Power Usage Reduction of Humanoid Standing Process using Q-Learning

Ercan Elibol[1,⋆], Juan Calderon[1,2], Martin Llofriu[1], Carlos Quintero[2], Wilfrido Moreno[1], and Alfredo Weitzenfeld[1]

[1] University of South Florida, Tampa, FL, USA
{ercan,juancalderon,mllofriu}@mail.usf.edu
{wmoreno,aweitzenfeld}@usf.edu
http://usf.edu/
[2] Universidad Santo Tomás, Colombia
{juancalderon,carlosquinterop}@usantotomas.edu.co
http://www.usta.edu.co/

Abstract. An important area of research in humanoid robots is energy consumption, as it limits autonomy, and can harm task performance. This work focuses on power aware motion planning. Its principal aim is to find joint trajectories to allow for a humanoid robot to go from crouch to stand position while minimizing power consumption. Q-Learning (QL) is used to search for optimal joint paths subject to angular position and torque restrictions. A planar model of the humanoid is used, which interacts with QL during a simulated offline learning phase. The best joint trajectories found during learning are then executed by a physical humanoid robot, the Aldebaran NAO. Position, velocity, acceleration, and current of the humanoid system are measured to evaluate energy, mechanical power, and Center of Mass (CoM) in order to estimate the performance of the new trajectory which yield a considerable reduction in power consumption.

Keywords: humanoid, dynamic modeling, energy analysis, optimization, Q-Learning

1 Introduction

Energy efficiency is a significant challenge of humanoid robots, and mobile robots in general. These robots contain many different components that consume energy, but a great portion is consumed by DC motors that transform direct current to mechanical energy to drive them. While all components should be analyzed for energy efficiency, DC motor activation and control consume most of the energy required by many dynamic and static motion tasks.

⋆ This work is funded by NSF IIS Robust Intelligence research collaboration grant #1117303 at USF and U. Arizona entitled "Investigations of the Role of Dorsal versus Ventral Place and Grid Cells during Multi-Scale Spatial Navigation in Rats and Robots," and supported in part by the Agencia Nacional de Investigacion e Innovación (ANII).

NAO robots are used in the Standard Platform League (SPL) of RoboCup. It is known that battery duration is one of the main constraints for longer humanoid robot autonomous performance. In order to use the SPL Robots for extended time, it is necessary to recharge batteries at least once during a single game.

Humanoids body weight, power needs and consumption of individual components play a significant role in energy utilization, balance and stability [1].

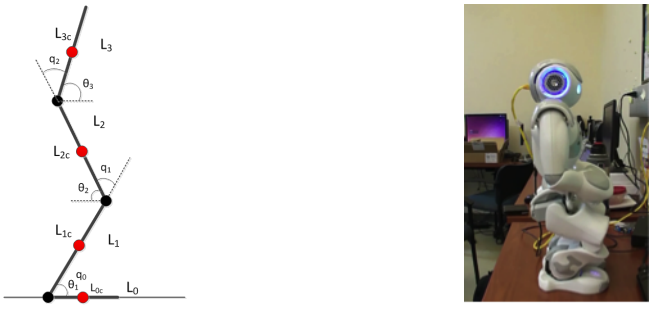
In terms of humanoid tasks, different approaches have been used to address the problem of stability. In [2], a humanoid robot stands up from sitting on a chair by using data previously collected from human demonstrations, where stable humanoid motion is accomplished by emulating human-like movements and speed. In [3], a three link simulated inverted pendulum learns to stand up using a tiered reinforcement learning method. A hierarchical architecture is applied on a three links two joints single legged robot during learning to stand up by trial and error. Our approach is dealing with a multiple goal settings; the robot has to learn the motion task while minimizing energy consumption. The study presented in [4] used a genetic algorithm fitness function to analyze the relationship between walking distance and energy consumption while keeping the knee joint on the supporting leg straight. In [5] a trajectory generation method for humanoid robots is proposed to achieve stable movement by using consumed energy as a condition, and generating a series of joint motions with a feedback technique to increase its stability. Reinforcement learning algorithms have been widely applied to other legged motions tasks [6]. Most of such work involves learning how to walk using biped robots [7], [8], [9], [10]. Other related work includes learning to perform a penalty kick with a biped robot and learning to keep robot balance with an inverted pendulum model [11]. The work by Kuindersma et al. [12] had energy consumption optimization explicitly coded as a learning goal. Their work focused on moving the robot arms to compensate for balance disturbances and coded for the energy utilization of their movements in the cost function. Our model however, deals with the energy required to accelerate the whole robot body upwards, which requires a dynamic humanoid model of motions of the produced torque at each joint.

