Eye-Head Stabilization Mechanism for a Humanoid Robot Tested on Human Inertial Data

Lorenzo Vannucci[1](✉), Egidio Falotico[1], Silvia Tolu[2], Paolo Dario[1],
Henrik Hautop Lund[2], and Cecilia Laschi[1]

[1] The BioRobotics Institute, Scuola Superiore Sant'Anna,
Viale Rinaldo Piaggio 34, 56025 Pontedera, Pisa, Italy
{l.vannucci,e.falotico,p.dario,c.laschi}@sssup.it
[2] Department of Electrical Engineering, The Center for Playware,
Technical University of Denmark, Elektrovej Building 326,
2800 Kongens Lyngby, Copenhagen, Denmark
{stolu,hhl}@elektro.dtu.dk

Abstract. Two main classes of reflexes relying on the vestibular system are involved in the stabilization of the human gaze: the vestibulocollic reflex (VCR), which stabilizes the head in space and the vestibulo-ocular reflex (VOR), which stabilizes the visual axis to minimize retinal image motion. Together they keep the image stationary on the retina.

In this work we present the first complete model of eye-head stabilization based on the coordination of VCR and VOR. The model is provided with learning and adaptation capabilities based on internal models. Tests on a simulated humanoid platform replicating torso disturbance acquired on human subject performing various locomotion tasks confirm the effectiveness of our approach.

Keywords: Head stabilization · VOR · VCR · Eye-head coordination · Humanoid robotics

1 Introduction

Several neuroscientific studies focus on eye-head behaviour during locomotion. Results about head and eyes during walking mostly come from two-dimensional studies on linear overground, turning, treadmill locomotion, running and walking on compliant surface [1–6]. These studies have shown that the body, head, and eyes rotate in response to the up-down and side-to-side motion to maintain stable head pointing and gaze in space. This is achieved through the joint effect of two main classes of reflexes, which rely on the output of the inertial system: 1. the vestibulo-ocular reflex (VOR), which stabilizes the visual axis to minimize retinal image motion; 2. the vestibulocollic reflex (VCR), which stabilizes the head in space through the activation of the neck musculature in response to vestibular inputs. The VOR compensates for head movements that would perturb vision

by turning the eye in the orbit in the opposite direction of the head movements [7]. Several approaches have been used to model the VOR depending on the aim of the study. In robotics literature we found some controllers inspired by the VOR [8–11]. The VCR stabilizes the head based on the inertial input space by generating a command that moves the head in the opposite direction to that of the current head in space displacement. When the head is rotated in the plane of a semicircular canal, the canal is stimulated and the muscles are activated. This stimulation produces a compensatory rotation of the head in the same plane. If more than one canal is activated, then an appropriate reflex response is produced. Unlike the VOR, the VCR controls a complex musculature. The VOR involves six extraocular muscles, each pair acts around a single rotation axis. On the other hand, the neck has more than 30 muscles controlling pitch, roll and yaw rotations.

In robotics, some head stabilization models already exist implemented on humanoid robots. Gay et al. [12] proposed a head stabilization system for a bipedal robot during locomotion controlled by the optical flow information. It is based on Adaptive Frequency Oscillators to learn the frequency and phase shift of the optical flow. Although the system can successfully stabilize the head of the robot during its locomotion, it does not take in consideration the vestibular inputs. The most close to the neuroscientific findings of the VCR are the works proposed by Kryczka et al. [13–15]. They proposed an inverse jacobian controller [13,14] based on neuroscientific results [16] and an adaptive model based on a feedback error learning (FEL) [15] able to compensate the disturbance represented by the trunk rotations. All the presented models try to reproduce specific aspects of the gaze stabilization behaviour, but none of them can provide a comprehensive model of gaze stabilization, integrating eye stabilization (VOR) together with head stabilization (VCR).

By considering the analysis of neuroscience findings, we can conclude that in order to replicate eye-head stabilization behaviours found in humans it is necessary to be able to replicate the joint effect of VCR for the head and VOR for the eye. This work goes in this direction by presenting a model that replicates the coordination of VCR and VOR and is suitable for the implementation on a robotic platform. We used, as a disturbance motion, inertial data acquired on a human subject performing various locomotion tasks (straight walking, running, walking a curved path on normal and soft ground) and replicated by the torso of a humanoid robot. The purpose of these tests is to assess the effectiveness of the stabilization capabilities of the proposed model rejecting the torso disturbance measured in real walking tasks through the joint stabilizing effect of head and eye of the simulated iCub robot.

