A spiking Bidirectional Associative Memory for modeling intermodal priming

David Meunier, Hélène Paugam-Moisy

A “spiking” Bidirectional Associative Memory for modeling intermodal priming

David MEUNIER & Hélène PAUGAM-MOISY
Institute for Cognitive Science, UMR CNRS 5015
Lyons - France
{dmeunier, hpaugam}@isc.cnrs.fr

Published in:
Proceedings of the 2^{nd} IASTED International Conference
Neural Network and Computationnal Intelligence
February 23-25, 2004, Grindelwald, Switzerland
Acta Press, ISBN 0-88986-389-1, p 25-30

ABSTRACT
Starting from a modular artificial neural system modelling the integration of several perceptive stimuli, this article proposes a new implementation of the central module performing a multimodal associative memory. A Bidirectional Associative Memory (BAM) has been emulated in temporal coding with spiking neurons. Since input patterns are dynamically encoded, the effects of the latency of evocation can be simulated with the “spiking BAM”, thus adding temporal properties to the model. For highlighting the contribution of the new module and the relevance for modelling cognitive processes, the “spiking BAM” has been tested in the context of an experimental protocol of cognitive psychology.

KEY WORDS
Neural Network Models, Cognitive Processes, Spiking Neurons, Intermodal Priming.

1 Introduction

Intrinsic temporal properties of biological neurons are not taken into account by common artificial neurons, currently derived from the mathematical model introduced by McCulloch and Pitts [1]. Hence usual artificial neural networks [2, 3], even recurrent networks, are too static to explain temporal interactions between assemblies of natural neurons. The modular network, simulating a multimodal associative memory for binding several modalities, developed by [4] and described in [5, 6], suffers from this drawback. The design and properties of this model are summarized in section 2.

Neuroscience experiments have shown that the average frequency of neurons is not meaningful to explain fast computation in the brain [7], or to account for a precise sequence of firing for each neuron participating in the synchronisation of an assembly [8]. Following these discoveries, a more complex model, the *"spiking neuron"*, has been proposed [9], where each action potential is simulated. Networks of spiking neurons have the property to emulate networks of sigmoidal neurons or threshold units [10], such as Hopfield networks [11]. The dynamic properties of a recurrent network better emerge from a spiking neuron emulation. When the dynamics of the network has converged, the attractor pattern is rebuilt at each new wave of spikes, allowing the network to integrate new inputs while computing. The present article proposes an emulation of a Bidirectional Associative Memory (BAM) [12] in temporal coding, with spiking neurons. More precisely, we emulate a "multiple BAM" [4] which is the central module of the multimodal associative memory model. The new BAM, called **"spiking BAM"**, is presented in section 3. The purpose is to observe the emergence of higher-scale dynamics in the multimodal associative memory model, as result of integrating intrinsic temporal properties at the neuron level in the BAM.

For highlighting the contribution of the new module and the relevance for modelling cognitive processes, the "spiking BAM" has been tested in the context of an experimental protocol which is usually applied to human beings by psychologists. The *priming* effect is a psychological phenomenon well-known in cognitive science [13]. Under certain temporal conditions, the presentation of a first stimulus, the *primer*, makes easier the processing of a similar second stimulus, the *target*. **Intermodal priming** consists in presenting a primer and a target in two different perceptive modalities. This specific aspect of priming effect, not yet widely developed in literature, is the subject of current research in Psychology [14]. Section 4 precises the protocol applied to the connectionist model and presents experimental results and discussion.

2 Cognitive Science inspiration

2.1 Multimodal associative memory model

Starting from a functional architecture for high-level visual perception, proposed by Kosslyn and Koenig [15], we assume with psychologists that similar architectures hold for other sensory modalities. Low-level processing, corresponding to the *recognition* phase, is modality-specific. Similar architectures of the low-level sub-systems are replicated in every modalities. High-level processing, leading to *identification*, is unique, multimodal, and realizes a data fusion from the pre-processed patterns. A modular neural network [4, 5] has been built, from several basic bricks, for modelling this multimodal architecture of associative memory (figure 1).

