Temporal Pattern identification using Spike-Timing Dependent Plasticity

Frédéric Henry[*], Emmanuel Daucé[*][†], Hédi Soula[‡]

August 31, 2006

Abstract

This paper addresses the question of the functional role of the dual application of positive and negative Hebbian time dependent plasticity rules, in the particular framework of reinforcement learning tasks. Our simulations take place in a recurrent network of spiking neurons with inhomogeneous synaptic weights. The network spontaneously displays a self-sustained activity.

A Spike-Timing Dependent Plasticity (STDP) rule is combined with its "opposite", the "anti-STDP". A local regulation mechanism moreover maintains the output neuron in the vicinity of a reference frequency, which forces the global dynamics to be maintained in a softly disordered regime.

This approach is tested on a simple discrimination task which requires short-term memory : temporal pattern identification. We show that such temporal patterns can be categorized, and present tracks for future improvements.

1 Introduction

Since the first observations of synaptic plasticity [3], the measurement techniques have considerably grown up. Important interest has recently come over the fine dependence on the timing of spike arrival in the synaptic potentiation or depression phenomena. Those time dependent mechanism have been popularized as "Spike-Timing Dependent Plasticity" (STDP), and various models and implementations have been proposed. It can be noticed however that both "positive" and "negative" spike-timing dependence have been observed, depending both on the animal and on the location. At the present time, too few measurements have been made for an exhaustive description of the spike-timing dependent rules taking place in the brain.

More generally, the biological mechanisms of knowledge acquisition and memory formation remain at a very early stage of understanding. We propose in the present paper to explore the mechanism of a dual application of STDP and anti-STDP for the realization of a classification task in an artificial neural network. The idea is to use the spontaneous capacity of random recurrent neural networks to form complex patterns of activity, and to use STDP and anti-STDP mechanisms as positive and negative reward to "shape" those patterns of activity in order to fulfill at best the external constraints.

[*]Movement & Perception (UMR6559), Faculty of Sport Science, University of the Mediterranean, 163, avenue de Luminy, CP910, 13288 Marseille CEDEX 9, France

[†]École généraliste d'ingénieurs de Marseille (EGIM), Technopôle de Château-Gombert, 38, rue Joliot Curie, 13451 Marseille Cedex 20, France

[‡]Laboratory of Biological Modeling, NIDDK, National Institutes of Health, Bldg 12 A Bethesda, MD 20892, USA

Our paper is organized the following way. The second section gives the model of neuron, the STDP plasticity rule, and the structure of the network we simulate. The third section presents some basic features on the effect of STDP and anti-STDP on the local and global neuronal dynamics. In the fourth section we present the main learning task we use: a discrimination task between several temporal sequences. We also make comparisons with other methods, and present some tracks in terms of biological plausibility. The fifth section presents our conclusions and discussions about the future work.

2 Neuron and network models

We are mainly interested in the group behavior of artificial neurons. For that, we simulate rather simple and classical models of integrate-and-fire neurons.

2.1 Neuron model

The model of neuron we use is the leaky integrate-and-fire [13]. This well-known model does not fulfill every biological constraint, but reasonably models the temporal behavior of spiking neurons. It is easy to implement, and thus allows the simulation large networks.

We actually use a discrete implementation of this model where a time step roughly corresponds to one millisecond. The membrane potential of neuron i at step t is given by:

$$u_i(t) = \gamma u_i(t-1) + \sum_{j=1}^{N} w_{ij}\delta(t - T_j) \qquad (1)$$

where γ is the neuron's leak, w_{ij} the synaptic weight from neuron j to neuron i, T_j the date of the last EPSP arrival from neuron j, and

$$\delta(x) = \begin{cases} 1 & \text{if } x = 0 \text{ ;} \\ 0 & \text{elsewhere.} \end{cases} \qquad (2)$$

If $u_i(t) > \theta_i(t)$, the neuron fires, and its voltage is reset to its resting potential 0.

In our model the threshold θ is given by a random gaussian draw of mean $\bar{\theta} = 1.0$ and standard deviation $\sigma_\theta = 0.2$.