2   Eye-Head Stabilization Model

In order to implement the VOR-VCR system, a bio-inspired feed-forward control architecture was used. The model uses classic feedforward controllers that generate motor commands purely based on the current error. Each controller

is coupled with a learning network that generates predictions based on internal models that are used to fine tune motor commands. An overview of the model can be seen in Fig. 1.
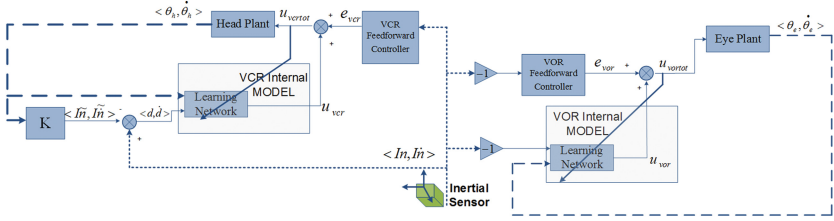
Fig. 1. The proposed model of eye-head stabilization. Dashed lines represent encoder readings, dotted lines show inertial values.

2.1 Head Stabilization System

Inside the head stabilization system, the output of the VCR internal model (u_{vcr}) is added to the output of the feedforward controller (e_{vcr}) in order to generate motor commands that stabilized the head against the disturbance originating from the torso movements. The VCR Feedforward Controller is implemented as a PD controller, and its output is computed as a function of the inertial readings (In, $\dot{In}$):

$$e_{vcr} = k_p \cdot In + k_d \cdot \dot{In}. \tag{1}$$

The inputs to the learning network are the current and desired position and velocity of the robotic head, and the network is trained with newly generated motor commands. In order to provide a proper reference to the VCR internal model, the current value of the external disturbance must be estimated. Using the readings coming from the inertial measurement unit and the encoder values, the disturbance vector (d) can be estimated using only direct kinematics functions, by computing $d = In - \tilde{In}$, i.e. by subtracting the expected angular rotations given by the encoder values ($\tilde{In}$) from the inertial readings (In). $\tilde{In} = [\varphi, \vartheta, \psi]$ are the Euler angles for the rigid roto-translation matrix $K(\theta_h)$ from the root reference frame to the inertial frame, computed as:

$$\varphi = atan2(-K(\theta_h)_{2,1}, K(\theta_h)_{2,2}), \tag{2}$$
$$\vartheta = asin(K(\theta_h)_{2,0}), \tag{3}$$
$$\psi = atan2(-K(\theta_h)_{1,0}, K(\theta_h)_{0,0}). \tag{4}$$

Likewise, the same procedure can be followed in order to estimate the velocity of the disturbance:

$$\dot{d} = \dot{In} - \dot{\tilde{In}} = \dot{In} - J(\theta_h) \cdot \dot{\theta_h}, \tag{5}$$

where J is the geometric Jacobian from the root reference frame to the inertial frame.

2.2   Eye Stabilization System

The eye stabilization system implements the VOR and, similarly to the head stabilization system, produces a motor command for the eyes that is the sum of the feedforward controller output (e_{vor}) and the VOR internal model one (u_{vor}). Given that the eye should stabilize the image against the relative rotation of the head, the error is computed as the difference between inertial measurements and the current eye encoders (θ_e, $\dot{\theta}_e$). Thus, the output of the VOR feedforward controller is computed as

$$e_{vor} = k_p \cdot (-In) + k_d \cdot (-\dot{In}). \tag{6}$$

The VOR internal model receives in input the head position and velocity signal as references, acquired through the vestibular system, along with the proprioceptive feedback, and uses the generated motor command as a training signal.

2.3   Learning Network

Prediction of the internal model is provided by a learning network that is implemented with a machine learning approach, Locally Weighted Projection Regression (LWPR) [17]. This algorithm has been proved to provide a representation of cerebellar layers that in humans are responsible for the generation of predictive motor signals that produce more accurate movements [18,19]. The LWPR spatially exploits localized linear models at a low computational cost through an online incremental learning. Therefore, the prediction process is quite fast, allowing real-time learning. LWPR incrementally divides the input space into a set of receptive fields defined by a centre c_k and a Gaussian area characterized by a positive definite distance matrix D_k. The activation on each receptive field k in response to an input x is expressed by

$$p_k(x) = exp\left(-\frac{1}{2}(x - c_k)^T D_k (x - c_k)\right), \tag{7}$$

while the output is $y_k(x) = w_k \cdot x + \epsilon_k$, where w_k and ϵ_k are the weight vector and bias associated with the k-th linear model. For each iteration, the new input, x, is assigned to the closest RF based on its weight activation, and consequently, the centre, the weights and the kernel width are updated proportionally to a training signal. Moreover, the number of local models increases with the complexity of the input space.

