

An Inhibition of Return Mechanism for the Exploration of Sensorimotor Contingencies

Abstract—The modelling of cognition is fundamental to designing robots that are increasingly more autonomous. Indeed, researchers take inspiration from human and animal cognition in order to endow robots with the ability to learn and adapt to their environment. In specific cases, the robot has to find the right compromise between exploring the environment, or exploiting its own experience to advance its knowledge of a skill. Our approach considers a neurally-inspired model to learning sensorimotor contingencies based on exploration and exploitation. For the exploration, an inhibition of return mechanism is implemented that generates new actions. In this work, we investigate how the tuning of the inhibition of return affects the exploratory behavior. To do so, we set up an experiment where a 3D printed humanoid robot arm GummiArm has to learn how to move a baby mobile toy with only a visual feedback. The results demonstrate that the tuning of the inhibition of return influences the exploratory behavior, leading to a faster learning of sensorimotor contingencies as well as the exploration of a reduced motor space.

Keywords : *Developmental Robotics, Cognitive Robotics, Embodiment, Exploration, Dynamic Neural Fields*

I. INTRODUCTION

The learning of new skills in robotics is a challenging task. The understanding of human cognition provides new insights on how to address these issues. For example, developmental robotics is a research field that focuses on reproducing developmental stages observed in infants [1]. The purpose of such a paradigm can help to develop new machine learning algorithms, but also provides a biological interpretation of infant cognition [2].

In this research, we propose a neural model for learning sensorimotor contingencies based on dynamic neural fields with a humanoid robotic arm. At first, the architecture proceeds an exploration stage by performing motor babbling [3], then an exploitation phase generates an optimal sequence of actions based on the knowledge acquired during the exploration. The model embodies the sensorimotor experience with the robot's body and takes inspiration from reinforcement learning. Exploring the environment is challenging since the decisions taken at that stage influence the representation and knowledge learned. Therefore, we study the ability for an inhibition of return mechanism to produce a "meaningful" choice regarding the selection of future actions.

Inhibition of return (IOR) is a neural effect involved in spatial attention [4]. Posner et al. measured the time to look at a specific peripheral location from a target, then asked the subject to look back at the center. They then measured the time needed to again look at the previous peripheral location. They discovered that directly after looking at the cued location (<250 ms), the subject took less time to look

again toward this one. But after a short delay (250-300ms), the time needed to look at this peripheral location rises. The IOR function can be interpreted as a visual foraging facilitator [5], avoiding already seen locations for visual search. Various robotic models of visual attention implement the IOR effect [6], [7], [8]. In this work, we use the inhibition of return effect to avoid selecting an action already performed during the motor babbling behavior.

Piaget was the first to introduce the notion of motor babbling by formulating the "primary circular-reaction hypothesis" [3]. Indeed, he noticed that children generate "reflexes" and these reflexes change (even slightly) when they produce an effect on the children's environment. This constructivism approach suggests that infants evolve and learn about the world by developmental stages. They first associate their actions with the perceptual outcomes before learning more complex tasks. This motor babbling behavior has been studied by researchers in robotics to predict future motor states to influence the exploration strategy [9]. Other work demonstrated that motor babbling could lead to the acquisition of more complex skills such as reaching with obstacles and grasping [10]. To develop our model of motor babbling with the IOR mechanism for action generation, we adopt the Dynamic Field Theory.

Dynamic Field Theory (DFT) is a new approach to understand cognitive and neural dynamics [11]. This is suitable to deliver homeostasis to the architecture and provide various ways of learning. The basic learning mechanism in DFT is the formation of memory traces of positive activation of a Dynamic Neural Field [12] as well as Hebbian Learning [13], [14].

In this work, we propose to implement a motor babbling behavior based on exploration and exploitation with Dynamic Field Theory. An inhibition of return mechanism supports the generation of actions during exploration. We set up an experiment where the robot is attached to a baby mobile toy with a rubber band, similar to the baby mobile experiment with infants [15]. The proposed architecture is self-regulated and uses Dynamic Neural Fields in a closed loop, meaning the actions influence future perceptions. In particular, we propose the following contributions:

- A dynamic exploration architecture by varying the strength of the IOR.
- A dynamic exploitation mechanism using new neural dynamics.
- The grounding of visual stimuli with motor actions in a memory field.
- Implementation and experimental results of the dynamic

exploration architecture.

