

Skill refinement through cerebellar learning and human haptic feedback: an iCub learning to paint experiment*

Jordi-Ysard Puigbò¹, Clément Moulin-Frier¹, Vasiliki Vouloutsi¹,
Martí Sanchez-Fibla¹, Ivan Herreros¹ and Paul FMJ. Verschure^{1,2},

Abstract—This article presents a model of the control of hand movements learned from human imprecise feedback. The setup consists of an iCub humanoid robot able to paint with a pencil on an interactive display. It learns to avoid an abstract boundary, which is drawn on the table and that the robot can perceive. Before learning, the robot does not know that it has to paint only inside the boundary. This is instead learned from human feedback provided whenever the pencil goes outside of the shape. We use a neurocomputational model of the cerebellum, which learns to anticipate the human feedback from the perception of the previously meaningless shape boundaries.

We show how this biologically-grounded adaptive mechanism, which is plausible in terms of human infant development, allows the learning of a precise painting behavior, using the perception of the shape boundaries as a predictive signal of aversive stimuli. This mechanism can be generalized to any kind of task where some aversive feedback can be considered correlated with available sensory cues. Consequently, the model allows a human teaching to a robot any kind of complex task, as long as the human knows how to provide consistent feedback and the robot has available the sufficient sensory cues. This becomes specially relevant for educating a robot on social behaviour or bringing it to a socially rich environment.

I. INTRODUCTION

Humanoid robots are of special interest when it comes to social interaction with non-expert human users [1]. Such robots need to behave in a transparent and natural way [2]. In general, there is a tendency to design robots that are more anthropomorphic, as it seems to be the most appropriate form for social robots [3]. Their anthropomorphic shapes foster natural interactions, as they provide a more intuitive interface to establish social expectations [4]. However, how such natural interactions can be used for social learning, i.e. for the acquisition of relevant skills by knowledge transfer from the user to the humanoid robot, is still considered a difficult problem despite the number of contributions in the field.

A particular problem is the learning of task constraints from user feedback. In this case, the user does not provide specific information about the way of performing a particular task (this is the domain of learning by demonstration which is outside the scope of this paper), but only about whether

the current behavior of the robot is considered as appropriate or not by the user to actually perform the task. Therefore, the problem faced by the robot is how to interpret the user feedback to adapt its own behavior in order to maximize the amount of rewarding stimuli and/or minimize the amount of aversive ones. Efficient models of social learning from feedback have been proposed, using reinforcement learning [5], [6] or schema-based action selection [7]. However, those studies focus on learning what action to choose in a discrete action set.

In this paper, we are rather interested on learning the timing of action triggering from user feedback and contextual sensory signals. We address the problem by adopting a biologically-grounded approach based on our previous work on neurocomputational models of reactive and adaptive control. We base this work on the classical conditioning paradigm [8]. Here, animals learn to anticipate aversive stimuli when it is presented shortly after consistent contextual information. This learning process has been precisely targeted inside the cerebellum [9], [10] and some authors of the present paper have recently proposed a neurocomputational model of this brain structure [11] to anticipate the aversive stimuli from the perception of contextual cues. Additionally, the cerebellum has been related also to motor refinement, where the aversive stimuli can be compared to an error in sensory prediction [12].

Neurophysiological models of the cerebellum are becoming popular in robotics, although to our knowledge, they have not been used in a complete developmental scenario including human feedback. In [13], a mobile robot learns to anticipate and avoid collisions. The same cerebellar model has been applied for anticipating the turning action of a robot in a rapid navigation task [14] (which addresses generalization properties of the cerebellar learning). Examples of postural and position control tasks are described in [15], in which spiking models of the cerebellum are used. [16] uses cerebellar based adaptation to refine the precision of an arm control task when exploring a surface of uncertain depth. In our case we enhance the precision of the arm movement with the goal of painting a predefined area without surpassing the borders.

