Associative memories and aftereffects

Federica Menghini

31st May 2006

Introduction

In this project we analyze the properties of attractor dynamics of a simple autoassociative network that can be thought of as a plausible model of a patch of cortex. We then use our model to reproduce the "cortical" adaptation aftereffect in the classification of emotional ambiguous faces.
In the first part, we implement an associative neural network provided with two layer structure. The first layer works as an input station and projects the activity to the output layer through feed-forward connections. The output layer is provided with recurrent collaterals that connect the units within the layer and make the network work as an associative memory.
The simulation is divided into a training and a testing phase. During the training phase the network stores p patterns through the learning algorithm based on recurrent weights modification. In the testing phase we test the ability of the network to correctly retrieve a pattern when it's presented at the first iteration and it's immediately removed. We measure the performance of the network in the testing phase in term of percent of correct decoding and we study how the performance varies with the modulations of the strength of the contribution from recurrent collaterals during learning. In testing phase the strength of the contribution of recurrent collaterals is always equal to the feed forward strength to allow pattern completion.
We consider two different cases: the first whenthe synapic weights modified through the learning process can assume positive and negative values, and the second when positive values only. We compare network performance in the two cases.
We also study the convergence of the network to a stable point and the ability of the network to distinguish between learnt patterns. We then study how the introduction of a geometrically organized connectivity affects network performance. We want in fact to investigate if such a simplified model can incorporate a plausible element such as a geometrically organized connectivity.
We also study the storage capacity of the system measuring how the performance of the network varies when we increase the number of patterns to be stored. We finally investigate if the learning process has an effect

also in unlearnt patterns comparing the "convergence" of dynamical process to a stable point when unlearnt patterns are presented before and after learning process. In the second part we try to reproduce the psychophysical adaptation aftereffect through our model implemented allowing positive and negative weights and recurrent collaterals not active during learning.
We then create morphs from the learnt patterns and we present a noisy version of them to the network that has to decide if the output activity is closer to the first or the second pattern generators. We investigate whether the presentation of a prime and a mask before the target stimulus (morph) induces a shift in the curve obtained presenting the morph only. An analogous shift in perception is experimentally found when subjects, that have to classify as emotive or neutral a face with ambiguous emotional expressions, are "primed" with an emotive or a neutral face. We find that the reproduction of the experimental results is related to the introduction of an adaptation term in firing rate that we model through two different shapes. We finally investigate if the shift is still present when morphs are created from unlearnt patterns.

Away from the prime

We present a simple autoassociative neural network model of cortical adaptation that replicates a **high level category aftereffect**, similar to that described in Webster and MacLin (1999) [1] and Webster (2004) [2]. We replicated this aftereffect in a series of psychophysical experiments. In these experiments, a target face image with an ambiguous emotional expression is made to seem more or less emotional or neutral by means of a prime face image preceding the target. The effect of the prime, in particular, is to push the perception of the target in the opposite direction to the prime. For example an ambiguous face with an emotional component will seem "more emotional" if the prime is neutral. We study in our model the comparable effect of a pattern on retrieval from a partial cue of short duration (relative to the number of cycles needed for retrieval). We also explore which learning conditions permit correct pattern retrieval.

Chapter 1

A model of a cortical patch

1.1 An associative network

Autoassociative networks are able to retrieve a pattern previously stored when a noisy or occluded version of it (partial cue) is provided as input. This ability is due to the formation of dynamical attractors [3, 4, 5] that capture network activity if an input is sufficiently close to one of the patterns stored as attractors in the network.
We can recognize the content of a visual stimulus even when it is severely degraded [6]; faces in particular are extremely robust, and a convincing way to explain this is by assuming that the cortex has similar pattern completion dynamics [10].
Our model simulates a hypothetical local network in a higher level area of the visual system, e.g. in the inferior temporal lobe. The network includes an input station which projects the activity to the "cortical" output layer through sparse feed-forward (FF) connections (from 68 out of 900 input units, per receiving units) [11]. The output layer is provided with recurrent collateral (RC) connections (342 out of 900 output units).

