



## Distributed processing for modelling real-time multimodal perception in a virtual robot

Sylvain Chevallier, H  l  ne Paugam-Moisy, Fran  ois Lema  tre

### ► To cite this version:

Sylvain Chevallier, H  l  ne Paugam-Moisy, Fran  ois Lema  tre. Distributed processing for modelling real-time multimodal perception in a virtual robot. PDCN'2005, Feb 2005, Innsbruck, Austria. pp.393-398. hal-00529527

**HAL Id: hal-00529527**

**<https://hal.science/hal-00529527>**

Submitted on 13 Apr 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin  e au d  p  t et    la diffusion de documents scientifiques de niveau recherche, publi  s ou non,   manant des   tablissements d'enseignement et de recherche fran  ais ou   trangers, des laboratoires publics ou priv  s.

## Distributed Processing for Modelling Real-time Multimodal Perception in a Virtual Robot

Sylvain CHEVALLIER & Hélène PAUGAM-MOISY  
 Institute for Cognitive Science, UMR CNRS 5015  
 Lyon - France  
 {schevallier,hpaugam}@isc.cnrs.fr

François LEMAÎTRE  
 Ecole Centrale de Lyon  
 Ecully - France  
 Francois.Lemaitre@ec12004.ec-lyon.fr

### ABSTRACT

Built from a need for modelling cognitive processes, a modular neural network is designed as the “brain” of a virtual robot moving in a prey-predator environment. The robot decides its path from the animals it identifies around. Both a parallel implementation of distributed processes and a temporal coding of spiking neurons allow the robot to develop multimodal perception with attentional mechanisms and to react in real-time to its dynamic environment.

### KEY WORDS

Distributed Real-time Systems, Neural Networks, Multimodal Perception, Attentional Mechanisms, Robotics

## 1 Introduction

On the one hand, parallel and distributed algorithms are usually designed to speed up scientific computations [1, 2] or artificial neural networks training [3, 4, 5, 6, 7]. On the other hand, cognitive processes are highly concurrent and distributed in the brain, but in a different way than in parallel computers. In this article, we propose to take advantage of parallel computing for simulating a brain-like integration of multimodal informations (image, sound, etc). Starting from a cognitive point of view, a modular neural network has been previously designed for modelling a multimodal associative memory simulating multisensory integration [8, 9]. The model has been implemented as the “brain” of a virtual robot moving in a prey-predator environment, as described in section 2. The parallel implementation of the virtual robot and the simulation of a dynamic environment are presented in section 3. Section 4 proposes the implementation of real-time attention shifting mechanisms. In section 5, we show how crossmodal interaction can be simulated, due to the combination of spiking neurons and distributed processing.

## 2 Virtual robot, prey-predator environment

Starting from a functional architecture designed for vision by cognitive psychologists [10] and the hypothesis stating that this architecture can be replicated for other sensory modalities, we have designed a modular neural network modelling a multimodal associative memory [11].

Built from several neural networks as basic bricks, the model computes the low-level perceptive processes (one prototype-based incremental classifier for each perceptive modality), the central data fusion (a unique BAM - Bidirectional Associative Memory - adapted for multiple input vectors coming from the different perceptive modules) and an output network (incremental classifier) computing the object identified by the whole model. An experimental platform has been developed for testing the behaviour of the model embedded in a virtual robot moving among static animals, in a virtual zoo [12, 13, 14]. For the robot, animals are either predators to be avoided, preys to be eaten or neutral animals.

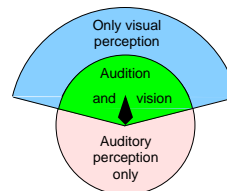


Figure 1. Perceptive fields of the virtual robot.

An image and a sound are associated to each animal. The robot has two sensory modalities, vision and audition (figure 1). The robot is designed as a head with a mouth and one eye. It has only a partial vision of the environment since its visual field is forward oriented, within an obtuse angle (from  $-75^\circ$  to  $+75^\circ$ ). The auditory field extends all around the robot, but with a smaller reach. Hence, the robot can sometimes see one animal and hear another one.

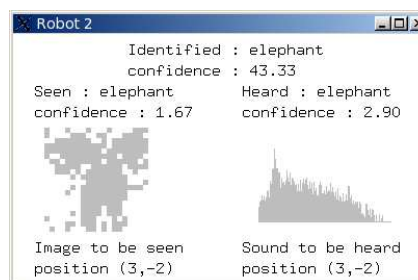


Figure 2. Auditory and visual recognition of noisy patterns, plus name of the animal identified by the robot.