The motion of standing up from crouch position seems like a simple and common motion for humans, but it is quite complex, dynamic, and can become challenging for biped robots. Calderon et al. [13], present a joint stiffness control algorithm with the aim of reducing energy usage during the standing up procedure of a NAO robot. The goal of our new research is to optimize the standing up motion focusing on energy usage. In order to reduce energy consumption, a simulated kinematic and dynamic model uses Q-Learning to improve joint angular trajectories and implement an optimized route on a physical robot.

The rest of the paper is composed of Section 2 - Humanoid Modeling, Section 3 - Q-Learning Power Optimization, Section 4 - Energy and Power Performance Evaluation, Section 5 - Experimental Setup, Section 6 - Results, and Section 7 - Conclusions.

2 Humanoid Modeling

For humanoid modeling we used a NAO robot having 25 degrees of freedom (DoF), including two legs, two arms, a trunk, and a head, as shown in Fig. 1.



(a) (b)

Fig. 1: Robot model and Humanoid Robot (NAO) in the sagittal plane.

A three DoF model in the x-z plane is used with the objective of reducing the complexity of the humanoid mathematical model (Fig. 1). The model has three joints and four links. The joints are ankle, knee, and hip pitch. The links represent foot, lower leg, thigh and trunk. The reduced dimensionality model is provided by the nature of the treated movement. This means that in a jumping movement both legs (left and right) are performing same action. The model is using the first two degrees of freedom (ankle and knee for both legs) and the last one corresponds to hip attached to the trunk. At the same time trunk is taken as a one mass, which includes arms, chest, and head.

Kinematic Model. The kinematical model is used to estimate the position of each joint and CoM for every link and the whole robot, see Fig. 1. L_i denotes the length of link i, θ_i is the absolute rotation of joint i, and L_{ic} shows the position of CoM for the corresponding link i.

Dynamic Model. The dynamic model is obtained using the Lagrangian formulation with the physical parameters of the NAO robot. The dynamic formulation has the following form in equation (1):

$$D(\theta)\ddot{\theta} + H(\theta, \dot{\theta})\dot{\theta} + G(\theta) = T_\theta \tag{1}$$

$D(\theta)$ is a 3×3 inertial matrix, $H(\theta, \dot{\theta})$ is a 3×1 vector of Coriolis and centrifugal forces, $G(\theta)$ is a 3×1 gravitational forces matrix. T_θ represents the torque vector and external applied forces at the joint. Finally the vectors θ, $\dot{\theta}$, $\ddot{\theta}$ represent rotational position, velocity, and acceleration of each joint.

Motor Model. The joint DC motor model used in simulations was estimated from collected actual motor responses to produce an approximation of each joint motor response during standup motions. A dynamic robot model used in simulations consists of three DoF. This simulation uses three different models, one for each joint. The transfer function model is shown in equation (2), parameters for the ankle are: $\omega = 0.0098$ rad/sec, $\zeta = 1.4$, for the knee are: $\omega = 0.0097$ rad/sec, $\zeta = 0.9$ and for the hip are $\omega = 0.0021$ rad/sec, $\zeta = 4.95$.

$$G_{(s)} = \frac{1}{\omega^2 s + 2\omega\zeta s + 1} \tag{2}$$

A second order system is used as the process model and the Predicting-Error Minimization (PEM) algorithm [14] is used as the estimation method. In order to increase the accuracy of the information given to the estimation algorithm, joint position and target position from the each joint motor was recorded previously while performing Aldebaran version of the stand-up process. The data was applied to PEM estimation method and the humanoid dynamic model to obtain an individual joint model as close to the real behavior of each motor as possible for the standing up motion.