In this paper, we address a robot painting training scenario via cerebellar learning with human tactile feedback. This setup is inspired from infant-caregiver interaction, where the adult provides an aversive stimuli (e.g saying "no" or by tactile feedback) when the child paints outside of the shape it is supposed to color on a paper sheet. From these interactions,

¹Synthetic, Perceptive, Emotive and Cognitive Systems group (SPECS), Center of Autonomous Systems and Neurorobotics, Department of Information and Communications Technologies, Universitat Pompeu Fabra, Barcelona, Spain

²Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

*Find a video demonstration at: <http://webservices.specs.upf.edu/SPECSExperiments/icub/ExperimentDrawingRobot.mp4>

the child can observe that the caregiver feedback is associated with the crossing of the shape boundary and learn how to time the drawing movements according to the perception of the shape, eventually avoiding the caregiver aversive stimuli. This paradigm currently applied in a drawing task but can be easily generalized for a context where social constraints are imprecisely given through human advice. In contrast to other work that center on the development of the drawing behavior [17], [18], we present a general mechanism grounded on brain theory that uses avoidance behaviors as a basis for acquiring more complex behavior.

We claim that (1) the described learning process is plausible to occur in the cerebellum through a classical conditioning paradigm and (2) that a neurocomputational model of that brain structure can be applied in a human-robot interaction setup to learn task constraints from imprecise user feedback. To support these hypotheses, we apply an existing model of the cerebellum [11] to a scenario where an iCub robot [19] learns the timing of painting movements from human aversive stimuli through tactile perception.

We thus adopt the view that robotics and neuroscience have to be in fruitful collaboration [20], and we exactly place our contributions in this intersection by proposing a complete system implementing the minimal components of a developmental scenario in which an iCub robot learns a skill through human feedback, by acquiring the adaptive motor responses using a biologically valid model of the cerebellum.

In our task we learn from a reactive controller, exploiting the anticipation provided by cerebellar learning, producing an additional adaptive component that refines and makes the execution of the task more precise. This architecture is framed in the the Distributed Adaptive Control architecture (DAC) [21], which explains that behaviour is organized in a layered structure building bottom up from a reactive component and adding, through learning, an adaptive contribution (contextual aspects of the task are out of the scope of the paper).

Our minimal real-world biologically valid implementation can benefit the field of training by human feedback (as in [22]) by adding the consideration of temporal constraints and proposing an alternative to reinforcement learning.

This article is organized as follows: section II provides details about the neurological substrate behind each part of the model, while in section III we provide the implementation and functional details, together with a description of the experimental setup. Finally, section IV details the results that are further discussed in section V.

II. BIOLOGICAL BACKGROUND

Our controller is based on an adaptive learned response that anticipates and modulates the activation of a reactive behavior. This structure nicely fits the biologically grounded Distributed Adaptive Control framework (DAC, [21], [23]). DAC proposes that cognition is organized in a number of hierarchical layers of increasing complexities: from reactive (reflex behavior pre-wired from evolution), to adaptive (involving prediction through sensorimotor association learn-

ing), to contextual (involving planning and memory). This paper focuses on the coupling of the reactive and adaptive layers.

A. Reactive Layer

The reactive layer is based on the biological concept of homeostasis, the biological mechanism by which self-regulation is maintained and allows animals to adapt to ever changing environment and body conditions. Homeostasis can be globally understood as the set of needs and drives that govern an animal's behavior. In biological terms, the reactive layer serves the goal of keeping and satisfying a series of these basic needs through pre-wired and reflexive behaviors. Each of these needs can be related to an homeostatic regulatory loop that the agent has to maintain within a comfort zone (from now on CZ). Each loop is dedicated to regulate a certain variable or state of the organism, e.g., hunger, breathing, safety, rest, etc.

Following the model that we proposed in [24] and that we implemented within a humanoid robotic setup in [25] each homeostatic loop has a predefined, pre-wired mechanism (also called regulator) to maintain its variable within the CZ. In the reactive layer, these correspond to predefined, automatic actions that can also be considered reflexes. Whenever the system gets out of the CZ, the homeostatic system can be regulated to achieve again a desired state.

In our model we define a need or drive as a CZ in some sensory dimension. The boundaries of this CZ are defined as thresholds on this sensory dimension. The reactive controller generates reflexes that react to surpassing the boundaries of this CZ. Surpassing the threshold triggers an action that, ideally, brings the measurement back to the CZ. Must be noted that, in the lowest implementation of the controller, the reflexes are not precise actions, but overreactions that ensure that the measure is back to safety. An example of an homeostatic drive could be the sense of contact on the skin of a robot, or its force/torque sensors. The reactive controller receives a signal when these sensors go out of the CZ and generates a general movement action on the opposite direction associated to the sensor.