The paper is organized as follows. Section II describes the methodological background, with the dynamic field theory and the associated related work. Section III presents the model design, that includes the action selection strategy and the exploration and exploitation stage that compose the learning mechanism. Following, Section IV presents the experimental setup and the results of the experiments. Finally, Section V discusses the limitations of our work, future efforts, and concludes the paper.

II. METHODOLOGICAL BACKGROUND

Dynamic Field Theory is a theoretical framework that provides a mathematically explicit way to model the evolution in time of neural population activity [11]. It was originally used to model reactive motor behavior [16] but demonstrated its ability to model complex cognitive processes [17]. The core elements of DFT are Dynamic Neural Fields (DNF) that represent activation distributions of neural populations. Stable peaks of activation form as a result of supra-threshold activation and lateral interactions within a field. A DNF can represent different features and a peak of activation at a specific location corresponds to the current observation. For example, a DNF can be used to represent a visual color space (Red, Green, Blue) and a peak at the "blue location" would mean a blue object is perceived. Neural Fields are particularly suitable to represent continuous space.

Dynamic Neural Fields evolve continuously in time under the influence of external inputs and lateral interactions within the Dynamic Field as described by the integro-differential equation :

$$\tau \dot{u}(x, t) = -u(x, t) + h + S(x, t) + \int f(u(x, t)) \omega(x - x') dx' \quad (1)$$

where h is the resting level ($h < 0$) and $S(x, t)$ is the external inputs. $u(x, t)$ is the activation field over feature dimension x at time t and τ is a time constant. An output signal $f(u(x, t))$ is determined from the activation via a sigmoid function with threshold at zero. This output is then convolved with an interaction kernel ω that consists of local excitation and surround inhibition [18]. The role of the Gaussian kernel is crucial since different shapes influence the neural dynamics of a field. For example, local excitatory (bell shape) coupling stabilizes peaks against decay while lateral inhibitory coupling (mexican-hat shape) prevents activation to spread out along the neural field. By coupling or projecting together several neural fields of different features and dimensions, DFT is able to model cognitive processes. If neural fields are the core of the theory, other elements are essential to our work.

Dynamic neural nodes are basically a 0-dimensional dynamic neural field and follow the same dynamics:

$$\tau \dot{u}(x, t) = -u(x, t) + h + c_{uu} f(u(t)) + \sum S(x, t) \quad (2)$$

The terms are similar to a Neural Field except for c_{uu} which is the weight of a local nonlinear excitatory interaction. A node can be used as a boost to another Neural Field. By projecting its activation globally, the resting level of the neural field will rise allowing to see the rise of peaks of activation.

Finally, the memory trace is an other important component of DFT:

$$\dot{v}(t) = \frac{1}{\tau_+} (-v(t) + f(u(t))) f(u(t)) + \frac{1}{\tau_-} (-v(t)(1 - f(u(t)))) \quad (3)$$

with $\tau_+ < \tau_-$. A memory trace in DFT has two different time scales, a build up time τ_+ that corresponds to the time for an activation to rise in the memory and a decay time τ_- which is the time decay of an activation. In our model, we use a 2-dimensional memory trace which keeps track of visual activation.

III. MODEL

In this work, we propose a cognitive architecture allowing a robot to learn a specific movement with a visual motion detector. The robot resembles a human arm, where the upper arm roll motor is used for exploration and exploitation. For clarity, we split our architecture according to the different phases: exploration with the inhibition of return as an action generation mechanism, and the exploitation of the motor babbling outcomes. Every 2-Dimensional field is divided by states and actions along respectively the horizontal and vertical axis. The horizontal axis corresponds to the state of the upper arm roll encoder. The vertical dimension gives the possible actions to be selected. All neural fields are defined in the interval $[0;100]$ and represent a motor angle of the upper arm motor within the interval $[-1;1]$. For example, If a peak emerges at position $[50;10]$, that means at motor state 50 (0 on the joint angle), the action 10 (-0.8) is taken.

