Can Active Impedance Protect Robots from Landing Impact?

Houman Dallali, Petar Kormushev, Nikos G. Tsagarakis and Darwin G. Caldwell

Abstract— **This paper studies the effect of passive and active impedance for protecting jumping robots from landing impacts. The theory of force transmissibility is used for selecting the passive impedance of the system to minimize the shock propagation. The active impedance is regulated online by a joint-level controller. On top of this controller, a reflex-based leg retraction scheme is implemented which is optimized using direct policy search reinforcement learning based on particle filtering. Experiments are conducted both in simulation and on a real-world hopping leg. We show that although the impact dynamics is fast, the addition of passive impedance provides enough time for the active impedance controller to react to the impact and protect the robot from damage.**

I. INTRODUCTION

Humanoid robots are required to perform various dynamic and explosive motions such as jumping. One of the interesting challenges in jumping is to have a safe landing phase. In this work we discuss various aspects of landing, and propose solutions for better dealing with the impacts.

The issue of safety during an impact can be approached from two points of view. The first is human safety, that is when the robot comes in contact with the human it must have low inertia to avoid injuring the humans [1]. It was shown that only link velocity and inertia-mass will dominate the impact with humans. The overall impact happens in 10 milliseconds or less. The second aspect of impacts is the robot safety. Most humanoid robots developed so far are quite fragile and can not have large impacts with the environment without breaking a part, such as the actuator transmission. This aspect of robot safety against impacts is dealt with from the hardware or software point of view. When considering robot safety the landing duration is about 200 millisecond as shown in [2], giving the time for active control to play a role in safety.

Introduction of elastic elements in the robots' actuation system is one of the hardware design solutions to protect the robots against impact [3], [4], [5]. In [6] it was shown how change of the passive stiffness of the actuator can improve the shock absorption properties of a bipedal robot. By using softer springs the robot was dropped from 25 cm instead of 5 cm without reaching critical torque peaks at the transmission. In [7] pneumatic actuators were used to provide a soft actuation system for a bipedal robot. This robot was dropped from 1 meter without breaking any parts on the robot.

The authors are with the Department of Advanced Robotics, (Fondazione) Istituto Italiano di Tecnologia, via Morego, 30, 16163 Genova, Italy. {Houman.Dallali; Petar.Kormushev; Nikos.Tsagarakis; Darwin.Caldwell}@iit.it

It is well known that stiff robots have good position control ability but they are quite fragile when it comes to interaction with environment, in particular in explosive motions such as jumping. On the other hand, robots with low stiffness are good for absorbing impacts but bad for position control.

From the software design point of view, [8] used a leg retraction method to protect the robot in simulation. It was shown that a rigid robot such as HRP-2 can be dropped from a 10 cm height. In addition, other robots such as ASIMO use leg retraction when jumping to reduce landing impacts. In both mentioned cases, the robots are rigid.

In this paper, we investigate the effects of rigid and compliant joints, sensor and actuator delays and landing prediction on the control software used for safe robot landing. We also investigate the use of reflexive leg retraction before impact and propose a reinforcement learning approach to optimize the reflex parameters.

Machine learning methods have been used before in the legged locomotion literature. In [9] reinforcement learning was used for obtaining energy efficient high jumping heights, while in [10] it was used for walking energy efficiency.

In bio-mechanical studies, it was also found that humans soften their joints while landing which can be used in robots for safe landing [2]. Also in [11] the resonance of elastic joints was used for bipedal jumping.

Another aspect of impact landing is the energetics. In [6], the energy stored and released in the series springs were monitored to avoid bouncing on the ground after impact. However, in this work energy storage is not the main focus and the energetics of impact are considered only for choosing the suitable values for the spring, mainly for robots' safety. The main novelty of this work is to show how active control can be used to increase the safe landing height of humanoid robots with fixed passive stiffness.

In this paper, we would like to answer the following questions. Why active control is interesting for impact absorption in robots? How can prediction be used for reducing impacts? How fixed elasticity can be chosen given the weight and the impact tolerance requirements of the robot. The dynamic model of a robotic leg is used for studying the effect of active and passive impedance on the safely landing, and protecting the robot.