The global output of the LWPR is given by the weighted mean of all the predictions y_k of the linear local models created:

$$u(x) = \frac{\sum_{k=1}^{N} p_k(x) y_k(x)}{\sum_{k=1}^{N} p_k(x)}. \tag{8}$$

3   Experimental Procedure

In order to collect human inertial data relative to locomotion tasks, experiments were conducted on a human subject with no visual and vestibular impairments. An inertial measurement unit (IMU) was placed on the back of the subject, near T10, the tenth vertebra of the thoracic spine, as depicted in Fig. 2.
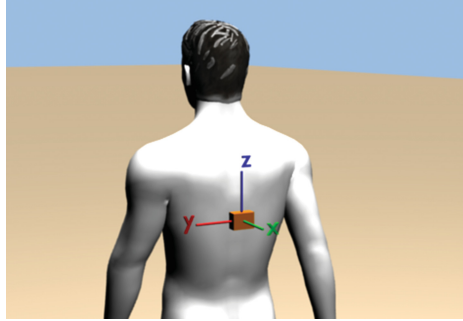
Fig. 2. Placement of the inertial measurement unit on the subject.

The IMU used was an Xsens MTi orientation sensor[1], that incorporates an on-board sensor fusion algorithm and Kalman filtering. The inertial unit is able to produce the current orientation of the torso at a frequency of 100 Hz.

Three different tasks were performed by the subject: straight walking (25 m), circular walking and straight running (25 m). The circular walking was carried out by asking the subject to walk with a circular pattern, without any indication of the pattern on the ground. Such task was executed both on normal and soft ground, provided by placing a foam rubber sheet on the ground. The foam had a density of $40 \, \mathrm{kg/m^3}$ and the sheet measured $103 \times 160 \times 15 \, \mathrm{cm}$. All tasks were performed with bare feet.

Due to the fact that the inertial readings relative to the yaw rotational axis (rotation around z) can often be inaccurate because of drifting, we decided not to use such readings. Moreover, in order to prevent drifts of the sensor measurements on the other two rotational axis (pitch and roll, rotations around y and x respectively), each trial lasted less than one minute with a reset of the rotational angle at the beginning of the trial [20].

4   Robotic Platform

The proposed model was implemented for the iCub robot simulator [21], a software included with the iCub libraries. The iCub head contains a total of 6 degrees of freedom: 3 for the neck (pan, tilt and swing) and 3 for the eyes (a common

---

[1] https://www.xsens.com/products/mti/.

tilt, version and vergence), while the torso has 3 degrees of freedom (pan, tilt and swing). The visual stereo system consists of 2 cameras with a resolution of 320×240 pixels.

In order to assess the repeatability of the experiments on the iCub simulator, first test were conducted to evaluate whether the measurements of the simulated robot inertial rotations were compatible with the collected data. Thus, the collected torso rotations were given as motor commands to the robot torso. A graphical comparison can be seen in Fig. 3, where the actual IMU data is shown alongside the robot one. It can be observed that the simulation is accurate enough to reproduce the data, even if with a delay of 50 ms. The error between the two signals was then computed after a temporal alignment and its Root Mean Squared value is 0.21 deg for the pitch rotational axis and 0.12 deg for the roll.
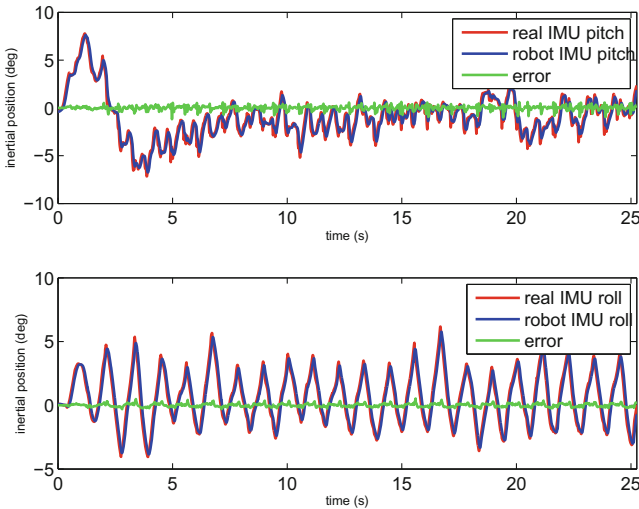
Fig. 3. Comparison between collected IMU readings and simulated ones, on pitch rotational axis (top) and roll (bottom). (Color figure online)