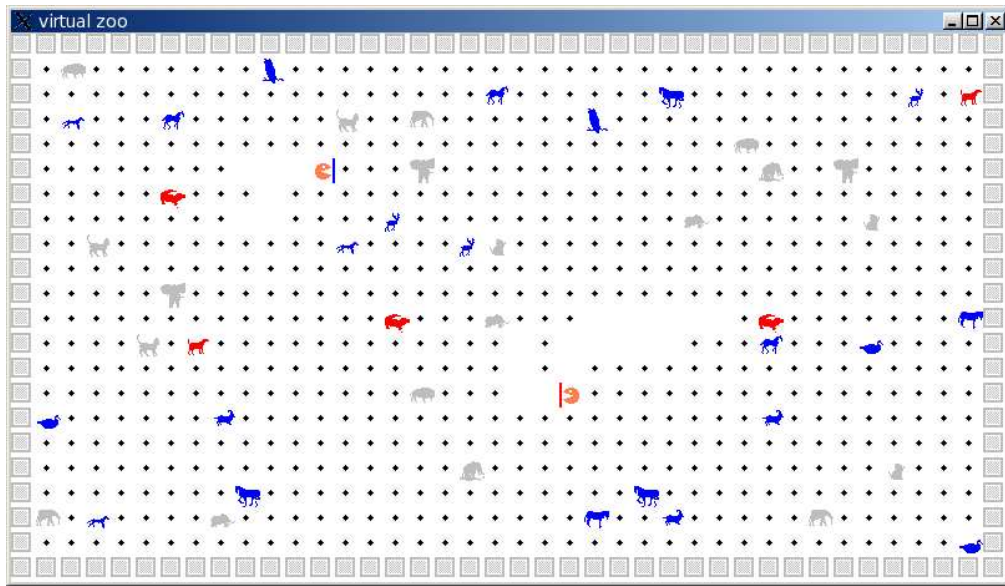


Figure 3. Two virtual robots moving in a virtual prey-predator environment.

The closer the robot is to an animal, the better the sensory information that could be perceived, on both modalities. This effect is simulated by adding noise to the image or the sound in relation with the distance between the robot and the animal, inside the receptive fields. The effective perceptions of the robot are the results of the low-level computations of the noised stimuli by the model. The output prototypes are the inputs of the BAM and the result of the computation of the whole model is either the animal identified by the robot (with a degree of confidence) or a non-answer in doubtful cases. On figure 2, an elephant has been identified, with a rather low confidence. Since an elephant is a neutral animal, the robot will continue moving in front of it (the direction of the mouth), with a low probability to change its direction randomly.

The purpose of this article is to present several improvements making the virtual robot closer than an actual robot evolving in a real-world environment, with a cognitively plausible behaviour and real-time reactivity to a dynamic environment.

### 3 Dynamic environment, distributed system

First, the dynamic feature of the environment results from giving life to all the animals in the zoo. The animals move according to their nature with regards to the robot. With a probability of moving less or equal to 0.1 and within a threshold distance to the robot, a predator (crocodile or wolf) becomes closer to the robot and a prey flees from it. Neutral animals (in light gray on figure 3), or preys and predators that are outside the threshold distance, can move randomly, with the same probability of moving [15].

Second, the robot becomes able to manage this dy-

namic environment by the means of a parallel implementation of its “brain”. The sequential execution of the model is not realistic, from a cognitive point of view. For instance, in sequential simulations, visual recognition is computed first, auditory recognition is computed afterwards, and then the BAM processes both resulting information. In parallel versions of the model, all the neural network modules can run continuously, e.g. visual and auditory recognitions are processed concurrently [16]. Moreover, the parallel model will make it possible to take advantage of temporal properties in the distributed neural architecture, even if, in this article, the distribution of processes is slightly different from the mapping proposed in [16]. The computing tasks are distributed according to cognitive concerns (see figure 4), but not in the aim of reaching an optimal speedup. The parallel implementation of the model has been developed for a network of PC-computers, considered as a virtual MIMD parallel machine with distributed memory. All the processors run a Linux operating system, using the LAM (Local Area Multicomputer) high-quality open-source implementation of the MPI (Message Passing Interface) specification.