This paper is organized as follows. Section II describes the leg prototype used in this research and it's dynamic model. Simulation results based on learned reflexes are presented in section V. The experimental results are presented in section VI and the conclusions are drawn in section VII.

II. Leg Model

In this section, the mechatronics of the newly developed jumping robot is described, followed by the developed dynamic model.

A. Mechatronics

The robot system used in this work is the single degree of freedom leg robot [12] shown in Fig. 1. It consists of a knee joint that is driven by motors M_1 and M_2. The first actuator M_1, is a high performance motor that provides power to the knee through a series elastic transmission. M_1 is placed close to the hip level to reduce the overall leg inertia and the center of mass distance. This helps to minimize the power requirements during dynamic motions with high acceleration profiles. The second actuator M_2 is also placed at the hip level and actuates the knee joint through a uni-directional acting elastic element that is realized using a rubber type elastic transmission (bungee cord). Using motor M_2 the pretension of the elastic element can be modulated permitting the generation of additional torque in the knee joint.

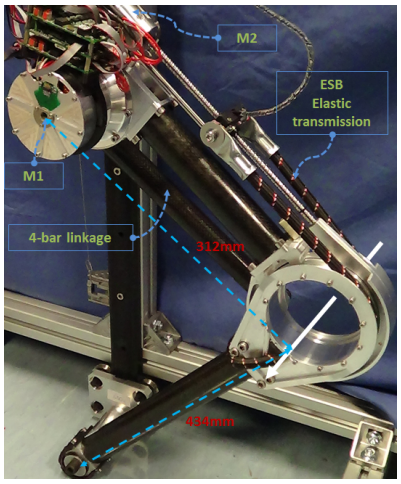

Fig. 1: Physical prototype of the joint

In terms of sensors, M_1 is equipped with three position sensors allowing full state feedback control of the actuator. In detail, an incremental 12-bit optical encoder is mounted at the motor side before gear box, one relative 19-bit magnetic encoder monitors the position after the harmonic reduction drive while an additional 19-bit absolute encoder measures the angle of the link after the flexible torsion bar. The M_1 torque is monitored through the measurement of the deflection of the torsional bar, which has stiffness of 940 Nm/rad.

In terms of actuator speed, the electrical time constant of the brushless DC motor is $\frac{L}{R} = \frac{0.9\times10^{-3}}{1.22} = 0.74\ ms$ and the mechanical time constant of the motor plus the gearbox is $\frac{J_m R}{K_t^2} = \frac{4.72\times10^{-5}\times1.22}{0.0855^2} = 7.87\ ms$. This time is less than the impact duration of 200 milliseconds which confirms that the motor can physically speed up faster than the impact propagation. The motor parameters are: L is the inductance, R is the resistance, J_m is the combined rotor and gearbox inertia, and K_t is the torque constant.

The control system update rate is $1\ kHz$, using Ethernet communication between the motor drivers and the central PC. Impedance control is used for position tracking of the joints, which is implemented using an inner PI torque loop and an outer PD position loop which sets the torque reference based on the input position command for the inner torque loop as shown in Fig. 2.

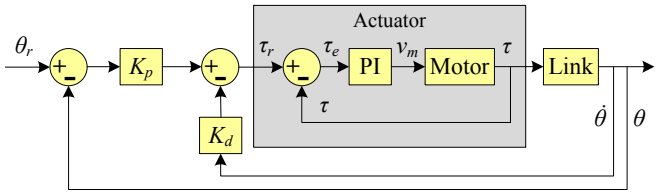

Fig. 2: Impedance Control Diagram

B. Dynamic Model

The dynamic model is developed using a vertical floating base attached to the top slider mass, and three joints of the leg. The dynamic equation is given as

$$M(\boldsymbol{\theta})\ddot{\boldsymbol{\theta}} + C(\dot{\boldsymbol{\theta}}, \boldsymbol{\theta}) + B\dot{\boldsymbol{\theta}} + G(\boldsymbol{\theta}) = S\,\tau + \tau_P + J^T\boldsymbol{\lambda}, \quad (1)$$