The network is composed of several modules. Low-level: In each modality, an incremental neural classifier [16] associates a prototype (result of recognition) to an input stimulus. High-level: A global associative memory receives all the output prototypes of the classifiers and associate them to an amodal representation. Finally, this representation is the input of another incremental classifier which realizes the identification, i.e. the answer of the whole model to the set of stimuli.

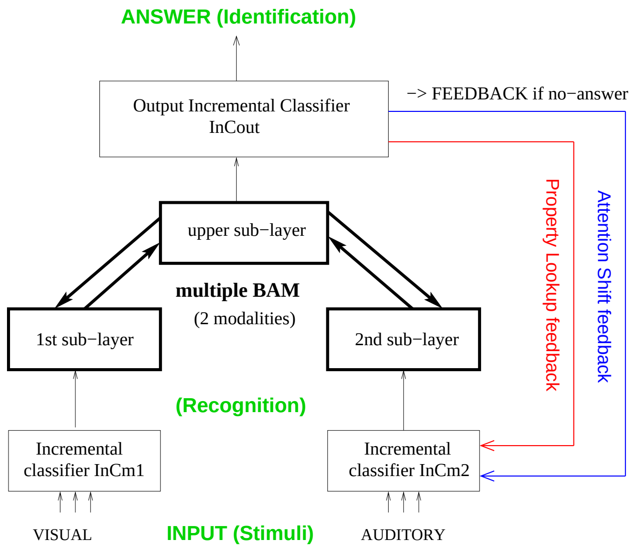

Figure 1: Modular neural network modelling a multimodal associative memory. The central module is a multiple BAM (thick boxes and arrows).

2.2 The central module, a multiple BAM

The central module, modelling the associative memory (cf. figure 1), is a "multiple BAM" [4], an adaptation of the classical Bidirectionnal Associative Memory (BAM) defined by Kosko [12]. In a multiple BAM, the lower layer is separated into several sub-layers, each of them receiving different inputs (figure 2). A *bimodal BAM* is a multiple BAM with two sub-layers.

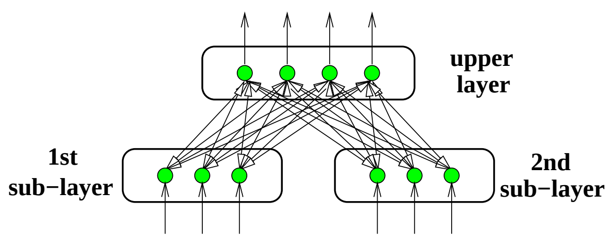

Figure 2: Connectivity of a bimodal BAM.

From initial states given in input to all the sub-layers, the dynamics starts and the network reaches a stable state after a finite number of iterations, from bottom to top and conversely. One of these up and down iterations is called a *reverberation*. The number of reverberations can be considered as a measurement of time.

A multiple BAM has the ability to simulate associative recall properties, modelling the phenomenon of mental image evocation in the context of cognitive processes [6].

Before reverberations After reverberations

Figure 3: The dynamics of the multiple BAM shows efficient associative recall properties.

Figure 3 illustrates the property on an artificial database, with letters and digits as modalities, learning associations between letters and digits of same rank, e.g. pairs (A,1), (B,2) or (H,8). In generalisation phase, the dynamics of the bimodal BAM is able to regenerate the image of a "8" from a missing input on the digit-modality and a prototype of "H" on the letter-modality (figure 3).

