Robotics and Autonomous Systems 164 (2023) 104400

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

Roboception and adaptation in a cognitive robot

Agnese Augello [a], Salvatore Gaglio [a,b], Ignazio Infantino [a], Umberto Maniscalco [a], Giovanni Pilato [a], Filippo Vella [a,*]

[a] High Performance Computing and Networking (ICAR) of National Research Council of Italy (CNR), via Ugo La Malfa 153, Palermo, 90146, Italy
[b] DIID, Department of Engineering, University of Palermo, Viale delle Scienze, ed 6., Palermo, 90128, Italy

ARTICLE INFO

Article history:
Received 25 April 2022
Received in revised form 3 March 2023
Accepted 5 March 2023
Available online 11 March 2023

Keywords:
Somatosensory system
Soft sensors
Cognitive architecture
Humanoid robot
Roboception
Reinforcement learning

ABSTRACT

In robotics, perception is usually oriented at understanding what is happening in the external world, while few works pay attention to what is occurring in the robot's body. In this work, we propose an artificial somatosensory system, embedded in a cognitive architecture, that enables a robot to perceive the sensations from its embodiment while executing a task. We called these perceptions *roboceptions*, and they let the robot act according to its own physical needs in addition to the task demands. Physical information is processed by the robot to behave in a balanced way, determining the most appropriate trade-off between the achievement of the task and its well being. The experiments show the integration of information from the somatosensory system and the choices that lead to the accomplishment of the task.

1. Introduction

Robots are composed of a set of different parts working together as a whole. They are usually provided with a series of sensors that may produce heterogeneous stimuli. In the perception process, it is essential to consider both the external environment's effects and its internal body conditions [1]. All living creatures have a sort of somatosensory system that defines their boundary with the external environment, helping also to monitor their body. It influences their behavior, protecting them from risks and dangerous situations. Robotic platforms typically include some self-integrity mechanisms such as avoidance collision that prevents possible bumps among body parts or automatic recharging that checks the battery level when it is below a certain threshold. This is a relevant issue in autonomous robots. They need to take proper decisions to safeguard also their integrity, in particular in situation where they cannot report immediately their status to humans. However, the signals are usually processed individually and they are used reactively. As it happens in living beings, dangerous situations usually raise an alarm "signal" that, producing, for example, a painful sensation, represents an effective penalty mechanism to modulate or avoid inappropriate actions. According to Damasio, fatigue, energy, wellness, and sickness can be considered background emotions and affective states experienced under certain body conditions [2]. Such conditions can strongly influence the possibility to perform a task and an unpleasant condition can lead to a withdrawal even if an urge requires to perform a given task [3,4].

Although it is not well determined if human beings are the best examples to shape an artificial entity's mechanism, human biology is still a great source of inspiration [5]. A robot, perceiving some sort of "artificial pain", can adapt its behavior to avoid or minimize damage to itself or to the environment [6–8]. In agreement with [5,9], we consider the robot embodiment keeping in mind the substantial differences between humans and robots. We propose a bio-inspired artificial somatosensory system that allows the robot to behave according to the perceptions of the states of its own body components. A simpler approach has been presented in [10], where preliminary results have been sketched. To underline the difference with physical sensations arising from the receptors in the human somatosensory system, we adopt here the term *roboceptions* referred to a robotic entity. In particular, the somatosensory system has been designed for the Aldebaran NAO robot, starting from the analysis of its built-in basic hardware sensors. Roboceptions are related to hardware measurements of this platform. The approach is general and it is possible to adapt it to other robotic platforms. Additional roboception can be added considering additional sensors for different platforms or even more sophisticated monitoring functions such as global well-being or motivation. In general, the somatosensory system allows the humanoid to perceive physical inputs or feedback, and to interpret them according to abstractions loosely inspired by human sensations. The robotic somatosensory system relies on a layer composed of different soft sensors, built on top of the robot's physical proprioceptive and exteroceptive

* Corresponding author.
E-mail address: filippo.vella@icar.cnr.it (F. Vella).

https://doi.org/10.1016/j.robot.2023.104400
0921-8890/

sensors [11–16]. Soft sensors, also known as virtual sensors, are mathematical models implemented as software tools and capable of calculating difficult or impossible quantities to measure. The models, that are biologically inspired, have been chosen on an empirical basis and provide an output according to a transduction function processing the hardware measurement. We consider the transduction function and the raw sensory data fusion as fully-fledged soft sensors since they estimate quantities that physical instruments cannot measure. They can evaluate quantities analogue to "muscular" pain caused by high current or the exertion caused by a rising temperature in actuators that has an effect on muscle ability to produce an effective force. They could also estimate pleasant sensations as a caress. In our framework, soft sensors emulate natural nerve cells and nerve fibers that lead to stimuli.

Both pleasant and unpleasant sensations can drive the behavior of the robot, which, paying attention to what happens in its body, will choose the most suitable modality to tackle a given task.

Let us consider, for example, a dancing robot executing an improvised set of movements such as a choreography. If such a robot is has a knowledge about its physical condition, it can adapt its dance by choosing movements from its repertoire that require less effort. Its state will then influence its dance, and the audience could perceive its sensations. The task's execution derives from balancing two concurrent parameters: the reward for the task execution and the cost, in terms of resources, required to achieve the task. An excessive resource consumption, such as energy, may produce an early stop of the task due to low battery. The somatosensory system collects multiple information coming from the robot parts, and, consequently, it can provide values that monitor its body wellness. A roboception can unveil a critical situation and let the robot choose a strategy to fulfill the task while maintaining roboception values within a wellness area.

Roboceptions provide the robot with some sort of knowledge of its own current physical condition. The robot can reallocate the resources and adapt its strategy to perform the task as properly as possible without compromising the accuracy of its performance.

To implement this mechanism, we employ a reinforcement learning approach, in particular SARSA algorithm [17,18], to properly act during a complex task, trying to take into account conflicting reward functions.

In Section 2 we describe techniques that are related to the present work. From Section 3 we describe our proposal in terms of a cognitive architecture involving a Somatosensory system, that is described in detail in Section 4. In Section 5 we show the behavior of the robot in different physical situations. It has to perform a given task while maintaining a good physical status. Finally, we draw some conclusions of this work in Section 6 and discuss the future direction of this research.

2. Related work

At present, robotics is mainly focused on analyzing the perceptions of external signals by a robot control system while few works investigate the importance of considering in robots the interactions with their internal part of the body [1]. Among these works, for example, Saegusa et al. [19] present a framework to let the robot recognizing its own body and generating proper actions accordingly. The scene's vision and proprioception, which is the robot parts' position, drives the acquisition of the knowledge of its own body. Thus, the framework starts from a basic level, and it learns an association between visual and tactile information with proprioceptive data.

In [20], a self-perception of a robotic manipulator is presented. The activities on the manipulator motors are the input signals, while the position of the links are monitored through visual markers. The connection among input signals and link position is learnt through bayesian networks with different topologies.

In [21] a robot, able to self percept its body, is used. The robot is endowed with accelerometers along its arms and can have information about the movement of its body. The sensor signals are processed with Sensory Motor Contingency [22] to connect motor signals and detected movements. A further connection is then performed between sensing and external objects.

Moreover, it is clear that a robot that always acts in the same manner and is repetitive in its behaviors, looks unattractive and trivial to the end-user after the first few interactions. Instead, if a robot acts autonomously and consciously, the interaction is more exciting and stimulating for the user. In this context, a fundamental role is played by motivation: for this reason, for example, in [23], a model was introduced that combines the presence of motivation with a Consciousness-based Architecture (CBA) in order to obtain proper and coherent changes in the behaviors of a robot, resulting in a more autonomous and effective behavior.

According to [24], different terms have been exploited to refer to the studies on using computational models to emulate an intelligent behavior in machines. A standard term used in the literature is that one of "machine consciousness". Many models of awareness and consciousness have been proposed in the literature, e.g.: [25–32]. A key role is played by sensory perception, both through proprioceptive and exteroceptive sensors, for establishing the state of self. The former sensors aim to monitor the robot's internal state, while the latter is oriented to perceive the external environment. This can lead to two kinds of attention: the inner and the sensor one [33]. In the literature, some studies have shown that pain can be related to the concept of "self-awareness", both in robots and humans. In particular, Steen and Haugli explored in [34] the correlation between musculoskeletal pain and increased awareness of self. Another study illustrated how there might be a correlation between affective self-awareness and pain [35]. The relationship between pain and self-awareness has also been reported in other more recent work, such as [36,37]. In this context, it is highlighted that pain sensation is among the most relevant aspects of self-awareness. As reported by [33], artificial pain can be computationally generated without actually being any real sensation of pain, designing only the functional aspects of pain itself.

Our work considers the use of a cognitive architecture that employs roboceptions for a better self-awareness approach. We aim to obtain higher-level information through roboceptions to modulate the robot activities in more complex tasks. We collect information about the robot body from a processing layer composed of soft sensors. Instead of binding the detected signals with actions, we investigate how the internal state can influence the deliberation of a humanoid robot, for example, how the Aldebaran NAO[1] can change how to tackle a task according to a proper cognitive architecture [38].