2.2 Learning rule

Our synaptic update rule is a particular implementation of the Spike-Timing Dependent Plasticity (STDP) [2], where the long-term potentiation is additive while long-term depression is multiplicative [14]. The weight change Δw depends of the temporal difference Δt between the presynaptic EPSP arrival and the postsynaptic spike. The weight change is given by $\Delta w = F(\Delta t)$ with

$$F(\Delta t) = \begin{cases} A_-\alpha e^{-\Delta t/\tau} & \text{if } \Delta t < 0 \text{ ;} \\ -A_+\alpha w e^{-\Delta t/\tau} & \text{if } \Delta t > 0. \end{cases} \qquad (3)$$

where A_- and A_+, and α are the learning coefficient, and τ is the relaxation rate. We set $\tau = 10$ and $A_+ = 1$; thus two parameters are still needed in order to characterize the rule: α and A_-. The "anti-STDP" simply corresponds to a STDP with a negative α.

2.3 Network structure

The network we simulate belongs to the category of random recurrent neural networks. All the synaptic weights are set according to a Gaussian draw (see figure 1 for the precise parameters). Those parameters are set in order to allow the internal self-sustained activity to compete with the external stimulation. It can be noticed that a precise analysis of the spontaneous activity of comparable random networks of integrate-and-fire neurons is given in [11].

In this particular setup, we use a three-layers network. The first layer is composed of input neurons, which receive the signal from the environment. Those neurons send connections toward every neuron of the internal layer. The internal layer is composed of a hundred of fully connected neurons. At last, some output neurons receive synapses from the internal layer. Those output neurons do the "read-out" of the internal activity. The output neurons are moreover mutually inhibitive.
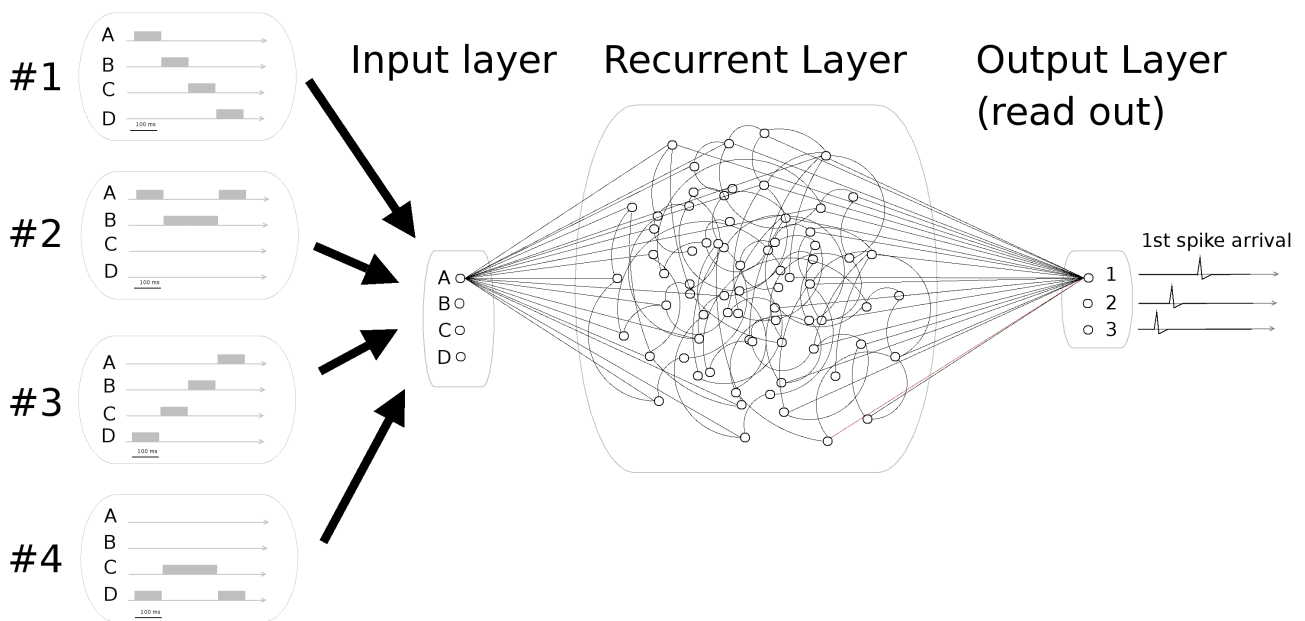

Figure 1: **Experimental setup**. N temporal patterns are to be presented to the network in order to be classified in K categories. The network is composed of three populations. The input layer is composed of 4 neurons (labeled A, B, C and D). The input connections follow a random Gaussian law of mean zero and standard deviation 0.04. The hidden layer contains 100 fully connected neurons. The recurrent connections follow a random Gaussian law of mean 0 and standard deviation of 0.02. The output layer is composed of K neurons, with lateral inhibition (not represented). The output connections follow a random gaussian law of mean 0.09 and standard deviation 0.01.