2.3 Lack of temporal behavior

The model is able to simulate cognitive behavior like improvement of performance by activation of feedback mechanisms (cf. figure 1) or mental image evocation [5]. Tested in a virtual robotic environment inhabited by animals (preys and predators), the model controls a creature able to move around safely, according to visual and auditory perceptions [17]. However, the model is too static to reproduce temporal behavior such as integrating visual and auditory informations incoming at slightly different times. One way to alleviate this drawback consists in distributing the main modules of the model on different processors of a parallel computer. The asynchrony of message passing is a solution for introducing temporal behavior in the model, at a macroscopic level (see [18]). The purpose of the present article is to study how temporal behavior can be introduced at a microscopic level, by means of spiking neurons.

The multiple BAM neurons are threshold units, with no intrinsic representation of time properties. We wish the BAM becoming able to integrate new inputs, at any time, while the dynamics is running. Therefore, we design a new version of BAM, starting from spiking neurons and following the construction of Maass [11] for emulating a Hopfield network in temporal coding. In the **"spiking BAM"**, the network rebuilds the stored patterns at each reverberation. Hence the patterns are *dynamically* encoded, and the *latency* of evocation can be simulated and analyzed. We prove that, with the integration of intrinsic temporal properties at the neuron level, higher-scale dynamics emerge in the model, through the mechanisms of the spiking BAM reverberations.

3 The model of "spiking BAM"

3.1 Spiking neuron model

Unlike low-level representations of the natural neuron properties, such as Hodgkin & Huxley neuron or Integrate & Fire neuron, the spiking neuron is a phenomenological model of the biological neuron [9]. A spiking neuron acts as a *coincidence detector*

[19], since it pulses whenever a large enough number of post-synaptic potentials (PSPs) hit simultaneously the soma of the neuron. In the SRM_0 **model** [20], the state of the neuron only depends on the last time of emission t_i^f.

The behavior of the neuron is described with two variables varying in time, the *membrane potential* $u(t)$ and the *threshold* function $\theta(t)$. The membrane potential depends on the last times of emission of the pre-synaptic neurons through the time of impact of the resulting PSP $\epsilon(t)$, whereas the threshold only depends on the last time of emission of the neuron:

$$u(t) = \sum_{j=1}^{n} \omega_{ij} \; \epsilon(t - t_j^f - d_{ij})$$

Whenever the membrane potential is higher than the threshold, the neuron sends an action potential, i.e. the neuron updates its last time of firing to the current time.

$$t_{i+1}^f = t \Longleftrightarrow u(t) = \theta(t - t_i^f)$$

Then the threshold highly increases, thus modelling an absolute refractory period. After that, the neuron goes back to the reference threshold level during a relative refractory period.

3.2 Maass construction in temporal coding

The Maass & Natschläger construction [10] is based on the fact that post-synaptic potential can *shift* the time of firing of a neuron. An excitatory PSP brings forward the time of firing, whereas an inhibitory PSP makes the neuron spiking later.

In this construction, each time of firing bears information relatively to a *reference time*, where the neuron would have spiked due to stimulation of auxiliary *pacemaker* neurons only. Since these auxiliary neurons fire independently of the other neurons of the network, with a fixed period, they define a *temporal coding interval* centered on the reference time. If it fires precocely (resp. lately), i.e. at the beginning (resp. the end) of the interval, the neuron is considered as active (resp. inactive), which corresponds to a value of +1 (resp. -1) in bipolar encoding. To emulate threshold units network, each spiking neuron has to fire once and only once at each iteration. Since time of firing can take any value in the temporal coding interval, the network achieves real values based computation.

For storing weights, we use the method of *projection encoding* [2], which is the classical way to compute weights for storing wished patterns:

$$\omega_{ij} = \frac{1}{P} * \sum_{p=0}^{P} \xi_i^p \xi_j^p$$

where P is the number of patterns. The weights are computed once, at the beginning of the simulation (learning phase). They remain fixed after that. The spiking BAM is emulated in generalisation phase, with spiking neurons in every sub-layers of a multiple BAM architecture.

3.3 BAM emulated by spiking neurons

For better understanding the dynamics of a spiking BAM, temporal diagrams of firings are displayed for each neuron of the network and for each reverberation, i.e. one bottom-up, and then top-down, iteration (figure 4). The dynamic process of the network appears as *waves of spikes*, translating in time from lower to upper layer and conversely, with a fixed transmission delay.