A. Exploration

As described earlier, the exploration phase performs motor babbling and observes the sensor outcomes of each action (Figure 1). At first, the action formation (AF) field serves as an action generation mechanism. The slow boost module slowly rises the resting level of the AF field and IOR Inhib field until a peak of activation emerges. These neural fields (AF, IOR excit, memory trace action, IOR Inhib field) form the inhibition of return mechanism. After an action is performed, the activation of that action is stored into the memory trace actions and projected back to action formation field as an inhibition. This prevents the selection of the same action in the future. Once a peak appears in AF field, the slow boost module stops rising the resting level and the peak spreads to the motor intention field. In some rare cases, multiple peaks could emerge from the AF field and the selective neural field allows the selection of only one of them [11]. The Condition of Satisfaction (CoS) field receives two inputs from the encoder and the motor intention field.

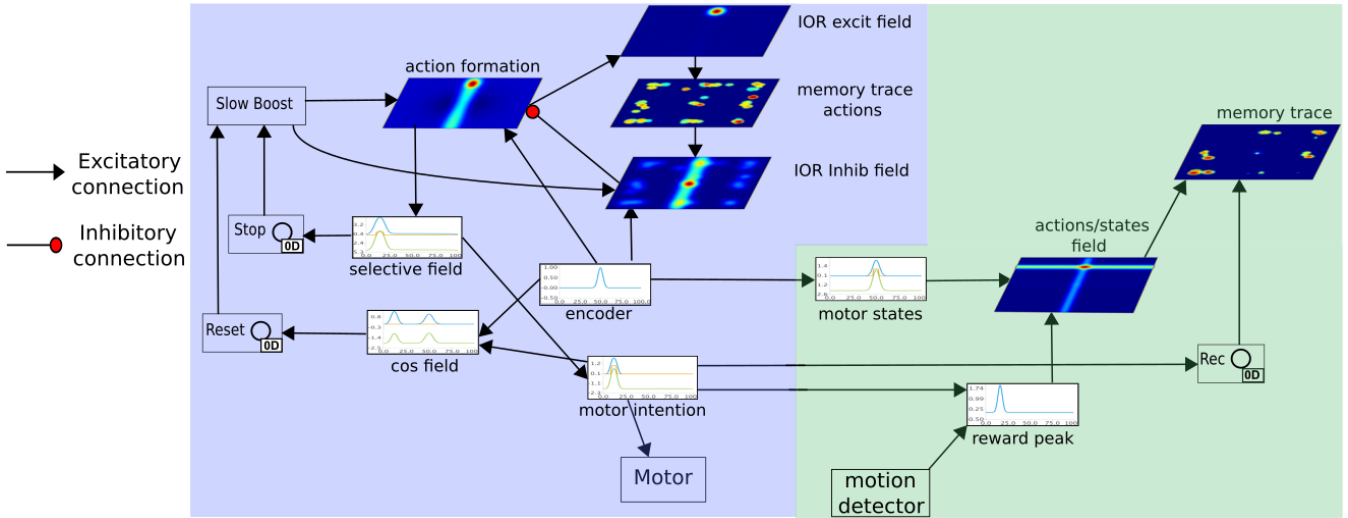


Fig. 1: Exploration stage divided by the action generation mechanism with the inhibition of return (blue) and the recording of the visual outcomes (green). A peak of activation from actions/states field spreads in memory trace only when the Rec node is active. This means storing the visual activation exactly while an action is performing. The decay of the memory trace τ_- is 10 seconds, and only happens when the Rec node is active. The motor module converts the neural field value to the desired angle position.

When both of them reach the same location within the CoS field, the activation goes beyond threshold and activates a node that will reset the slow boost module.