3 Effects of STDP and Anti-STDP

We present first in this section some effects of the dual application of the STDP and the anti-STDP rules. Since the STDP increases the synaptic weight when the postsynaptic neuron fires shortly after the presynaptic one (in equation 3, $\Delta t < 0$), and, conversely, decreases it when it fires before ($\Delta t > 0$), the most noticeable effect of this rule is thus to reinforce the sequential co-activation of the pre and post-synaptic neurons. On the contrary, no straightforward interpretation of the anti-STDP rule can be given, since the depression of a co-activation path and the potentiation of post-spike EPSP tends to radically modify the spike timing of the target neuron (i.e. to decorrelate to the most EPSP arrivals and spike emission). To the most can we say that the anti-STDP tends to "blur" the output neuron response.

We present in the following two simulations examples where the positive STDP rule is followed by the anti-STDP rule.

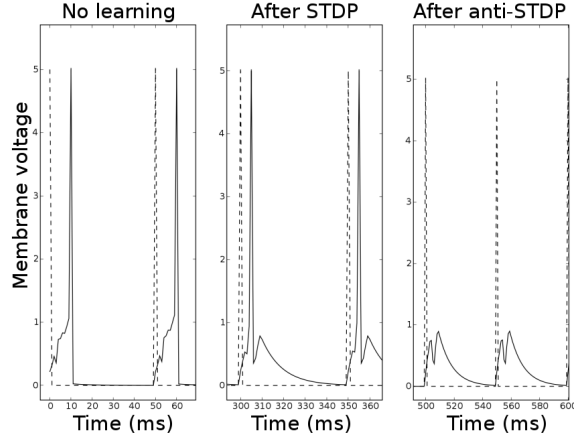## 3.1   Reduction of response delay

Figure 2: **Effect of the STDP and anti-STDP rules on the response of an output neuron**.   The left figure represents the membrane potential of one input neuron (dashed line) and of the output neuron (solid line). The central figure shows the same thing after application of the STDP, which causes the output neuron to fire earlier, whereas the right figure shows the membrane potentials after anti-STDP, which in that case prevents the output neuron to fire.

In the most simple cases, the interpretation of the STDP and anti-STDP rules is rather straightforward. Let us consider a network only composed of 20 input neurons and one output neuron with random connections from the input layer to the output one (gaussian draw of mean $\bar{w} = 0$ and standard deviation $\sigma_w = 0.1$; in this application delays are also added to the synapses and correspond to a poisson draw of mean $\bar{d} = 5$). We force the input neurons to fire, and then externally stimulate the output neuron during 10 steps. Figure 2 gives the evolution of the output neuron membrane potential after the synaptic adaptation. After several time steps, the neuron is found to fire earlier. The neuron has learned to become more reactive to its post-synaptic firing. Conversely, the application of the anti-STDP tends to delay the answer or even, in this case, prevents the output neuron to fire.
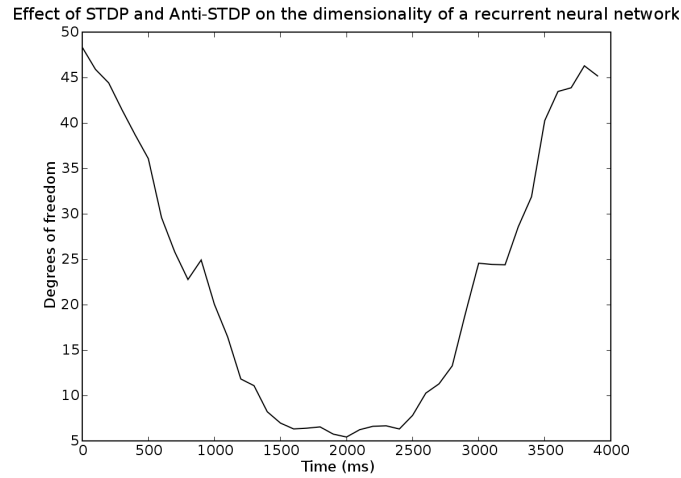
3.2   Recurrent network