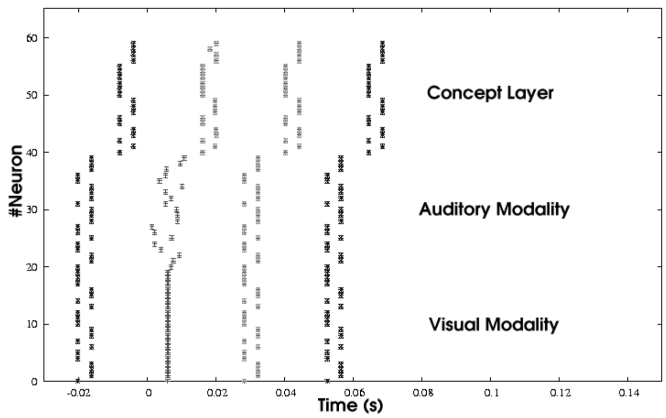

Figure 4: Emulation of bimodal BAM

For sake of clarity, all the graphical diagrams (figures 4 and 5) display the behavior of a small network. The network is composed of two lower sub-layers (#0-19 and #20-39), each receiving inputs from a different modality, and one upper layer (#40-59). The temporal coding interval appears as the difference between *precoce* and *late* spikes within each reverberation. However, some of the spikes take intermediate values, which might be responsible for faster convergence than in the original model of multiple BAM with threshold units.

On figure 4, the pattern on the left is an exact version of the full pattern (both modalities and abstract concept) in input. On the very right, a model of the stable state to be recognized (target pattern) is reproduced, for helping to test the convergence. The input of the network (starting at time 0) is an absence of information on the visual modality and a noisy version of the pattern on the left, injected in auditory modality only. The noise corresponds to a random value added to the pattern spikes, this value being taken as a fraction of the temporal coding interval (here, we add $\pm 0.8\times$temporal coding interval). Next waves of spikes show the evolution of the dynamics, in the upper layer (abstract representation of the concept) and in the lower layers (internal representations of perceptive stimuli) alternatively. A pattern is considered as retrieved when an error of less than 10% with target pattern is reached on the upper layer. For the present simple example, two reverberations are sufficient to retrieve the target.

The spiking BAM reproduces all the classical BAM properties (e.g. mental image evocation, robustness to noise). The main advantage of the spiking BAM is that the dynamics never stops running. Even after the network has converged, the attractor pattern is rebuilt continually, at every new reverberation. Hence new inputs can be integrated while the network is still computing and the influence of a discordant information (i.e. a pattern corresponding to another attractor basin) can be observed on the dynamics.

In order to experiment the new temporal abilities of the model, the spiking BAM has been tested in the context of intermodal priming, involving both temporal properties and multimodal interactions.

4 Intermodal priming test of the network

4.1 Experimental protocol, as defined by cognitive psychlogists

Priming consists in presenting two items to the system, a *primer* i_p and then a *target* i_t and to measure performance and processing time of i_t, according to i_p. To study intermodal priming, we present i_t on one perceptive modality and i_p on the other one. In the case of congruent priming, i_p and i_t correspond to the same concept (e.g. i_p is a miaowing and i_t is a cat's picture) whereas in the case of non-congruent priming, i_p and i_t correspond to different concepts (e.g. i_p is a barking and i_t is a cat's picture).

The primer (resp. target) *latency* is the interval of time during the presentation of the primer (resp. target). The ISI (Inter-Stimuli Interval) is defined as the time left between the end of the primer latency and the beginning of the target stimulus presentation. Results of intermodal priming, on human beings, in cognitive psychology, show significant differences between congruent and non-congruent conditions for the time of response and for the identification performance (percentage of right identification of the target). Other results involving intermodal priming make evidence of an optimal ISI, i.e. an ISI where the difference between the two opposite conditions is optimal [14].

