

# A state-action neural network supervising navigation and manipulation behaviors for complex task reproduction

F. D’halluin\*\*

A. de Rengervé\*

M. Lagarde\*

P. Gaussier\*

A. Billard\*\*

P. Andry\*

\*ETIS, CNRS ENSEA University Cergy-Pontoise F-95000 Cergy-Pontoise

\*\*LASA, EPFL, CH-1015 Lausanne, SWITZERLAND

## Introduction

In this abstract, we combine work from [Lagarde et al., 2010] and [Calinon et al., 2009] for learning and reproduction of, respectively, navigation tasks on a mobile robot and gestures with a robot arm. Both approaches build a sensory motor map under human guidance to learn the desired behavior. With such a range of actions possible at the same time, the selection of action becomes a real issue and needs a higher level controller.

Several solutions exist to this problem : hierarchical architecture, parallel modules including architectures or even a mix of both [Bryson, 2000]. In navigation, a temporal sequence learner or a state-action association learner [Lagarde et al., 2010] enables to learn a sequence of directions in order to follow a trajectory. These solutions can be extended to action sequence learning. The main challenge we tried to solve in this work is having a simple architecture based on perception-action that is able to produce complex behaviors from the *incremental* learning *from demonstration* of a combination of different simple tasks, by combining two different learning sub-systems. Then we discuss advantages and limitations of this architecture, that raises many questions.

## 1. Description of the Task

The desired behavior is that the robot navigates to a fixed point, where it must wait until an object is placed in its gripper. According to the size of the object, the robot must navigate to different places and perform different actions. An overview of the task is shown on Figure 1.

The purpose of this experiment is to mix several simple tasks items in a complex sequence of behaviors, the combination of those simple tasks being taught to the robot by interaction with a human teacher.

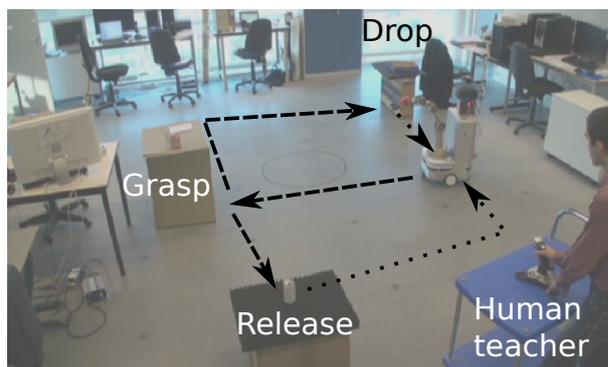


Figure 1: Overview of the task. At the grasping point, the robot chooses one side given the size of the grasped object. The self-localization of the robot is based on vision which enables a robust navigation.

## 2. A simple reactive architecture for action-selection

At any moment, the robot has to choose from a set of possible actions. The relevance of each action depends on the context of the robot which corresponds to the global state of the robot. Each sensory modality (ultrasound sensor, obstacle proximity, aperture of the gripper) has been categorized. The localization of the robot and the last action performed by the arm are also categorized. The context of the robot is the conjunction of all these categorizations. Therefore, each context neuron  $C_i$ , where  $i$  is the neuron index, corresponds to a unique state of the robot.

In a given context, the robot must select an action to perform. The implemented strategy is a simple reactive architecture (Figure 2) based on a gradient descent derived from Least Mean Square minimization [Widrow and Hoff, 1988]. Interaction with humans generates an unconditional stimulus  $D_j$  that supervises the learning so that the predicted output  $O_j^P$  for  $j$ th neuron in the conditioning group converges to the desired output. When the same context neuron  $C_i$  as the one activated during learning is active, the system is able to predict the correct

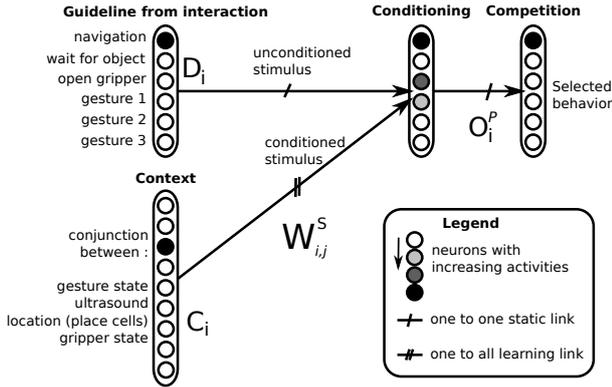


Figure 2: Model of the neural network for action selection. A context is the conjunction of all the categorized sensory inputs. The guidelines are determined using some sensory inputs such as joystick actions, or sensors on the gripper (for grasp or release guidelines). A classical conditioning enables to associate a selected behavior with a sensory context.

behavior.