5   Results

The stabilization model was tested on the data coming from the three different locomotion tasks (straight walking, circular walking and straight running). Due to the fact that the collected inertial data related to the yaw rotational axis was not considered, the eye-head stabilization model has been simplified, so that no stabilization on the yaw axis was performed. Moreover, given that the robot eyes cannot influence stabilization on the roll rotational axis, due to the fact that only tilt and pan motors are present, only disturbance on the pitch axis was compensated by the VOR model.

The main measure of error during a stabilization task is the movement of the camera image. In particular, human vision is considered stable if the retinal slip (the speed of the image on the retina) is under $4\,\mathrm{deg/s}$ [22]. In order to compute the error from the camera image, a target was placed in front of the simulated robot and its position was tracked from the camera images via a colour filtering algorithm during the execution of the task. Another measure of performance considered is the inertial orientation and speed of the head. As already stated before, no movement on the yaw rotational axis was considered, thus only the camera error on the vertical axis is relevant for the evaluation.

For each task, a comparison between the same task performed with and without the stabilization model will be presented. The values of the gains of the PD controllers were set to $k_p = 5.0, k_d = 0.1$ for the VCR model and to $k_p = 1.0, k_d = 0.1$ for the VOR model, for all trials.

5.1 Straight Walking

Results for the compensation of the disturbance of straight walking inertial data can be found in Table 1, where the Root Mean Square (RMS) values for inertial readings and target position and speed are presented. In this and subsequent tables, $In_p, \dot{In}_p$ are the inertial readings for rotation (deg) and rotation speed (deg/s) on the pitch axis, $In_r, \dot{In}_r$ are the inertial readings for rotation (deg) and rotation speed (deg/s) on the roll axis, $v, \dot{v}$ are the position of the target on the camera image (deg) and its speed (retinal slip, deg/s).

Table 1. Results for straight walking data.

| Trial	In_p	$\dot{In}_p$	In_r	$\dot{In}_r$	v	$\dot{v}$
No stabilization	2.06	11.36	2.05	11.56	3.66	8.88
Stabilization	0.84	14.71	0.22	3.45	2.71	3.30

Figures 4, 5 and 6 show the behaviour of the task, showing the target position and retinal slip, inertial data for the pitch rotational axis and inertial data for the roll axis, respectively. From these results it can be noticed that while the roll disturbance is almost completely compensated by the VCR model, the magnitude of the rotational velocity on the pitch axis is too high to be fully compensated by the said model, that only provides an improvement in the position space. Nevertheless, the VOR subsystem is still able to maintain the camera image stable, with a mean vertical retinal slip lower than $4\,\mathrm{deg/s}$. Moreover, Fig. 4 also shows a comparison between the full stabilization model and a simplified model with only the PD controllers. While the PD only implementation is able to reduce the error on the camera, it is outperformed by the complete model, thus proving the effectiveness of the latter.