4.2 Test of the network

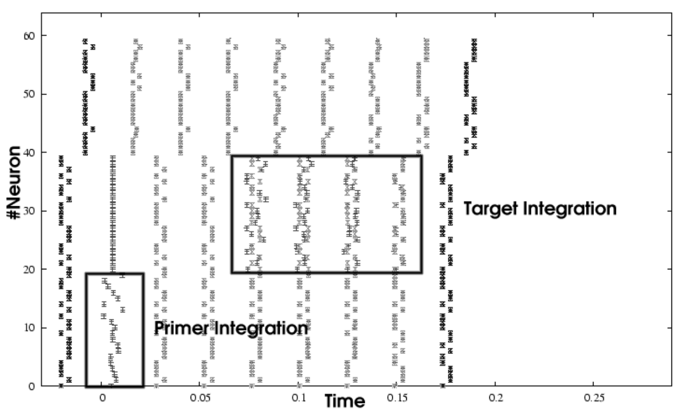

Figure 5: Protocol of an intermodal priming test

Tests have been realized with a network of 3*64 neurons, storing 9 different patterns, corresponding to letters (1st modality) and digits (2nd modality). At each test, the patterns corresponding to primer and target are randomly chosen within the stored patterns. Figure 5 presents an example of network dynamics evolution, with a primer latency equal to 1 reverberation, and an ISI equal to 2 reverberations. The target latency always lasts for 5 reverberations, but the network usually retrieves the target pattern before the fifth presentation.

Successively, the pattern corresponding to the primer is presented in the first modality, the network dynamics goes on during a time corresponding to the ISI, and the pattern corresponding to the target is integrated into the second modality. The integration of new inputs while the network is computing is realized by randomly mixing reverberated spikes with values corresponding to inputs, in the second modality, whereas the first modality is not updated.

As the BAM is part of a larger modular network (see Section 2), the influence of other modules has been simulated by presenting several times the same patterns (for primer and target). The first presentation is the most noisy ($\pm 0.5\times$ temporal coding interval), then the noise decreases linearly towards a perfect presentation for simulating the convergence of the other networks. On figure 5, the target has been retrieved correctly after 4 reverberations, when the state of the upper layer has reached the target pattern. Before the third presentation, the state of the upper layer has not changed enough to switch from the primer to the target, which illustrates the interest of repeating the presentation of patterns.

4.3 Priming Results

Figure 6 presents the performance (identification rate) and figure 7 presents the latency of the answer, i.e. the number of reverberations required to retrieve the target pattern.

