

Exploration and Exploitation of Sensorimotor Contingencies for a Cognitive Embodied Agent

Quentin Houbre, Alexandre Angleraud and Roel Pieters

Cognitive Robotics Group, Automation Technology and Mechanical Engineering, Tampere University, Tampere, Finland

Keywords: Cognitive Robotics, Embodiment, Sensorimotor Contingencies, Dynamic Neural Fields.

Abstract: The modelling of cognition is playing a major role in robotics. Indeed, robots need to learn, adapt and plan their actions in order to interact with their environment. To do so, approaches like embodiment and enactivism propose to ground sensorimotor experience in the robot's body to shape the development of cognition. In this work, we focus on the role of memory during learning in a closed loop. As sensorimotor contingencies, we consider a robot arm that moves a baby mobile toy to get visual reward. First, the robot explores the continuous sensorimotor space by associating visual stimuli to motor actions through motor babbling. After exploration, the robot uses the experience from its memory and exploits it, thus optimizing its motion to perceive more visual stimuli. The proposed approach uses Dynamic Field Theory and is integrated in the GummiArm, a 3D printed humanoid robot arm. The results indicate a higher visual neural activation after motion learning and show the benefits of an embodied babbling strategy.

1 INTRODUCTION

The role of robotics in society is increasing, and with it comes the problem of modelling intelligence. With more complex tasks to perform, robots need to adapt to their environment. To address these issues, researchers are focusing on cognition, autonomy, their development in humans and how to model them in robots.

Developmental approaches in robotics try to reproduce experimental results observed in infants to understand cognition. The notion of Sensorimotor Contingency ties together perceptions and motor actions in a situated agent. For example, the motor babbling behavior provides an explanation for the learning of sensorimotor contingencies by associating actions with their outcomes. Piaget (Piaget and Cook, 1952) formulated the "primary circular-reaction hypothesis" where children generate "reflexes" and these reflexes change (even slightly) when they produce an effect on the children's environment. Later, the hypothesis was experimentally confirmed (Von Hofsten, 1982). This behavior led researchers to investigate such early behavior as a learning mechanism in robotics. One cognitive architecture implements this mechanism with Bayesian Belief Networks (Demiris and Dearden, 2005) where a robot learns to associate motor commands with their sensory consequences and how the inverse association can be used for imitation. Other research (Saegusa

et al., 2009) applied motor babbling with neural networks to predict future motor states to influence the exploration strategy, avoiding to learn all the motor states and perception associations. Around the same time, researchers proposed a model (Caligiore et al., 2008) using motor babbling to support the learning of more complex skills such as reaching with obstacles and grasping. The problem was mainly to demonstrate that motor babbling is suitable to generate action sequences in time. More recent work (Mahoor et al., 2016) proposed a neurally plausible model of reaching by encoding the trajectory of the movements within three interconnected neuron maps. These interesting but non-exhaustive works provide interesting ways on how motor babbling could be implemented in robotics, allowing us to propose our own model based on Neural Dynamics with an enactive approach.

To understand enactivism, the notion of embodiment needs to be defined. Embodiment (Francisco J. Varela and Rosch, 1991) is an approach where the body is an interface that shapes the development of cognition through its interactions with the environment. More than an interface, the body is a structured living organism and so must be considered embodiment in robotics (Chrisley and Ziemke, 2006; Morse et al., 2011; Ziemke, 2016). For example, in (Lafraquière and Hemion, 2015) researchers proposed an architecture to ground object perception from a robot's sensorimotor experience. As a simple form

of embodiment, proprioceptive signals from the robot to improve the learning to count, was demonstrated in (Ruciński et al., 2012). To go beyond embodiment, some researchers proposed a new multi-disciplinary approach named computational grounded cognition (Pezzulo et al., 2013), where every aspect of cognition is grounded through sensorimotor experience. Even if such approach is beneficial to help the design of cognitive architectures, the concept of autonomy needs to be addressed and integrated with grounding. Enactivism is an answer to this issue, even if here we distinguish sensorimotor enactivism, which puts perceptual abilities at the center, with autopoietic enactivism where there is a necessary link between conscious experiences and autopoietic processes (Degeenaar and O'Regan, 2017). This means the grounding of experience depends on the internal system's organization and the ability to change it, but also on the regulation of the system itself (Barandiaran, 2017). The first statement remains problematic and complex to deal with in robotics, but the second one can be addressed. In this case, we talk about homeostasis (Cannon, 1929), the process of self-regulation. The design of a cognitive architecture must take into account the circular causality of the sensorimotor experience if this exhibits autonomy with an enactive approach (Vernon et al., 2015). In order to produce these self-regulated dynamics, the use of neural fields is promising.