One process is devoted to the management of the zoo: Image and sound that the robot can perceive, animals moves, graphical windows providing information to the experimenter, etc.... Five other processes implement the different cognitive modules of a robot: Vision, audition, multimodal associative memory, output network computing the model answer, and a motor module computing the next action of the robot. Figure 4 clarifies the information exchanged by communications (non-blocking message passing) between the different processes. Several robots can move simultaneously in the environment (five concurrent processes for each robot).

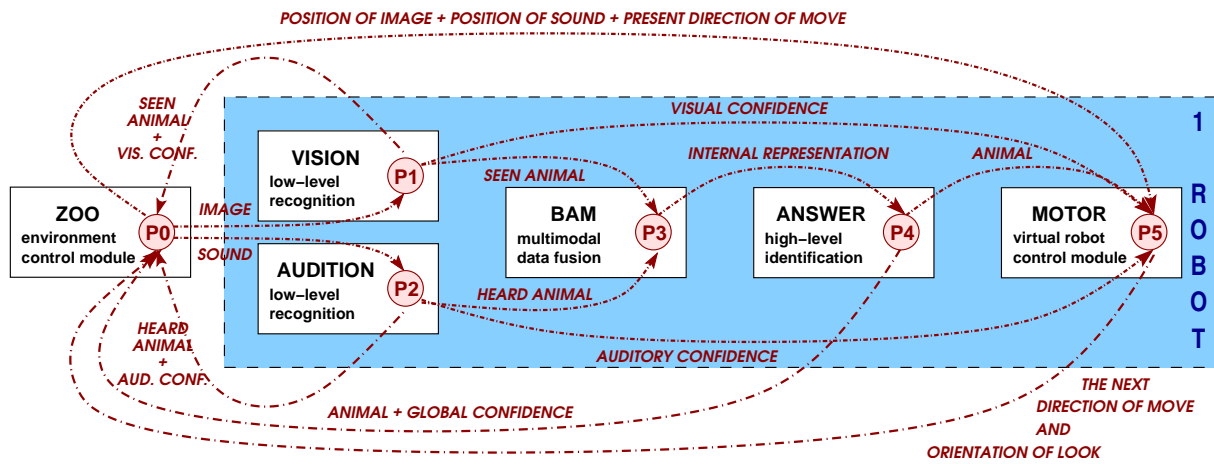


Figure 4. Distributed system: Mapping of cognitive processes and message passing.

## 4 Visual attention

In computer vision, the importance of modelling attentional mechanisms has been highlighted, for time saving [17]. For instance, selecting a region of interest reduces the processing complexity. A model of visual attention, proposed by Itti and Koch [18], has motivated part of our work.

### 4.1 Orientation of the look

In the preliminary version of the zoo [12], the robot could see only in a conical space in front of it, i.e. in the direction pointed by its mouth. A first improvement has been to disconnect the direction of move and the orientation of the look. In order to precise this additional information on the graphical representation, a bar has been added to the robot head, in the direction of the look. Moreover, in case of several robots in a same environment, the colour of the bar is an indication of the nature of one robot for the others. On figure 5, top right, a prey robot (light blue bar) is looking at a predator robot (dark red bar) in its back, and it is running away. Bottom left, the predator robot is going after the prey robot that it has just identified at the previous step.



Figure 5. Robots with bars for the orientation of their look.

Thanks to this improvement, the robot is more able to follow a prey it has just detected. Performance augmentation has been measured on experiments. Starting near the center of the environment, the robot moves around, according to the animals it identifies and how much it trusts its decisions. If eaten by a predator, a robot can start a new life from a place defined by the experimenter. On a run of

275 moves, we observe that a robot with directed visual attention usually eats twice as many preys as a robot with a fixed look in front of it (means on 10 experiments in randomly evolving environment). Note that the efficiency of the device cannot be measured for the predators because the robot runs away as soon as it has identified a vicious animal, hence the predator soon disappears from the visual field.

### 4.2 Modelling attentional mechanisms

Even if the robot is able to point its look according to its previous observations, there can be several animals inside its visual field. In the preliminary version of the zoo [12], only the image (or sound) of the closest animal in the visual (auditory) field was sent to the robot. However, Wolfe [19] explains that a region of interest can be defined quickly, from a few specific features only and at a low level of cognitive processing. We have exploited the knowledge of psychologists for modelling attentional mechanisms in the robot behaviour:

- the nature (prey or predator) of an animal can be distinguished at a pre-attentive level,
- predators can be made more salient than all other animals, and preys can be more salient than neutral animals (e.g. colour as selective feature),
- this selectivity is possible only inside a small limit of proximity (threshold distance).