Our contribution can be also related to the papers that forecast the maintenance for industrial machines under the name of predictive maintenance, for example [39,40]. This task can be synthesized as: collection of signals through sensors installed on the most critical components; signal processing techniques to extract relevant information and detection of relationships between the extracted parameters and the health condition or the Remaining Useful Life (RUL) of the analyzed components. The two approaches have many points in common; furthermore, mutual techniques can be shared among them. One of the main differences is that the maintenance task is used to assure that a component can continue its own function during time, while,

[1] https://www.ald.softbankrobotics.com

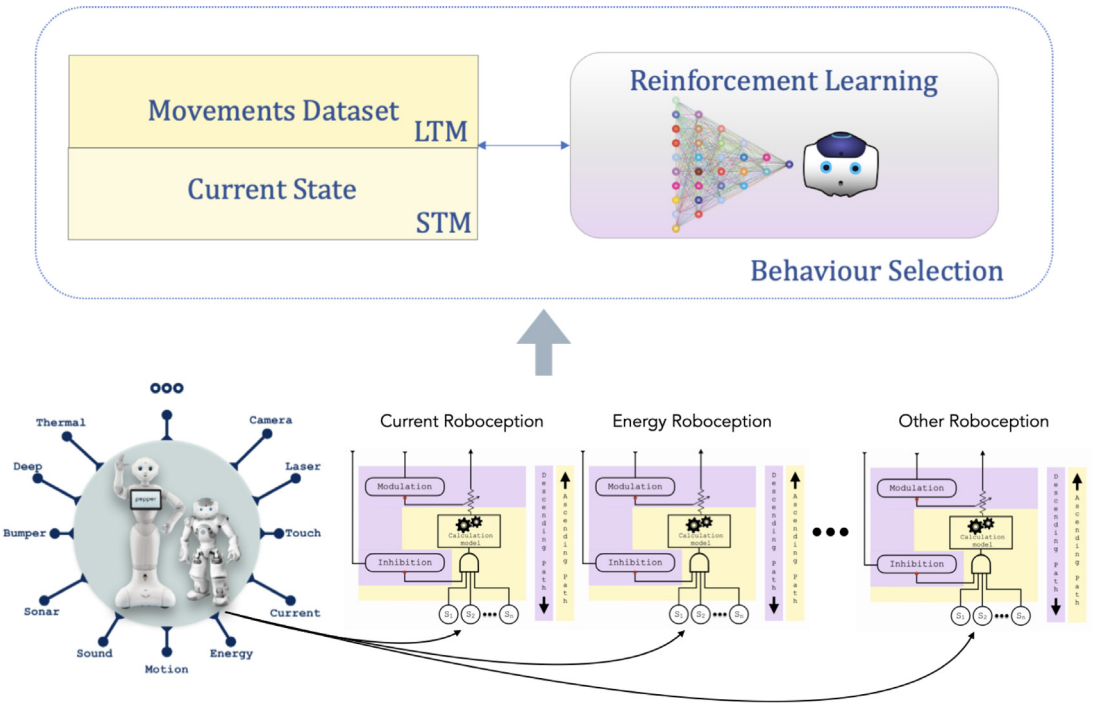


Fig. 1. PSI-inspired Architecture at the basis of the Robot's Behavior. The somatosensory system allows the robot to collect its *roboceptions*. The robot, through a RL process, learns how to choose, according to the specific situation, the modes of execution of a task.

in the proposed system, the information coming from the embodiment can drive the system actions preserving all the robot resources.

The work proposed here is the evolution of a set of contributions that have been presented in previous papers. We exploit a reinforcement learning approach to improve both self-awareness and effectiveness of the execution, even in critical situations. The robot tries to determine a trade-off between the tasks to be accomplished and the available resources.

3. PSI-inspired robot architecture

The use of Cognitive Architectures (CA) enriches the robots capabilities and enables the emulation of behaviors typical of human beings. CAs constitute the infrastructure managing the processes of perception, recognition, categorization, reasoning, planning and decision-making [41]. The different CA models, proposed in literature, attempt to establish the necessary modules to mimic the complex interactions among perception, memory, learning, planning, and action execution [42]. A cognitive activity has to be influenced not only by the input coming from the external environment but also by the stimuli in the body. Some behaviors can be triggered by physical sensations and perceptions, conveyed by the somatosensory system. In previous contributions, we have successfully exploited a cognitive architecture in human–robot interaction [38,43,44], taking inspiration from the Psi model [45,46]. In the current formalization of the robot's CA, shown in Fig. 1, we modeled the awareness of the robot of its physical condition, by introducing a process of roboception, related to the bodily dimensions which influence the robot's behavior.

A bio-inspired somatosensory system (described in Section 4), is at the basis of such roboception process. It relies on a layer composed of different soft sensors [12,13], built on top of the robot's physical sensors. A Behavior Selection module, based on a reinforcement learning process (described in Section 5), is used by the robot to choose a modality to accomplish a task. In such a process, the robot exploits the knowledge about possible movements to perform and their associated costs and rewards and keeps into account its needs (which can have a physiological, cognitive, or social nature). This information is stored in its Long Term Memory (LTM). Indeed, the activity of the robot is influenced by urges, and an urge arises when there is a considerable difference between the current need and its target value [45]. In particular, in this work we mainly focus on physiological urges arising from the discrepancy between a roboception and a desiderata physical condition. The behavior is then selected by considering the current situation, stored in the Short Term Memory (STM), consisting of information about the physical condition of the robot, evaluated by means of the somatosensory system, the ongoing task, the occurrence of an urge. The robot, aware of its physical state and of the costs associated with the different movements, will execute a task activating different modes according to the specific situation and can decide to stop a task if the perceived physical conditions will not allow it to complete the activity. A description of the system, that is inspired by Nilsson architectures of intelligent agents [47], is given in Fig. 2. The system is composed of perception, decision and action modules. The information, coming from sensing devices, is filtered by the somatosensory system and constitutes the input to the system. Through a learning process, here implemented with Reinforcement Learning, the actions are chosen or modulated with the aim to obtain the highest reward while preserving the well being of the robot. The green blocks in the figure represent the somatosensory system, the learning process and the modulation of the actions that are detailed in the next sections.

In the next sections, we will describe in detail the modules of the robot's architecture and the performed experiments.

4. Soft somatosensory system

The human beings' somatosensory system is a complex system of nerve cells and receptors that react to changes on the surface or in the inner part of the body. It plays a key role, driving humans

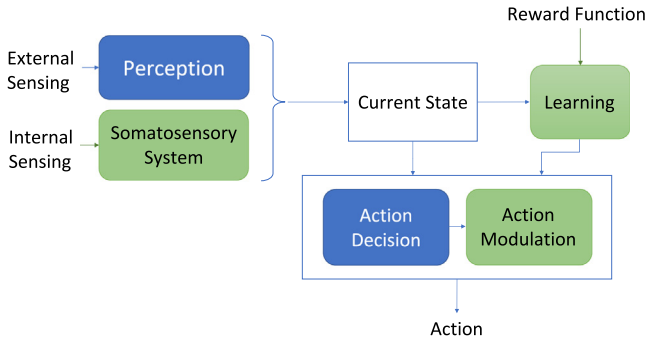


Fig. 2. The proposed architecture with the learning module that maximizes the reward function.

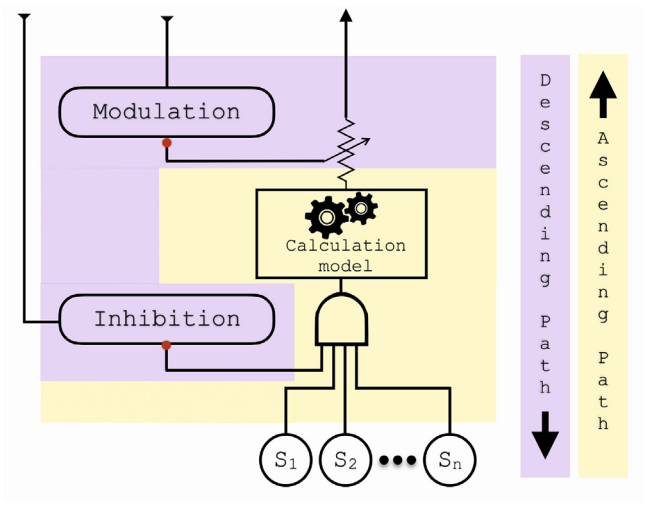


Fig. 3. The schema of a generic soft sensor in which the ascending path and the descending path are highlighted with different colors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and, in general, living creatures' behaviors and protecting them from taking not appropriate actions. On the contrary, this system can also increase the wellness level when positive stimuli are perceived, encouraging actions that could be judged as potentially risky in standard conditions. An artificial somatosensory system can replay the characteristic of the biological ones endowing the robot with concurring information flows that improve the management of the body and the execution of the tasks. With this aim, we have designed our artificial somatosensory system. We started from the built-in basic hardware sensors of a robot, in particular a consumer robot, and we defined a set of soft sensors able, albeit with simple functions, to provide information relevant for the well-being of the robot [48–52].

The soft sensors loosely reproduce and, somehow replace, the natural nerve cells, the nerve fibers that conduct stimuli and the somatosensory cortex. For each kind of sense or stimulus, we tried to emulate a biological model. Whenever it was not possible, we have used a model oriented at guaranteeing the robot's safety (e.g., avoiding the robot falling). In doing so, the robot can get its roboceptions computing the signals coming from its receptors.

Each soft sensor is responsible for the computation of a specific roboception. This approach somewhat resembles the layers of Brooks' subsumption architecture [53] where each layer implements a competence and higher levels subsume the lower ones (reading, inhibit or suppress the signals of lower layers). However, our source of inspiration was the biological somatosensory system, with its complex set of receptors, neural pathways, and inhibition mechanisms.

The generic schema of a soft sensor constituting the artificial somatosensory system is shown in Fig. 3. The sensors, referred to hardware sensors, directly correspond with the receptors of the biological model. The set of sensors depicted in the lower part of the image express that any soft sensor can use one or more sensors as an input. The soft sensor input can be homogeneous (i.e., all temperature sensors), or heterogeneous and can include different types of sub-sensors (i.e., gyroscopes, accelerometers and pressure sensors). By taking inspiration by the biological model, it was considered an ascending path that brings information from the sensors (receptors) to the computation unit (somatosensory cortex), through calculation models represented by the gears drawn in Fig. 3. At the same time, an opposite descending path implements a modulation of the perception or even a stimulus suppression. It is analogous to the production of endogenous substances or the intake of exogenous substances that, in biological systems, affect the perception process. The AND gate at the soft sensor input represents the inhibitory function, while the variable resistor at the soft sensor output represents the modulating function (see Fig. 3).

Each category of information can be either enabled or utterly disabled by using a gate port. Following this approach, we can set up the soft somatosensory system according to a specific task that the robot must execute. For some tasks, the robot can consider the stimuli coming from all categories of sensors, and in particular situations, the robot can neglect some specific category of stimuli.

As an implementation test-bench for our artificial somatosensory system, we focused on the humanoid NAO robot. However, the illustrated methodologies are not bound to a specific robot model and can be generalized to any robot. We chose the NAO robot since this robot fits particularly well to the purposes illustrated in this paper. It is equipped with many basic sensors capable of measuring different parameters, many states (or changes of state) can be identified and several events can be managed.