The application of STDP/anti STDP on a recurrent network with self-sustained activity is a much more challenging issue. We simulate here a network with no input, such that all the activity comes from the reverberated activity of the internal layer (see section 2.3). In our experiment, the STDP is applied for 2000 time steps, followed by 2000 steps of anti-STDP (see Figure 2). During the STDP application, a progressive increase of the neurons regularity can be observed. This effect can be interpreted as the strengthening of an internal co-activation path, resulting in some neurons taking part in the dynamics at a high firing rate, while others remain silent (see figure 3a). This effect is distinct from the previous one (section 3.1), and is specific to the case of recurrent dynamics. Interestingly, the opposite effect happens while the anti-STDP is applied. The neurons are found to de-correlate their activity and tend to fire more aperiodically, finally resuming the initial nearly random activity.

Those varriations in regularity can be measured in a practical way using an approximation of the effective number of degrees of freedom using a straightforwrd linear annalysis [17], [10].
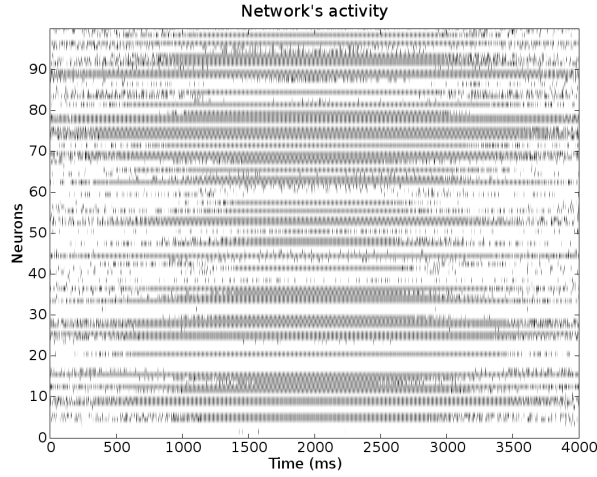
(a) Activity

(b) Dimensionality

Figure 3: **Application of STDP and anti-STDP on a recurrent network with self-sustained activity**. The network is here a full connected network. Weights are set with a gaussian draw of mean zero and standard deviation 0.25. STDP is applied in the first 2000 steps ($\alpha = 0.001, A_- = 2.5$), and anti-STDP in the last 2000 ($\alpha = -0.001$). a) Activity of the network. Each neuron is represented by a black bar when it fires, and by a white bar else. STDP raises the activity and synchronises the neurons, while anti-STDP do the opposite. b) Evolution of the number of effective degrees of freedom ("dimensionality") across the time. This number is calculated every 100 steps with the values of neurons' membrane potentials in the last 100 steps. STDP reduces the dimensionality, thus increasing the "order" of the system, while anti-STDP does the opposite.

Our data set is composed of the membrane potentials of all the neurons over sliding windows of 100 time steps. A Principal Components Analysis[1] is first applied to the data set, followed by a calculation of the entropy of the normalized principal values p_i of the transformation matrix :

$$S = -\Sigma_{i=1}^{N} p_i ln(p_i) \tag{4}$$

This value considered [10] is considered as an approximate log count of significant principal components weighted by their respective size, so that

$$D = e^S \tag{5}$$

is an approximation of the effective number of degrees of freedom. This measure of the effective dimension is plotted on figure 3b. The almost symmetrical effect of the two opposite rules is clearly exhibited. Interestingly, a comparable duration for the two rules exposition allows to restore the initial disordered activity, i.e. to erase of the initial memory trace.

This example clearly illustrates the complementarity of the two rules. Their concurrent application may allow to control the degree of complexity of the internal dynamics, and thus the nature of the network's response.

4 Application

4.1 Reinforcement learning

Reinforcement learning (also called reward learning or operant conditioning) is a class of learning problems where an agent attempts to adapt its behaviour to maximize its rewards. In biology, it is often assumed that such rewards occur through the release of particular neurotransmitters.

Reinforcement is suggested to be one of the most primitive nervous adaptation mechanisms, though it does not need any explicit model or consign. It can be noticed, for instance, that some forms of operant reward learning have been shown to take place on very simple invertebrate animals [4]. It is thus of real interest to understand the basis of reward operant learning in order to allow deeper and anatomically founded analysis.

