the PRTS (that is oriented to the study of steady-state behavior). This kind of stimulus represents a "softer" and safer version of a step function input and it was used in previous experiments with human subjects (Lippi et al., 2014).

4 DATA AND CODE

This performance indicator and the data-set used to define the standard for the benchmark are available through the EUROBENCH project http://eurobench2020.eu/contact-us/. The EUROBENCH framework includes several tests and performance indicators provided with the aim of testing the performance of robots at any stage of development.

ACKNOWLEDGMENT

REFERENCES

Abel, M., Kuz, S., Patel, H. J., Petruck, H., Schlick, C. M., Pellicano, A., and Binkofski, F. C. (2020). Gender effects in observation of robotic and humanoid actions. *Frontiers in psychology*, 11:797.

Akçay, M. E., Lippi, V., and Mergner, T. (2021). Visual Modulation of Human Responses to Support Surface Translation. *Frontiers in Human Neuroscience*, 15:98.

AlBakri, M. (2008). Development of a mathematical model and simulation environment for the postural robot (PostuRob II). Technical report, Uniklinik Freiburg.

Alexandrov, A. V., Frolov, A. A., Horak, F. B., Carlson-Kuhta, P., and Park, S. (2005). Feedback equilibrium control during human standing. *Biological cybernetics*, 93(5):309–322.

Alexandrov, A. V., Lippi, V., Mergner, T., Frolov, A. A., Hettich, G., and Husek, D. (2017). Human-inspired eigenmovement concept provides coupling-free sensorimotor control in humanoid robot. *Front. Neurorobot.*, 11:22.

Assländer, L., Hettich, G., and Mergner, T. (2015). Visual contribution to human standing balance during support surface tilts. *Hum. Mov. Sci.*, 41:147–164.

Atkeson, C. G., Babu, B. P. W., Banerjee, N., Berenson, D., Bove, C. P., Cui, X., DeDonato, M., Du, R., Feng, S., Franklin, P., et al. (2015). No falls, no resets: Reliable humanoid behavior in the darpa robotics challenge. In *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*, pages 623–630. IEEE.

Atkeson, C. G., Benzun, P. B., Banerjee, N., Berenson, D., Bove, C. P., Cui, X., DeDonato, M., Du, R., Feng, S., Franklin, P., et al. (2018). What happened at the DARPA Robotics Challenge finals. In *The DARPA Robotics Challenge Finals: Humanoid Robots to the Rescue*, pages 667–684. Springer.

Bayon, C., Emmens, A., Afschrift, M., Van Wouwe, T., Keemink, A., Van Der Kooij, H., and Van Asseldonk, E. (2020). Can momentum-based control predict human balance recovery strategies? *IEEE transactions on neural systems and rehabilitation engineering*, 28(9):2015–2024.

Colasanto, L., Van der Noot, N., and Ijspeert, A. J. (2015). Bio-inspired walking for humanoid robots using feet with human-like compliance and neuromuscular control. In *2015 IEEE-RAS 15th international conference on humanoid robots (humanoids)*, pages 26–32. IEEE.

Conti, R., Giovacchini, F., Saccares, L., Vitiello, N., Pons, J. L., and Torricelli, D. (2018). What do people expect from benchmarking of bipedal robots? Preliminary results of the EUROBENCH survey. In *International Symposium on Wearable Robotics*, pages 132–136. Springer.

Goodworth, A. D. and Peterka, R. J. (2018). Identifying mechanisms of stance control: a single stimulus multiple output model-fit approach. *Journal of Neuroscience Methods*, 296:44–56.

Guizzo, E. and Ackerman, E. (2015). The hard lessons of DARPA's robotics challenge [News]. *IEEE Spectrum*, 52(8):11–13.

Hettich, G., Assländer, L., Gollhofer, A., and Mergner, T. (2014). Human hip—ankle coordination emerging from multisensory feedback control. *Hum Mov Sci*, 37:123–146.

Hettich, G., Lippi, V., and Mergner, T. (2013). Human-like sensor fusion mechanisms in a postural control robot. In Londral, A. E., Encarnacao, P., and Pons, J. L., editors, *Proceedings of the International Congress on Neurotechnology, Electronics and Informatics. Vilamoura, Portugal*, pages 152–160.

Hettich, G., Lippi, V., and Mergner, T. (2015). Human-like sensor fusion implemented in the posture control of a bipedal robot. In *Neurotechnology, Electronics, and Informatics*, pages 29–45. Springer.

Jilk, D. J., Safavynia, S. A., and Ting, L. H. (2014). Contribution of vision to postural behaviors during continuous support-surface translations. *Experimental Brain Research*, 232(1):169–180.

Kashiri, N., Ajoudani, A., Caldwell, D. G., and Tsagarakis, N. G. (2016). Evaluation of hip kinematics influence on the performance of a quadrupedal robot leg. In *Proceedings of the 13th International Conference on Informatics in Control, Automation and Robotics - Volume 1: ICINCO,*, pages 205–212. INSTICC, SciTePress.

Kim, S., Kim, C., You, B., and Oh, S. (2009). Stable whole-body motion generation for humanoid robots to imitate human motions. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2518–2524. IEEE.