$$J_m\ddot{\theta}_m + B_m\dot{\theta}_m + N^{-1}\tau = K_t i, \quad (2)$$

$$L\dot{i}_m + Ri_m = v_m - K_\omega\dot{\theta}_m, \quad (3)$$

$$\tau = K_h\left(N^{-1}\theta_m - \theta\right) \quad (4)$$

where $M(\boldsymbol{\theta})$ is the mass inertia matrix, $C(\dot{\boldsymbol{\theta}}, \boldsymbol{\theta})$ is the Coriolis, $\boldsymbol{\theta}$, θ_m, $\dot{\boldsymbol{\theta}}$, $\dot{\theta}_m$ are the angular positions and velocities of the load and motor in relative coordinates; $\boldsymbol{\theta}$ is a vector of floating base position along z-axis, unactuated hip position and actuated knee position; $S = [0, 0, 1]^T$, τ denotes the knee joint torque; i_m and v_m are the motors current and voltage, respectively. We assume that the knee joint torque τ, the load and the motor positions are measured. Equation (1) models the brushless DC motor dynamics including the current dynamics. In (1) and (2) B, B_m and G are viscous damping on the link, motor damping and gravity terms. J is the Jacobian from the inertial frame to the end-effector and $\boldsymbol{\lambda}_{2\times1}$ is the ground reaction force (GRF) vector. N is the gear ratio, K_ω is the back EMF constant and K_h is the combined transmission stiffness. τ_P is the bungee torque applied to the knee joint which is zero in this study (the bungee is relaxed).

The model is developed in the Robotran software [13] using symbolic equations in C language and then compiled as a binary executable to allow fast simulations to speed up the reinforcement learning, described in section IV. The execution time of one full dynamic simulation is more than 10 times faster than real-time speed.

III. Impact Transmissibility

In this section, we use the force transmissibility ideas from vibration analysis [14] and apply them to a one degree of freedom mass-spring-damper model to draw some conclusion regarding the choice of spring and dampers for the prototype leg. Consider the diagram shown in Fig. 5, where

the robot mass m is connected to a spring k and damper b. The excitation force $f(t)$ is applied on the mass and the transmitted force f_s is derived after the spring-damper (e.g. the force felt by the impact). The force transmissibility transfer function is defined in (5).

$$T(s) = \frac{F_s(s)}{F(s)} = \frac{bs + k}{ms^2 + bs + k} \tag{5}$$

The effect of stiffness and damping on the force transmissibility can be studied using the bode diagram of (5). The effect of increasing stiffness is shown in Fig. 3 for fixed damping ratio of $\zeta = 0.2$. It is shown that increasing the stiffness

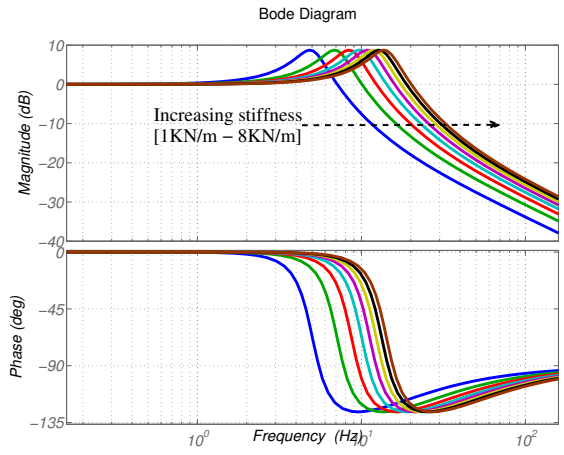

Fig. 3: Bode diagram of $T(s)$ when varying the stiffness, with leg mass of $m = 1\ kg$, damping ratio of $\zeta = 0.2$.

increases the natural frequency of the system $\omega_n = \sqrt{\frac{k}{m}}$ and increases the force transmissibility magnitude in high frequencies. Also in terms of phase, which is often neglected in analysis one can see that the amount of phase lag is decreasing as the stiffness increases. This means that the control system has more time for reacting when using softer springs than stiffer spring. Quantitatively speaking, for the first graph of stiffness (in dark blue, $k_1 = 1000\ Nm/rad$) the amount of phase lag at 10 Hz is about $\phi = 120$ degrees which corresponds to $\Delta t = \frac{120}{360} = 333\ ms$. That is the active controller with a suitable bandwidth can modify the stiffness suitably to lower values, providing a protection mode.