Dynamic Field Theory (DFT) is a new approach to understand cognitive and neural dynamics (Schöner et al., 2016). This is suitable to deliver homeostasis to the architecture and provides various ways of learning. The most basic learning mechanism in DFT is the formation of memory traces of positive activation of a Dynamic Neural Field (Perone and Spencer, 2013). Hebbian Learning is possible (Luciw et al., 2013) and the learning of sequences could be done via a structure involving elementary behaviors, intentions and conditions of satisfaction (Sandamirskaya and Schöner, 2010).

In this work, we propose a new mechanism of exploration and exploitation with Dynamic Field Theory. 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 (Watanabe and Taga, 2006). We investigate how memory is shaping the experience of the robot and thus how this helps to optimize the robot's motion. 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 based on mo-

tor babbling.

- The grounding of visual stimuli with motor actions in a memory field.
- A dynamic exploitation mechanism using new neural dynamics and taking inspiration from Reinforcement Learning (Q-Learning).
- Implementation and experimental results of the dynamic exploration architecture.

The paper is organized as follows. Section 2 describes the methodological background, with the dynamic field theory and the associated related work. Section 3 presents the model design, that includes the action selection strategy and the exploration and exploitation stage that compose the learning mechanism. Following, Section 4 presents the experimental setup and the results of the experiments. Finally, Section 5 discusses the limitations of our work, future efforts, and concludes the paper.

2 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 (Schöner et al., 2016). It was originally used to model reactive motor behavior (Kopecz and Schöner, 1995) but demonstrated its ability to model complex cognitive processes (Spencer et al., 2009). 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 convoluted with an interaction kernel ω that consists of local excitation and surrounding inhibition (Amari, 1977). 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 from spreading 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 neural field and follow the same dynamic:

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

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 (Figure 4).

Finally, the memory trace is another 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.

2.1 Q-Learning

Q-Learning algorithm (Watkins and Dayan, 1992; Sutton et al., 1998) is a model-free reinforcement learning algorithm that learns a policy in order to choose the best action according to a given state. The learned action/value function Q is defined by :

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha [R_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)], \quad (4)$$

where s_t and a_t are, respectively, a state and an action at time t , α is the learning rate, r_{t+1} the reward at time $t+1$ and γ the discount factor. In practice, the learning rate determines to what extent newly acquired information overrides old information and the discount

factor γ determines the importance of future rewards. The Q-Values are stored and updated in a table (Q-Table) divided along a state and action dimension. After learning, and given a state s_t , a Q-Value represents a probability of obtaining a certain reward in time after performing an action. The next section introduces our model and draws the parallel of the Q-Learning inspiration with Dynamic Neural Fields.

3 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 simplicity, we split our architecture according to the different phases: exploration with an action generation mechanism and exploitation of the motor babbling outcomes.

3.1 Action Generation for Motor Babbling

As described earlier, motor babbling consists of associating motor actions with their perceptual outcome. In this work, we show the possibility to generate actions directly from a neural field. To do so, we combine two different neural mechanisms: A slow boost of the resting level and an Inhibition of return (Figure 1). With DFT, the tuning of the resting level is an essential component that leads to express different neural behavior (Schöner et al., 2016). It is well-known that the resting membrane potential of neurons can vary under different conditions (Wilson and Kawaguchi, 1996)(Franklin et al., 1992). Here, instead of defining a static resting level, we choose to dynamically vary the resting level of two neural fields to generate a new action. The slow boost module increases the resting level of the action formation (AF) field until a peak of activation emerges. The module ceases to increase the activation when the stop node is active and resets the activation to zero when the reset node is active. The peak within the AF field is then projected to a set of neural fields reproducing an inhibition of return. This mechanism is well studied, especially regarding visual attention (Posner et al., 1985), (Tipper et al., 1991), where immediately following an event at a peripheral location, there is facilitation for the processing of other stimuli near that location. Here, we use this effect to avoid generating the same action twice given a motor state. When a peak emerges from the AF field and is projected to the Inhibition Of Return excitatory field (IOR excit),