Lippi, V. (2018). Prediction in the context of a human-inspired posture control model. *Robotics and Autonomous Systems*.

Lippi, V., Assländer, L., Akcay, E., and Mergner, T. (2020a). Body sway responses to pseudorandom support surface translations of vestibular loss subjects resemble those of vestibular able subjects. *Neuroscience Letters*, 736:135271.

Lippi, V., Brands, K. G., and Seel, T. (2020b). Real-time implementation and evaluation of magnetometerless tracking system for human and humanoid posture control benchmarking based on inertial sensors. In *Proceedings of the 17th International Conference on Informatics in Control, Automation and Robotics - Volume 1: ICINCO*, pages 675–680. INSTICC, SciTePress.

Lippi, V., Hettich, G., and Mergner, T. (2014). Modeling postural control of support surface translations. In *IEEE Humanoids, Workshop on cognition, perception and postural control for humanoids Madrid, Spain*.

Lippi, V. and Mergner, T. (2017). Human-derived distur-
bance estimation and compensation (DEC) method
lends itself to a modular sensorimotor control in a hu-
manoid robot. *Front. Neurorobot.*, 11:49.

Lippi, V., Mergner, T., Maurer, C., and Seel, T. (2020c).
Performance Indicators of Humanoid Posture Control
and Balance Inspired by Human Experiments. In *2020
International Symposium on Wearable Robotics and
Rehabilitation (WeRob)*.

Lippi, V., Mergner, T., Seel, T., and Maurer, C. (2019).
COMTEST project: A complete modular test stand
for human and humanoid posture control and balance.
In *2019 IEEE-RAS 19th International Conference on
Humanoid Robots (Humanoids) Toronto, Canada. Oc-
tober 15-17*.

Lippi, V., Mergner, T., Szumowski, M., Zurawska, M. S.,
and Zielińska, T. (2016). Human-inspired humanoid
balancing and posture control in frontal plane. In
*ROMANSY 21-Robot Design, Dynamics and Control:
Proceedings of the 21st CISM-IFToMM Symposium,
June 20-23, Udine, Italy*, volume 569, pages 285–292.
Springer.

Mergner, T., Maurer, C., and Peterka, R. J. (2003). A multi-
sensory posture control model of human upright stance.
Progress in Brain Research, 142:189–201.

Ott, C., Henze, B., Hettich, G., Seyde, T. N., Roa, M. A.,
Lippi, V., and Mergner, T. (2016). Good posture, good
balance: comparison of bioinspired and model-based
approaches for posture control of humanoid robots.
IEEE Robotics & Automation Magazine, 23(1):22–33.

Oztop, E., Franklin, D. W., Chaminade, T., and Cheng, G.
(2005). Human–humanoid interaction: is a humanoid
robot perceived as a human? *International Journal of
Humanoid Robotics*, 2(04):537–559.

Pasma, J. H., Assländer, L., van Kordelaar, J., de Kam, D.,
Mergner, T., and Schouten, A. C. (2018). Evidence
in support of the independent channel model describ-
ing the sensorimotor control of human stance using a
humanoid robot. *Frontiers in Computational Neuro-
science*, 12:13.

Peterka, R. (2002). Sensorimotor integration in human pos-
tural control. *Journal of neurophysiology*, 88(3):1097–
1118.

Peterka, R. J. (2003). Simplifying the complexities of main-
taining balance. *IEEE Engineering in Medicine and
Biology Magazine*, 22(2):63–68.

Singh, N. K. (2016). Detection of hesitant dynamic postural
control. In *2016 IEEE 12th International Colloquium
on Signal Processing & Its Applications (CSPA)*, pages
36–40. IEEE.

Torricelli, D., Mizanoor, R. S., Gonzalez, J., Lippi, V., Het-
tich, G., Asslaender, L., Weckx, M., Vanderborght,
B., Dosen, S., Sartori, M., et al. (2014). Benchmark-
ing human-like posture and locomotion of humanoid
robots: A preliminary scheme. In *Conference on
Biomimetic and Biohybrid Systems*, pages 320–331.
Springer.

Torricelli, D., Mizanoor, R. S., Lippi, V., Weckx, M., Mathi-
jssen, G., Vanderborght, B., Mergner, T., Lefeber, D.,
and Pons, J. L. (2020). Benchmarking human likeness

of bipedal robot locomotion: State of the art and fu-
ture trends. In *Metrics of Sensory Motor Coordination
and Integration in Robots and Animals*, pages 147–166.
Springer.

Villalobos, J. and Zielińska, T. (2017). Study of postural
adjustments for humanoidal helpmates. *Journal of
Automation Mobile Robotics and Intelligent Systems*,
11.

von Zitzewitz, J., Boesch, P. M., Wolf, P., and Riener,
R. (2013). Quantifying the human likeness of a hu-
manoid robot. *International Journal of Social Robotics*,
5(2):263–276.

Zebenay, M., Lippi, V., and Mergener, T. (2015). Human-like
humanoid robot posture control. In *Proceedings of the
12th International Conference on Informatics in Con-
trol, Automation and Robotics - Volume 2: ICINCO*,
pages 304–309. INSTICC, SciTePress.