A visual buffer has been added to the vision module of the distributed system (cf. figure 4). Among all the animals present at a given time in the visual field of the robot, only one is processed by the visual recognition module, in order to preserve a real-time processing. The image is selected with respect to the principles above. For instance, for a prey and a neutral animal at equal distances, the attention will focus on the prey (figure 6, left). Moreover, even if a prey

is in front of the robot, the robot can perceive a predator at a short distance behind the prey (figure 6, right). In the graphical platform of the zoo, a square box has been added to circle the animal taken into account by the robot, at each step.

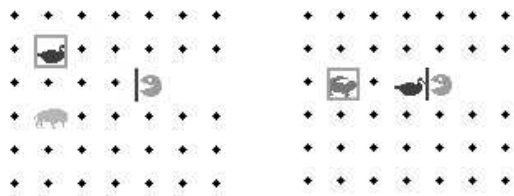


Figure 6. Selection of the more salient animal in the visual buffer.

The effects of modelling attentional mechanisms have been measured with several experiments. The influence of the attentional capacities on the robot behaviour is clearly positive. At each run, several variables are recorded:

1. the number of successive lives of the robot,
2. the total number of moves,
3. the number of moves forward,
4. the number of different places visited by the robot,

In a first experiment, the total number of moves has been fixed. The program is started 10 times, for 200 moves each run, at first for a robot with attentional mechanisms, and then for a robot with a plain orientation of look, without selection of a salient animal. The mean number of lives is significantly lower (Student statistical test,  $p < 0.05$ ) for the robot attentive to the surrounding animals (figure7), proving that the attentive robot is able to survive for longer.

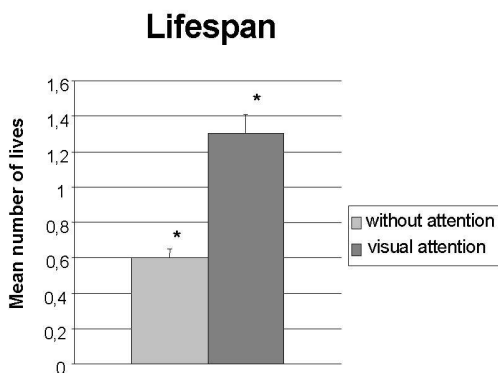


Figure 7. A robot with attentional mechanisms stays alive for longer.

Second experiments have been carried out, with a fixed number of lives for the robot. The numbers of

moves, both all around and forward moves, and the number of visited places are significantly higher (Student test,  $p < 0.05$ ) for the robot with attentional mechanisms (figure 8).

### Quality of exploration

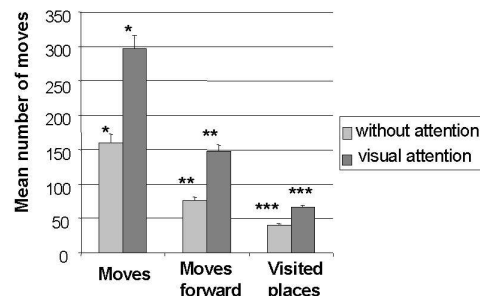


Figure 8. A robot with attentional mechanisms explores better the environment.

So, we have improved the capacity of the robot to manage a dynamic environment and to avoid the traps. Moreover, the simplicity of how attentional mechanisms are modelled (even if based on cognitive processing) preserves the real-time feature of the system and does not seriously unbalances the computational load of the concurrent processors.

## 5 Real-time crossmodal interactions

Recent advances in neuroscience prove that early cross-modal interactions take place in multisensory processing, both in animal and human brain [20, 21, 22]. More precisely, the neuroscientists observed that the brain has a capacity to redirect the visual attention towards a peripheral sound source. Our distributed system is a good candidate for modelling such interaction, especially if a “spiking-BAM” [23] replaces the initial BAM module (cf. figure 4). The spiking-BAM is an emulation of BAM in temporal coding, based on spiking neurons [24, 25] instead of classical threshold neurons. Spiking neural networks can integrate information through time. Each spiking neuron integrates the spikes of upstream neurons as soon as they are communicated. In the distributed system, the combination of non-blocking message passing and spiking neurons in the BAM module gives way to an actual real-time processing of variable information flows, since the BAM receives on-line the patterns processed by the visual and auditory modules.