As shown in Fig. 4, from the embedded sensors, we can get basic information from the battery, the CPU, the force sensitive resistors, the inertial sensors, the joints, and actuators, together with LEDs, sonars, switches, touch sensors and other sensors. In our implementation we have taken into account as basic stimuli, acquired directly from the NAO embedded sensors, the value of the current and the temperatures of twenty-five actuators

Head: HeadPitch, HeadYaw.
Arms: RShoulderRoll, RShoulderPitch, RElbowYaw, RElbowRoll, RWristYaw, RHand, LShoulderRoll, LShoulderPitch, LElbowYaw, LElbowRoll, LWristYaw, LHand.
Legs: RHipPitch, RHipRoll, RKneePitch, RAnklePitch, RAnkleRoll, LHipYawPitch, LHipPitch, LHipRoll, LKneePitch, LAnklePitch, LAnkleRoll.

Other information gathered by the robot is: the pressure of four switches located at the tip of each foot; two (one for each foot) weight values based on force sensitive resistors; nine (ON/OFF) touch sensors are located on the head (*front, rear, middle*) and the hands (*back, left, right*); three distances measurements are achieved by sonars situated at the robot's chest; two pairs of values about the angles, the accelerations and the inertia of the robot along X and Y axes.

All the above mentioned primary stimuli are associated with the design of the soft sensors. In the following subsections, we will illustrate some soft sensor designed for our somatosensory system.

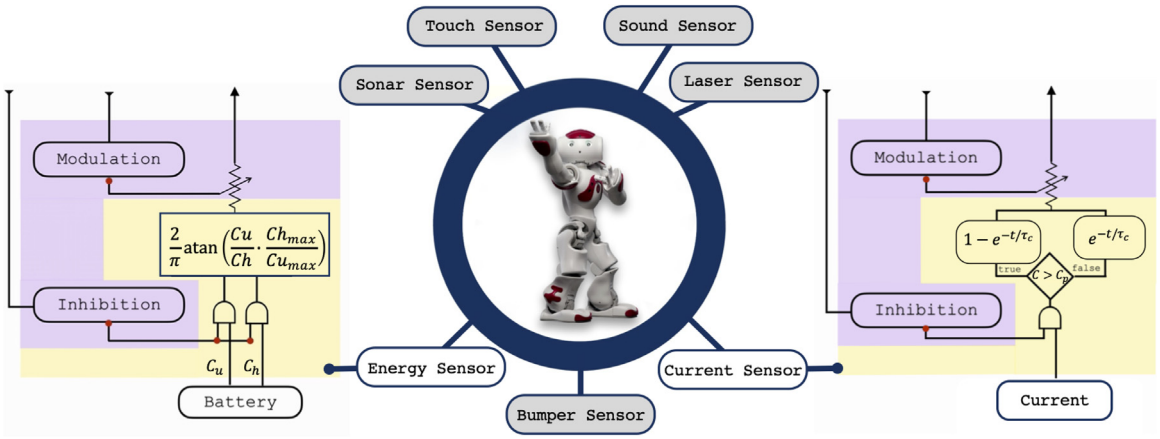


Fig. 4. The humanoid robot NAO and some of the main sensors involved in our soft somatosensory system.

4.1. Energy roboception

A fundamental capability of human beings and, in general of living beings, is the assessment of the sufficient energy quantity to complete a task. In alternative, they have to consider additional sources of energy they can employ during the task's performance.

When the physical energies run out, a discomfort is suffered and the chances of completing the task considerably decrease. Living beings, thanks to the possibility of perceiving fatigue, can organize a plan and perform their tasks by managing the energies in their possession.

If a humanoid robot had an analogous capacity to perceive fatigue, it could manage its residual energy to complete its task. Alternatively, it could rearrange the plan to maximize the result reachable with the residual energy. In other words, the knowledge of residual energy and the corresponding roboception are the essential prerequisites for modifying the robot's behaviors to maximize the desired results.

The artificial somatosensory system we have designed includes a specific soft sensor to generate a roboception bound to energy. This roboception employs some parameters linked to the battery as input. In particular, the amount of remaining battery power and the amount of instantaneous current supplied by the battery. The Energy roboception is computed in (1) as the arctangent of the ratio between the instantaneous current supplied Cu (normalized) by the battery and the instantaneous charge of the battery Ch.

$$Ex(t) = \mathcal{M}(1-\mathcal{I})\frac{2}{\pi}atan\left(\frac{Cu}{Ch}\cdot\frac{Ch_{max}}{Cu_{max}}\right) \tag{1}$$

$\mathcal{M}$ represents the modulation coefficient of the soft sensor output. $\mathcal{I}$ indicates the inhibition. If $\mathcal{I}$ is equal to one, the whole roboception is inhibited. To normalize the value of the currents and the value of the battery charge, they are divided by their maximum value. The output of the arc-tangent is brought in the range from 0 to 1, dividing by $\pi/2$. If the level of the battery is low, the argument of the arc-tangent function is very high and the result of the soft sensor is near one. If the charge of the battery is higher, the argument of the arctangent is lower. The provided current acts inversely, the lower is the current Cu, the lower is the output of the soft sensor. A higher value of the current give higher outputs. An additional sensor, indicating if the charger is plugged, can be integrated. We avoided this addition in favor of simplicity. If the charger is plugged, the hunger roboception returns the value 0. The reason is that, in this case, the charge of the battery is not relevant since a power outlet is providing all the necessary energy. Whenever the charger is not plugged into the robot, this roboception is computed, as stated before, according to (1).

4.2. Current roboception

The actuators of the NAO robot are composed of stepper motors. Each one of them is driven by a current signal with a value between zero and C_{max} value. Different actuators may have different maximum operating currents.

The operating current normally remains under the 80% of its C_{max} value, although some current spike may occur from time to time.

When a robot's movement is hampered, the currents in the involved motors increases and can exceed the operating value reaching, in some case, the C_{max} current. We consider the raising of the current over the operating value as an unusual condition that can potentially damage the robot itself and can be associated with discomfort. This condition can be considered analogous to the condition of pain in living beings, therefore, also considering its physical origin, we call this "Current Roboception" and can resemble a "Pain" sensation

This roboception can originate either in a single actuator or, simultaneously, in a set of actuators depending on the kind of the movement and if the movement is hampered or just it witnesses the perception of a large current.

We have therefore identified a stimulus, the pilot current in the actuator, and a type of receptors. The sensor model processes the stimulus and transforms it into a roboception. Considering a biological reference model, analogue to pain sensation, we considered two aspects. The first one is related to the intensity threshold that the stimulus must overcome before the living being begins to feel of pain. The other one is the exponential character of the sensation of pain that tends to saturate.

The model adopted to implement this soft sensor replicates the exponential trend and the character of saturation of this sensation. Thus, the soft sensor implements this bio-inspired feature adopting the model of charge and discharge of an RC (Resistor–Capacitor) circuit.

This type of circuit is characterized by two phases. A charging phase in which, thanks to the applied potential, the capacitor is charged through a current that flows into a resistor. The second, in which the capacitor is discharged by dissipating the charge accumulated on the resistor and transforming it into heat. Both phases show an exponential trend and they are compliant to represent the roboception of the current. This model is also characterized by a time constant τ making the charging and the discharging phase faster or slower. The value of τ allows us to create a soft sensor that is more or less reactive.

$$P_{curr}(t) = \begin{cases} \mathcal{M}(1-\mathcal{I})(1-e^{-t/\tau_c}) & if\ C > C_p \\ \mathcal{M}(1-\mathcal{I})e^{-t/\tau_c} & if\ C <= C_p \end{cases} \tag{2}$$

The current roboception is calculated as shown in (2). The $\mathcal{I}$ parameter can be 0 or 1 and it opens or closes the AND gate operating the inhibition action. The $\mathcal{M}$ parameter is in the range [0, 1] and it is represented by the variable resistor in the soft sensors figure. It performs the modulation function. The rise or the reduction of the current roboception, due to the current, are calculated as in (2). Above is calculated the roboception when the stimulus is present, the second when the stimulus is missing. The values of the pilot current in all step motors are measured by the current sensors. For each current sensor, the sampled value of current is transported through a fiber that can act as a localization function. The sampled current value is compared with a threshold value C_p, if the pilot current overcomes this value, the charging phase starts. Otherwise, a discharging phase begins. The C_p is empirically set, it is the value of current that, if applied for a long time, can cause damage to robot hardware.

The developed artificial somatosensory system is highly parameterized. A simple XML configuration file can customize the character of a robot. In this way, it is possible to let robots behave differently although it is subject to the same stimuli.

An implementation of an artificial somatosensory system for a humanoid robot [54] is ASS4HR, a software implementation for ROS (Robot Operating System). The ROS implementation of the system is extremely scalable. In fact, it is possible to add new "roboceptions" to the system, leaving the overall architecture unchanged. The software package is distributed under a free Apache License 2.0 and can be downloaded at the link https://github.com/crss-lab/ASS4HR

5. Implementation of somatosensory-aware robot behavior through reinforcement learning

A useful way to exploit the information from the somatosensory system is to evaluate the state coming from the set of the soft sensors, and perform the tasks in a way that lets the robot in a well being state. In order to estimate the best action to be undertaken, we train the robot to select the best policy through Reinforcement Learning [17]. This technique is inspired by behaviorist psychology and is based on the cumulative reward an agent collects when it acts in the environment. The model is a Markov Decision Process (MDP) that is represented as a tuple $\langle S, A, P, R\rangle$. S is the set of states, A is a finite set of actions, P is the transition probability and R is the reward [55–57].

The goal of solving the MDP is to find a policy, $\pi : S \rightarrow A$, that creates a correspondence between states and actions such that the agent can maximize the cumulative future reward selecting, time by time, the best action. If the parameters of P and R are known, then the optimal control policy, for the agent, can be efficiently determined using techniques such as value iteration.

The value of a state is evaluated according to the reward that it is possible to collect and the policy used to select an action according to its current state.

The value of the Q-function is given by solving the Bellman equation:

$$Q^{\pi}(s, a) = R(s, a) + \gamma_r \sum_{s'} P(s'|s, a) Q^{\pi}(s', \pi(s')) \tag{3}$$

where $R(s, a)$ is the reward when action a is performed in state s, $P(s'|s, a)$ is the transition probability of reaching the state s' after executing action a when the system is in state s. γ is the parameter that focuses the algorithm on short term reward (γ near zero) or to long term rewards (higher values of γ).

There are multiple algorithms that have been used to solve this problem. In this case, we use active learning since we want that the system can learn its policy with multiple trials on the field. We did not use MonteCarlo methods since we desire to improve the policy step by step, avoiding to learn at the end of each training episode. We focused on a policy gradient technique (instead of a value gradient), setting an arbitrary decision rule and letting the evolution of the algorithm adjust the most suitable decision for any state.