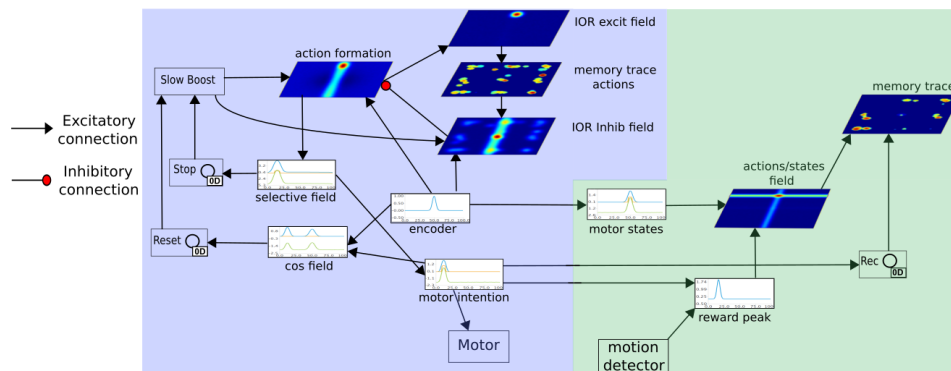


Figure 1: Exploration stage divided by the action generation mechanism with the inhibition of return (blue) and the recording of the visual outcomes (green). A peak of activation from actions/states field spreads in memory trace only when the Rec node is active. This means storing the visual activation 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.

a memory trace stores this activation. This memory trace is recording all the actions taken during the motor babbling with a slow decay τ_{-} . The memory trace then projects all the activation into an Inhibition Of Return field (IOR inhib). This last neural field closes the loop of the Inhibition of Return mechanism by projecting an inhibitory connection to the AF field. The kernel interaction within that field allows the rise of peaks of activation. When the slow boost module begins to rise the resting level of both action formation and IOR Inhib field, the activation within the action field generates a stable peak and projects it to the selective field. In some rare cases, the neural dynamics generate more than one action within the action formation field. Thus, the selective field assures the emergence of a single peak. The dynamics observed within the AF field are influenced by the speed of the increasing boost. If we increase the resting level of the fields too quickly, they begin to oscillate between a supra-threshold and a below-threshold states. Nevertheless, this mechanism allows the generation of a unique single action at a given state.

All neural fields are defined in the interval $[0;100]$ and represent a motor angle position within the interval $[-1;1]$. The action formation field is divided along the state space on the horizontal dimension and the action space along the vertical dimension. If a peak emerges at position $[50;90]$, that means at motor state 50, the action 90 is taken. The encoder module corresponds to the motor value from the upper arm roll motor (e.g motor angle with the interval $[-1;1]$).

The Condition of Satisfaction (CoS) field receives inputs from the encoder and the motor intention field. 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.

3.2 Exploration

The first stage of our model consists of exploring the sensorimotor space (Figure 1) with the action generation mechanism. The purpose of exploration is quite simple: generate an action to perform and store the visual outcome into a 2-dimensional memory trace. Thus, the architecture performs an action, stores the neural activation within a memory trace while executing the action, then stops storing the activation when the action is over.

The actions/states field, memory trace and record (Rec) node are the core components of the exploration. As stated earlier, the condition of satisfaction field (CoS Field) is a one dimensional neural field representing the motor space and basically indicates when an action is over. It receives activation from both the motor intention field and the encoder. When a new action is selected, the CoS field receives an activation from the intention field. The motor module performs the action and the encoder's new value is updated. This causes a peak to rise within the CoS field and activates a node that resets the slow boost component.

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 the motion detector's value.