3 Q-Learning Power Optimization

Q-Learning is a machine learning algorithm capable of learning a policy based solely in spurious feedback [15]. We used the canonical tabular version, in which the Q-value is updated according to equation (3), where α is a learning rate parameter.

$$Q(s,a) = Q(s,a) + \alpha\big(r + max_{a'}Q(s',a') - Q(s,a)\big) \tag{3}$$

A policy Π of which action to perform at each state can be derived from the Q table, as shown in equation (4).

$$\Pi\colon s \rightarrow a :: \Pi(s) = argmax_a Q(s,a) \tag{4}$$

The fact that this algorithm does not rely on a labeled training set, and does not rely on a model of action outcomes, make it suitable for its application to robotics. In fact, reinforcement learning algorithms including Q-Learning have been widely applied to robotics [6].

3.1 Power Aware Stand Up Learning Algorithm

The proposed solution to the problem of learning to stand up with the minimum possible energy consumption is implemented as a QL algorithm. A planar humanoid robot is modeled with three joints: ankle, knee and heap, as explained in section 2. The QL algorithm controls the ankle and knee joints only, whereas the hip joint position was set so the robot remains with the torso vertical to the ground.

The angular velocities and positions of the ankle and knee joints determine the state. Each joint state-space was discretized using a fixed length discretization step of $\pi/20$ rad. The same fixed length discretization was performed for velocities, with a discretization step of 0.1 rad/s.

The agent was allowed to perform one of three possible actions. Each of them performed a change on the ankle and knee velocities: decrement it, leave it unmodified or increment it. The decrements and increments were done by a fixed predefined value. Equation (5) shows how the reward is computed. A negative reward is given whenever the humanoid performs a motion that leaves a joint in an invalid position (jointOutOfConstraints), according to the NAO robot limits. A negative reward is also given if the humanoid falls down (robotFell). It is considered to have fallen when the hip displacement along the sagittal plane is beyond a non-return point.

A positive reward is given if the humanoid reaches a target stand up position within some error tolerance (standingUp) and all joint angular velocities are below a threshold (notMoving). The position requirement is necessary for the humanoid to learn the task of standing up. The velocity constraints, on the other hand, ensures that the final inertia of the standup motion does not make the robot fall or force it to make a big energy effort to lower it. The average torque produced is subtracted from the positive reward value. This promotes solutions that minimize torque application, which in turn minimizes energy consumption. Only one non-zero reward is given in each episode, right at episode termination.

$$r = \begin{cases} -10, & \text{if jointOutOfConstraints or robotFell} \\ 3 - averageTorque, & \text{if standingUp and notMoving} \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

Calibration. The ability of the learning algorithm to find a good solution depends on the value of a set of initial parameters. It was decided to perform a calibration process for the three parameters which we considered are the most important and they are: the learning rate α, the eligibility traces decay parameter γ and the exploration half-life decay parameter ϵ (ϵ-greedy with exponential decay). A coarse parameter sweep of 5 different values per parameter was performed. For each set of parameters, the algorithm was executed 5 times and the average reward was taken as a score. The set of parameters with the highest score was picked as the definitive set of parameter values.

4 Energy and Power Performance Evaluation

In order to evaluate power performance, we are assessing the average mechanical power, standard deviation and energy lost in every joint. Given joint j of the leg i, the mechanical power is the product of the motor torque τ and the angular velocity $\dot{\theta}$. The overall average power is obtained by averaging the mechanical absolute power delivered over a period T for all joints by equation (6):

$$P_{avm} = \frac{1}{T} \sum_{i,j} \int_0^T |\tau_{ij} \dot{\theta}_{ij}| dt \tag{6}$$

For some dynamic motions performed by humanoid robots, a sudden very high power demand can occur at the joints. Even though the average value of power usage can be small, the peak can actually be very high. The standard deviation measure is used to evaluate the distribution of power around the mean absolute power as seen in equations (7) and (8):

$$P_{sd} = \sqrt{\frac{1}{T} \int_0^T \big(\sum_i \tau_{ij} \dot{\theta}_{ij} - P_{avm} \big)^2 dt} \tag{7}$$

For a humanoid robot, it is also necessary to consider the energy lost in the electric motors [16]. This can be defined as shown in equation (8):

$$E_{Lost} = \frac{1}{T} \int_0^T \tau^\intercal \tau dt \tag{8}$$

5 Experimental Setup

5.1 Learning Cycle

In order to be able to perform offline learning, a simulator was programmed using the motor model and the humanoids kinematic and dynamic models previously described in section 3. Fig. 2 shows the flow of events of a single iteration of Q-Learning episode. First, an action is selected by determining the state and querying the Q-Value table. Then, the motor models are used to compute the motor response to the required velocities. After that, kinematic models are applied to find joint positions, velocities and accelerations. This data is used by the dynamic model to compute the performed torques. Those torques, along with the kinematic information, are in turn used to compute the reward and update the Q-Value table.