In particular, we applied the SARSA algorithm, that is a method with temporal differences (does not need that the training episode ends) and is "on policy". SARSA algorithm differs from Q-learning, that is instead an "off policy" because the correction of the value of the couple (state, action) is made according to the transition between s and s' with action a, chosen according to the current policy [18]. The Q-learning chooses the best policy but tends to be less robust. An example of this characteristic is shown in the Cliff walking problem [58]. The SARSA algorithm provides a good balance between exploration (evaluate the rewards from the environment) and exploitation (maximizing the reward).

Considering training episodes lasting T seconds, the value of an action, when the system is in a given state, is evaluated as:

$$\begin{aligned} Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha_t(R_{t+1} + \\ +\gamma(Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t))) \end{aligned} \tag{4}$$

the value of t is between 0 and $T - 1$. If a teacher assesses the global performance, an additional reward is assigned at time T, the evaluation of Q is, at time T as shown in (5).

$$\begin{aligned} Q(s_{T-1}, a_{T-1}) \leftarrow Q(s_{T-1}, a_{T-1}) + \alpha_T(R_T + \\ +\gamma(Q(s_T, a_T) - Q(s_{T-1}, a_{T-1}))) + \beta_T \mathcal{R}_T^M \end{aligned} \tag{5}$$

where R_T is the value of the reward as in previous instants (e.g. Table 3), while $\mathcal{R}_T^M$ is the reward considering the global evaluation added at the end of the task or of the performance. This computation is performed only when the teacher is present. The training can be carried on with rare teacher rewards while the self-improving learning is performed in all the other cases. This learning strategy resembles the learning from a teacher, when guided trials are a limited part of the global training and a conspicuous part of learning is assigned to the student, or the practitioner, that has to learn by himself. This autonomous learning, moreover, fits very well with RL technique where a reward, in the last instants of the trial, is propagated to previous steps through multiple iterations.

For the convergence of the algorithm, it has been proven that SARSA algorithm converges to the optimal action-value function $Q(s, a) \rightarrow Q^*(s, a)$ if the training is "Greedy in the Limit with Infinite Exploration" (GLIE) [59,60]. In particular, a learning policy is GLIE if it satisfies two conditions:

- all the state–action pairs are explored infinitely many times

$$\lim_{k\rightarrow\infty} N_k(s, a) = \infty \tag{6}$$

- the policies converge on a greedy policy

$$\lim_{k\rightarrow\infty} \pi_k(a|s) = 1(a = argmax_{a' \in A} Q_k(s, a')) \tag{7}$$

According to the work of Robbins–Monro, the property of being *greedy in the limit with infinite exploration* is granted if the values of α are under the conditions that the step-seizes $\alpha(r_t)$ [61]:

$$\sum_{t=1}^{\infty} \alpha_t = \infty \tag{8}$$

$$\sum_{t=1}^{\infty} \alpha_t^2 < \infty \tag{9}$$

In the case of a robot with a somatosensory system, the choice of the action can be driven by the robot's state and the collected

rewards, through learning, can drive the choice of the best action, based on the specific situation. In the following sections, we show how we modeled a choice mechanism that, according to the state, can select the best behavior to execute a task.

For example, we considered the behavior of the robot as the way it executes a task. Instead of focusing on a choice of single actions that are strongly dependent from the robot activity, we considered a set of modes that the robot can adopt to carry on its tasks.

In Section 5.1, we describe the training of the most adequate modes during the execution. Modes are composed of a set of parameters and drive the execution of operations. In Section 5.1.1, the selection of three modes is done considering the Energy soft sensor. In Section 5.1.2, the same roboception is used to trigger the robot behavior, when a cognitive urge is present. An additional experiment is described in Section 5.2, where the robot behavior is focused on the Current Roboception and, beyond to perform a sequence of movement, it chooses the mode to avoid a strong current flow. The code for the experiments described above is available under Apache License 2.0 at https://github.com/filippovella/Somatosensory_RL.

5.1. Behavior mode selection through reinforcement learning

The complexity in the management of composite hardware, such as the robot hardware, can be reduced by selecting a set of modes used to address the robot's behavior. The *action modes* are introduced to simplify the planning activities of the robot and to insert a layer aimed at synthesizing the operating modes, whose choice is indirectly influenced by the global sensations of the robot.

The modes, influenced by the bottom-up *sensations* processed by the soft sensors, are selected considering the aim and the accomplishment of the task. The modes selection can be seen as an equilibrium point between the lower level instances and the higher level demands.

We illustrate a set of experiments aimed at linking bottom-up instances with, both the environment and task information. A first experiment is related to Energy roboceptions, while a second experiment deals with the soft sensor referring to Current roboceptions. In the first case, a set of states have been properly defined to drive the robot behavior. In the second case, the robot has to choose among a set of movements with similar characteristics. The aim of the experiments is to show how a robot can be provided with the capability of behavior selection that modulates its actions while the task is accomplished.

5.1.1. Action modes driven by energy soft sensor

In this section, we consider the robot behavior as depending on the state provided by the quantized value of the Energy soft sensor. Four states have been considered: *Normal, Hungry, Starved, Out of Charge*. They depend on the output of the sensor. The state as function of the Energy roboception is listed in Table 1.

The action to be performed, and that is learned through Reinforcement Learning, is the choice of one the working modes. The robot executes the task with the *Full* mode or with *Economy* mode. The third option, *Recharge*, activates the recharge process postponing the execution of the task.The energy consumption is higher in the *Full* mode and it is reduced in the *Economy* mode. The mode does not affect the single action that is being performed but it affects the speed and the required energy. The modes that the robot can activate are shown in Table 2.

For each action, we considered a different reward that also depends on the robot state. An example of reward is shown in Table 3. To model the reward values according to a function, we used the Poisson function in (10). The values of the function have been sampled for n equal to 2, 4, 8. The value of λ has been set to 3 for the **Normal** state (red plot), to 5 for the **Hungry** state (green plot)and to 7 for the **Starved** (blue plot) (Fig. 5(a)). To manage integer values, the Poisson values have been multiplied for one thousand and the value of one hundred has been subtracted from any value of the reward. The plot of the function is given in Fig. 5(a). The subplots shows two different setting, when is present or not a cognitive Urge. The Urges are derived from Psi architecture [45] and are detailed in Section 5.1.2.

Table 1
States according the value of energy roboception.

Energy roboception	State
[100%, 75%[	**Normal** : The energy roboception is high, all activities can be carried on
[75%, 40%[	**Hungry** : The current roboception is medium, a prolonged activity cannot be carried on
[40%, 15%[	**Starved**: The current roboception is low, activity in this state can bring to a sudden stop
[15%, 0%]	**Out of Charge**: The current roboception is near to zero,no activity can be performed, it is needed an external intervention to continue any activity.

Table 2
Working modes for energy roboception.

Mode	Description
Full	the execution is unchanged with respect to the planned task;
Economy	the robot continues the execution of the task changing (reducing) the set of movements and slowing the speed of the execution
Recharge	The robot stops the task execution and activates a procedure to charge the battery and restoring the energy roboception to **Normal**

Table 3
Rewards $R(s)$ for the working modes and the energy states, generated with Poisson function (Fig. 5(a)).

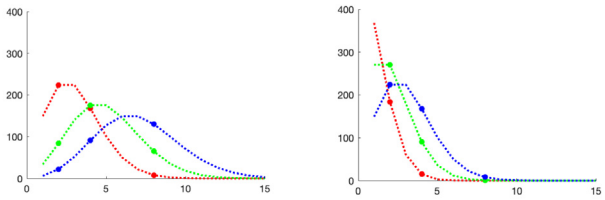
State	Full	Economy	Recharge
Normal	124	68	−92
Hungry	−18	75	−35
Starved	−78	−9	30
Out of Charge	−200	−200	−200

The function has the form:

$$\mathcal{P}_\lambda(n) = \frac{\lambda^n}{n!}e^{-\lambda} \tag{10}$$

It is derived from the Poisson distribution, that is a discrete distribution, measuring the probability of a given number of events occurring in a specified time interval. The function has been chosen for the possibility of changing the distribution shape varying one single parameter. This kind of distribution is commonly used to determine the probability of the number of events taking place in unit time. In our case, it can be interpreted as the motivation that the robot has in the accomplishment of the task. A lower motivation allows the robot to stop the current operations and activate the recharge procedure whenever it needs more charge. On the other side, a stronger motivation overcomes the physical needs and, as result, the robot tends to continue its activity.

In Fig. 5(b), the Poisson function, with a boost in values, has been plotted. In this case the increased values witness that a different motivation is present. The values are sampled in the same points as above. The values of λ has been set to 1 for the



(a) values for the reward without an urge

(b) values for the reward with an urge

Fig. 5. Plot of the Poisson Reward Values. Red is for **Normal** State, Green is for the **Hungry** State, Blu is for the **Starved** State. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Transition Matrix for the *Full* option.

	Nrml	Hngr	Stvd	OoC
Normal	0.4	0.5	0.1	
Hungry		0.4	0.5	0.1
Starved			0.4	0.6
Out of Charge				1.0

Table 5
Transition Matrix for the *Economy* option.

	Nrml	Hngr	Stvd	OoC
Normal	0.5	0.4	0.1	
Hungry		0.5	0.4	0.1
Starved			0.7	0.3
Out of Charge				1.0

Table 6
Average and sigma of $Q^{\pi}(s, a)$ of the work modes according to the energy states in a training with 1000 epochs.

State	*Full*	*Economy*	*Recharge*
Normal	197.7 ± 47.2	150.4 ± 34.0	41.8 ± 17.9
Hungry	−4.3 ± 66.3	6.3 ± 69.3	81.5 ± 30.5
Starved	−120.7 ± 64.8	−18.3 ± 83.5	171.5 ± 26.1
Out of Charge	0.0	0.0	0.0

Normal state, to 2 for the **Hungry** state and to 3 for the **Starved**. This set of values will be used in Section 5.1.2.

Considering a learning constant α equal to 0.5 and a discount rate γ_r equal to 0.7. The action selection has been carried on with a ϵ-greedy policy with a value of ϵ equal to 0.2. The transition probabilities when the execution of the task is performed are given in Table 4 and in Table 5 and reflect the change between the roboception during the execution of the tasks.