B. Adaptive Layer

The adaptive layer develops on top of the reactive one from the agent-environment interaction to refine action through learning mechanisms, allowing e.g. the anticipation of action to improve the reactive control.

On top of the reactive controller we use a model of the cerebellum, which acts as an adaptive controller. The cerebellum has been related to sensory-motor prediction, anticipation and adaptive motor refinement for a long time [12].

The most studied paradigm in which cerebellum plays a crucial role is Classical Conditioning. In Classical Conditioning, an aversive Unconditioned Stimulus (US), like an air-puff delivered to the eye, generates a reflexive Unconditioned Response (UR) of closing of the eyelid and is preceded by a neutral Conditioned Stimulus (CS), e.g. a tone. Subsequent

CS-US presentations will build up a Conditioned Response (CR), learnt in the cerebellar circuitry, preceding and anticipating the US. The CR will then avoid (completely or partially) the US [26].

We use the model of the cerebellum presented in [11] which has been already generalized to a robot navigation, collision anticipation/avoidance task in [14]. A relevant feature of this model is the presence of a negative feedback gain called k_{noi} (of neurophysiological relevance, see [11] for details) which balances precisely the reactive and adaptive components of the CR. This gain controls up to which extent the CR will completely or partially avoid the US, and, if partially, up to which degree. Some faint presence of the US has to remain present, because if not, the model extinguishes the response.

III. MODEL IMPLEMENTATION

This section describes the experimental setup and neuro-computational model we use in our experiment, our modeling choices being directly derived from the biological background we have exposed in the previous section.

A. Experimental Setup

Our experimental setup consists in a humanoid iCub robot, a 3D-printed pen, and an interactive table called the Reactable [27], which is pictured on Figure 1. The iCub stands in front of the table and uses a pre-existing inverse kinematics module to control the position of its right hand on an horizontal plane just above the table. The hand is able to move on a $15 \times 30\text{cm}$ rectangle that we call the working area. A 3D-printed pencil is attached to the hand of the robot. This pencil ends with a white tip allowing the Reactable to detect its (x, y) position in real time. A closed shape inside the working area is permanently displayed on the Reactable, e.g. a circle or a triangle, and the robot can perceive the distance between the pencil tip and the closest shape boundary. The iCub is also equipped with a tactile skin allowing to perceive a strong grasp on its arms (the US). We instruct human subjects to provide feedback to the robot by grasping its left arm (i.e. the arm without the pen) whenever it paints outside of the displayed shape.

The neuro-controller we are going to describe outputs (x, y) coordinates inside the working area that the robot can reach by moving its whole body as instructed by the inverse-kinematics module. It consists in two coupled control loops: a reactive and an adaptive ones.

B. Reactive Controller

The reactive controller provides the robot with two needs. The first one corresponds to a basic exploratory behavior which drives the robot to move the pencil in random directions. A direction is defined by a random angle sampled uniformly in $[0, 2\pi]$ and a random distance sampled from a normal distribution with mean 0.3 and standard deviation 0.1. The second need is a reactive behavior which drives the robot hand to the starting position. This behavior is triggered by an input $h \in [0, 1]$ to the reactive controller, such that the robot's

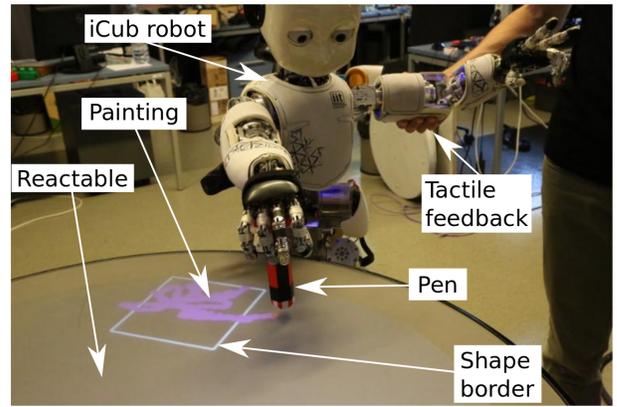


Fig. 1. Our experimental setup consists in a humanoid iCub robot, a 3D-printed pen, and an interactive table called the Reactable.

hand moves to the center with a distance proportional to h , i.e. completely to the center when $h = 1$ and at equidistance of the current hand position and the center when $h = 0.5$. h is computed as the normalized intensity returned by the robot's skin sensor. The *go_home* behavior is only triggered when h is above a small threshold ϵ , otherwise the exploratory behavior operates in isolation.