A decision process is carried out to determine whether the episode has failed, succeeded, or it should continue. In the latter case, the cycle starts all over again.

Finally, the best standing routine was obtained by executing the calibrated algorithm 50 times. Then, the route with the highest reward was chosen.

5.2 Robot Execution

The obtained route was interpolated to a 4 second routine. This was done in order to be able to compare it with Aldebaran's stand-up routine, which was also set so as it would stand-up in 4 seconds. Fig. 3 shows the robot at different points of the standing up routine. Then, the angleInterpolation function of the naoqi API was used. Twenty-five repetitions of the experiment were carried out

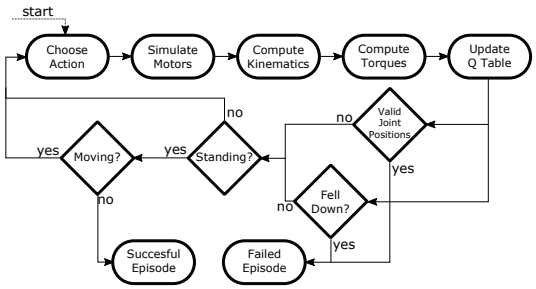


Fig. 2: The flow of events of a single iteration inside an episode.

for each routine. A custom made NAO local module *mllofriu/getSensorValues* (available in Github) was used to sample position and electric current values. The sample time was set to 10ms, which is the minimum loop latency allowed by the robot.

(a) (b) (c)

Fig. 3: The robot performing a stand up motion using the data obtained from Q-Learning.

6 Results

In this section, the results of both routines (Aldebaran routine and Q-Learning routine) are presented and discussed. The discussion is mainly based on the electric current consumption and position of each joint (ankle, knee, and hip), since the greater power consumption is located on these joints as discussed by Elibol et al. in [17]. The current consumption will be used to calculate electrical performance (motor input power) and the position will be used to calculate the location of Center of Mass, angular velocity, acceleration, produced torque, mechanical power and energy lost due to produced torque. The trajectories followed by joints in each routine are shown in Fig. 4. The Q-Learning trajectory is very different from Aldebaran's, which partially explains the difference in performance, as it will be discussed later in this section.

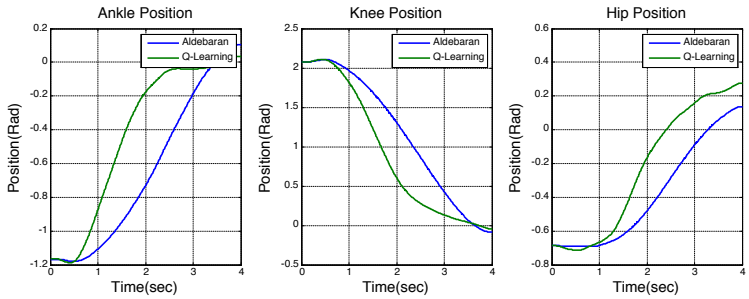
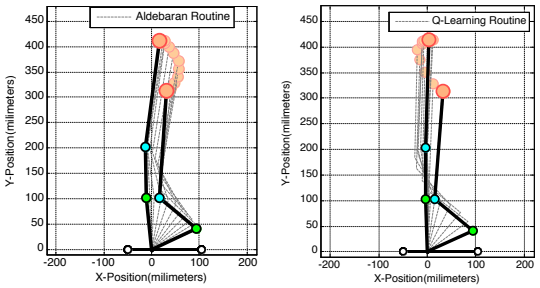


Fig. 4: Trajectories followed by every joint, Q-Learning trajectories follow different paths to reach the target positions.

Fig. 5 shows the humanoid robot executing the Aldebaran routine and the Q-Learning one. In this figure, the difference between the trajectories can be seen more clearly. Notice how Aldebaran trajectory tends to keep the body on top of the middle of the foot, while the Q-Learning trajectory leans all weight directly on top of the ankle joint. See also videos 1 (https://youtu.be/qsdiczXCBSQ) and 2 (https://youtu.be/YbnB6dx9cII) of the additional multimedia material for a sample learning iteration and a real robot testing iteration, respectively.