5.2   Circular Walking

Two sets of data were collected for circular walking tasks: one for normal ground and one for soft ground. Results for both cases are presented in Table 2, where it can be observed that walking on soft ground produces a greater disturbance, especially in the velocity space. Despite the higher disturbance the model is still able to stabilize the head and the camera image, achieving stable vision in both cases. As in the straight walking case, the disturbance on the pitch axis cannot be
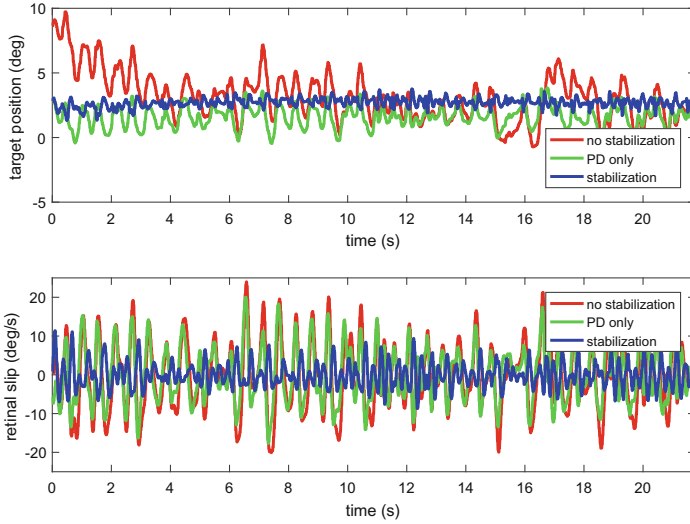
Fig. 4. Stabilization task with data from a straight walking task, target position (top) and retinal slip (bottom). (Color figure online)
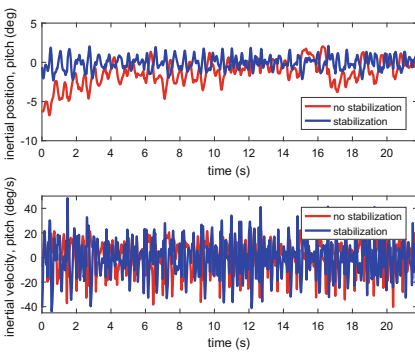
Fig. 5. Stabilization task with data from a straight walking task, inertial position data (top) and inertial velocity (bottom), pitch axis. (Color figure online)
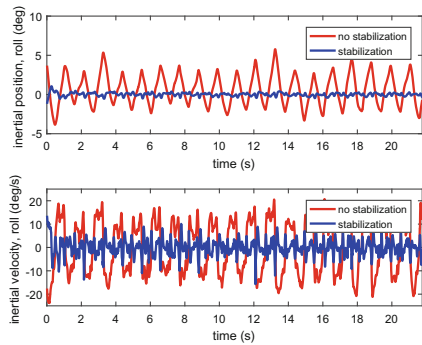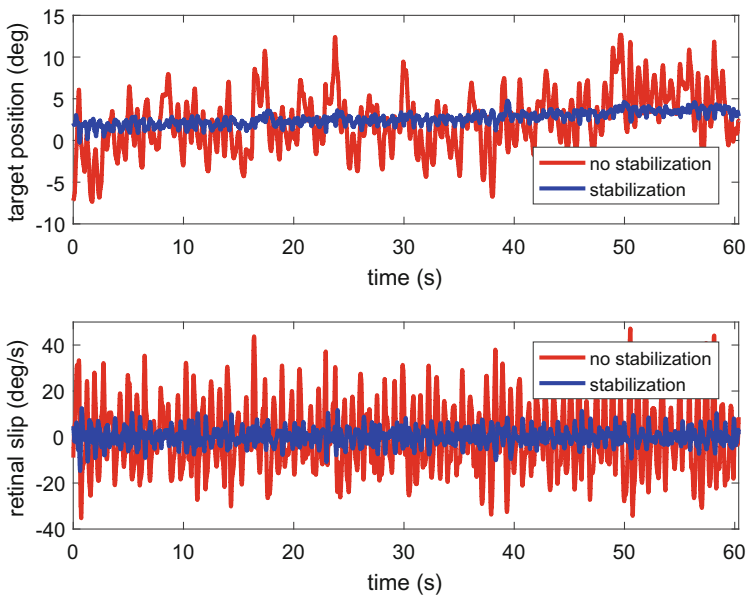
Fig. 6. Stabilization task with data from a straight walking task, inertial position data (top) and inertial velocity (bottom), roll axis. (Color figure online)

fully compensated by the VCR alone, but thanks to the VOR module, the vision remains stable. The behaviour on the soft ground task can be observed in Fig. 7.

Table 2. Results for circular walking data on normal and soft ground.

| Ground	Trial	In_p	$\dot{In}_p$	In_r	$\dot{In}_r$	v	$\dot{v}$
Normal	no stabilization	3.35	14.03	2.11	6.36	3.61	12.86
Normal	stabilization	0.49	6.34	0.17	2.84	0.88	1.85
Soft	no stabilization	3.09	16.00	3.89	10.95	4.23	15.15
Soft	stabilization	0.46	6.35	0.24	3.07	2.74	3.56

Fig. 7. Stabilization task with data from a circular walking task on soft ground, target position (top) and retinal slip (bottom). (Color figure online)

5.3 Straight Running

During the last experiment, data from the straight running was used to move the robot torso. From Table 3 it can be observed that the model is not able to achieve a complete compensation of the disturbance, due to the high rotational velocities on the two axes. Nevertheless, the mean retinal slip is reduced to a quarter of the one of the trial with no stabilization. Thus, the model provides a viable solution even for disturbances of this magnitude, as it is also shown in Fig. 8.