In a situation like the one shown in figure 9, the simulation of crossmodal interactions is realized as follows:

- the robot receives the image of an animal V, in its visual field,
- simultaneously, the robot receives the sound of a very close animal A,





Figure 9. The robot sees a bison (animal V) and hears a crocodile (animal A). The robot will redirect its look immediately towards the crocodile.

- recognition procedures are started simultaneously by the vision and audition modules,
- the robot redirects its look towards the animal A (salient for audition),
- the image of animal A is sent to the vision module,
- the BAM module integrates the newly seen animal during the on-going process of data fusion.

The new information is taken into account in real time, as presented in figure 10 showing activity of the neurons through time. For each neuron, a point is plotted each time a postsynaptic spike is emitted by the neuron. The diagram represents the variation of the activity patterns of the BAM neural network, during a short time.

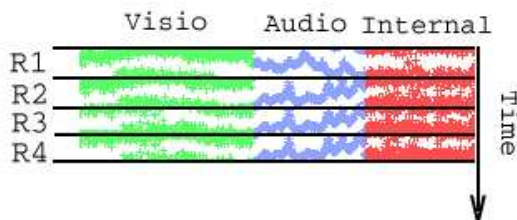


Figure 10. Diagram of spikes emitted through time, in the spiking-BAM. The new visual input modifies the patterns.

From left to right, the first 400 neurons (“visio”) represent the visual input of the spiking-BAM, the next 256 neurons (“audio”) represent the auditory input and the last neurons (“internal”) are for the internal representation built by the BAM in four consecutive recurrent iterations. At iteration R1, the robot sees and hears animal V. At iteration R2, the sound of animal A is perceived. The robot redirects its look: The image of animal A is perceived as soon as iteration R3. The BAM processing goes on (iteration R4), with modified patterns. Hence, we can see that the input pattern is changed before the BAM stabilisation, since the newly seen animal is received during the BAM processing, thanks to the new message communicated by the incoming visual process. The robot can perceive, in real-time, the danger of being close to a predator, even if it was just

looking elsewhere, since the patterns are modified on-line before the BAM converges towards a stable state. Hence the animal identified by the whole system at this step will be a crocodile and the robot will be able to run away from this predator.

## 6 Conclusion and perspectives

We have presented a distributed system capable of simulating the “brain” of a virtual robot moving in a dynamic prey-predator environment, with a cognitive behaviour. The parallel processing of the cognitive modules and the on-line message passing of the information are the ingredients of the real-time and clever reactivity of the robot to the traps of the virtual zoo. This article has presented several improvements to the preliminary version of the virtual robot in the zoo:

- Since all the animals can move, the environment is dynamic.
- The orientation of the look is independent of the direction of move.
- A salient image can be selected in the visual buffer by implementing attentional mechanisms.
- A crossmodal interaction between audition and vision enhances the identification of traps.

All these improvements make the robot more efficient and its behaviour becomes more realistic, from the cognitive point of view. Moreover, the speed of the robot moving is totally compatible with real-time constraints.

The system is still far from being directly implementable on a real-world robot. However, in the long path of research required to realize a performant cognitive robot, we addressed a part of the problem that is not studied usually. Mixing our results with the advances of researchers working on computer vision, for instance, could lead to a very efficient real robot in the short term. For instance, the model developed by Machrouh, Liénard and Tarroux [26], or other work based on the Itti & Koch’s model, could be used for replacing the vision buffer of the virtual robot by a device selecting an animal in a real-world visual scene.

## References

- [1] G. Fox, S. Otto, and A. Hey. Matrix algorithm on a hypercube: Matrix multiplication. *Parallel Computing*, 4:17–31, 1987.
- [2] D.P. Bertsekas and J.N. Tsitsiklis. *Parallel Distributed Computation: Numerical methods*. Prentice-Hall, 1989.
- [3] J. Ghosh and K. Hwang. Mapping neural networks onto message-passing multicomputers. *Journal of Parallel and Distributed Computing*, 6(2):291–330, 1989.