At the level of this work, homeostasis is presented not as a basic, primary need, but as a drive to return to the origin when aversive feedback is provided in the form of excessive pressure on the robot's skin. Going out of the comfort zone produces two outcomes: triggers a *go_home* reactive behavior and provides the error signal to the adaptive controller. The first acts as a reactive measure to return to the comfort zone. This way, the robot performs random drawing when the pencil is inside the shape and each time it goes outside of it, the human subject grasps the robot's left arm, which triggers in turn the *go_home* behavior returning in the center of the working area.

C. Adaptive Controller

The adaptive controller learns to anticipate the tactile feedback based on the perception of the distance between the pencil tip and the closest shape boundary. For this it simply acts as a linear regressor whose target signal is the negative stimuli provided by the human. As basis of this approximation it uses a repertoire of time-varying signals that code contextual information, the analogue of CS in classical conditioning, which in this setup is provided by the *distance between the pencil and the closest boundary*. How the adaptive controller is coupled with the reactive one is shown on Figure 2.

The generation of this context-coding signal comprises two steps. First 5 bases output a 1 when the distance to the boundary is less than a given threshold value and 0 otherwise, where each basis has a different threshold. This can then be interpreted as the controller receiving a set of CSs. Once these signals enter the adaptive controller, they are replicated n -times and passed through a series of double-convolutional filters with different temporal profiles

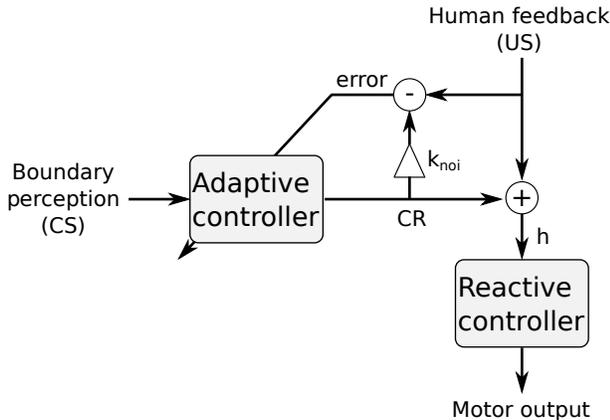


Fig. 2. The robot’s cognitive architecture is based on two coupled control loop: a reactive and an adaptive one. The reactive controller inputs h , which is induced by human tactile feedback (US) and/or by its anticipated prediction from the adaptive controller, using contextual sensory cues (CS). The learning signal of the adaptive controller is the error given by $US - k_{noi}CR$, see text for details.

(the details can be found in [14]) to generate a series of diverse time-varying signals. We refer to these signals as *cortical bases*. At time t , the vector $\mathbf{p}(t)$ contains the state of these bases. The output of the adaptive controller, that by analogy with the conditioning paradigm we refer to as *CR* is equal to $\mathbf{p}(t)^T \mathbf{w}(t)$, where $\mathbf{w}(t)$ are the regression weights.

The controller regresses its teaching signal $e(t)$ to the context at time $t - \delta$. The regression weights $\mathbf{w}(t)$ are learned on-line using the decorrelation learning rule [28]:

$$\dot{\mathbf{w}}(t) = -\eta \mathbf{p}(t - \delta) e(t)$$

where η is a sufficiently small learning rate. Finally, the error signal is computed as follows

$$e(t) = US(t) - k_{noi}CR(t - \delta)$$

This implies that the adaptive controller requires a degree of feedback to confirm the correctness of the anticipatory action, otherwise a negative error would be generated, proportional to the k_{noi} factor, resulting in the extinction of the acquired behavior. In our setup, since the feedback is binary, this will lead to a non-zero error rate at learning asymptote. Note that the human reinforcement is compared not to the current output of the controller but to a preceding one at time $t - \delta$, denoting that if an error occurs at time t it should have been prevented not by an action simultaneous to the error, but by one anticipated by δ seconds. In short, this δ sets the extent of the anticipation.