While the transition matrix for the action *Economy* are in Table 5 When the *Economy* mode is chosen, there is a higher probability to remain in the same state of charge, while the more demanding *Full* mode is characterized by a higher probability to reach the lower charge category.

The Q value has been calculated applying a SARSA greedy policy. The mean value of $Q^{\pi}(s, a)$ and the σ value for a training with one thousand epochs, are shown in Table 6.

According to these values, when the robot is in a **Normal** state, the preferred action is to work in a *Full* mode, that is to act normally, at full pace, and trying to finish the task. In the **Hungry** state the actions to be preferred are to continue to work in an *Economy* mode, preserving the consumption of some resources or, even, to pass in *Recharge* mode. When the robot is in a **Starved** state, the action with the best value is *Recharge* that allows returning in the **Normal** state to continue the task when the battery is recharged.

Table 7
Rewards $R(s)$.

State	*Full*	*Economy*	*Recharge*
Normal with a CognitiveUrge	84	−85	−100
Hungry with a CognitiveUrge	170	−10	−99
Starved with a CognitiveUrge	124	68	−92
Out of Charge	−200	−200	−200

Table 8
Average and sigma of Q(s, a) of the work modes according to the energy states when a Cognitive Urge is present, after 1000 epochs training.

State	*Full*	*Economy*	*Recharge*
Normal w C.Urge	141.87 ± 54.26	−29.44 ± 42.88	−23.84 ± 29.59
Hungry w a C.Urge	42.73 ± 99.40	−48.68 ± 57.91	−46.16 ± 33.27
Starved w a C.Urge	−133.84 ± 52.79	−96.78 ± 50.71	−19.39 ± 42.23
Out of Charge	0.0	0.0	0.0

5.1.2. Robot behavior when a cognitive urge is present

Beyond the previous case, where the decision is taken according to the state of the robot's somatosensory system, we also considered the case when a cognitive urge is present. The urges are formalized in Psi theory by D'orner and used to characterize the robot behavior according to physical and social needs [46,62]. The urges are used to identify several well-defined motives that drive the human or robotic agent. Some of them are the need for food, water, the avoidance of pain, certainty, competence and affiliation. The authors of this paper have previously adopted this architecture in [63,64] to drive a robot behavior. For the current work, we consider that when an urge is present, not only the bottom-up signals are considered but also a cognitive demand is relevant. The robot behaves differently since the task has to be completed. An example of such a scenario is when the robot is performing a task that has a high priority and its accomplishment is relevant.

A case of such sort could happen when the robot is performing a task and the battery level is running low. The residual energy is decreasing according to the normal evolution of the power consumption but the task is so absorbing that the behavior of the robot is modified to follow what the higher cognitive level is demanding.

In this case, it is not only important to cope with the physical data coming from the soft sensor but, in general, priority has to be changed. The behavior of the robot can be sensibly modified and the robot continues its work at a high speed instead of reducing the power consumption and limit the work performances. The presence of the cognitive urge changes the reward values for the activities, according to the Poisson function, and therefore modifies the Q value for the modes that should be chosen in each state. A modified reward table can change the robot behavior according to a new setting (Table 7). The value are generated with the Poisson function (see Fig. 5(b)) with values of λ equal to 1 for the **Normal** state, to 2 for the **Hungry** state and to 3 for the **Starved**. The values are sampled at n equal to 2, 4 and 8 as in the previous case. Following the previous interpretation, the function, with lower values of λ, increases the rewards witnessing the raised relevance of the task.

The reward is higher when the normal work is performed and the recharge has a higher cost. The recharge should be avoided and the other action should be chosen since the main push is to complete the task. The new value for the actions in the states is shown in Table 8

According to the above results, the preferred mode is the *Full* mode both in **Normal** and in **Hungry** state. In this case, although the battery level is not as full as at the beginning, the preferred mode is to work at full speed and with higher power consumption. This choice does not allow a long duration

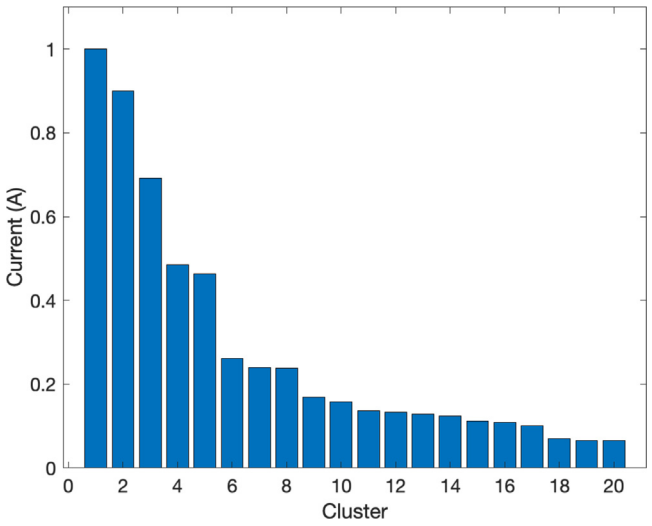


Fig. 6. Examples of the current value for the clusters.

of the performance but, standing a cognitive urge that asks to perform as best as possible, all the given energy is focused on the performance while the power consumption falls in a second level.

5.2. Current roboception driven behavior

The possibility to endow a robot with dancing capabilities is raising much interest at the moment and is an appealing application in human robot interaction [65–68] .

In previous papers we discussed how the robot can acquire dance movements by observing a human teacher and how can associate proper sequences of movements with music [63,64,69–71]. The teacher expresses a judgment on the goodness of the robot's dance influencing the underlying genetic algorithm. We aim to show how the robot's physical conditions, inferred by the somatosensory system, influence the dance balancing the physical demands and teacher's judgment.

In this case, the humanoid ability to create and execute a choreography relies on the computational creativity process proposed in [72]. The creative process is exploited by the robot to choose, during the execution of the dance, the most appropriate movements according to the current situation. In this work, we consider the perception of the music and the robot's internal state. The dance can be generated to fit a more general context considering, for example, also people interacting with the robot.

During the dance's execution, the robot can choose among various sets of movements. The movements are grouped in clusters according to the current employed to execute them. Each movement is characterized by a reward, suggesting the beauty of the single movement and a global reward can be given by the teacher at the end of the task.

The somatosensory system can have an effect on the robot behavior and preserve its body. Since there are some movements that require less current while other are more demanding in term of current to operate the engines, the selection of proper movements can let the robot collect high rewards while preserving the engine from an excessive amount of current.

We run the robot's behavior using 20 clusters of movements that can be used to pick a movement in the dance composition. The clusters gather the movements according to their current usage as shown in Fig. 6. The movements were acquired using an RGB-Depth device and mapping the detected movement to the cinematic chain of the Nao robot in a similar way to what is done in [73].

Table 9
Working modes for the execution of movements chosen from 20 clusters, the working mode is implemented with transition matrices that change to most used clusters.

Mode	Description
Mode 1 *High Current*	This transition matrix prefers the movements from the clusters with lower indices, that correspond to high current. A statistic of occurrence of movements for this mode is shown in Fig. 7(a)
Mode 2 *Mixed*	The transition matrix selects the movements from any clusters the value of the current is neither too low or too high. A statistic of occurrence of movements for this mode is shown in Fig. 7(b)
Mode 3 *Relaxing*	The transition matrix selects the movements from clusters with medium or low current, that are in the center and in the right part of the histogram. A statistic of occurrence of movements for this mode is shown in Fig. 7(c)

An experiment with three modes is performed, the modes characterize how the movements are chosen and have a tight bond with flowing current (see Table 9).

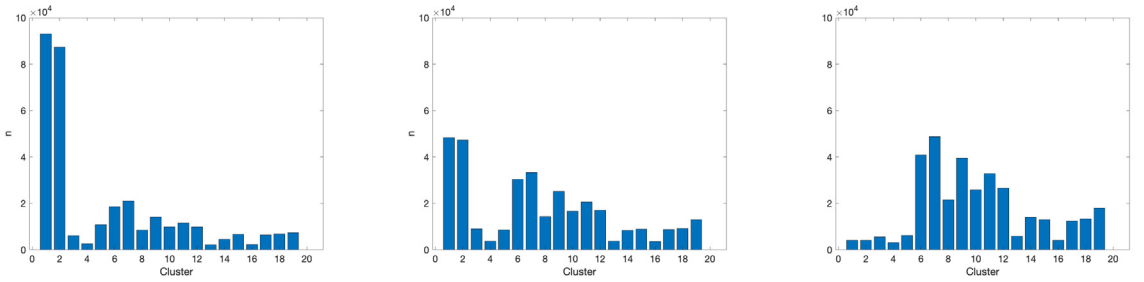
The modes variation is accomplished with different transition matrices used for each of the three modes. The statistic of occurrences of the number of clusters is shown in Fig. 7 and is shown how the number of choices of the first cluster is higher in the first mode. In the second mode, the clusters have been selected in the left and in the middle part, meaning a reduced current. In the third mode, most of the chosen clusters are in the central part, while a few are chosen from the left, employing a minimum current amount. In Fig. 8, additional plots are given. In Fig. 8(a), the profile of the probability in the transition matrices is shown, highlighting the differences between the three modes. In Fig. 8(b), the value of current between the three phases is shown. The three modes are employed in ten thousand iterations. Each mode is activated for a third of the iterations. The vertical lines show the change in the working modes.

We consider a free evolution according to the transition matrices in the three different modes. The agent can choose which mode is more suitable to limit the unpleasant roboception, in this case resembling a painful sensation, and get the best reward.

The possible states, in this case, are three: **Normal**, **Tired**, **Aching**. They are related to how unpleasant, in terms of task execution, they are. While in the **Normal** state the robot will complete the task, **Tired** represents an intermediate state. The permanence in the state **Aching** will stop the execution of the task.

Reinforcement learning is used to choose the best transition matrix (corresponding to a given mode) when a different state is reached. For each performed action it is associated a value, according to the output of the current roboception soft sensor, and a reward that is bound to the quality of the gesture. The robot behavior, after the training phase, will be the result of a trade-off between the pleasantness of the performed movements and states driven by the current values. Since the active state is strongly dependent on the previous actions, the movement is chosen according to the instant reward and according to the past actions given the rewards attributed by the teacher in the evaluated performances.