The actions/states field is a 2-dimensional neural field where the horizontal axis represents the motor states and the vertical axis the motor actions. When a new action is executed, the grounding of vision/action

peak field projects along the horizontal axis of the action/state field while the current motor state projects along the vertical axis. This creates a 2 dimensional peak of activation depending of the strength of the visual input. Finally, the memory trace field stores activation from the action state neural field. A convolution (gaussian kernel) is applied to the output of the actions/states field to smooth the peak of activation in the memory trace. The rec node plays the role of trigger for the storage of neural activation since it allows them to happen only when the node is active. In our design, the rec node is active only when an action is generated (peak within the motor intention field). This way, the memory trace accepts input from the actions/states field only when an action is currently being executed (Figure 2).

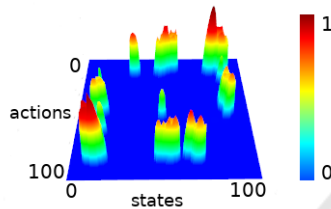


Figure 2: 3D view of the Memory Trace after exploration with 100ms of build up activation (τ_+) and 10 seconds decay (τ_-).

The inspiration from Q-Learning comes from the memory trace storing the visual outputs. Since the actions/states field is divided along, respectively, the vertical and horizontal space, the memory trace stores peaks with the amplitude of a visual activation given a specific action taken at a specific state. Similarly to a Q-Value that represent a probability to get a reward from a state/action pair, the amplitude of a peak within the memory trace reflect the probability of getting a high visual neural activation. The memory trace is then analogous to a Q-Table where the highest peak along the current state dimension represents the action with the highest visual outcome. Contrary to Q-Learning, the actions/states neural field is updated with the current visual activation, without a discount factor and a learning rate. The goal of exploration is purely to observe all sensorimotor outcomes after performing an action.

3.3 Exploitation

The last step of our architecture is the exploitation phase (Figure 3). The architecture runs by following the different activations within the memory trace. The goal is to follow the "path" left by every high activation until reaching an optimal sequence of actions.

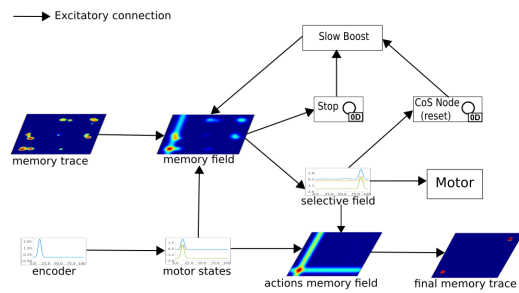


Figure 3: Exploitation phase. The slow boost node slowly rises the resting level of the memory field provoking a peak at the location of the best action taken during exploration.

The memory trace from the exploration serves as input to a 2-dimensional neural field. This memory field is the core component of the exploitation mechanism. First, a motor state field spans the current motor state over the horizontal dimension of the memory field which rises the neural activation of taken actions during the exploration stage at that particular state. The neural activation within the memory field remains below the threshold of activation. To see the rise of the highest activation, we apply a slow boost. This component remains the same seen during the action generation dynamics. This slow boost stops to increase the resting level when an activation appears (Stop node). The 2-dimensional memory field then projects the output activation to a 1-dimensional selective action field. Finally, the selective field represents the best action taken (the best visual outcome) at that state (Figure 4).

The activation within the selective action field is projected into the motor module that will perform the corresponding action. The selective field will emit the action's ending to a Condition of Satisfaction (CoS) node then to the slow boost. The boost already receives an activation from the Stop node to block the increase of the resting level. The signal coming from the CoS node resets the iterative boost to its initial resting level. The CoS node is mainly a trigger informing the boost module that an action has been performed and the selection of a new action from the new state can take place.

Finally, the actions memory field forms a peak of activation at each action taken if the visual outcomes are strong enough. It then stores it into a final memory trace field that represents the optimal sequence of actions. During the exploitation phase, the motion takes some time to reach the stabilized sequence of motor state transitions. At the beginning, it might go through some motor states that will not be visited again and this is why we use a final memory trace. With an arbitrary chosen activation decay of 4 seconds (τ_-), the motor states that are visited only once