Also the effect of increased damping ratio ζ from 0.1 to 2, for fixed stiffness of $k = 2000\ Nm/rad$ is shown in Fig. 4. It is shown that increasing the damping from under-damped case to critical damped and then the over-damped has reduced the magnitude of the resonant peak in the force transmissibility. In terms of phase lag, increased damping also reduces the phase lag. These graphs can be used to decide on the values of the spring and damper for safe landing. In practice we can modify the active damping part of the leg to better provide the protection control mode against impacts while the fixed physical damping is provided by the torsional bar as explained in the mechatronics of the prototype leg.

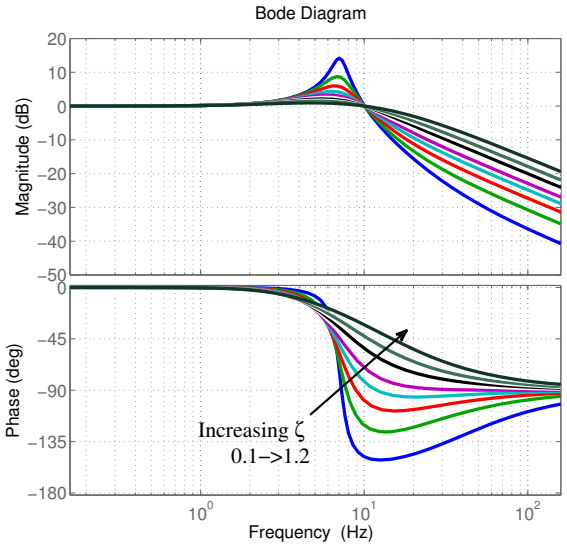

Fig. 4: Bode diagram of $T(s)$ when varying the damping ratio, for stiffness of 2 kN/m.

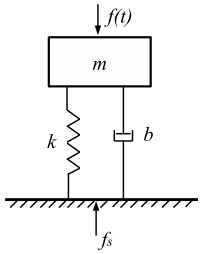

Fig. 5: Single degree of freedom system. The mass is considered to be the leg shank $m = 1\ kg$ and the mechanical ground is the motor holding it's position. The impact is $f(t)$ applied at the shank and the transmitted value is $f_s(t)$.

Moreover, from a leg drop simulation, as shown in Fig. 8, we can derive the frequency power spectrum of the impact signal, to look for which frequencies are excited when the impact occurs. Then based on the frequency range appropriate stiffness and damping can be selected. In simulations, frequencies up to 30 Hz are excited by the impact. This section provides an analytical method for designing the passive impedance given the weight and the impact tolerance requirements of the robot. Tuning the elasticity of the robotic leg is studied in [15].

IV. Reinforcement Learning

This section gives a brief description of the optimization method used, namely, reinforcement learning. It is structured in two parts: introduction of the learning algorithm, and explanation of the parameterization used for representing the policy.

A. Learning Algorithm

Reinforcement learning (RL) is a field that is progressively becoming mature [16]. Among the many potential applications for RL, the field of robotics stands out in particular,

due to its numerous challenges that RL could help solve, such as learning motor skills for difficult-to-control complex systems [17].

Some of the most commonly used RL algorithms in robotics nowadays are direct policy search methods, which are similar to black-box optimization algorithms. They avoid the curse of dimensionality of the state space by directly searching for an optimal policy in a lower-dimensional parameterized policy space. Examples for such approaches are the PI^2 and PoWER algorithms [18]. These methods can be characterized as being "local search" methods, because they start from an initial policy and converge to a single locally-optimal policy. However, in many applications of RL, including the one presented here, it is important to explore broadly the policy space and find multiple near-optimal policies [19]. This valuable information helps to discover the structure of the policy space and to analyze the properties of the optimal policies. In addition, it allows multi-objective and decision-making approaches to be used for choosing the final optimal policy parameters among all other discovered optimal policies in the Pareto front. For these reasons, in this study we selected a direct policy search method that has the capabilities of a "global search" method, namely, Reinforcement Learning based on Particle Filtering (RLPF) [20].