Because these trajectories are different, different current consumption are expected for each joint. Fig. 6 shows the comparison of current profile between both routines for each joint.



(a) (b)

Fig. 5: Planar humanoid model shows Aldebaran (a) and Q-Learning (b) routines. Both routines show the initial and final position. Green circle is knee, blue circle is hip, and orange circle is head.

The current required by ankle joint is far less for the Q-Learning routine than Aldebaran routine as shown by Fig. 6 and Table 1. This current saving is because of the decreased demand of produced torque at the ankle joint, which is effected by hip vertical position. A similar result is found when analyzing the hip current. Hip position is kept vertically during QL routine. The reduction of current of the hip and ankle joints responds to the Q-Learning reward schema, where lower torques are considered better. The knee, on the other hand, shows a slight increase in current consumption. Since hip position is moved to vertical position quicker in standing up motion (see Fig. 6 b), this creates an increase torque consumption on the knee joints, which demands more current.

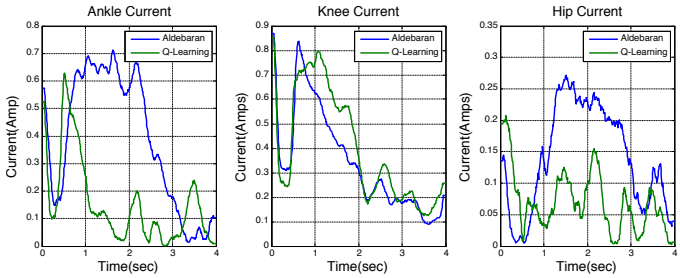


Fig. 6: Current consumption results averaged over 25 trials.

Table 1: Electrical Current and Power Consumption Comparison.

	Average Current		**Electrical Input Power**		
	Amp		Watt		
Joints	QL	Aldebaran	QL	Aldebaran	Saved energy per link
Ankle	0.193	0.425	4.7864	10.54	54.59%
Knee	0.401	0.348	9.9448	8.6304	-15.23%
Hip	0.084	0.167	2.0832	4.1416	49.70%
Total			16.814	23.312	27.87%

Table 1 shows the average current consumption for each joint. A decrease of 54.59% and 49.7% in current consumption was achieved for the ankle and hip joints respectively. Also, it can be seen that the increase of the knee current consumption (15.23%) is not as high as the amount saved by the other two joints together. This is also aligned with the Q-Learning reward schema, in which the overall torque is used, as opposed to minimizing each joint separately. By doing so, the algorithm was able to find a better tradeoff point between ankle, knee and hip joints applied torques minimizing current consumption. Other two important aspects of link movement are velocity and acceleration, since they directly affect the humanoids dynamics. This is shown in equation (3), where inertial matrix D and Coriolis and centrifugal matrix H depend directly on these variables.

Fig. 7 shows the angular velocity and acceleration profile for each joint. The Q-Learning routine shows higher values of velocity and acceleration at the beginning of the movement. This increase is producing different effects on the performance of the routine. First, the trajectories of the joints were affected, which in turn affected the required torques needed to accomplish the routine. Secondly, the dynamics of the system were affected by quickly building up more inertia at the beginning and then reducing current and torque later. Fig. 8 shows the required torque by each joint. The torque is reduced in the ankle and hip joints and slightly increased in the knee joint, which is in accordance with our analysis and supported by the experimental results.

The trajectories of Centers of Mass of both routines are shown in Fig. 9. The difference between them is highlighted. The Q-Learning location of CoM is causing a reduction of ankle and hip torques and an increase in knee torque as

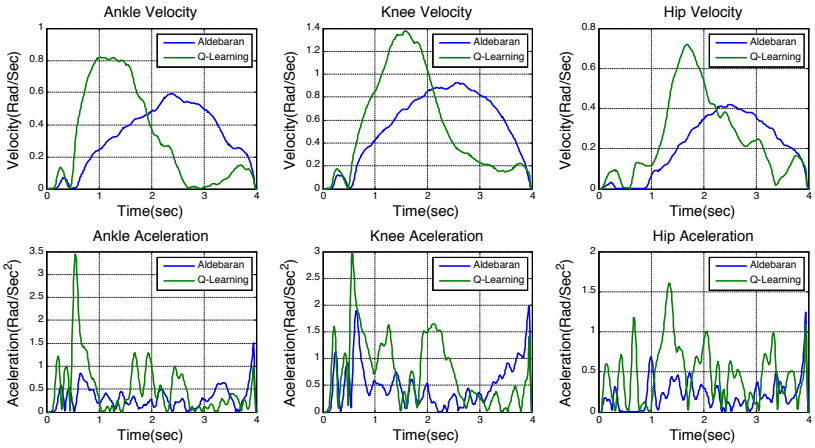


Fig. 7: Comparison of velocity and acceleration for: ankle(a), Knee(b), and Hip(c).