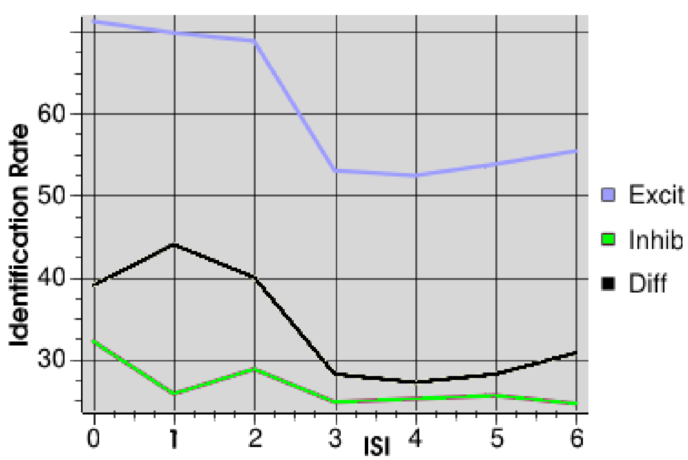

Figure 6: Intermodal priming effect on identification rate

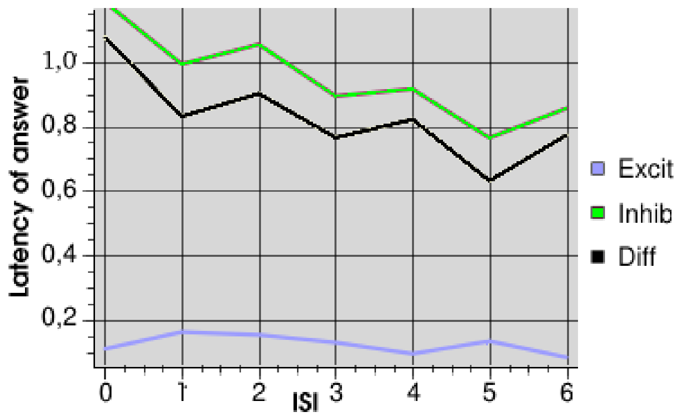

Figure 7: Intermodal priming effect on latency of answer

Both measurements are presented as functions of ISI, in congruent (Excit) and non-congruent (Inhib) conditions, for a primer latency of 2 reverberations. The third curve is the difference (Diff) between the two conditions.

Each variable set (Primer latency = 1...4, ISI = 0...6) has been tested for 5 sessions of 100 tests. The "Diff" curve stands for the effect of priming (congruent vs. non-congruent), and has a tendency to become smaller for higher values of ISI. However, this curves shows an intermediate peak, corresponding to an optimal value of ISI, where the priming effect is most visible.

ANOVAs have been performed on the 5 sessions. Each session is considered as the results of a single subject, tested for all the variable sets. First we find a significant difference between congruent and non-congruent conditions for both latency ($F(1,4)=1003$, $p \leq 0.0001$) and identification rate ($F(1,4)=11258$, $p \leq 0.0001$). Second we find a significant effect for ISI ($F(6,24)=8.8$, $p \leq 0.0001$ for latency, $F(6,24)=35.7$, $p \leq 0.0001$ for identification rate), proving that the variations observed in figures 6 and 7 are consistent.

4.4 Discussion

There is no significant effect of the primer latency. However an optimal ISI clearly emerges only in the case where the latency of the primer is long enough (≥ 2 reverberations). In congruent condition, the network has quite better results when the primer first drives the network to the primer pattern (what happens quite more often with repeated presentations), and then converges to the corresponding target pattern. Conversely, in non-congruent condition, the dynamics of the network has to change from one attractor towards a different one, which is more difficult for small values of the ISI. Hence the priming effect is increased, specially for ISI = 2 reverberations.

The results reported in this article reproduce qualitatively the same behavior as observed on human subjects [14]. It should be noticed that the model is a mean to get quite complete results, including a systematic study of the ISI effect. Similar experiments, on human subjects, have not been reported in litterature, for intermodal priming. A model can be experimented on a large number of tests more easily. However quantitative comparisons are not directly possible between the behavior of the model and human beings. Priming experiments on human beings usually involve time $\sim$ 1s for each presentation, whereas a simulation on the network lasts $\sim$ 100ms (the delay between two successive reverberations is 24.4ms).

Moreover, the spiking BAM is a quite simple model of associative memory, with "directly recurrent" connections, as opposed to "polysynaptic loops" [21], that are supposed to induce long-term dynamics in the brain [22].

5 Conclusion

A BAM network has been emulated in temporal coding, with spiking neurons instead of usual threshold units. The spiking BAM reproduces the properties of a classical BAM, but with faster convergence, as was reported by Maass & Natschläger for the emulation of an Hopfield network. Furthermore, a spiking BAM has the property to integrate new inputs while its dynamics is running. Adding such intrinsic temporal properties at the neuron level induces temporal effects of higher scale in a modular neural network modeling a multimodal associative memory. This article has proved that improvement from experiments testing the model in intermodal priming conditions, inspired from the protocols applied to human subjects. The behavior of the model is qualitatively similar to the behavior of human beings. The hypothesis of existence of an optimal value of ISI has been validated on the model.