A lot of actor-critic models[2] have been proposed in the recent years, few of which being consistent with the known anatomy of the basal ganglia [8]. The main problem with such "high level" models is the lack of knowledge of the real anatomy of the implied structures.

Unlike actor-critic models, we hypothesise that the reinforcement mechanisms do not need any explicit model of the environment (so that one single structure may be involved). Our model thus falls in the category of "direct Policy learning" methods[16], which are much coarser but also more relistic than the highly sophisticated TD-learning [12] and Q-learning [15] methods.

We suggest here :

- To model the pattern generation mechanism with the use of a random recurrent neural network with self-sustained activity. This endogenous activity is seen as a basis for short term memory (STM) capacity, as already noticed by [9] and [7].

- To model the selection process with a balanced Hebbian/anti-Hebbian mechanism. A Hebbian weight reinforcement is supposed to take place when a positive reward occurs, while a anti-Hebbian reinforcement would take place when a negative reward occurs. This kind of model has been realized for instance in [1] with stochastic neurons.

---

[1]which is a linear transformation to a new coordinate system where the greatest variance of the transformed data lies on the first axis, the second greatest on the second axis, and so on.

[2]which separate the controllers in two part, one of which modelling the world ("critic") and the other part choosing the appropriate actions ("actor").

In our application (see figure 1), we measure the reactivity of the output neurons to various temporal input patterns. If the first neuron to spike belongs to the pre-defined category, a positive reward is sent. In the other case, a negative reward is sent. We associate the positive reards with the application of the STDP rule while a negative reward is associated with the application of the anti-STDP rule. We hope the STDP to increase the reactivity of the output neuron in the case of a correct answer, and the anti-STDP to do the opposite (see figure 2).

4.2 Network's dynamics

In this application, all the delays are set to one and there is no refractory period. A light background random activity is added to the system. The spontaneous dynamics is weak, very irregular, and the neurons fire asynchronously. The standard deviation of the weights distribution belongs to the lower limit of the parameter domain which allows such self-sustained activity, *provided significant input is sent on the input layer.* Such a network thus lies in the parametric domain described by [9] and [7].

The figure 3 has shown that a balanced application of positive and negative STDP may maintain the network initial regime. The idea is thus to explore the parameter space through this balanced synaptic process, for the system to improve its behavior. The use of both STDP and anti-STDP is supposed to maintain the network activity in such a "viable" domain.

However, keeping this balance is not easy in practice. Despite our initial observations, the positive STDP tends to dominate in the long term the anti-STDP. This small lack of symmetry gives rise in the long term to a highly correlated internal activity which causes the dynamics to become stereotypical and the network to ignore its inputs (blindness situation).

In order to prevent this, we add a regulation principle in order to increase the negative part of STDP when a neuron's frequency starts to raise (and on the contrary does the opposite when the frequency starts to decrease). Our mechanism is a local one which operates on parameter A_- in order to maintain the firing frequency "close" to the target frequency f_{target}:

$$A_- = \frac{f_j(t)}{f_{target}} \qquad (6)$$

where $f_j(t)$ is the trace of the activity of the postsynaptic neuron :

$$f_j(t) = \gamma_f f_j(t-1) + (1 - \gamma_f)x_j(t) \qquad (7)$$

where $\gamma_f = 0.999$ is a leak factor and $x_j(t) = 1$ if neuron j has fired at step t and zero else.

4.3 Temporal patterns classification

Our method is tested here on a temporal patterns classification task. The network must learn to classify N different temporal patterns in K categories. The number of potential categories is given by the number of output neurons. In this experiment we take $N = 4$ and $K = 3$. The total pattern duration is 400 steps (that is, 400ms). In a given sequence, each letter (from A,B,C and D) corresponds to the stimulation of a particular input neuron for 100ms at a rate of 100Hz. The four different input sequences are the following:

- A, B, C, D

- A, B, B, A

- D, C, B, A

- D, C, C, D

The expected category for each sequence is chosen for the answer not to rely at any time on a single active input neuron. The network must thus develop its short-term memory capacity in order to give a proper answer. We consider that the network's answer is the output neuron that fires first after the input presentation (the most "reactive" neuron to that particular input). Notice that the network is not supposed to answer too soon (namely not after the third element of the sequence has been presented). In our setup, we introduce an unbalance in the category repartition : the two sequences (A, B, C, D) and (D, C, C, D) correspond to the same output neuron (neuron 1), so that the probability of a reinforcement occuring on neuron 1 is twice stronger than on the other ones.
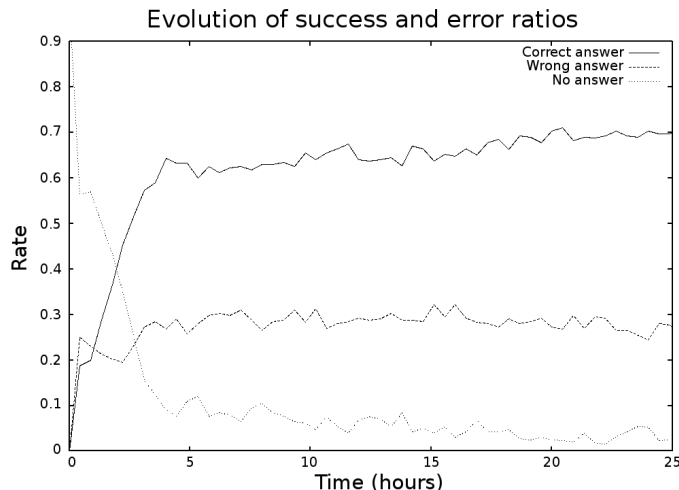

Figure 4: **Temporal patterns identification.** This figure gives the rate of correct pattern identification during the learning process. The rates are computed on a sliding window of the last 100 presentations of patterns.

A significant improvement of the network response is obtained during the whole learning process (see figure 4): the STDP/anti-STDP mechanism is found to allow to associate a decrease (resp. increase) of the response delay for the output neurons when the correct (resp. wrong) answer is given. Due to the regulation mechanism presented above (equation 6), the synaptic saturation (and thus the catastrophic collapse of the performance) is avoided (or at least postponed for a very long learning time).

In the given example, the network learns to classify some of the patterns but never manages in learning all of them. If we look more closely on the success ratios for the four different patterns (figure 5), we see that the network manages to classify (A,B,C,D) and (D,C,C,B) in category 1, and (A,B,B,A) in category 2, but fails to classify (D,C,B,A) in category 3. The reason of this misfit lies in the statistics of the rewards : two success for one failure in the "category 1" response. This "two against one" configuration consolidates the actual response, and prevents the weight to adapt to the specific failure case.

The given simulation is representative of typical network performances. The success rates remain of the order of 60/70%, which is still good for a task that requires active short term memory. This mixed success is thus a first step in order to validate the STDP/anti STDP mechanism as possible implementation of the direct policy learning methodology in realistic neurons. Its simplicity make of it a good candidate for plausibility. However, the poor capacity to compensate unbalanced rewards on the output category also shows its limits. A simple adaptation mechanism at the level of the neurons may possibly compensate this, giving more credit to the most "rare" events (in the case of disbalanced positive or negative rewards).
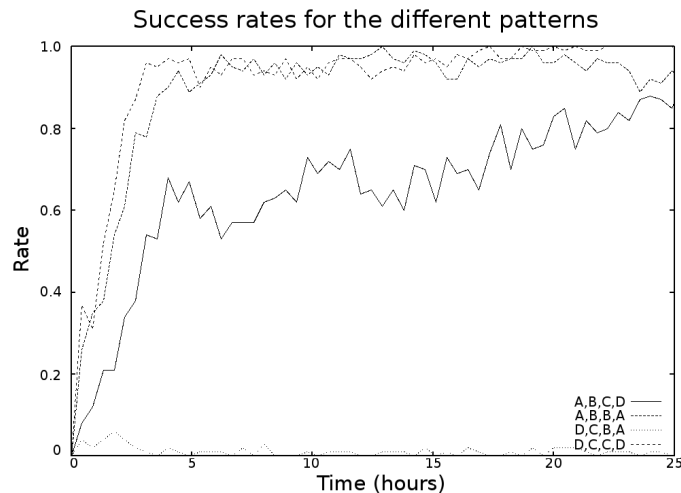
Figure 5: **Identification rates for each of the four different patterns.** The rates are computed on the last 100 presentations for a given pattern and correspond to the percent of correct answer for this pattern. While three patterns are correctly learned, the network fails to give the appropriate answer for D,C,B,A.