Training phase

The p patterns stored in the network are generated randomly by assigning a value derived from a common truncated exponential distribution to each

input unit. Differently from the Hopfield model that includes one layer only, in our model the attractors do not correspond to the generated patterns, but to their projection through feedforward connections in the output layer.
Feedforward connections are defined at the beginning and kept fixed during the simulation, while recurrent weights are updated according to a modified "Hebbian" rule

$$\delta J_{ij}^{rc} = k r_i^{\mu}(r_j^{\mu} - \langle r \rangle), \tag{1.1}$$

as a result of a training phase, where r_i^{μ} and r_j^{μ} are the pre and post synaptic firing rate of the input pattern, $\langle r \rangle$ the activity averaged over all the units and k the learning rate, which quantifies the strength of the learning process. We set the value of the learning rate to $k = 0.006$.
In the training phase, the p patterns are presented to the network one by one. The activity circulates in the network for 60 time steps and at each time step every unit receives an input comprised of the contribution of FF and RC connections plus a global inhibitory term b

$$h_i = \sum_j J_{ij}^{ff} r_j^{inp} + M \sum_j J_{ij}^{rc} r_j^{out} + b, \tag{1.2}$$

where r_j^{inp} and r_j^{out} are the activities of the jth unit in the input and in the output layer respectively. The response of a unit is characterized by a threshold linear function

$$r_i^{out} = \begin{cases} g(h - T_{thr}) & h > T_{thr} \\ 0 & h < T_{thr} \end{cases} . \tag{1.3}$$

where g is the gain amd T_{thr} is the threshold. They are updated at the end of the dynamical process (60 reverberations); then the next pattern is presented.

Connection weights

The connections are not organized according to a geometrical structure but they randomly defined. They are initially defined by the equation

$$J_{i,j}^{in} = J_{i,j}^{0} + J_{i,j}^{ran}, \tag{1.4}$$

where $J^0_{i,j}$ is a constant term and $J^{ran}_{i,j}$ is a small random asymmetrical component described by the equation

$$J^{ran}_{i,j} = -\frac{1}{2}\beta \log(1 - \frac{y}{\alpha}), \tag{1.5}$$

with α and β constant terms and y statistical variable uniformly distributed between 0 and 1. The value of the weights is sincronously updated at the end of each pattern presentation according to the equation (1.1); weight values are then given by

$$J_{i,j} = J^{in}_{i,j} + \delta J_{i,j}. \tag{1.6}$$

Two cases are considered: positive and negative weights and positive weights only.

In the first case we set $J^0_{i,j}$ to 0 and we let $J_{i,j}$ assume positive and negative values, corresponding, collectively to EPSPs from pyramidal cells, mediated by synapses with modifiable efficacies, and IPSPs from inhibitory interneurons.

In the second case, we consider excitatory synapses only and $J_{i,j}$ is cut off at 0 if it attains negative values. We set $J^0_{i,j} = \frac{1}{N_{rc}}$ to avoid that the too many synapses are set to 0 as a result of learning process.

Testing phase

In the testing phase a pattern is provided as input to the network and removed after the first iteration; the parameter M is set to 1 to allow the dynamics to proceed through the activity of recurrent collaterals.

The output activity at the end of 60 iterations is then compared in term of overlap with all the patterns as they are represented in the output layer. The representation of each pattern is obtained by projecting the activity of each unit in the pattern to the output layer through the FF connections (i.e. output activity after 1 iteration).

The overlap is calculated with the equation

$$Ov^{1,t} = \frac{\sum_i r^{out}_i(1) r^{out}_i(t)}{\sqrt{\sum_i (r^{out}_i(1))^2}\sqrt{\sum_i (r^{out}_i(t))^2}}, \tag{1.7}$$

where $r^{out}_i(1)$ is the activity of the ith unit in pattern representation after 1 time step, and $r^{out}_i(t)$ the activity of the ith unit at the end of the dynamical

process of 60 iterations.
The pattern retrieved is defined as the one with highest overlap with the result of the test.

1.1.1 What is the role of recurrent connections in learning?

We investigate the role of the parameter M in the training phase to see how the activity of recurrent collaterals affects learning. During testing phase the parameter M is set to 1 to allow the dynamicsto proceed through the activity of recurrent collaterals We vary M between 0 (*RC not active* during learning) and 1 (*RC active* during learning with the same average strength as FF weights) and measure the performance of the network as percent of correct decoding. This measure indicates whether the network is able to correctly retrieve the input pattern when it is removed after the first iteration.

Fig. 1.1 shows the percent of correct decoding as a function of the parameter M for posive only and positive and negative weights. In both cases the network performance is better when recurrent collaterals are not active during learning ($M = 0$), and it rapidly decreases to an asymptotic value with increasing of *RC* strength. The result is in agreement with Hasselmo's hypothesis [7] according which recurrent connections are not strongly activated during learning, but they are widely in the retrieving phase. when recurrent collaterals are active during learning.
Fig. 1.1 also shows that the network performance is improved for any vaule of M by the introduction of negative weights.