Table 3. Results for straight running data.

Trial	In_p	$\dot{In}_p$	In_r	$\dot{In}_r$	v	$\dot{v}$
No stabilization	3.14	45.36	1.77	16.28	3.49	25.96
Stabilization	1.06	43.03	0.53	10.78	1.61	6.36
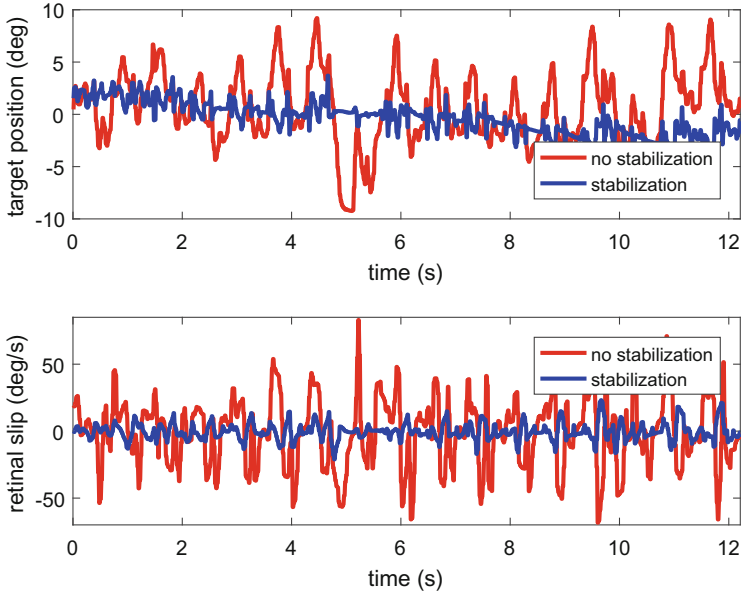
Fig. 8. Stabilization task with data from a straight running task, target position (top) and retinal slip (bottom). (Color figure online)

6   Conclusions

In this work we present the first complete model of gaze stabilization based on the coordination of VCR and VOR and we validate it through an implementation on a simulated humanoid robotic platform. We tested the model using, as a disturbance motion, inertial data acquired on a human subject performing various locomotion tasks that we replicated with the torso of the simulated iCub robot. Results show that the model is able to perform well in almost all trials, with the exception of the straight running task, by reducing the retinal slip below 4 deg/s, thus achieving stable vision. In the running task, the model was still able to improve the stabilization by reducing the retinal slip to a quarter of the one from the task were no stabilization was present. As such, this model has proven suitable to be used on humanoid robotic platforms, where it could help during visually guided locomotion tasks by stabilizing the camera view against the disturbance produced by walking.