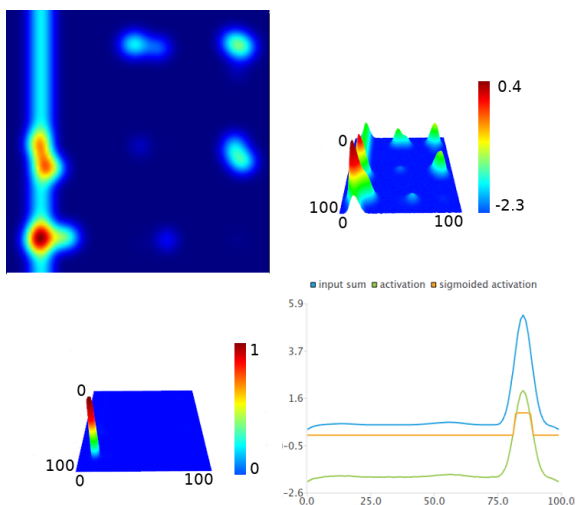


Figure 4: Top panel: projection of the current motor state on the memory field, the slow boost node is active and stopped. Bottom left: results of the sigmoid activation when the boost is active and stopped. Bottom right: projection of the action dimension to a 1-dimensional selective field. In the rare case where two peaks have exactly the same activation, only one of them will remain above threshold.

during the whole exploitation step disappear, while the most used motor states are kept active. One could see the final memory trace as a clean (for a given exploration) sequence of actions since the unused activations are pulled off.

The exploitation phase in our architecture is analogous to exploitation with Q-Learning. Indeed, the choice of the next action to perform within the memory field is always the highest peak given a motor state. With Q-Learning, this would mean choosing the action with the highest Q-value.

4 EXPERIMENT AND RESULTS

In order to validate our approach, a set of 10 explorations and 10 exploitations are performed with a robot arm that mimics a human arm and torso.

4.1 Setup

The gummiArm robot (Stoelen et al., 2016) 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 rubber band is attached from the palm of the hand to one of the moving toys in the baby mobile (Figure 5). The motor space of the upper arm roll joint is an angle position situated within the interval $[-1;1]$ where -1 and 1 represent respectively the extreme left and extreme right

position of the end-effector. The motor space interval is scaled along the Motor Intention Field from 0 to 100.

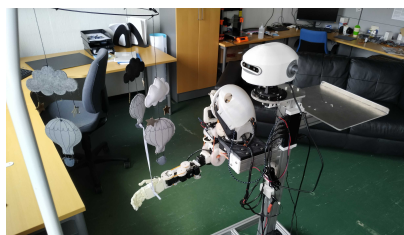


Figure 5: GummiArm robot in initial position, with the palm of the hand attached to the baby mobile (grey balloon).

The camera mounted inside the head (Intel RealSense D435), is used for 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 3 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. Despite the noisiness of the motion detector, the exploration mechanism allows the emergence of a pattern during motor babbling.

The exploration phase begins with the robot's arm at 0 position (50 over the motor intention field). As already mentioned, the memory field applies the decay time τ_- only when the trigger node associated is also active. The exploration and exploitation stages run for 210 seconds each. The length of these two stages were chosen according to the memory trace's decay (τ_-) and the slow boost mechanism. For exploration, we apply a decay τ_- of 10 seconds. This is enough to see a pattern emerging and thus support the exploitation of the sensorimotor contingencies. As mentioned earlier, the slow boost module slowly increases the resting level of the memory field, strongly influencing the time needed to choose an action. Since we use the same parameters to increase the resting level for both stages, the time needed to generate an action remain the same.

For exploring, we record the visual neural activation happening during an action (when the memory node is active) every 50ms. In the meantime, we record the activation within the motor intention field module to keep track of the actions taken with the same rate (50ms). We apply the same procedure for the exploitation (a vision module is used in the architecture to record the visual activation the exact same way as for the first phase). We run 10 explorations, then apply the exploitation stage to each of these runs.