B. Policy Parameterization

A major advantage of any direct policy search method, and RLPF in particular, is that it allows for a natural integration of existing prior expert knowledge. This is done through the structure and the initialization of the policy. In our experiment, each policy generates a different pulse-shaped reflex signal. The parameterized policy uses three scalar parameters t_1, t_2 and θ_r which parameterize the reflex, where t_1 is the starting time of the reflex, t_2 is the ending time, and θ_r is the magnitude of the reflex as shown in Fig. 6

V. Simulation Results

In the dynamic simulations, an input position reference (reflexive retraction) is generated by the learning algorithm which is then passed to the impedance controller, as shown in Fig. 6. The learning objective is to minimize the transmitted impact to the gearbox, by changing the input reference. As shown in Fig. 6 the timing and size of these pulses are varied in hundreds of simulations (trials) to find the best reflex to achieve this objective.

In this work, ground clearance (or drop height) is referred to the hight of the robot's end effector from the ground. In each simulation, the robot is maintaining a 90 deg knee angle, using the impedance controller. The ground contact is modelled with a linear spring damper with stiffness of $K = 30\ kN/m$, Damping of $D = 1\ kN.s/m$ and cofefficent of friction of $\mu = 0.9$. The PI torque controller used is $v_m = k_p\ \tau_e(t) + k_i \int_0^t \tau_e(t).dt$, where $k_p = 1,\ k_i = 2.5$, and $\tau_e(t)$ is shown in Fig. 2. The initial ground clearance of the robot is chosen as 0.2 meter. The total mass of the robot is 12.68 kg (124.4 N), which is the steady state value during standing

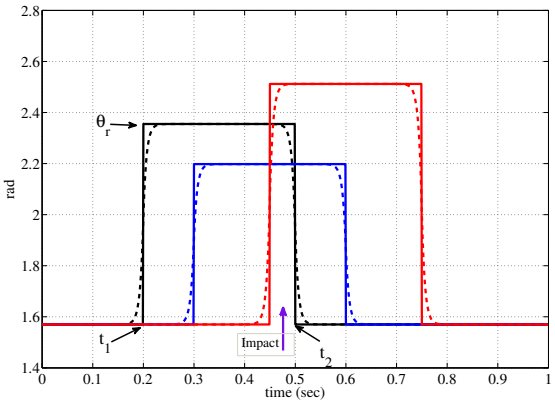

Fig. 6: Retraction reflexes generated by the learning algorithm. The figure shows three different examples generated from the proposed policy parameterization. The indicated time of impact is just for illustration and in simulation it is not known. The ideal pulses are shown with solid lines which are smoothed using a first order low pass filter as shown with dashed lines.

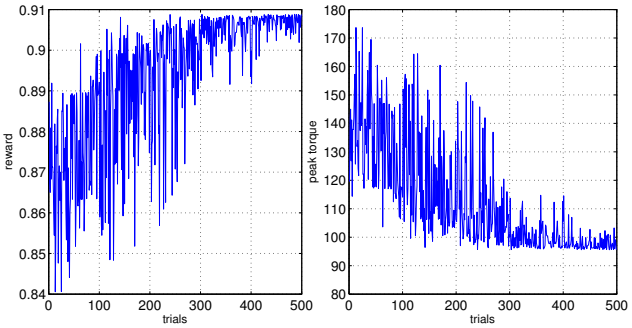

Fig. 7: The learning convergence in the rigid joint case.

on the ground. The peak impact force for the given drop height can reach 1600 N. Moreover, the landing time can be predicted from the initial drop height using $x_0 = \frac{1}{2}gt^2$, where $g = 9.81$ is the gravitational constant and the time can be obtained as $t = \sqrt{\frac{2x_0}{g}}$, i.e. if $x_0 = 0.1$, then $t = 0.14\ sec.$ In other words if the robot has a height estimation sensor with an accuracy of $\pm 0.01\ m$ (which is specification of common laser sensors), this results in an error of $\sqrt{\frac{0.22}{9.81}} - \sqrt{\frac{0.2}{9.81}} = 7\ ms$ in estimating the impact time. Moreover, the velocity at the time of impact can be derived as $v_i = gt_i$, where t_i is the estimated impact time.