- [4] N. Dodd. Graph matching by stochastic optimization applied to the implementation of multi-layer perceptrons on transputer networks. *Parallel Computing*, 10:135–142, 1989.
- [5] H. Paugam-Moisy. Multiprocessor simulation of neural networks. In M. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 605–608. The MIT Press, 1995.
- [6] V. Demian, F. Desprez, H. Paugam-Moisy, and M. Pourzandi. Parallel implementation of RBF neural networks. In *EURO-PAR'96, Parallel Processing, Lecture Notes in Computer Science*, volume 1124, pages 243–250. Springer, 1996.
- [7] P.A. Estévez, H. Paugam-Moisy, D. Puzenat, and M. Ugarte. A scalable parallel algorithm for training a hierarchical mixture of neural networks. *Parallel Computing*, 28:861–891, 2002.
- [8] E. Reynaud. *Modélisation connexionniste d'une mémoire associative multimodale (in French)*. PhD thesis, Institut National Polytechnique de Grenoble, Oct 2002.
- [9] H. Paugam-Moisy, D. Puzenat, E. Reynaud, and J.-P. Magué. Neural networks for modeling memory: Case studies. In M. Verleysen, editor, *ESANN'2002, 10th Europ. Symp. on Artificial Neural Networks*, pages 71–82. d-side, 2002.
- [10] S. M. Kosslyn and O. Koenig. *Wet Mind: The new cognitive neuroscience (2nd ed.)*. The Free Press, New-York, 1995.
- [11] A. Crépet, H. Paugam-Moisy, E. Reynaud, and D. Puzenat. A modular neural model for binding several modalities. In H. R. Arabnia, editor, *IC-AI'2000, Int. Conf. on Artificial Intelligence*, pages 921–928, 2000.
- [12] E. Reynaud and D. Puzenat. A multisensory identification system for robotics. In *IJCNN'2001, Int. Joint Conf. on Neural Networks*, pages 2924–2929, 2001.
- [13] C.-H. d'Adhémar. Amélioration de l'interface graphique du programme de robot virtuel (in French). Rapport de stage de 2ème année, Ecole Centrale de Lyon et Institut des Sciences Cognitives, 2003.
- [14] F. Lemaître. Réécriture orientée objet du programme de robot virtuel (in French). Rapport de stage de 2ème année, Ecole Centrale de Lyon et Institut des Sciences Cognitives, 2003.
- [15] F. Lemaître. Animation graphique dans un environnement robotique virtuel (in French). Rapport de projet de 3ème année, Ecole Centrale de Lyon et Institut des Sciences Cognitives, 2004.
- [16] Y. Bouchut, H. Paugam-Moisy, and D. Puzenat. Asynchrony in a distributed modular neural network for multimodal integration. In *PDCS'2003, IASTED Int. Conf. on Parallel and Distributed Computing and Systems*, pages 588–593. ACTA Press, 2003.
- [17] J.K. Tsotsos. Analysing vision at the complexity level. *Behavioral and brain sciences*, (13):423–469, 1990.
- [18] L. Itti and C. Koch. Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203, 2001.
- [19] J. Wolfe. *Seeing (2nd ed.)*. Academic Press, 2000. (see Chapter: Visual Attention).
- [20] M.-H. Giard and F. Perronet. Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *J. of Cognitive Neurosc.*, 11(5):473–490, 1999.
- [21] A. Fort, C. Delpuech, J. Pernier, and M.-H. Giard. Early auditory-visual interactions in human cortex during nonredundant target identification. *Cognitive Brain Research*, 14:20–30, 2002.
- [22] A. Falchier, S. Clavagnier, P. Barone, and H. Kennedy. Anatomical evidence of multimodal integration in primate striate cortex. *Journal of Neuroscience*, 22(13):5749–5759, 2002.
- [23] D. Meunier and H. Paugam-Moisy. A “spiking” bidirectional associative memory for modeling inter-modal priming. In *NCI'2004, IASTED Int. Conf. on Neural Networks and Computational Intelligence*, pages 25–30. ACTA Press, 2004.
- [24] W. Gerstner and W. Kistler. *Spiking Neuron Models*. Cambridge University Press, 2002.
- [25] W. Maass and Natschläger. Networks of spiking neurons can emulate arbitrary hopfield nets in temporal coding. *Network: Computation in Neural Systems*, 8(4):355–372, 1997.
- [26] Y. Machrouh, J.-S. Liénard, and P. Tarroux. Multi-scale feature extraction from visual environment in an active vision system. In *Proc. of Int. Workshop on Visual Form*, Lecture Notes in Computer Science, pages 388–397. Springer-Verlag, 2001.