5 Discussion

We have shown with this model that the dual application of a STDP/anti-STDP mechanism allows to extract from an active recurrent layer the information necessary to achieve an classification task that requires short term memory.

This achievement is of course dependent of a significant set of parameters (weights distribution, thresholds, coefficients for STDP, etc.). Since the model uses a very sensitive mechanism to keep the balance between two dynamics, the parameters must be set very precisely. An interesting approach to resolve the parameter problem would be to use optimization mechanisms, such as genetic algorithms, not to directly calculate the "good" weights, but the good parameters. This approach has already given some interesting results with Hebbian rules [6] [5], and we expect it to be operant with the dual STDP/anti-STDP approach.

In terms of biological modelling, two points remain under consideration. The first concern is about the functional role of the various dynamical regimes we observe in simulation. We did not fully elucidate whether a change in the regimes (synchrony/disorder) did take part in the production of the correct answer. We can only suppose that the maintenance of a softly disordered regime helps the system to explore various answers and to select and consolidate the most appropriate ones.

A second question remains about the plausibility of such STDP/anti-STDP mechanisms. Could they be triggered by the release of different neurotransmitters, for instance dopamin and serotonin? The release of such transmitters taking place on rather long duration in comparison with the neuronal integration time, they may not be associated to a single spike response but more realistically to a firing pattern. This numerical experiment is thus a first step toward more realistic and biologically-founded models, using refractory periods, delays and behavior-based reinforcement tasks.

References

[1] P. L. Bartlett and J. Baxter. Hebbian synaptic modifications in spiking neurons that learn. Technical report, Research School of Information Sciences and Engineering, Australian National University, 1999.

[2] G.-Q. Bi and M.-M. Poo. Synaptic modifications in cultured hippocampal neurons : Dependence on spike timing, synaptic strength, and postsynaptic cell type. *The Journal of Neuroscience*, 1998.

[3] T. Bliss and T. Lomo. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J. Physiol*, 232:331–356, 1973.

[4] B. Brembs. Operant conditioning in invertebrates. *Current Opinion in Neurobiology*, 13:710–717, 2003.

[5] E. A. Di Paolo. Evolving spike-timing-dependent plasticity for single-trial learning in robots. *Phil. Trans. R. Soc. Lond. A.*, 361:2299–2319, 2003.

[6] D. Floreano and J. Urzelai. Evolutionary robots with on-line self-organization and behavioral fitness. *Neural Networks*, 2000.

[7] H. Jaeger. Adaptive nonlinear system identification with echo state networks. *NIPS*, 2002.

[8] D. Joel, Y. Niv, and E. Ruppin. Actor-critic models of the basal ganglia: new anatomical and computational perspectives. *Neural Networks*, 15-4:535–547, 2002.

[9] W. Maass, T. Natschläger, and H. Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation*, 2002.

[10] A. Penn. Steps towards a quantitative analysis of individuality and its maintenance : A case study with multi-agent systems. In *Fifth German Workshop on Artificial Life: Abstracting and Synthesizing the Principles of Living Systems*, 2002.

[11] H. Soula, G. Beslon, and O. Mazet. Spontaneous dynamics of asymmetric random recurrent spiking neural networks. *Neural Computation*, 18:60–79, 2006.

[12] R. Sutton. Learning to predict by the method of temporal differences. *Machine learning*, 3:9–44, 1988.

[13] H. Tuckwell. *Introduction to Theoretical Neurobiology*. Cambridge University Press, Cambridge UK, 1988.

[14] M. C. W. van Rossum, G. Q. Bi, and G. G. Turrigiano. Stable hebbian learning from spike timing-dependent plasticity. *The Journal of Neuroscience*, 20(23):8812–8821, December 2000.

[15] C. Watkins and P. Dayan. Q-learning. *Machine learning*, 8:279–292, 1992.

[16] R. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8:229–256, 1992.

[17] W. A. Wright, R. E. Smith, M. Danek, and P. Greenway. A generalisable measure of self-organistion and emergence. In G. Dorffner, H. Bischof, and K. Hornik, editors, *Artificial Neural Networks - ICANN 2001*, 2001.