The unconditional stimulus  $D_j$  also modulates the learning rate  $\mu(t)$  so that there is only conditioning when there is interaction. Learning is made on the weights  $W_{i,j}$  connecting the sensory context  $C_i$  to the predicted output  $O_j^P$ .

$$dW_{i,j}^S = \mu(t) \cdot (D_j - O_j^P) \cdot C_i \quad (1)$$

When interacting with the robot, the human provides a guideline  $D_j$  to the robot, in order to associate the desired behavior with the current context. For instance, noticing that an object which was hold firmly has been removed is associated to the guideline “Release the object”. 6 possible guidelines are currently pre-wired as variations of specific sensory inputs : Navigation, Waiting until an object is given or Opening the hand, and the 3 learned manipulation gestures.

A competition on the output predicted by the conditioning determines which behavior is activated. Other modules are thus inhibited. In the case of gesture reproduction, each gesture has been taught separately to the system using the approach described in [Calinon et al., 2009].

### 3. Discussion And Conclusion

In this paper, we developed a neural architecture that integrates two complementary task-solving modules into a unique controller : arm manipulation and navigation. The system has been tested on the described task. It was successfully taught, with on-line learning and correction to grasp and dispatch objects at different location based on their size. In further work we plan to measure the performance of the learning system in term of ease of use and required number of required demonstrations.

Learning is made on context neurons which en-

codes a singular state of the robot. If another context neuron is activated, another learning round is required. Hence, generalization abilities of this system are limited. An alternative could consist in recruiting context neurons on demand (when the teacher gives a directive), so that the number of neurons will be reduced. Then, with a competition between most plausible contexts, the robot could then present better generalization abilities. Another possibility is direct connection between sensory cues and the conditioning neural group, instead of using the conjunction. It would enable more generalization at the expense of a slower convergence and stability issues for learning. Furthermore, this may lead to non-linearly separable problems, requiring an additional hidden layer.

With the presented system, the robot is immediately able to demonstrate the learned behavior, allowing the teacher to directly correct it. The teaching process is then less tedious for the human, as he/she can see directly the current knowledge of the system. However, such a fast learning does not ensure the stability and the consistency of the learning.

In [Ogata et al., 2004], the author notices that in interaction, especially when continued over a long time, the lack of stability and generalization in learning is noticeable by the human which interacts with the robot. Memory consolidation appears to be essential in the learning processes. Memory consolidation is present in human beings : During sleep short term memory is transferred and re-encoded in neo-cortex to become long term memory. Such a system in our architecture will enable reusing past experiments to generalize, while being more robust to noise disturbances in demonstrations.

### Acknowledgments

This work was supported by the French Region Ile de France, the Institut Universitaire de France (IUF), the FEELIX GROWING European project and the INTERACT French project referenced ANR-09-CORD-014.

### References

- Bryson, J. J. (2000). Hierarchy and sequence vs. full parallelism in action selection. In *From Animals to Animats 6: Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior*, pages 147–156, Cambridge, MA.
- Calinon, S., D’halluin, F., Caldwell, D. G., and Billard, A. (2009). Handling of multiple constraints and motion alternatives in a robot programming by demonstration framework. In *Proceedings of 2009 IEEE International Conference on Humanoid Robots*, pages 582 – 588.
- Lagarde, M., Andry, P., Gaussier, P., Boucenna, S., and Hafemeister, L. (2010). Proprioception and imitation: On the road to agent individuation. In *From Motor Learning to Interaction Learning in Robots*, volume 264, chapter 3, pages 43–63.
- Ogata, T., Sugano, S., and Tani, J. (2004). Open-end human robot interaction from the dynamical systems perspective: Mutual adaptation and incremental learning. In *Innovations in Applied Artificial Intelligence*, pages 435–444.
- Widrow, B. and Hoff, M. E. (1988). *Adaptive switching circuits*, pages 123–134. MIT Press, Cambridge, MA, USA.