IV. RESULTS

We instructed a human subject to grasp the robot’s upper left arm every time the pencil crossed the boundary of the shape, and release it after it entered back. For the first round of experiments, the shape was always a square. The subject was presented 10 times with the reactive controller and 10 with the adaptive controller on top of it.

Figure 3 shows a sample of both conditions. One can observe an increase in both, the number of times the robot

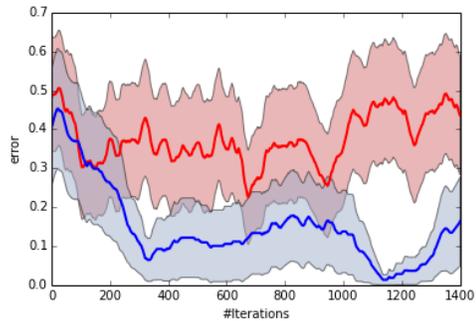


Fig. 4. The number of errors decreases with time in the adaptive controller, but not in the reactive. This Figure shows the mean and standard error of the feedback received for the ten executions of both the adaptive (blue) and reactive (red) controllers.

received feedback and the amount of time spent on average with every feedback for the reactive condition with respect to the adaptive. Comparing Figures 3 B and D show how, for a task that was only considering not going out of the borders of the shape, the adaptive controller (D) keeps the painting far from the borders, while the reactive (B) goes out numerous times. Figure C shows how, after a few trials of learning, the adaptive controller learns to produce an output that anticipates human feedback, thus avoiding it most of the times.

Furthermore, we defined a measure of the error as the human feedback. As the human was instructed to provide feedback to the robot whenever the pencil was painting outside of the shape, we consider that the feedback (the US) is a measure of the error perceived by the human. We use this measure of error because, for this HRI task, where a human is teaching a robot, the task can be considered successful as soon as the teacher is satisfied with the performance. Figure 4 shows that this measure of error is noisy but constant for the purely reactive controller, while it is reduced in the case of the adaptive controller.

Finally, Figure 5 shows a single experiment, where after a short period of training, the shape was changed from a circle to a square. The Figure shows similar performance for the different shapes, showing that the model allowed the successful learning of an abstract concept, as is a drawn border of a shape, from human teaching. To be more specific, the learning mechanism has associated a previously neutral stimulus (the border) to an aversive stimulus (the strong grasp), what now is triggering an anticipatory behavior, avoiding future punishment.

V. CONCLUSIONS

In this paper, we have proposed a biologically-grounded cognitive architecture based on a reactive control of hand movements from human feedback and an adaptive controller based on the prediction of that feedback from contextual sensory cues. The adaptive controller is based on an existing model of the cerebellum which learns in an on-line manner to predict an unconditioned stimuli US (the human feedback in