After the generation of an action, the action/states field records the visual outcomes of the action performed. Concerning the reward peak module, it receives input from the motion detector and the motor intention field. This is where the grounding of visual perception is happening. The implementation gathers the motor state position and the visual perception value to form a Gaussian curve centered on the motor’s position with an amplitude corresponding to the motion detector’s value. If the visual outcomes are strong enough, a peak forms in the action/states field and is then stored in the memory trace. A high neural activation within the memory trace represent an action with a high visual outcome.

After exploring all possible sensorimotor outcomes, the exploitation stage harnesses the neural activation to optimize the robot’s motion.

B. Exploitation

The purpose of the exploitation phase (Figure 2) is rather simple : follow the highest neural activation until reaching a stable sequence of actions. Indeed, the 2-dimensional memory trace is divided along horizontal and vertical axis by state and action spaces respectively, a high activation represents an action with a high visual value.

The slow boost module remains the same as in the exploration phase and slowly rises the resting level of the memory field. For a given state, rising the resting level sees the emergence of the highest peak corresponding to the action with the highest neural activation. The action is then performed and the state position of the upper arm

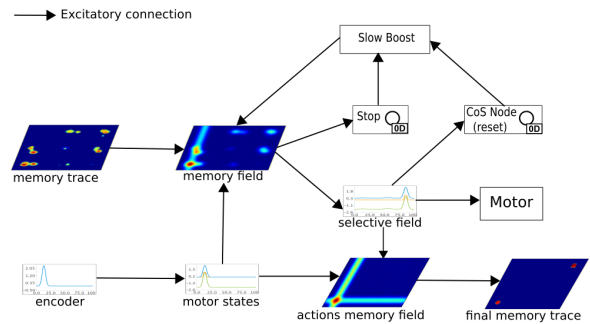


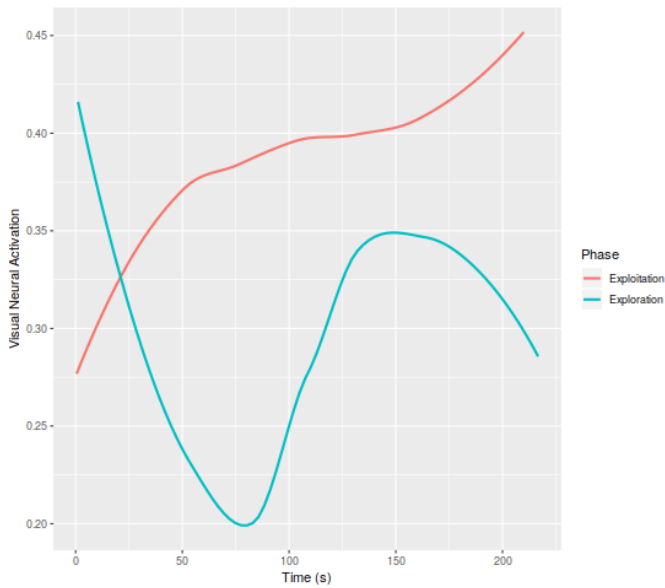
Fig. 2: Exploitation architecture. Some of the neural fields and memory trace are present in the exploration mechanism.

roll motor is updated. While an action is performed, it is stored into a final memory trace. The final memory trace gathers the actions performed the same way as seen in the exploration mechanism with the memory trace actions. The only difference is the time decay τ_- fixed to 4 seconds. This means that the motor states which are visited only once during exploitation decay and disappear while the most visited remain active. The activation within the final memory trace represents the actions with the highest visual value, and thus form a sequence.