The training process has been iterated for one thousand epochs. Each epoch is started with state **Normal** and Mode 1. The execution is run considering the transition matrices of the different modes. The epoch is stopped when a convergence in State value is obtained (if the variance in values is less than 0.005) or when the state **Aching** is experienced for a long period, for this experiment

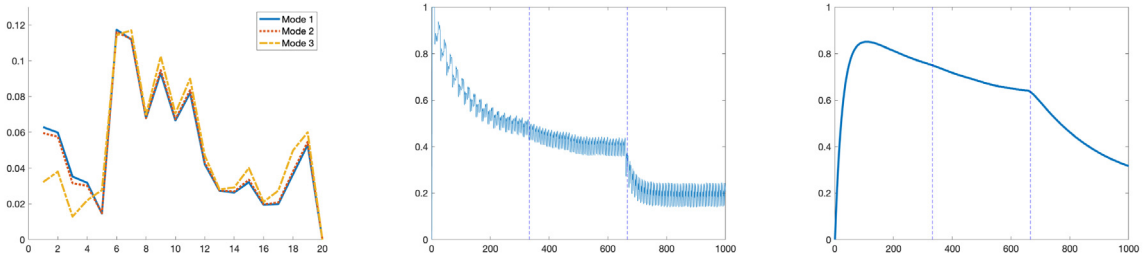


(a) Number of movement chosen in the clusters when mode 1 is activated

(b) Number of movement chosen in the clusters when mode 2 is activated

(c) Number of movement chosen in the clusters when mode 3 is activated

Fig. 7. Statistics for Cluster occurrence in the three modes, a set of 10000 iterations has been performed. For each mode a third of the iterations have been considered.

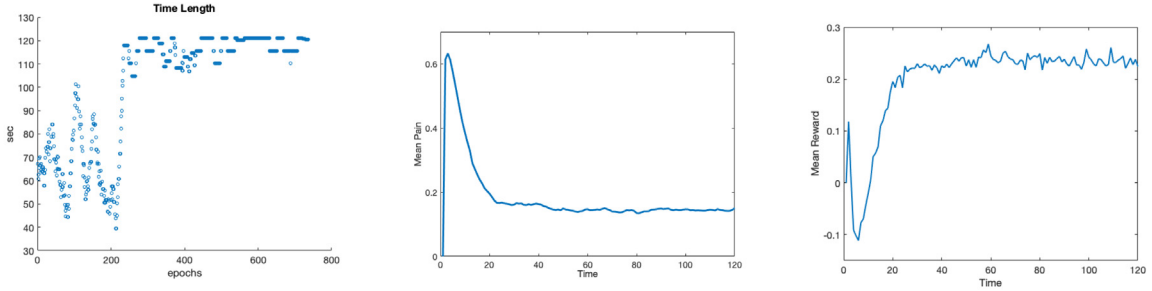


(a) Average Value for Transition Matrix for the three modes, the third mode show a lower profile for clusters one and two, showing a reduced transition towards clusters with higher current

(b) Average Current values vs time (vertical lines indicate the modes). The current flowing in the first cluster is sensibly higher than the values in the two other modes

(c) Average Value of Current Roboception value vs time (vertical lines indicate the modes). The value is function of the current and is sensibly higher in the first mode

Fig. 8. Statistics of Transition Matrix, Current and Pain in the three modes.



(a) Average time of tasks during train process

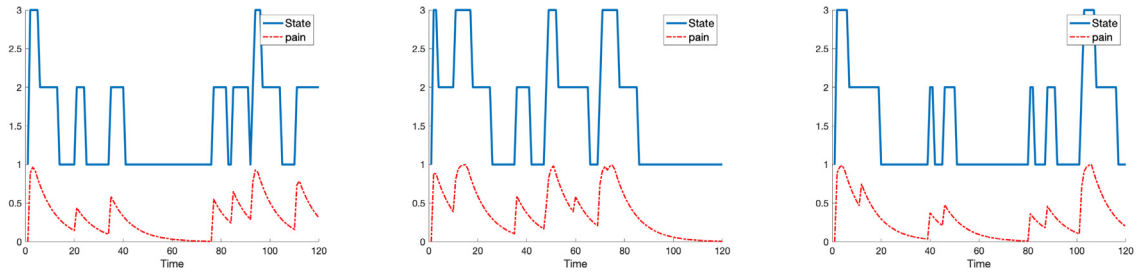
(b) Current roboception value averaged among the training episodes

(c) Reward value averaged among the training episodes

Fig. 9. Statistics of reinforcement learning.

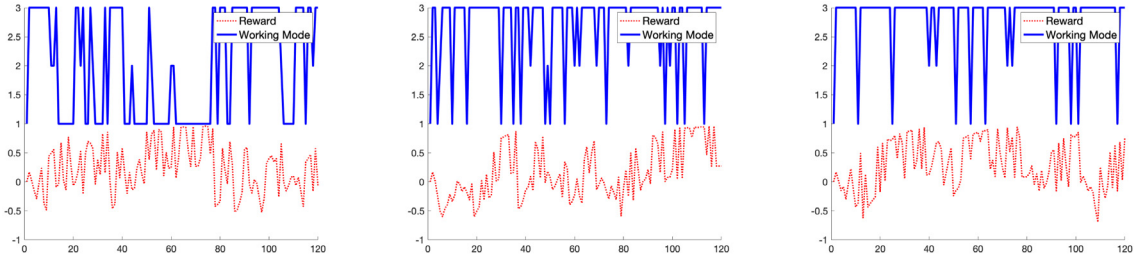
setup ten seconds. In this latter case, the execution is abandoned and a reward of −100 is attributed. The strong negative reward let the robot learn that it cannot stand a strong unpleasant roboception for a long time and it has to act in alternative ways. In Fig. 9, the evolution of some parameters during the train is shown. In Fig. 9(a) the value of time length of a training episode is shown. The first episodes have a lower a time length while the duration is increased during the train, the plot shows how the robot adjusts the mode to avoid an early stop for the task due to a prolonged negative state. In Fig. 9(b), it is shown the current roboception during the episode, averaged with all the values, in the training process. The first portion of the episode is typically characterized by an unpleasant roboception that is avoided with the actions that the system learns to apply. The end of the task shows a lower current roboception value. In Fig. 9(c), is shown the reward during the episodes. There is a first peak related with high rewards at the initial step of the training, these rewards also require a strong effort in terms of current. Then a valley is present when the unpleasant roboception is reduced, choosing instead lower current movements. At the end, an increasing reward is shown when the selected movements well balance negative roboception and collected rewards.

An evolution of how current roboception is processed during train is shown in Fig. 10. The figures are related to episodes during the training, the state of the robot is plot in blue and can assume three possible values, the current roboception is plot in red and ranges from zero and one. The states have labels reported in Table 10 where a lower current roboception corresponds to a **Normal** state (state 1); a higher value of current roboception corresponds to a **Tired** state (state 2) and a higher pain is the **Aching** state that is state 3. The red plot's value shows how the roboception can rise with exponential growth. The higher values



(a) Evolution of current roboception in episode 79 (b) Evolution of current roboception in episode 234 (c) Evolution of current roboception in episode 623

Fig. 10. Sample of current roboception evolution during an episode.



(a) Working Modes in episode 79 (b) Working Modes in episode 234 (c) Working Modes in episode 623

Fig. 11. Working modes during training episodes.

Table 10
States according to the value of current roboception.

Current roboception	Current State
[0.0, 0.3[	**Normal** : The current roboception is not high and the activity can be carried on
[0.3, 0.8[	**Tired** : The current roboception is neither too low or too high, a prolonged flow of current can be critical
[0.8, 1.0]	**Aching**: The current roboception is too high, if this value is unchanged for a long time, the robot can suffer permanent damage.

of the roboception are avoided since, for the experimental setup, a long permanence with the strongest values of the roboception would produce an immediate stop of the task and a global negative reward. While in 10(a) the maximum is not reached, in Figs. 10(b) and 10(c) a saturation of the roboception is obtained, the change in the working mode, according to the training phase, allows the reduction of the roboception along the task execution.

In Fig. 11 are shown the same episodes of Fig. 10, where the working mode is plotted. The working modes are related to the use of electrical current, thereof more painful (1), mixed mode (2) and a relaxed mode (3). Typically, the working mode 1 corresponds to the usage of actions with higher current. When the plot has a peak downward (reaching the value 1) the pain is, typically, going to increase. The other modes, two and three, labeled as *Mixed* or *Relaxing* are, in general, bound to a decrease in the value of the pain and, therefore, to an increase in the reward. With blue color are plotted the modes, the switching among these modes allows to modulate the current and, therefore the roboception. In red is plotted the reward that changes according to the selected mode.

The values are initialized to zero and the value is corrected according to the training evolution. The value changes and, as shown in Fig. 12 tend to converge towards a limit value. The training is stopped either when the number of iteration is reached or when the maximum variance of the three values, in the last twenty samples, is less than one above one thousand. State **Normal** has a value that is positive, state **Tired** has a value that is negative and then slightly positive. The value of the **Aching** state is strongly negative at the beginning, given by the fact that a set of the episode are terminated with a -100 reward. During training, the change of mode allows the robot to escape from states with a negative reward and complete the task within one of the other states. In general, since this state is the most painful, the value is negative.

Table 11
Average and σ value of $Q^{\pi}(s, a)$ for the working modes and the pain state. The values are evaluated across 1000 epochs.

	Mode 1 (*High current*)	Mode 2 (*Mixed*)	Mode 3 (*Relaxing*)
Normal	1.40 ± 0.27	1.41 ± 0.27	1.49 ± 0.24
Tired	0.22 ± 0.37	0.24 ± 0.34	0.41 ± 0.22
Aching	-7.38 ± 3.88	-7.39 ± 3.90	-3.13 ± 3.96

The value of average and σ of the $Q^{\pi}(s, a)$ in one thousand training epochs, is shown in Table 11. For the obtained values, the best choice is to act according to the Mode 3, that is relaxing and the value of pain is limited. In any case, the actions' selection allows the robot to choose also other modes that will provide different evolution and different rewards. The preferred choice with the *Relaxing* mode brings to a limitation in pain and in the recovery after painful roboceptions.

The values of the single states are: Normal has a value of 1.47 with a sigma equal to 0.25, the **Tired** state has a value of 0.38 with a sigma of 0.24 and the **Aching** state has a value of -3.67 with a sigma of 3.47.