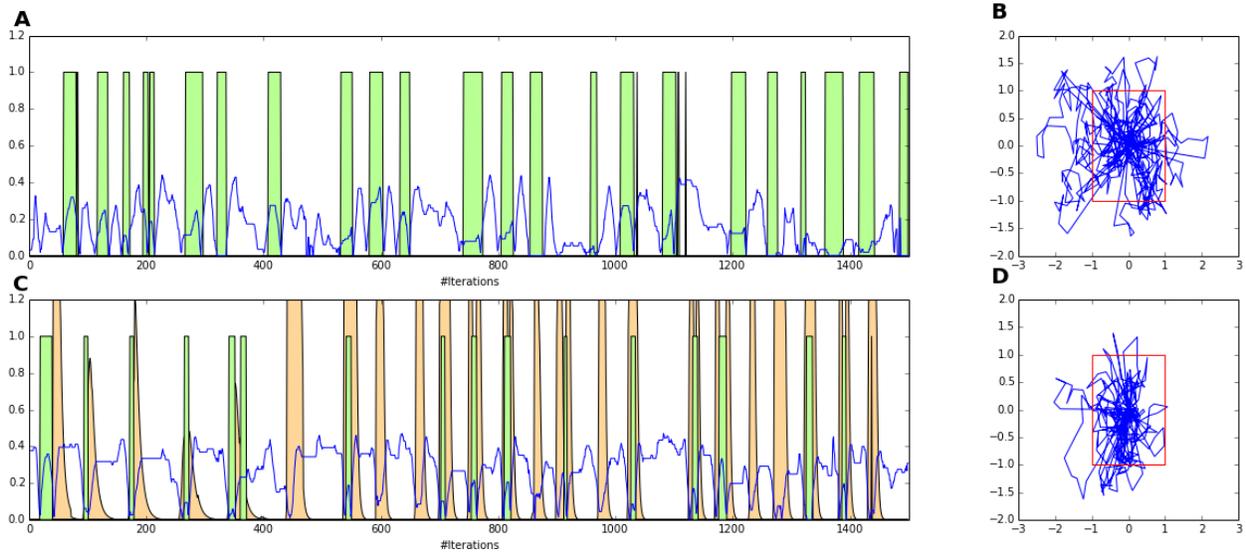


Fig. 3. This figure shows a comparison of the reactive (A,B) and the adaptive (C,D) behaviors. On A and C, the blue line shows the cs as detailed in section III. Green stripes show human feedback (US) as unit steps for the duration of the contact. For the adaptive behaviour, orange shows the cerebellum output. From iteration 250, anticipation of human feedback can be observed. Difference in widths (time span) of US between the reactive and adaptive controller has been measured for ten executions of the experiment ($\mu = 40.4$, $\sigma = 33.0$ iterations the reactive controller; ($\mu = 14.0$, $\sigma = 2.6$ iterations the adaptive). Figures B and D show the commands at the output of the adaptive and reactive controllers.

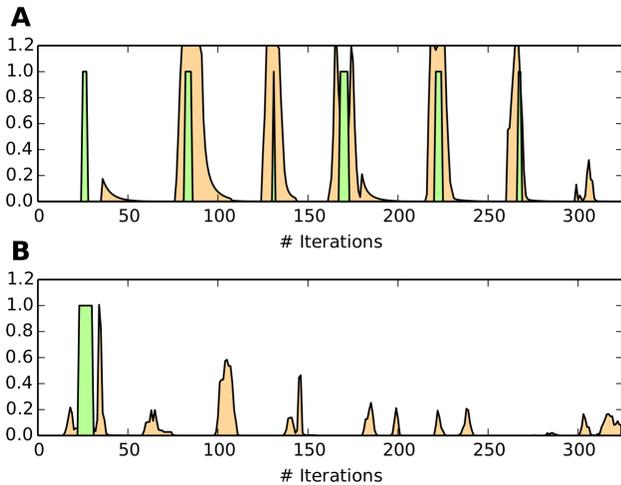


Fig. 5. Learning is independent of shape and position. Figure A shows an execution of the experiment for a circular shape for 325 iterations. After this, the shape and position of the reactable were changed. Figure B shows the successful outcome of the adaptive controller, after learning.

our experiment) from a conditioned stimuli CS (the distance between the pencil and the shape border).

The results show that the model is able to correctly anticipate the human feedback, allowing to react before crossing the shape border and to reduce or cancel the feedback. Moreover, the model displays interesting generalization properties, where the behavior learned on painting a particular shape can be successfully reused for painting other kind of shapes. This is thanks to the contextual sensory cues that are provided to the robot which are not shape-dependent but rather based on the distance between the pencil tip and the

closest shape border. Altogether, these results support the hypothesis that the cerebellum can play a role in a social learning context when the learning is focused on the precise timing of actions, and that neurocomputational models of this brain structure can be successfully used in a human-robot interaction context.

This model has a number of limitations that encourage further extensions. First, the reactive controller is quite simple, relying on only two predefined behaviors: exploring randomly or going back toward the center for a variable distance. We are considering to extend this reactive controller to trigger more complex movement primitives related to drawing movements or more elaborated exploration policies. Second, we observe that the model is subject to forgetting what it has learned before, which is a known issue but is actually coherent with animal behavior data (a phenomenon call "extinction" in the classical conditioning literature). However, in a human-robot social learning context, as the one proposed, it is an important issue that requires further investigation. In the Distributed Adaptive Control framework from which the current cognitive architecture is derived, a contextual layer resides on top of the adaptive. This layer allows the storage and retrieval of acquired behaviors in a long-term memory, what could solve the forgetting.