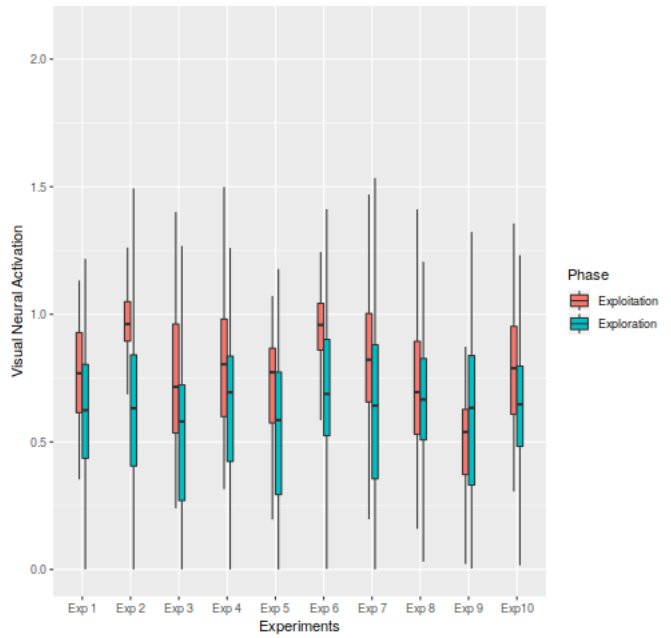
In the next section, we will tune the strength of the inhibition of return and evaluate its influence during exploration.

IV. EXPERIMENT AND RESULTS

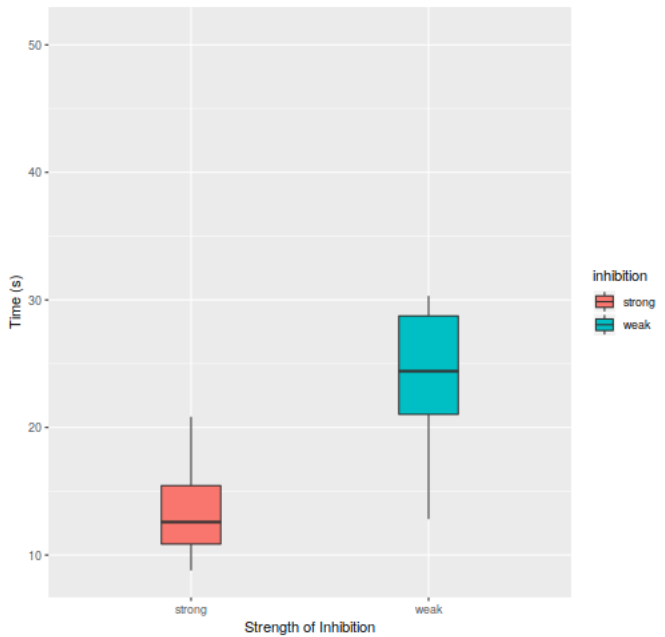
The gummiArm robot [19] is a 7 degrees of freedom (+2 for the head) 3D printed arm. In our case, only the upper arm roll joint will be used for demonstration of the architecture.



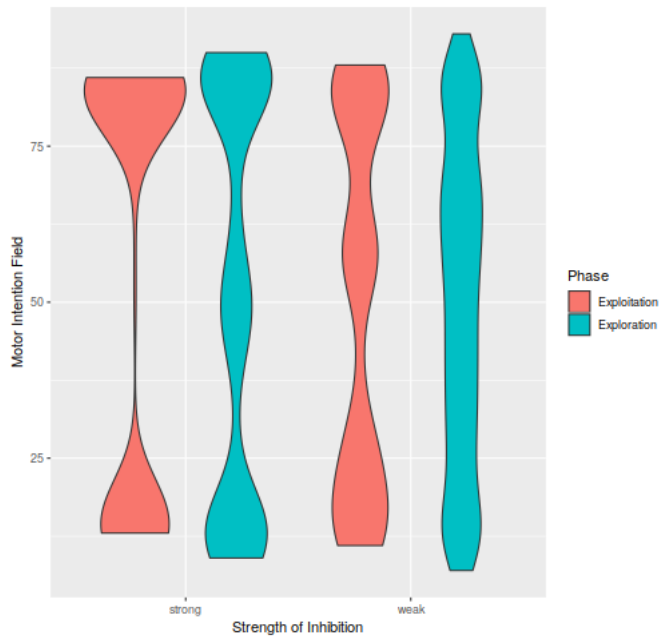
(a)



(b)



(c)



(d)

Fig. 3: (a.) linear regression for 10 runs of exploration with a weak inhibition. The figure show an improvement over time of visual activation during the exploitation of the exploratory behavior with a weak inhibition of return (b.) Distribution of visual neural activation for 10 runs of exploration with a strong inhibition. Each experiment demonstrates a higher visual neural activity while harnessing the sensorimotor contingencies (c.) Elapsed time before reaching a stable sequence of action during the exploitation of 10 explorations with a weak IOR as well as the exploitation of 10 exploration with a strong IOR. For both exploratory behavior, the exploitation stage follow a steady sequence of action over time. (d) Motor distribution during exploration and exploitation for both exploratory behavior.