The learning algorithm is run for 500 trials, although the learning converges much earlier to the optimum values. The reward function used for the learning was chosen as:

$$r = e^{-c.\tau^*},$$

where τ^* is the maximum joint torque during landing and $c = 0.001$ is a scaling factor. The evolution of the reward over the trials is shown in Fig. 7.

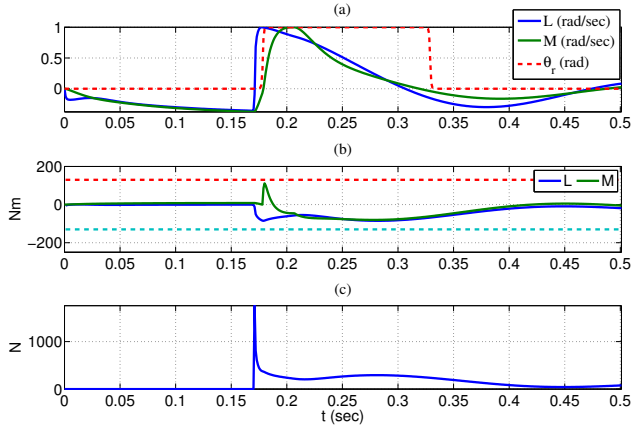

Fig. 8: The results of leg simulation with actual compliant joint including: (a) Normalized reflex produced by the learning overlapped with the motor and link velocities, (b) Motor and Link torques (solid lines), with safety regions (dashed lines), and (c) the ground reaction force.

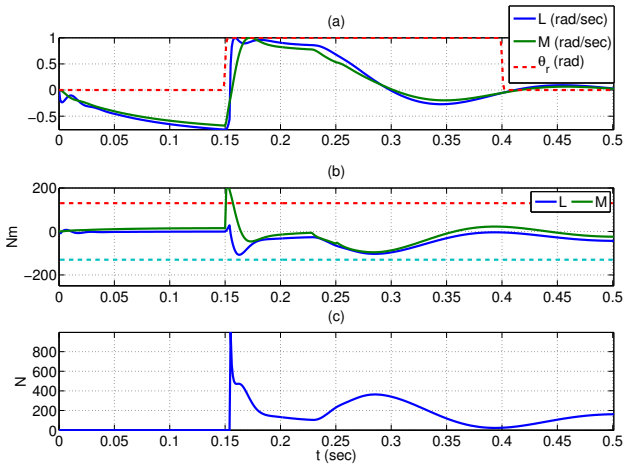

Fig. 9: The results of leg simulation with rigid joint including: (a) Normalized reflex produced by the learning overlapped with the motor and link velocities, (b) Motor and Link torques (solid lines), with safety regions (dashed lines), and (c) the ground reaction force.

A. The Compliant Joint Case

In this case, the real value of the passive stiffness (930 Nm/rad) from the robotic leg is used. The active joint impedance values (as shown in Fig. 2) were $K_p = 1000\ Nm/rad$ and $K_d = 10\ Nm.s/rad$, chosen comparable to the series passive stiffness. After 500 iterations, the best reflex value is derived and shown in Fig. 8. In part (a) of the figure the reflex is shown which is a retraction (i.e. a pulse with $\theta_r > 0$). Part (a) also shows the normalized link and motor velocities together with the normalized reflex shape. The robot initial posture is chosen as 90 deg. This is subtracted from the reflex in Figs. 8a and 9a for better illustration. Also the section (b) of the figure shows the joint torque (in blue) and motor input torque (in green) which are kept within the safe area shown with dashed lines as upper and lower threshold which should not be exceeded. Section (c) of this figure shows the ground reaction force during landing. In the case of compliant joint robots it can be seen that the maximum torque is not necessarily at the impact point but at the deceleration of the whole body, which is the part that the bungee can help and provide a safe landing and easier deceleration.