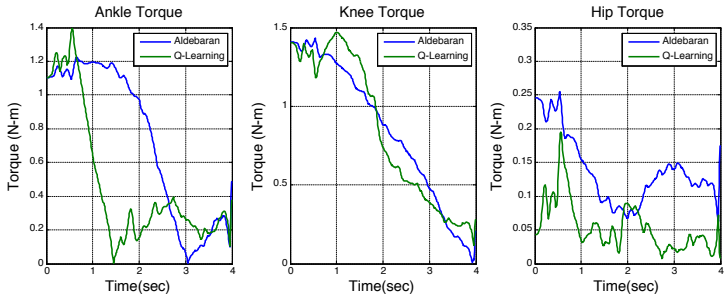


Fig. 8: Produced joint torque comparison between Aldebaran and Q-Learning routines.

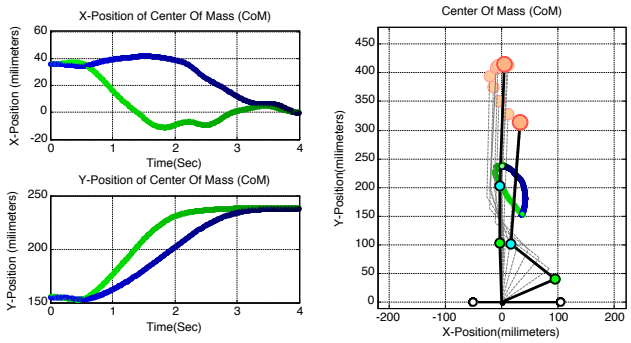


Fig. 9: CoM trajectories, x(t) & y(t), comparison between Aldebaran (blue), Q-L (green) routines (a), and spatial, y(t) vs. x(t), trajectory (b).

shown in Fig. 9. The mean and standard deviation of the CoM for both routines were calculated to have an idea about CoM Performance. The mean in the x axis for Aldebaran routine is 11.8 mm. and 3.3 mm for QL. This means that the QL route kept the Center of Mass closer to the ankle joint than Aldebaran route. This is one of the reasons why this QL trajectory is saving energy. The

Table 2: Produced Mechanical Power.

	Mechanical Power Average $N.m.\frac{rad}{s}$		**Mechanical Power Standard Deviation** $N.m.\frac{rad}{s}$		**Lost Energy** Joule		**Changes in Lost Energy**
Joints	QL	Aldebaran	QL	Aldebaran	QL	Aldebaran	
Ankle	0.133	0.204	0.233	0.195	0.614	0.86	-28.6%
Knee	0.555	0.421	1.26	0.888	0.963	0.955	+0.8%
Hip	0.012	0.0237	0.012	0.05	0.065	0.147	-55.7%

standard deviation is 16.8 mm. for Aldebaran and 11.8 mm for QL. This result suggests that QL is better at keeping balance.

The mechanical power performance of both routines was evaluated by calculating mechanical average power, standard deviation of mechanical power and lost energy due to required torque. Those parameters are shown in Table 2. The QL routine is producing less mechanical power and at the same time is losing less energy according with equations (6-8) previously presented. These results are consistent with the torque and velocity reduction shown above.