A rubber band is attached from the palm of the hand to one of the moving toys in the baby mobile. The camera mounted inside the head (Intel RealSense D435), is used as the motion detector that subtracts two consecutive images and applies a threshold to observe the changed pixels. The result returns the sum of changed pixels which is scaled from 0 to 1 and represents the visual neural activation. The toys hanging from the baby mobile are within the visual field of the camera whereas the arm itself is out of sight.

In order to tune the inhibition of return, we vary the strength of the synapses between IOR Inhib field and action formation. In practice, this means applying a convolution of a negative gaussian kernel to the output activation of IOR Inhib field. We run 10 explorations with a weak IOR (convolution with a gaussian kernel of amplitude 1.5 and standard deviation σ of 3.0) and 10 explorations with a strong IOR (gaussian kernel of amplitude 3.5 and σ 15.0). For each exploration, we run an exploitation stage. In total, we have 20 explorations (10 with a weak IOR, 10 with a strong IOR) and 20 exploitations (10 from the 10 exploration with a weak IOR and 10 from the exploration phase with a strong IOR).

A. Exploration with a weak Inhibition of return

The results demonstrates a general gain of visual neural activation during exploitation over exploration (Figure 3a-b). However, the visual neural activation are lower for an exploratory behavior with a weak IOR. These results can be explained by the actions selected during exploration (Figure 3-d). Indeed, the motor distribution for an exploration with a weak IOR is more uniform than with a strong IOR (Figure 4 and 5). When a peak projects a weak inhibition from IOR Inhib to action formation, the next action selected is located close to that inhibition. This explains the low visual neural activation during exploration, because almost the complete motor space has been explored even with no visual feedback. Regarding the exploitation of the sensorimotor contingencies, the distribution of motor intention is mostly reduced to [10;25],[50;62] and [75;87]. These motor spaces correspond respectively to the extreme left, center-right and extreme right location of the arm.

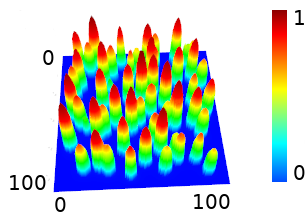


Fig. 4: Memory Trace actions of all the actions selected during an exploration with a weak IOR.

B. Exploration with a strong inhibition of return

If we look into the results of the exploration with a strong IOR (Figure 3b), we can see a higher visual neural

activation during exploitation over exploration for almost all the experiments. The motor intention distribution is more restricted than for the exploration with a weak IOR (Figure 3d, Figure 5 and 4). This means that only the extreme left, center and extreme right locations of the arm are explored. This time, when a peak projects a strong inhibition to the action formation field, the position of the next generated action is distant from that inhibition. This results in an even more restricted motor space during exploitation where only the extreme left and extreme right position are visited. We can wonder why the exploitation stage does not lead to the same motor intention distribution after both exploratory behaviors. This motor intention difference is due to the representation of an action. In our architecture, we choose to perform an action by selecting the next position of the upper arm roll. For example, this means that the extreme left location can be visited multiple times, but not from the same motor angle origin, leading to different visual stimuli. Therefore, it explains the motor intention distribution of the exploitation with a strong IOR. The actions selected were wide (extreme left to extreme right and vice versa) during exploration, thus providing strong visual neural activation and future sensorimotor contingencies to exploit.

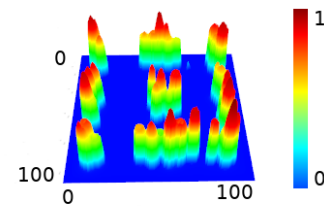


Fig. 5: Memory Trace actions for an exploratory behavior with a strong IOR.