In other words, active compliance by itself (without passive compliance) cannot significantly reduce the landing impacts unless the impact time can be predicted in advance, which is difficult in practice. Only in combination with passive compliance, the active impedance control has enough time to react and reduce the impact propagation to the transmission.

B. The Rigid Joint Case

In this case, the joint stiffness is increased form 930 Nm/rad to 6000 Nm/rad which is close to the harmonic drive stiffness without series passive element. The active joint impedance values were chosen as $K_p = 2000\ Nm/rad$ and $K_d = 10\ Nm.s/rad$. The best reflex value is shown in Fig. 9 where the reflex is a retraction and happens just before impact (prediction is needed). Moreover, the convergence of the learning is shown in Fig. 7. It is interesting to compare the ground reaction forces between the compliant and rigid cases. The compliant case has minimized the joint torque more than the rigid case while the ground reaction force is larger. Hence, depending on the joint stiffness reducing the impact at the foot or at the joint can be related or not completely related. The robot with rigid tranmission has low GRF and low transmitted impact as was studied in [8] for HRP robot, while a simulation with soft transmission shows higher GRF, while lower transmitted impact. The results presented in [6] also confirms this, where dropping a robot with soft passive stiffness experience large ground reaction forces while the drives are fully protected.

Another interesting point is depending on the joint stiffness, the motion reflex can be applied before or after the impact making a significant difference in terms of the sensory system of the robot, delays and the need for prediction.

Having done the simulation studies on the dynamic model of the leg, the next section presents experimental results on impact test characterization and time scales requirements.

VI. Experiments

This section explains the impact experiment results on the real platform. Several drop tests were done on the prototype leg to characterize the timings during the impact. Fig. 10 illustrates the normalized knee link velocity and the joint torque before and after the impacts. The robot is dropped from 10 cm above the ground on a soft rubber pad. The accompanying video shows the leg operation with slow and fast motions as well as drop tests.

It is shown that there is 125 ms time before the impact (time difference between the black and the red diamonds).

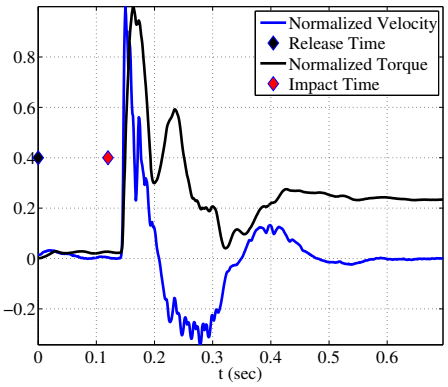

Fig. 10: Detection of impact dropping the robot from 10 cm. The velocity and the torque are measured at the knee joint.

This time computed from the free fall formula $t = \sqrt{\frac{2x_0}{g}} = 143\ ms$, but due to some uncertainties, the impact has happened a few milliseconds earlier. Synchronized audio signal (sampled at $44\ kHz$) was used in the experiment to detect the impact before the force is transmitted to the knee and the impact is felt at the knee joint. This is shown in Fig. 10 with the red diamond, marking the impact time. The audio signal has detected the event $25\ ms$ before it is sensed by the joint torque sensor. This is particularly useful when landing on uncertain environments, such as rough terrain. Also thanks to the passive compliance of the leg and the simulation results, the reflex can start just after the impact which confirms the feasibility of the proposed method. In case of more certain environments or rigid joints, learning based prediction can be combined with this sensor information to realize higher jumps with less damage to the robot.

VII. CONCLUSIONS

In this work we studied the role of active and passive compliance to protect the robot against landing shocks. we showed that when the tranmission is rigid or has high stiffness, active control cannot reduce the impacts unless impact time is predicted. In addition, if a suitable soft transmission is chosen, the active controller will have enough time to reduce the impact effect on robot's sensitive elements such has harmonic drives. In terms of the choice for the passive elements, impact transmisibility graphs were presented to show the stiffness and damping effect on the transmitted impacts.

It was shown that active impedance control has good features for impact absorption on robots leg with suitable passive stiffness.

Future work will focus on real-world experimental validation of the proposed reflexive control on the robotic leg.

ACKNOWLEDGMENT

This work is supported by the WALK-MAN European Commission projects (FP7-ICT-2013-10).