From a computational point of view, further improvements can be considered for the model of spiking BAM. Fixed synaptic weights have been used in the present work. Since a spiking neuron can locally update its weights while it is computing, according to the correlation of emissions between pre- and post-synaptic neurons, further research will focus on the integration of Spike-Time Dependent Plasticity (STDP) in the network, in order to realize on-line learning in the BAM.

References

[1] W. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115–137, 1943.

[2] J. Hertz, A. Krogh, and R.G. Palmer. *Introduction to the Theory of Neural Computation*. Reading MA : Addison-Wesley, 1991.

[3] S. Haykin. *Neural Networks: A comprehensive foundation, 2nd edition*. Prentice Hall, 1999.

[4] E. Reynaud. *Modélisation connexionniste d'un mémoire associative multimodale*. PhD thesis, Institut National Polytechnique de Grenoble, 2002.

[5] A. Crépet, H. Paugam-Moisy, E. Reynaud, and D. Puzenat. A modular network for binding several modalities. In H. R. Arabnia, editor, *IC-AI'2000*, pages 921–928. International Conference on Artificial Intelligence, 2000.

[6] H. Paugam-Moisy and E. Reynaud. Multi-network system for sensory integration. In *IJCNN'2001*, pages 2343–2348. International Joint Conference on Neural Networks, 2001.

[7] S.J. Thorpe and M. Imbert. Biological contraints on connectionist modelling. In R. Pfeifer, Z. Schreter, F. Fogelman-Soulie, and L.Steels, editors, *Connectionism in Perspective*, pages 63–92. Amsterdam : Elsevier, 1989.

[8] C.M. Gray and W. Singer. Stimulus specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Sciences*, 86:1698–1702, 1989.

[9] W. Gerstner and J.L. van Hemmen. Associative memory in a network of 'spiking' neurons. *Network*, 3:139–164, 1992.

[10] W. Maass. Fast sigmoidal networks via spiking neurons. *Neural Computation*, 9(5):279–304, 1997.

[11] W. Maass and T. Natschläger. Networks of spiking neurons can emulate arbitrary Hopfield nets in temporal coding. *Network: Computation in Neural Systems*, 8(4):355–372, 1997.

[12] B. Kosko. Bidirectional associative memories. *IEEE Transactions on Systems, Man & Cybernetics*, 18:42–60, 1988.

[13] L. Nadel, editor. *Encyclopedia of Cognitive Science*. nature publishing group, 2002.

[14] D. Cernis and R. Versace. Laboratoire d'études des méchanismes cognitifs, Univ. Lumière Lyon II, personnal communication.

[15] S. M. Kosslyn and O. Koenig. *Wet Mind: The new cognitive neuroscience (2nd edition)*. The Free Press, New-York, 1995.

[16] D. Puzenat. Priming an artificial neural classifier. In *IWANN'95, From Natural to Artificial Neural Computation, Lecture Notes in Computer Science*, volume 930, pages 559–565. Springer, 1995.

[17] E. Reynaud and D. Puzenat. A multisensory indentification system for robotics. In *IJCNN 2001*, pages 2924–2929. International Joint Conference on Neural Networks, 2001.

[18] Y. Bouchut, H. Paugam-Moisy, and D. Puzenat. Asynchrony in a distributed modular neural network for multimodal integration. In *PDCS 2003, Parallel and Distributed Computing and Systems*, page (to appear). IASTED, 2003.

[19] W. Maass. Networks of spiking neurons: the third generation of neural network models. *Neural Networks*, 10:1659–1671, 1997.

[20] W. Gerstner and W. Kistler. *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge : Cambridge University Press, 2002.

[21] E. D. Lumer, G. M. Edelman, and G. Tononi. Neural dynamics in a model of the thalamocortical system II. the role of neural synchrony tested through perturbations of spike timings. *Cerebral Cortex*, 7:228–236, 1997.

[22] G.M. Edelman. *Neural Darwinism: the theory of neuronal group selection*. New York : Basic Books, 1987.