C. Global Interpretation

Despite the exploratory behavior (strong or weak IOR), the exploitation stage always reaches a stable sequence of actions. We compare the time needed by the exploitation to follow an unchanged sequence of actions in Figure 3(c). The results demonstrate a faster convergence toward a stabilized sequence of actions if the exploration uses a strong inhibition of return. This is particularly true for the settings of our experiment (head of the robot centered on the baby mobile toy) but might not be the case otherwise. The main findings about tuning the Inhibition of return are :

- 1) The motor space explored with a weak inhibition of return is wide and uniform.
- 2) A more restricted motor actions space with a strong IOR. Due to the continuous nature of the Dynamic Neural Fields, a strong inhibition at a certain location would favor the emergence of an action far-off that location.

- 3) A faster exploitation of the sensorimotor contingencies for an exploratory behavior with a strong inhibition of return.

We will summarize our findings in the next section and discuss the implications for future work.

V. CONCLUSION AND DISCUSSION

This work proposes a cognitive architecture with an embodiment approach that allows a robotic arm to optimize its motion based on the neural activation coming from a motion detector. As such, the approach is grounding the sensorimotor experience within Dynamic Neural Fields. In experiments, a GummiArm robot is moving a baby mobile and observes the outcome of the actions taken to optimize its motion. An Inhibition of return mechanism selects the next action to perform, then the model records the visual outcome in a visual memory trace. Indeed, the sensorimotor contingencies can be encoded as neural activation within neural fields and explored through motor babbling. Then, an exploitation mechanism optimizes the motion of the robot, following the path left by high neural activation. Exploiting the high neural activations means choosing actions leading to the best visual reward. Finally, the tuning of the inhibition of return for generating a new action demonstrates two interesting exploratory behaviors. A wide and uniform exploration of the motor space happens when the IOR is weak. If the strength of the inhibitory connection is strong, the motor space explored is reduced and gathers extreme positions, but could accelerate the exploitation of the sensorimotor contingencies.

The influence from the inhibition of return leads to interesting future developments. For example, we currently apply it to model low level sensorimotor cognition such as motor babbling. Since motions are mostly goal directed [20], [21], we will modify our model to use the complete arm kinematics and the inhibition of return to influence the selection of the next goal to pursue. In future work, we plan to adapt the architecture to learn how to interact with objects (e.g. touch, push aside, push forward). With a strong inhibition, we could generate goals that would prefer wide motions and therefore explore extreme locations. On the contrary, a weak inhibition would generate a movement close to a certain area of interest and allows to explore the complete motor space.

Finally, we intend to extend our model to support an autonomous exploration of the environment. In that case, the synaptic inhibition should be able to adapt itself regardless of the environment and the intended goal. Indeed, the strength of the inhibition of return should facilitate the exploration when learning a new task. To do so, we look forward to developing an Hebbian learning mechanism that will adapt the strength of the inhibition of return.

VI. APPENDICES

The complete architecture, the source code and all the parameters are available at (*anonymous content but will be shared*).