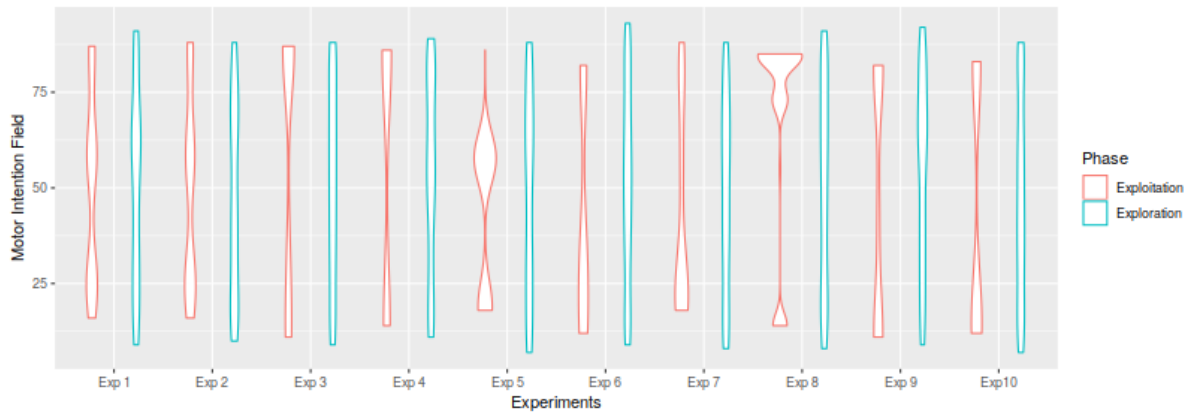


Figure 6: Violin plots for the distribution of Motor Intention. The plots show a specific motor space distribution during exploitation, while the exploration stage motor distribution is more uniform. The distribution of motor intention demonstrate a focus at three particular interval corresponding to the extreme left, center and extreme right position of the gummiArm.

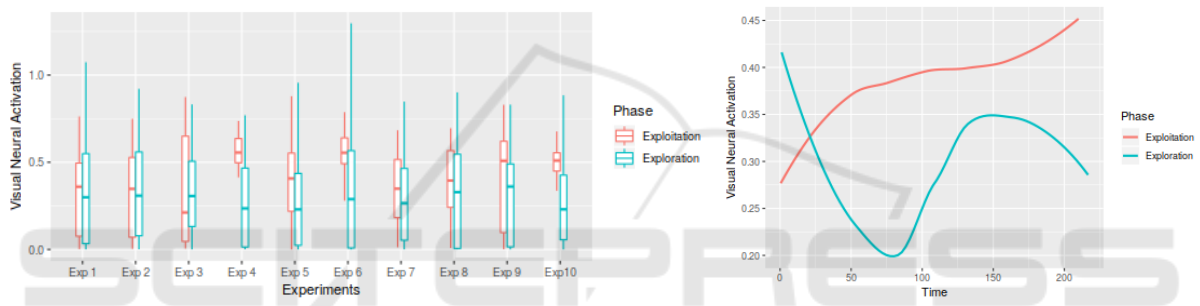


Figure 7: The figure shows the visual activation for the 10 experiments, for both exploration and exploitation. Left figure depicts that, for each experiment, a gain of visual neural activation during exploitation can be seen. Right figure depicts the sum of visual activation in time for each experiment and represented by a linear regression. This indicates a general higher neural activation during the exploitation of the sensorimotor contingencies.

4.2 Results

Figure 6 depicts the motor distribution of actions during the 10 experiments. These reflect the particular setting of the experiment, however, the motor intentions during exploitation show a preferred motor space to three different intervals: [10;30],[48;65] and [70;85]. This corresponds respectively to the extreme left, center and extreme right location of the arm. These motor positions represent the actions with a high neural visual activation.

For only one action performed there are many visual activations recorded. The distribution of the visual neural activation per experiment are presented in Figure 7-left. It can be seen that the visual activation is higher during exploitation than during exploration for most of the experiments. For the first three experiments, there is no clear gain of visual activation during exploitation. This is mostly due to the noisiness of the motion detector.

The average neural activation of the 10 experi-

ments in time is shown in Figure 7-right and depicts the benefits of exploitation. It is difficult to analyse why the linear regression of visual activation for exploration follows such dynamics. The experiment's dynamics represent this particular set up (head centered on the baby mobile) and would evolve differently with a different setting. If these results represent the dynamics of this particular settings, they still show an improvement of the visual activation when exploiting the "knowledge" gathered during motor babbling. To conclude, experimental demonstration shows that there is a gain of visual activation during the exploitation stage.

5 DISCUSSION AND CONCLUSION

This work proposes a cognitive architecture with an embodiment approach that allows a robotic arm to op-

timize 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. The intrinsic properties of the latter provides a certain level of homeostasis, meaning that a self regulation of the system is realized.