7 Conclusions

A dynamic model for a planar 4 link 3 joint robot is used with a Q-learning algorithm to learn how to stand up while reducing power consumption. Good quantitative and qualitative results are shown. The electrical power consumption was reduced for the ankle and hip joints, while the knee joint power consumption increased slightly. Mechanical energy loss is shown using different performance metrics as average mechanical power, standard deviation of mechanical power and energy lost due to required torque. By using new trajectories found by Q-learning for each joint, 28% of the electrical input power is saved for a single standing up routine. The Q-Learning strategy showed a better placement of the CoM over the ankle joint, greatly reducing the torque applied to it. In addition, a better management of inertia was observed, as the Q-Learning routine performed higher accelerations at the initial phases of the routine, lowering the torque required by both the hip and ankle joints later on. Additionally, a learning simulation platform was developed by integrating a motor model, dynamic and kinetic model of robot with a Q-Learning algorithm. Future work includes the use of this platform to optimize power consumption on more complex movements such as walking or jumping, which would increase the learning problem dimensionality.

References

1. M. Gonzalez-Fierro, C. Balaguer, N. Swann, and T. Nanayakkara, “A humanoid robot standing up through learning from demonstration using a multimodal re-

ward function," Humanoid Robots (Humanoids), 2013 13th IEEE-RAS International Conference on IEEE, 2013, pp.74,79
2. M. Mistry, A. Murai, K. Yamane, and J. Hodgins, "Sit-to-stand task on a humanoid robot from human demonstration, in Humanoid Robots (Humanoids), 2010 10th IEEE-RAS In-ternational Conference on. IEEE, 2010, pp. 218223
3. J. Morimoto and K. Doya, Acquisition of stand-up behavior by a real robot using hierar-chical reinforcement learning, Robotics and Autonomous Systems, 2001, vol. 36, no. 1, pp. 3751
4. Fuminori Yamasakitt, Ken Endot, Hiroaki Kitanots and Minoru Asada, "Acquisition of Humanoid Walking Motion Using Genetic Algorithm - Considering Characteristics of Servo Modules", Proceedings of the 2002 IEEE, International Conference on Robotics 8 Automation, Washington, DC, 2002
5. Xu-Sheng Lei, Jing Pan, Jian-Bo Su, "Humanoid Robot Locomotion", Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 2005
6. J. Kober, J. A. Bagnell, and J. Peters, "Reinforcement Learning in Robotics: A Survey, Int. J. Robot. Res., 2013
7. R. Tedrake, T. W. Zhang, and H. S. Seung, "Learning to walk in 20 minutes, in Proceedings of the Fourteenth Yale Workshop on Adaptive and Learning Systems, vol. 95585, 2005
8. G. Endo, J. Morimoto, T. Matsubara, J. Nakanishi, and G. Cheng, "Learning CPG-based Biped Locomotion with a Policy Gradient Method: Application to a Humanoid Robot, Int. J. Robot. Res. 2008, vol. 27, no. 2, pp. 213228
9. T. Geng, B. Porr, and F. Wrgtter, "Fast biped walking with a reflexive controller and real-time policy searching, in Advances in Neural Information Processing Systems, 2005, pp. 427434
10. E. C. Whitman and C. G. Atkeson, "Control of Instantaneously Coupled Systems applied to humanoid walking, in 2010 10th IEEE-RAS International Conference on Humanoid Robots (Humanoids), 2010, pp. 210217
11. T. H. and M. Q. and P. Stone, "Generalized Model Learning for Reinforcement Learning on a Humanoid Robot, International Conference on Robotics and Automation, 2010
12. S. Kuindersma, R. Grupen, and A. Barto, "Learning dynamic arm motions for postural recovery, in 2011 11th IEEE-RAS International Conference on Humanoid Robots (Hu-manoids), 2011, pp. 712
13. Juan M. Calderon, Ercan Elibol, Wilfrido Moreno, Alfredo Weitzenfeld, "Current Usage Reduction Through Stiffness Control in Humanoid Robot" in 8th Workshop on Humanoid Soccer Robots, IEEE-RAS International Conference on Humanoid Robots, 2013
14. L. Ljung. "System Identication - Theory for the User". Prentice-Hall, Upper Saddle River, N.J., 2nd edition, 1999.
15. R. S. Sutton and A. G. Barto, "Reinforcement learning: an introduction." Cambridge, Mass: MIT Press, 1998.
16. F. M. Silva and J. A. T. Machado, "Energy analysis during biped walking, in Proc. IEEE Int. Conf. Robot. Autom., 1999, vol. 14, pp. 5964
17. Ercan Elibol, Juan M. Calderon, Alfredo Weitzenfeld, "Optimizing energy usage through variable joint stiffness control during humanoid robot walking" RoboCup 2013: Robot Soccer World Cup XVII . Holland, 2013.