REFERENCES

- [1] M. Asada, K. Hosoda, Y. Kuniyoshi, H. Ishiguro, T. Inui, Y. Yoshikawa, M. Ogino, and C. Yoshida, "Cognitive developmental robotics: A survey," *IEEE transactions on autonomous mental development*, vol. 1, no. 1, pp. 12–34, 2009.
- [2] E. Thelen, G. Schöner, C. Scheier, and L. B. Smith, "The dynamics of embodiment: A field theory of infant perseverative reaching," *Behavioral and brain sciences*, vol. 24, no. 1, pp. 1–34, 2001.
- [3] J. Piaget and M. Cook, *The origins of intelligence in children*. International Universities Press New York, 1952, vol. 8, no. 5.
- [4] M. I. Posner, R. D. Rafal, L. S. Choate, and J. Vaughan, "Inhibition of return: Neural basis and function," *Cognitive Neuropsychology*, vol. 2, no. 3, pp. 211–228, 1985.
- [5] R. M. Klein, "Inhibition of return," *Trends in Cognitive Sciences*, vol. 4, no. 4, pp. 138 – 147, 2000. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1364661300014522>
- [6] A. Ude, V. Wyart, L.-H. Lin, and G. Cheng, "Distributed visual attention on a humanoid robot," in *5th IEEE-RAS International Conference on Humanoid Robots, 2005*. IEEE, 2005, pp. 381–386.
- [7] F. Orabona, G. Metta, and G. Sandini, "Object-based visual attention: a model for a behaving robot," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops*. IEEE, 2005, pp. 89–89.
- [8] J. Ruesch, M. Lopes, A. Bernardino, J. Hornstein, J. Santos-Victor, and R. Pfeifer, "Multimodal saliency-based bottom-up attention a framework for the humanoid robot icub," in *2008 IEEE International Conference on Robotics and Automation*. IEEE, 2008, pp. 962–967.
- [9] R. Saegusa, G. Metta, G. Sandini, and S. Sakka, "Active motor babbling for sensorimotor learning," in *2008 IEEE International Conference on Robotics and Biomimetics*, Feb. 2009, pp. 794–799.
- [10] D. Caligiore, T. Ferrauto, D. Parisi, N. Accornero, M. Capozza, and G. Baldassarre, "Using motor babbling and hebb rules for modeling the development of reaching with obstacles and grasping," in *International Conference on Cognitive Systems*, 2008, pp. E1–8.
- [11] G. Schöner, J. Spencer, and D. F. T. R. Group, *Dynamic Thinking: A Primer on Dynamic Field Theory*. Oxford University Press, 2016, google-Books-ID: ySexCgAAQBAJ.
- [12] S. Perone and J. P. Spencer, "Autonomy in action: Linking the act of looking to memory formation in infancy via dynamic neural fields," *Cognitive Science*, vol. 37, no. 1, p. 1–60, 2013.
- [13] M. Luciw, K. Lahkman, S. Kazerounian, M. Richter, and Y. Sandamirskaya, "Learning the perceptual conditions of satisfaction of elementary behaviors," p. 6, 2013.
- [14] Y. Sandamirskaya and J. Conradt, "Learning sensorimotor transformations with dynamic neural fields," in *Artificial Neural Networks and Machine Learning – ICANN 2013*, V. Mladenov, P. Koprinkova-Hristova, G. Palm, A. E. P. Villa, B. Appollini, and N. Kasabov, Eds. Springer Berlin Heidelberg, 2013, vol. 8131, pp. 248–255.
- [15] H. Watanabe and G. Taga, "General to specific development of movement patterns and memory for contingency between actions and events in young infants," *Infant Behavior and Development*, vol. 29, no. 3, pp. 402–422, July 2006.
- [16] K. Kopecz and G. Schöner, "Saccadic motor planning by integrating visual information and pre-information on neural dynamic fields," *Biological cybernetics*, vol. 73, no. 1, pp. 49–60, 1995.
- [17] J. Spencer, S. Perone, and J. Johnson, *Dynamic Field Theory and Embodied Cognitive Dynamics*, 06 2009.
- [18] S.-i. Amari, "Dynamics of pattern formation in lateral-inhibition type neural fields," *Biological Cybernetics*, vol. 27, no. 2, pp. 77–87, June 1977.
- [19] M. F. Stoelen, F. Bonsignorio, and A. Cangelosi, "Co-exploring actuator antagonism and bio-inspired control in a printable robot arm," in *From Animals to Animats 14*, E. Tuci, A. Giagkos, M. Wilson, and J. Hallam, Eds. Cham: Springer International Publishing, 2016, pp. 244–255.
- [20] D. Premack and A. J. Premack, "Infants attribute value to the goal-directed actions of self-propelled objects," *Journal of Cognitive Neuroscience*, vol. 9, no. 6, pp. 848–856, 1997, pMID: 23964604. [Online]. Available: <https://doi.org/10.1162/jocn.1997.9.6.848>
- [21] S. Biro and A. M. Leslie, "Infants' perception of goal-directed actions: development through cue-based bootstrapping," *Developmental science*, vol. 10, no. 3, pp. 379–398, 2007.