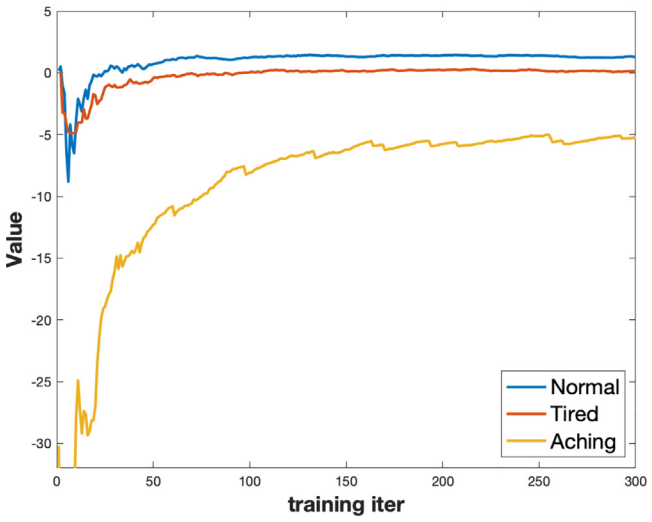


Fig. 12. Evolution of the State Value during a training epoch, the initial negative minimum value is due to episodes that are interrupted with a strongly negative reward, during the training the mode selection allows to increase the collected reward and, consequently, the states value.

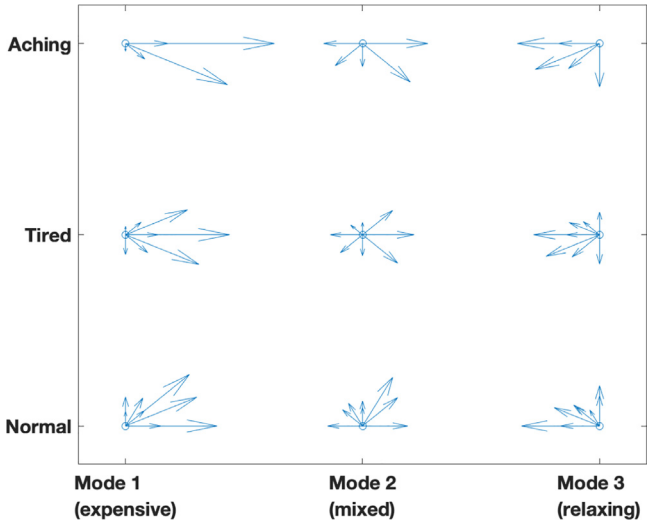


Fig. 13. Quiver Plot of the change in Mode and State in the training, the values have been normalized for each couple (State, Action).

To synthesize the evolution of the state, according to the chosen action, it is shown the quiver plot in Fig. 13. It can be seen that there is a tendency to migrate from the mode with *High current* and the *Mixed* mode, towards the *Relaxing* mode. It is a general tendency to reach the right part of the graph. At the same time, from the Aching state there is a tendency to go towards the state with a higher value, such as **Tired** and **Normal**. In general, since greedy policy allows any action, a generic action is allowed, although the training enables the system to escape to the most painful situation and continue the task with acceptable levels of unpleasant roboception while maximizing the reward.

6. Conclusion

The assumption of this work is that a robot, aware of its body and able to interpret physical sensations, can be more effective in the accomplishment of tasks while maintaining its well being. Loosely inspired by human beings' biology, we proposed an artificial somatosensory system to synthesize the robot's body information and make it improve its behavior selecting good choices that take into account task aims and robot embodiment. The system was modeled focusing on a specific robot, the NAO Aldebaran, even if the model is easily adaptable to different robotic platforms and the architecture can be enriched with hardware monitoring functions.

The behavior of the robot depends on a cognitive architecture. The robot's motivation is influenced by its cognitive and physiological urges and the latter are tightly bound to the specific physical status of the robot.

The experimental results summarize the costs associated to different movements and the motivation that influences the choices of an artificial dancer in different "physical" conditions. We analyzed the roboception of motor current, that resemble a pain "sensation" and the values of the "energy" values that can be bound to a "hunger" sensation.

Substantial differences between a human being somatosensory system a robotic entity have been highlighted, especially for low-cost consumer robot, equipped with elementary sensors designed to monitor several basic parameters. A relevant point for the robot is the perception of their internal state that, together with other processes tied to a cognitive dimension, can play a key role in the emergence of high-level emotions.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

References

[1] D. Parisi, Internal robotics, Connect. Sci. 16 (4) (2004) 325–338.
[2] A.R. Damasio, Looking for Spinoza: Joy, Sorrow, and the Feeling Brain, Houghton Mifflin Harcourt, 2003.
[3] K. Wiech, I. Tracey, Pain, decisions, and actions: a motivational perspective, Front. Neurosci. 7 (2013) 46.
[4] R. Melzack, K.L. Casey, Sensory, motivational and central control determinants of pain: a new conceptual model, Skin Senses 1 (1968).
[5] C. Torras, L. Garcıa, P.L. Jackson, Robot pain: a speculative review of its functions, in: Pain and Conscious Brain, Wolters Kluver, 2016, pp. 235–246.
[6] C. Bagnato, A. Takagi, E. Burdet, Artificial nociception and motor responses to pain, for humans and robots, in: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE, 2015, pp. 7402–7405.
[7] D. Parisi, G. Petrosino, Robots that *have* emotions, Adapt. Behav. (2010) 453–469.
[8] S. van Rysewyk, Robot pain, Int. J. Synth. Emot. (2014) 22–33.
[9] I. Rodriguez, J.M. Martínez-Otzeta, E. Lazkano, T. Ruiz, B. Sierra, On how self-body awareness improves autonomy in social robots, in: 2017 IEEE International Conference on Robotics and Biomimetics, ROBIO, 2017, pp. 1688–1693, http://dx.doi.org/10.1109/ROBIO.2017.8324661.
[10] A. Augello, I. Infantino, S. Gaglio, U. Maniscalco, G. Pilato, F. Vella, An artificial soft somatosensory system for a cognitive robot, in: 2020 Fourth IEEE International Conference on Robotic Computing, IRC, IEEE, 2020, pp. 319–326.
[11] G. Soter, A. Conn, H. Hauser, J. Rossiter, Bodily aware soft robots: Integration of proprioceptive and exteroceptive sensors, in: 2018 IEEE International Conference on Robotics and Automation, ICRA, 2018, pp. 2448–2453, http://dx.doi.org/10.1109/ICRA.2018.8463169.
[12] U. Maniscalco, R. Rizzo, Adding a virtual layer in a sensor network to improve measurement reliability, Adv. Math. Comput. Tools Metrol. Test. X 86 (2015) 260–264.
[13] U. Maniscalco, G. Pilato, G. Vassallo, Soft sensor based on E-αNETs, in: Proceedings of the 2011 Conference on Neural Nets WIRN10: Proceedings of the 20th Italian Workshop on Neural Nets, IOS Press, 2011, pp. 172–179.