In experiments, a GummiArm robot is moving a baby mobile and observes the outcome of the action taken to optimize its motion. After the selection of an action, 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.

Furthermore, the robot observes the outcomes of every action taken during exploration. The exploitation phase then selects only the actions with the highest visual outcomes. The results validate our approach by showing a restricted motor space during exploitation and demonstrates higher neural activation. The purpose of this work serves as a proof of concept and justifies further investigation on dynamical learning in a closed-loop fashion, with an embodied approach. However, few issues remain.

Firstly, the exploration does not stop independently. It was expected that the action selection mechanism will stop by itself when all motor positions were selected given a motor state. This was not conclusive since the IOR inhib field does not project enough inhibition to stop the emergence of a new peak within the action formation field. This is therefore a current limitation of the model. Future work will investigate the tuning of the kernel interaction of the IOR Inhib field as well as the excitatory connections coming from the memory trace actions field to resolve this. During experiments, it was noted that influence from the inhibition of return and tuning it could lead to different exploratory behavior. Future work will determine how the strength of the inhibition could help the exploitation stage to converge faster toward an optimal sequence of actions.

Secondly, the motor babbling behavior remains divided between the exploration and the exploitation stage. The exploration of the sensorimotor space directly influences the exploitation process. The general architecture will be further investigated to allow a switch from exploration to exploitation and vice versa. In general, the switch from exploration to exploitation is a major issue in unsupervised learning or reinforcement learning. This would mean inves-

tigating how such switch mechanism could be implemented dynamically, and more specifically, decide when the exploration ceases to improve the exploitation stage.

In this work, only a single degree of freedom was utilized to drive the baby mobile. This demonstrates the work as a proof of concept, with a focus on the motor babbling strategies. In a future work, we intend to use the whole robotic arm with an inverse kinematics model. Practically, this will mean dealing with 3D Cartesian space instead of motor position. In terms of embodiment, this will result in grounding the 3D position of the robot's end-effector with the neural activation from the motion detector. With the necessary cognitive transformation and gain modulation (Schöner et al., 2016), the different motor spaces will be reduced to be explored with motor babbling.

Finally, the embodiment of the sensorimotor experience within neural fields is promising for the learning of skills. Here, the task is only to shake a baby mobile toy with the feedback of a motion detector, but a certain pattern of neural activation can still be observed and exploited. In the future, we will investigate the grounding of more complex stimuli such as the orientation or the movement of an object and see if it supports the learning of higher cognitive tasks such as reaching, pushing or pulling objects.

REFERENCES

- Amari, S.-I. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27(2):77–87.
- Barandiaran, X. E. (2017). Autonomy and Enactivism: Towards a Theory of Sensorimotor Autonomous Agency. *Topoi*, 36(3):409–430.
- Caligiore, D., Ferrauto, T., Parisi, D., Accornero, N., Capozza, M., and Baldassarre, G. (2008). Using motor babbling and hebb rules for modeling the development of reaching with obstacles and grasping. In *International Conference on Cognitive Systems*, pages E1–8.
- Cannon, W. B. (1929). Organization for physiological homeostasis. *Physiological reviews*, 9(3):399–431.
- Chrisley, R. and Ziemke, T. (2006). Embodiment. In *Encyclopedia of Cognitive Science*. American Cancer Society.
- Degenaar, J. and O'Regan, J. K. (2017). Sensorimotor Theory and Enactivism. *Topoi*, 36(3):393–407.
- Demiris, Y. and Dearden, A. (2005). From motor babbling to hierarchical learning by imitation: a robot developmental pathway. In *International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*, volume 123, pages 31–37. Lund University Cognitive Studies.