References

1. Imai, T., Moore, S.T., Raphan, T., Cohen, B.: Interaction of the body, head, and eyes during walking and turning. Exp. Brain Res. **136**(1), 1–18 (2001)
2. Hirasaki, E., Moore, S.T., Raphan, T., Cohen, B.: Effects of walking velocity on vertical head and body movements during locomotion. Exp. Brain Res. **127**(2), 117–130 (1999)
3. Pozzo, T., Berthoz, A., Lefort, L., Vitte, E.: Head stabilization during various locomotor tasks in humans. Exp. Brain Res. **85**(1), 208–217 (1991)
4. Nadeau, S., Amblard, B., Mesure, S., Bourbonnais, D.: Head and trunk stabilization strategies during forward and backward walking in healthy adults. Gait Posture **18**(3), 134–142 (2003)
5. Hashimoto, K., Kang, H.J., Nakamura, M., Falotico, E., Lim, H.O., Takanishi, A., Laschi, C., Dario, P., Berthoz, A.: Realization of biped walking on soft ground with stabilization control based on gait analysis. In: IEEE International Conference on Intelligent Robots and Systems, pp. 2064–2069 (2012)
6. Kang, H.J., Hashimoto, K., Nishikawa, K., Falotico, E., Lim, H.O., Takanishi, A., Laschi, C., Dario, P., Berthoz, A.: Biped walking stabilization on soft ground based on gait analysis. In: Proceedings of the IEEE RAS and EMBS International Conference on Biomedical Robotics and Biomechatronics, pp. 669–674 (2012)
7. Barnes, G.: Visual-vestibular interaction in the control of head and eye movement: the role of visual feedback and predictive mechanisms. Prog. Neurobiol. **41**(4), 435–472 (1993)
8. Shibata, T., Schaal, S.: Biomimetic gaze stabilization based on feedback-error-learning with nonparametric regression networks. Neural Netw. **14**(2), 201–216 (2001)
9. Viollet, S., Franceschini, N.: A high speed gaze control system based on the vestibulo-ocular reflex. Robot. Auton. Syst. **50**(4), 147–161 (2005)
10. Porrill, J., Dean, P., Stone, J.V.: Recurrent cerebellar architecture solves the motor-error problem. Proc. Roy. Soc. Lond. B **271**(1541), 789–796 (2004)
11. Franchi, E., Falotico, E., Zambrano, D., Muscolo, G., Marazzato, L., Dario, P., Laschi, C.: A comparison between two bio-inspired adaptive models of vestibulo-ocular reflex (VOR) implemented on the iCub robot. In: 2010 10th IEEE-RAS International Conference on Humanoid Robots, Humanoids 2010, pp. 251–256 (2010)
12. Gay, S., Santos-Victor, J., Ijspeert, A.: Learning robot gait stability using neural networks as sensory feedback function for central pattern generators. In: 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 194–201, November 2013
13. Kryczka, P., Falotico, E., Hashimoto, K., Lim, H., Takanishi, A., Laschi, C., Dario, P., Berthoz, A.: Implementation of a human model for head stabilization on a humanoid platform. In: Proceedings of the IEEE RAS and EMBS International Conference on Biomedical Robotics and Biomechatronics, pp. 675–680 (2012)

14. Kryczka, P., Falotico, E., Hashimoto, K., Lim, H.O., Takanishi, A., Laschi, C., Dario, P., Berthoz, A.: A robotic implementation of a bio-inspired head motion stabilization model on a humanoid platform. In: IEEE International Conference on Intelligent Robots and Systems, pp. 2076–2081 (2012)

15. Falotico, E., Cauli, N., Hashimoto, K., Kryczka, P., Takanishi, A., Dario, P., Berthoz, A., Laschi, C.: Head stabilization based on a feedback error learning in a humanoid robot. In: Proceedings - IEEE International Workshop on Robot and Human Interactive Communication, pp. 449–454 (2012)

16. Falotico, E., Laschi, C., Dario, P., Bernardin, D., Berthoz, A.: Using trunk compensation to model head stabilization during locomotion. In: IEEE-RAS International Conference on Humanoid Robots, pp. 440–445 (2011)

17. Vijayakumar, S., Schaal, S.: Locally weighted projection regression: incremental real time learning in high dimensional space. In: ICML 2000: Proceedings of the Seventeenth International Conference on Machine Learning. Morgan Kaufmann Publishers Inc., San Francisco, pp. 1079–1086 (2000)

18. Tolu, S., Vanegas, M., Luque, N.R., Garrido, J.A., Ros, E.: Bio-inspired adaptive feedback error learning architecture for motor control. Biol. Cybern. **106**(8–9), 507–522 (2012)

19. Tolu, S., Vanegas, M., Garrido, J.A., Luque, N.R., Ros, E.: Adaptive and predictive control of a simulated robot arm. Int. J. Neural Syst. **23**(3), 1350010 (2013)

20. Bergamini, E., Ligorio, G., Summa, A., Vannozzi, G., Cappozzo, A., Sabatini, A.: Estimating orientation using magnetic and inertial sensors and different sensor fusion approaches: accuracy assessment in manual and locomotion tasks. Sens. (Switz.) **14**(10), 18625–18649 (2014)

21. Tikhanoff, V., Cangelosi, A., Fitzpatrick, P., Metta, G., Natale, L., Nori, F.: An open-source simulator for cognitive robotics research: the prototype of the icub humanoid robot simulator. In: Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems, PerMIS 2008, pp. 57–61. ACM, New York (2008)

22. Collewijn, H., Martins, A., Steinman, R.: Natural retinal image motion: origin and change. Ann. N. Y. Acad. Sci. **374**(1), 312–329 (1981)