Being the error nonspecific, in the sense that it only signals the aversive state but it does not contain information about how to correct it, one could argue that a Reinforcement Learning (RL) paradigm would have been more adequate; but we justify in the following why this is not the case. Our problem here is to generate rapid anticipatory well-timed responses, which is precisely the role of a cerebellar like structure. The cerebellar model used [11]

is a supervised learning system that in our case is driven by a nonspecific error from human feedback. The trick is that learning is anticipating the triggering of the reactive behaviour in advance which in itself contains information on the movement to make. Although the error is nonspecific, the contextual information needed for the reflexive action to generate the appropriate reflex directed to the center of the shape is contained in the reactive layer. Our aim is to model using a complete developmental scenario, how the cerebellum is involved in adaptively modulating reflexive behaviour. Moreover with the cerebellar architecture, we get generalization to different execution speeds for free since the first trials [14]. In the future we will consider of making the human feedback more specific, by for example being able to translate a tactile pattern, while holding the forearm of the iCub, into a correcting direction.

In contrast, another argument against the task proposed could be that the sensory cues are already complex and specific to the task. In this sense, it must be remarked that the purpose of this study is to provide a proof of concept showing that neutral features of the environment can be identified as relevant and used to accomplish a certain task, through the anticipation of noisy, imprecise human feedback.

Our future work aims at enhancing the existing model and introducing an extra non-specific learning phase grounded on the interactions between the amygdala and the sensory cortices. Such interactions allow the acquisition of *relevant* features from a high dimensional sensory environment and make them accessible to other areas of the brain, as detailed in [29]. With this, the distance could be extracted and learned from the relation between the pencil tip and the border of the shape using vision and not given by the Reactable.

Finally, we note that the proposed model can be generalized to any kind of task where an aversive stimuli can be predicted consistently by contextual sensory information and thus be used for robot learning. This advances in the direction of preparing service or assistant robots for acting in complex environments.

ACKNOWLEDGMENT

This work is supported by the EU FP7 project WYSIWYD (FP7-ICT-612139).