- Francisco J. Varela, E. T. and Rosch, E. (1991). *The Embodied Mind | MIT CogNet*. MIT Press.
- Franklin, J., Fickbohm, D., and Willard, A. (1992). Long-term regulation of neuronal calcium currents by prolonged changes of membrane potential. *Journal of Neuroscience*, 12(5):1726–1735.
- Kopecz, K. and Schöner, G. (1995). Saccadic motor planning by integrating visual information and pre-information on neural dynamic fields. *Biological cybernetics*, 73(1):49–60.
- Laflaquière, A. and Hemion, N. (2015). Grounding object perception in a naive agent’s sensorimotor experience. In *2015 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, pages 276–282.
- Luciw, M., Kazerounian, S., Lakhmann, K., Richter, M., and Sandamirskaya, Y. (2013). Learning the perceptual conditions of satisfaction of elementary behaviors. In *Robotics: science and systems (RSS), workshop “active learning in robotics: exploration, curiosity, and interaction”*.
- Mahoor, Z., MacLennan, B. J., and McBride, A. C. (2016). Neurally plausible motor babbling in robot reaching. In *2016 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, pages 9–14.
- Morse, A. F., Herrera, C., Clowes, R., Montebelli, A., and Ziemke, T. (2011). The role of robotic modelling in cognitive science. *New Ideas in Psychology*, 29(3):312–324.
- Perone, S. and Spencer, J. P. (2013). Autonomy in action: Linking the act of looking to memory formation in infancy via dynamic neural fields. *Cognitive Science*, 37(1):1–60.
- Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K., and Spivey, M. (2013). Computational Grounded Cognition: a new alliance between grounded cognition and computational modeling. *Frontiers in Psychology*, 3.
- Piaget, J. and Cook, M. (1952). *The origins of intelligence in children*, volume 8. International Universities Press New York.
- Posner, M. I., Rafal, R. D., Choate, L. S., and Vaughan, J. (1985). Inhibition of return: Neural basis and function. *Cognitive Neuropsychology*, 2(3):211–228.
- Ruciński, M., Cangelosi, A., and Belpaeme, T. (2012). Robotic model of the contribution of gesture to learning to count. In *2012 IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL)*, pages 1–6. IEEE.
- Saegusa, R., Metta, G., Sandini, G., and Sakka, S. (2009). Active motor babbling for sensorimotor learning. In *2008 IEEE International Conference on Robotics and Biomimetics*, pages 794–799.
- Sandamirskaya, Y. and Schöner, G. (2010). Serial order in an acting system: A multidimensional dynamic neural fields implementation. In *2010 IEEE 9th International Conference on Development and Learning*, pages 251–256.
- Schöner, G., Spencer, J., and Group, D. F. T. R. (2016). *Dynamic Thinking: A Primer on Dynamic Field Theory*. Oxford University Press. Google-Books-ID: ySex-CgAAQBAJ.
- Spencer, J., Perone, S., and Johnson, J. (2009). *Dynamic Field Theory and Embodied Cognitive Dynamics*.
- Stoelen, M. F., Bonsignorio, F., and Cangelosi, A. (2016). Co-exploring actuator antagonism and bio-inspired control in a printable robot arm. In *From Animals to Animats 14*, pages 244–255, Cham. Springer International Publishing.
- Sutton, R. S., Barto, A. G., et al. (1998). *Introduction to reinforcement learning*, volume 135. MIT press Cambridge.
- Tipper, S. P., Driver, J., and Weaver, B. (1991). Short report: Object-centred inhibition of return of visual attention. *The Quarterly Journal of Experimental Psychology Section A*, 43(2):289–298.
- Vernon, D., Lowe, R., Thill, S., and Ziemke, T. (2015). Embodied cognition and circular causality: on the role of constitutive autonomy in the reciprocal coupling of perception and action. *Frontiers in Psychology*, 6:1660.
- Von Hofsten, C. (1982). Eye–hand coordination in the newborn. *Developmental psychology*, 18(3):450.
- Watanabe, H. and Taga, G. (2006). General to specific development of movement patterns and memory for contingency between actions and events in young infants. *Infant Behavior and Development*, 29(3):402–422.
- Watkins, C. J. and Dayan, P. (1992). Q-learning. *Machine learning*, 8(3-4):279–292.
- Wilson, C. and Kawaguchi, Y. (1996). The origins of two-state spontaneous membrane potential fluctuations of neostriatal spiny neurons. *Journal of Neuroscience*, 16(7):2397–2410.
- Ziemke, T. (2016). The body of knowledge: On the role of the living body in grounding embodied cognition. *Biosystems*, 148:4–11.

APPENDIX

We would like to thank Mathis Richter and Jan Tekülve from the Institut für Neuroinformatik. The code, parameters and architecture files are available at <https://github.com/rouzinho/DynamicExploration/wiki>