[14] U. Maniscalco, R. Rizzo, A virtual layer of measure based on soft sensors, J. Ambient Intell. Humaniz. Comput. 8 (2016) 1–10.
[15] U. Maniscalco, G. Pilato, Multi soft-sensors data fusion in spatial forecasting of environmental parameters, Adv. Math. Comput. Tools Metrol. Test. X 84 (2012) 252–259.
[16] P. Ciarlini, U. Maniscalco, Mixture of soft sensors for monitoring air ambient parameters, in: Proceedings of the XVIII IMEKO World Congress, IMEKO, Hungary, 2006, pp. 1981–1986.
[17] A.G. Sutton, Reinforcement Learning: An Introduction, MIT Press, 1998.
[18] G.A. Rummery, M. Niranjan, On-Line Q-Learning using Connectionist Systems, vol. 37, University of Cambridge, Department of Engineering Cambridge, UK, 1994.
[19] R. Saegusa, G. Metta, G. Sandini, L. Natale, Developmental perception of the self and action, IEEE Trans. Neural Netw. Learn. Syst. 25 (1) (2013) 183–202.
[20] J. Sturm, C. Plagemann, W. Burgard, Unsupervised body scheme learning through self-perception, in: 2008 IEEE International Conference on Robotics and Automation, IEEE, 2008, pp. 3328–3333.
[21] P. Lanillos, E. Dean-Leon, G. Cheng, Yielding self-perception in robots through sensorimotor contingencies, IEEE Trans. Cogn. Dev. Syst. 9 (2) (2016) 100–112.
[22] C. Nabeshima, M. Lungarella, Y. Kuniyoshi, Timing-based model of body schema adaptation and its role in perception and tool use: A robot case study, in: Proceedings. the 4th International Conference on Development and Learning, 2005, IEEE, 2005, pp. 7–12.
[23] E. Hayashi, T. Yamasaki, K. Kuroki, Autonomous behavior system combing motivation with consciousness using dopamine, in: 2009 IEEE International Symposium on Computational Intelligence in Robotics and Automation, CIRA, IEEE, 2009, pp. 126–131.
[24] D. Gamez, Progress in machine consciousness, Conscious. Cogn. 17 (3) (2008) 887–910.
[25] A. Chella, A. Pipitone, A. Morin, F. Racy, Developing self-awareness in robots via inner speech, Front. Robot. AI 7 (2020) 16, http://dx.doi.org/10.3389/frobt.2020.00016.
[26] G. Tononi, M. Boly, M. Massimini, C. Koch, Integrated information theory: from consciousness to its physical substrate, Nat. Rev. Neurosci. 17 (7) (2016) 450–461.
[27] H.M. Gray, K. Gray, D.M. Wegner, Dimensions of mind perception, Science 315 (5812) (2007) 619.
[28] A.K. Seth, Measuring autonomy and emergence via granger causality, Artif. Life 16 (2) (2010) 179–196.
[29] M. Oizumi, L. Albantakis, G. Tononi, From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0, PLoS Comput. Biol. 10 (5) (2014) e1003588.
[30] G. Tononi, G.M. Edelman, Consciousness and complexity, Science 282 (5395) (1998) 1846–1851.
[31] J.A. Reggia, The rise of machine consciousness: Studying consciousness with computational models, Neural Netw. 44 (2013) 112–131.
[32] A. Chella, A. Cangelosi, G. Metta, S. Bringsjord, Consciousness in humanoid robots, Front. Robot. AI 6 (2019) 17.
[33] P.O. Haikonen, P.O. Haikonen, Consciousness and Robot Sentience, vol. 2, World Scientific, 2012.
[34] E. Steen, L. Haugli, From pain to self-awareness—a qualitative analysis of the significance of group participation for persons with chronic musculoskeletal pain, Patient Educ. Couns. 42 (1) (2001) 35–46.
[35] M.C. Hsu, H. Schubiner, M.A. Lumley, J.S. Stracks, D.J. Clauw, D.A. Williams, Sustained pain reduction through affective self-awareness in fibromyalgia: a randomized controlled trial, J. Gen. Intern. Med. 25 (10) (2010) 1064–1070.
[36] S. Koos, A. Cully, J.-B. Mouret, Fast damage recovery in robotics with the t-resilience algorithm, Int. J. Robot. Res. 32 (14) (2013) 1700–1723.
[37] M. Anshar, M.-A. Williams, Evolving synthetic pain into an adaptive self-awareness framework for robots, Biol. Inspired Cogn. Archit. 16 (2016) 8–18.
[38] A. Augello, I. Infantino, G. Pilato, R. Rizzo, F. Vella, Introducing a creative process on a cognitive architecture, Biol. Inspired Cogn. Archit. 6 (2013) 131–139.
[39] T. Zonta, C.A. da Costa, R. da Rosa Righi, M.J. de Lima, E.S. da Trindade, G.P. Li, Predictive maintenance in the industry 4.0: A systematic literature review, Comput. Ind. Eng. 150 (2020) 106889, http://dx.doi.org/10.1016/j.cie.2020.106889.
[40] H.M. Hashemian, State-of-the-art predictive maintenance techniques, IEEE Trans. Instrum. Meas. 60 (1) (2010) 226–236.
[41] P. Langley, J. Laird, S. Rogers, Cognitive architectures: Research issues and challenges, Cogn. Syst. Res. 10 (2) (2009) 141–160.
[42] B. Goertzel, R. Lian, I. Arel, H. de Garis, S. Chen, A world survey of artificial brain projects, Part II: Biologically inspired cognitive architectures, Neurocomputing 74 (1) (2010) 30–49.
[43] A. Augello, I. Infantino, G. Pilato, R. Rizzo, F. Vella, Creativity evaluation in a cognitive architecture, Biol. Inspired Cogn. Archit. 11 (2015) 29–37.
[44] A. Augello, I. Infantino, A. Lieto, G. Pilato, R. Rizzo, F. Vella, Artwork creation by a cognitive architecture integrating computational creativity and dual process approaches, Biol. Inspired Cogn. Archit. 15 (2016) 74–86.
[45] C. Bartl, D. Dörner, PSI: A theory of the integration of cognition, emotion and motivation, in: Proceedings of the 2nd European Conference on Cognitive Modelling, DTIC Document, 1998, pp. 66–73.
[46] J. Bach, D. Dorner, V. Vuine, Psi and MicroPsi: a novel approach to modeling emotion and cognition in a cognitive architecture, in: Proceedings of the 7th International Conference on Cognitive Modeling, Trieste, 2006, pp. 20–25.
[47] N.J. Nilsson, N.J. Nilsson, Artificial Intelligence: A New Synthesis, Morgan Kaufmann, 1998.
[48] A. Augello, I. Infantino, U. Maniscalco, G. Pilato, F. Vella, The effects of soft somatosensory system on the execution of robotic tasks, in: Robotic Computing (IRC), IEEE International Conference on, IEEE, 2017, pp. 14–21.
[49] U. Maniscalco, I. Infantino, An artificial pain model for a humanoid robot, in: International Conference on Intelligent Interactive Multimedia Systems and Services, Springer, Cham, 2017, pp. 161–170.
[50] A. Augello, G. Città, M. Gentile, I. Infantino, D. La Guardia, A. Manfrè, U. Maniscalco, S. Ottaviano, G. Pilato, F. Vella, et al., Improving spatial reasoning by interacting with a humanoid robot, in: International Conference on Intelligent Interactive Multimedia Systems and Services, Springer, Cham, 2017, pp. 151–160.
[51] A. Galipò, I. Infantino, U. Maniscalco, S. Gaglio, Artificial pleasure and pain antagonism mechanism in a social robot, in: International Conference on Intelligent Interactive Multimedia Systems and Services, Springer, Cham, 2017, pp. 181–189.
[52] I. Trifirò, A. Augello, U. Maniscalco, G. Pilato, F. Vella, R. Meo, How are you? How a robot can learn to express its own roboceptions, in: Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 24th International Conference KES2020, Procedia Comput. Sci. 176 (2020) 480–489.
[53] R. Brooks, A robust layered control system for a mobile robot, IEEE J. Robot. Autom. 2 (1) (1986) 14–23.
[54] U. Maniscalco, A. Messina, P. Storniolo, ASS4hr — An artificial somatosensory system for a humanoid robot. The ROS package, SoftwareX 11 (2020) 100501, http://dx.doi.org/10.1016/j.softx.2020.100501.
[55] R. Bellman, A Markovian decision process, Indiana Univ. Math. J. 6 (1957) 679–684.
[56] M.L. Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming, John Wiley & Sons, 1994.
[57] D.P. Bertsekas, C.C. White, Dynamic programming and stochastic control, IEEE Trans. Syst. Man Cybern. 7 (10) (1977) 758–759, http://dx.doi.org/10.1109/TSMC.1977.4309612.
[58] N.D. Nguyen, T. Nguyen, S. Nahavandi, System design perspective for human-level agents using deep reinforcement learning: A survey, IEEE Access PP (2017) 1, http://dx.doi.org/10.1109/ACCESS.2017.2777827.
[59] S. Singh, T. Jaakkola, M.L. Littman, C. Szepesvári, Convergence results for single-step on-policy reinforcement-learning algorithms, Mach. Learn. 38 (3) (2000) 287–308.
[60] H. van Hasselt, M.A. Wiering, Convergence of model-based temporal difference learning for control, in: 2007 IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning, IEEE, 2007, pp. 60–67.
[61] H. Robbins, S. Monro, A stochastic approximation method, in: Herbert Robbins Selected Papers, Springer, 1985, pp. 102–109.
[62] Z. Cai, B. Goertzel, N. Geisweiller, OpenPsi: Realizing dörner's "psi" cognitive model in the OpenCog integrative AGI architecture, in: Artificial General Intelligence, Springer, 2011, pp. 212–221.
[63] I. Infantino, G. Pilato, R. Rizzo, F. Vella, Humanoid introspection: A practical approach, Int. J. Adv. Robot. Syst. 10 (2013).
[64] G. Pilato, R. Rizzo, F. Vella, I. Infantino, Human-robot interaction based on introspective capability, in: Complex, Intelligent and Software Intensive Systems (CISIS), 2012 Sixth International Conference on, IEEE, 2012, pp. 461–468.
[65] G. Sun, Y. Wong, Z. Cheng, M.S. Kankanhalli, W. Geng, X. Li, DeepDance: Music-to-dance motion choreography with adversarial learning, IEEE Trans. Multimed. (2020) 1.
[66] S. Gao, Inspiration mechanism of dance creation based on brain subconsciousness theory, Rev. Argent. Clín. Psicol. 29 (1) (2020) 453–460.
[67] T. Tang, J. Jia, H. Mao, Dance with melody: An LSTM-autoencoder approach to music-oriented dance synthesis, in: Proceedings of the 26th ACM International Conference on Multimedia, 2018, pp. 1598–1606.

[68] A. Manfrè, I. Infantino, A. Augello, G. Pilato, F. Vella, Learning by demonstration for a dancing robot within a computational creativity framework, in: Robotic Computing (IRC), IEEE International Conference on, IEEE, 2017, pp. 434–439, Electronic ISBN: 978-1-5090-6724-4, Print on Demand(PoD) ISBN: 978-1-5090-6725-1.
[69] A. Augello, I. Infantino, A. Manfrè, G. Pilato, F. Vella, A. Chella, Creation and cognition for humanoid live dancing, Robot. Auton. Syst. 86 (2016) 128–137.
[70] A. Augello, E. Cipolla, I. Infantino, G. Pilato, A. Manfrè, F. Vella, Creative robot dance with variational encoder, in: International Conference on Computational Creativity, 2017, Available at https://deeplearn.org/arxiv/10380/creative-robot-dance-with-variational-encoder.
[71] I. Infantino, A. Augello, A. Manfrè, G. Pilato, F. Vella, ROBODANZA: Live performances of a creative dancing humanoid, in: Proceedings of the Seventh International Conference on Computational Creativity, 2016, pp. 388–395.
[72] A. Manfrè, I. Infantino, F. Vella, S. Gaglio, An automatic system for humanoid dance creation, Biol. Inspired Cogn. Archit. 15 (2016) 1–9.
[73] A. Manfrè, A. Augello, G. Pilato, F. Vella, I. Infantino, Exploiting interactive genetic algorithms for creative humanoid dancing, Biol. Inspired Cogn. Archit. 17 (2016) 12–21.

Agnese Augello is a staff research scientist at ICAR-CNR. She has received a Ph.D. degree in Computer Science from the University of Palermo. Her main research interest is the modeling of virtual agents and social robots with a particular focus on cognitive architectures.

Salvatore Gaglio is Full Professor of Artificial Intelligence at the University of Palermo. He is author of more that four hundreds contributions in international conferences and journals. He is associated with ICAR-CNR. He is director of Sicily Section of the Italian association of Automatic Computing (AICA).

Ignazio Infantino obtained master degree and Ph.D. at University of Palermo (Italy). Since 2001 he is a research scientist at CNR. His research activities deal with computer vision, cognitive robotics, image processing, human–computer interfaces.

Umberto Maniscalco received his Ph.D. in Electrical, Communication and Computer Engineering from the University of Palermo in 1998. He currently is a technologist at ICAR CNR heading the Human-Robot Interaction Group. He has also responsible for the ICAR of the ''robotics and automatics''' and ''ICT Devices and Systems'' project areas of CNR-DIITET.

Dr. Giovanni Pilato received his ''cum laude'' Ph.D. degree in computer science from the University of Palermo, Italy, in 2001. He is currently a staff research scientist at the ICAR-CNR. He is also lecturer at the University of Palermo, Italy. His research interests include knowledge representation, web data mining and natural language processing.

Filippo Vella took the degree in Electronic Engineering at the University of Palermo. He worked at STMicrolectronics, in the AST division. He took his Ph.D. in Computer Science at University of Palermo. Since 2006 he is staff researcher at CNR - ICAR Palermo. His current research activities deal with artificial vision, robotics and quantum computing.