REFERENCES

[1] S. Haddadin, a. Albu-Schaffer, and G. Hirzinger, "Requirements for Safe Robots: Measurements, Analysis and New Insights," *The International Journal of Robotics Research*, vol. 28, no. 11-12, pp. 1507–1527, Aug. 2009.

[2] A. Lees, "Methods of impact absorption when landing from a jump," *Engineering in Medicine*, vol. 10, no. 4, pp. 207–211, 1981.

[3] D. Hobbelen, T. de Boer, and M. Wisse, "System overview of bipedal robots flame and tulip: Tailor-made for limit cycle walking," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2008, pp. 2486–2491.

[4] K. Radkhah, T. Lens, and O. V. Stryk, "Detailed Dynamics Modeling of BioBiped's Monoarticular and Biarticular Tendon-Driven Actuation System," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012, pp. 4243–4250.

[5] J. W. Hurst, J. E. Chestnutt, and A. A. Rizzi, "The actuator with mechanically adjustable series compliance," *IEEE Trans. On Robotics*, vol. 26, no. 4, pp. 597–606, 2010.

[6] K. Radkhah and O. V. Stryk, "A study of the passive rebound behavior of bipedal robots with stiff and different types of elastic actuation," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 5095–5102.

[7] R. Niiyama, Y. Kuniyoshi, L. Robot, and E. Movements, "Design of a Musculoskeletal Athlete Robot : A Biomechanical Approach," in *Proceedings of the 12th International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines*, 2009, pp. 1–8.

[8] S. Sakka, "Motion Pattern for the Landing Phase of a Vertical Jump for Humanoid Robots," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006, pp. 5477–5483.

[9] P. Fankhauser, M. Hutter, C. Gehring, M. Bloesch, M. a. Hoepflinger, and R. Siegwart, "Reinforcement learning of single legged locomotion," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nov. 2013, pp. 188–193.

[10] P. Kormushev, B. Ugurlu, S. Calinon, N. G. Tsagarakis, and D. G. Caldwell, "Bipedal walking energy minimization by reinforcement learning with evolving policy parameterization," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2011, pp. 318–324.

[11] B. Ugurlu, J. A. Saglia, N. G. Tsagarakis, and D. G. Caldwell, "Hopping at the resonance frequency: A trajectory generation technique for bipedal robots with elastic joints," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2012, pp. 1436–1443.

[12] N. G. Tsagarakis, S. Morfey, H. Dallali, G. A. Medrano-Cerda, and D. G. Caldwell, "An asymmetric compliant antagonistic joint design for high performance mobility," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2013, pp. 5512–5517.

[13] J. C. Samin and P. Fisette, *Symbolic modeling of multibody systems*. Springer, 2004, vol. 112.

[14] C. W. De Silva, *Mechatronics: an integrated approach*. CRC press, 2004.

[15] N. G. Tsagarakis, H. Dallali, F. Negrello, G. A. Medrano-Cerda, and D. G. Caldwell, "Compliant antagonistic joint tuning for gravitational load cancellation and improved efficient mobility," in *IEEE International Conference on Humanoids*, 2014.

[16] M. Wiering and M. van Otterlo, *Reinforcement Learning: State-of-the-art*. Springer, 2012.

[17] P. Kormushev, S. Calinon, and D. G. Caldwell, "Reinforcement Learning in Robotics: Applications and Real-World Challenges," *Robotics*, vol. 2, no. 3, pp. 122–148, 2013.

[18] J. Kober, A. Bagnell, and J. Peters, "Reinforcement learning in robotics: A survey," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1238–1274, 2013.

[19] P. Kormushev and D. G. Caldwell, "Simultaneous discovery of multiple alternative optimal policies by reinforcement learning," in *6th IEEE International Conference Intelligent Systems (IS)*, 2012, pp. 202–207.

[20] P. Kormushev and D. G. Caldwell, "Direct policy search reinforcement learning based on particle filtering," in *The 10th European Workshop on Reinforcement Learning (EWRL 2012)*, Edinburgh, UK, Jun. 2012.