REFERENCES

- [1] M. A. Goodrich and A. C. Schultz, "Human-robot interaction: a survey," *Foundations and trends in human-computer interaction*, vol. 1, no. 3, pp. 203–275, 2007.
- [2] C. Breazeal, "Role of expressive behaviour for robots that learn from people," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 3527–3538, 2009.
- [3] C. F. DiSalvo, F. Gemperle, J. Forlizzi, and S. Kiesler, "All robots are not created equal: the design and perception of humanoid robot heads," in *Proc. 4th conference on Designing interactive systems: processes, practices, methods, and techniques*. ACM, 2002, pp. 321–326.
- [4] B. R. Duffy, "Anthropomorphism and the social robot," *Robotics and autonomous systems*, vol. 42, no. 3, pp. 177–190, 2003.
- [5] B. Blumberg, M. Downie, Y. Ivanov, M. Berlin, M. P. Johnson, and B. Tomlinson, "Integrated learning for interactive synthetic characters," *ACM Transactions on Graphics (TOG)*, vol. 21, no. 3, pp. 417–426, 2002.
- [6] C. Isbell, C. R. Shelton, M. Kearns, S. Singh, and P. Stone, "A social reinforcement learning agent," in *Proc. 5th international conference on Autonomous agents*. ACM, 2001, pp. 377–384.
- [7] F. Kaplan, P.-Y. Oudeyer, E. Kubinyi, and A. Miklósi, "Robotic clicker training," *Robotics and Autonomous Systems*, vol. 38, no. 3, pp. 197–206, 2002.
- [8] I. Pavlov and G. Anrep, "Conditioned reflexes," *Dover Publ*, 1927.
- [9] K. M. Christian and R. F. Thompson, "Neural substrates of eyeblink conditioning: acquisition and retention," *Learning & memory*, vol. 10, no. 6, pp. 427–455, 2003.
- [10] C. H. Yeo and G. Hesslow, "Cerebellum and conditioned reflexes," *Trends in cognitive sciences*, vol. 2, no. 9, pp. 322–330, 1998.
- [11] I. Herrerros and P. F. Verschure, "Nucleo-olivary inhibition balances the interaction between the reactive and adaptive layers in motor control," *Neural Networks*, vol. 47, pp. 64–71, 2013.
- [12] M. E. Houk J, *Cerebellum*. Fundamental neuroscience. Academic Press, 2003.
- [13] C. Hofstoetter, M. Mintz, and P. F. Verschure, "The cerebellum in action: a simulation and robotics study," *European Journal of Neuroscience*, vol. 16, no. 7, pp. 1361–1376, 2002.
- [14] I. Herrerros, G. Maffei, S. Brandi, M. Sanchez-Fibla, and P. F. Verschure, "Speed generalization capabilities of a cerebellar model on a rapid navigation task," in *Intelligent Robots and Systems (IROS)*. IEEE, 2013, pp. 363–368.
- [15] R.-D. Pinzon-Morales and Y. Hirata, "A realistic bi-hemispheric model of the cerebellum uncovers the purpose of the abundant granule cells during motor control," *Frontiers in neural circuits*, vol. 9, 2015.
- [16] H. Barron-Gonzalez, J. Porrill, N. F. Lepora, E. Chinellato, G. Metta, and T. J. Prescott, "Cerebellum-based adaptation for fine haptic control over the space of uncertain surfaces," in *World Haptics Conference (WHC)*, 2013. IEEE, 2013, pp. 353–358.
- [17] S. Kudoh, K. Ogawara, M. Ruchanurucks, and K. Ikeuchi, "Painting robot with multi-fingered hands and stereo vision," *Robotics and Autonomous Systems*, vol. 57, no. 3, pp. 279–288, 2009.
- [18] S. Nishide, K. Mochizuki, H. Okuno, and T. Ogata, "Insertion of pause in drawing from babbling for robots developmental imitation learning," *Proc. of IEEE International Conference on Robotics and Automation (ICRA)*, 2014.
- [19] G. Metta, G. Sandini, D. Vernon, L. Natale, and F. Nori, "The icub humanoid robot: an open platform for research in embodied cognition," in *Proc. 8th workshop on performance metrics for intelligent systems*. ACM, 2008, pp. 50–56.
- [20] D. Floreano, A. J. Ijspeert, and S. Schaal, "Robotics and neuroscience," *Current Biology*, vol. 24, no. 18, pp. R910–R920, 2014.
- [21] P. F. Verschure, T. Voegtlin, and R. J. Douglas, "Environmentally mediated synergy between perception and behaviour in mobile robots," *Nature*, vol. 425, no. 6958, pp. 620–624, 2003.
- [22] W. B. Knox, P. Stone, and C. Breazeal, "Training a robot via human feedback: A case study," in *Social Robotics*. Springer, 2013, pp. 460–470.
- [23] P. F. Verschure, C. M. Pennartz, and G. Pezzulo, "The why, what, where, when and how of goal-directed choice: neuronal and computational principles," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 369, no. 1655, p. 20130483, 2014.
- [24] M. Sanchez-Fibla, U. Bernardet, E. Wasserman, T. Pelc, M. Mintz, J. C. Jackson, C. Lansink, C. Pennartz, and P. F. Verschure, "Allostatic control for robot behavior regulation: a comparative rodent-robot study," *Advances in Complex Systems*, vol. 13, pp. 377–403, 2010.
- [25] V. Vouloutsi, S. Lallée, and P. F. Verschure, "Modulating behaviors using allostatic control," in *Biomimetic and biohybrid systems*. Springer, 2013, pp. 287–298.
- [26] I. Gormezano, W. F. Prokasy, and R. F. Thompson, *Classical conditioning*. Psychology Press, 1987.
- [27] S. Jordà, "On stage: the reactable and other musical tangibles go real," *International Journal of Arts and Technology*, vol. 1, no. 3-4, pp. 268–287, 2008.
- [28] M. Fujita, "Adaptive filter model of the cerebellum," *Biological cybernetics*, vol. 45, no. 3, pp. 195–206, 1982.
- [29] J.-Y. Puigbo, I. Herrerros, C. Moulin-Frier, and P. Verschure, "Towards a two-phase model of sensor and motor learning," in *Biomimetic and Biohybrid Systems*. Springer, 2015, vol. 9222, pp. 453–460.