1.1.2 Attractor formation

We first check that the learning rate k is appropriate for the convergence of the dynamical process to stable points (Fig. 1.2), measuring the Euclidean distance between two successive time steps

$$DT_{\mu}^{t,t+1} = \sqrt{\sum_{i=1}^{N} \left(r_i^{\mu}(t) - r_i^{\mu}(t+1) \right)^2}. \tag{1.8}$$

Since the network performance decreases monotonically from $M = 0$ to $M = 1$, we perform the next measures for the "extreme" conditions $M = 0$

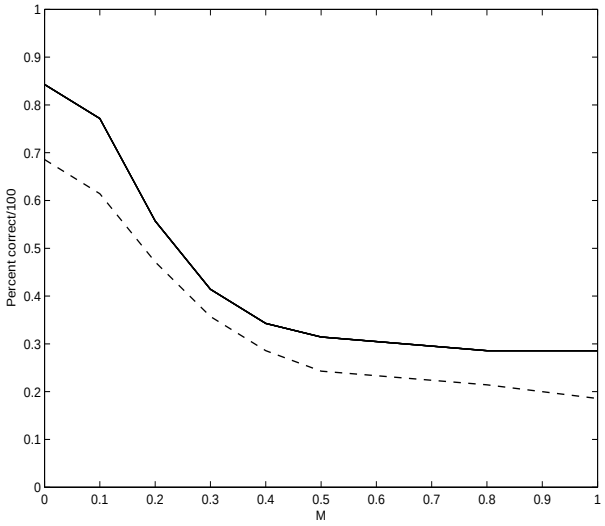

Figure 1.1: Percent correct as a function of the parameter M that modulates the strength of recurrent collaterals during learning, positive weights only (bottom curve), positive and negative weights (top curve).

(RC not active during learning), and $M = 1$ (RC active with the same strength as FF). Fig. 1.2 shows that with both values of M the network converges to a stable point in time. The largest state change is from the first (when the input is applied) to the second iteration (after input removal) and this change is much larger when the network is trained with *RC* active. However in *RC* active condition the convergence is quicker. Nevertheless in the RC not active condition the oscillation around the stable point are larger than in the RC active condition.

We then ask whether the network performance is related to the degree of "attractor separation". We compute the Euclidean distance between the network activities of any pattern pair at a certain time step

$$DR_t^{\mu,\mu'} = \sqrt{\sum_{i=1}^{N} \left(r_i^{\mu}(t) - r_i^{\mu'}(t) \right)^2}. \tag{1.9}$$

This measure expresses how much the activities of two patterns overlap after a fixed number of iterations. If the Euclidean distance is close to 0 the pattern pair falls into the same basin of attraction and the network is not

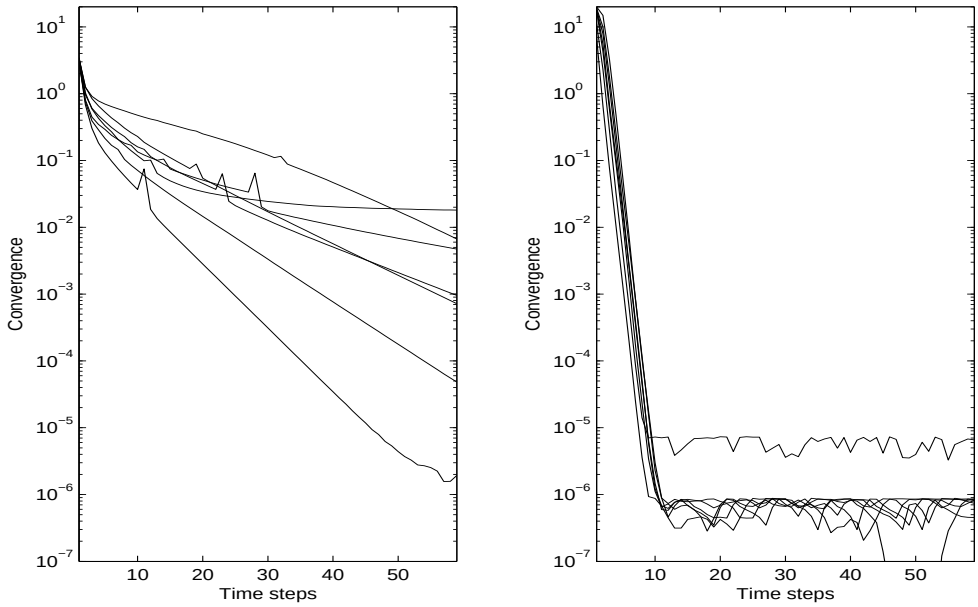

Figure 1.2: Semilogarithmic scale; Euclidean distances between two successive time steps as a function of time steps for $M = 0$ (left panel) and $M = 1$ (right panel) for the 7 patterns . The results shown are computed allowing negative weights, but the convergence for positive weights only is qualitative the same.

able to distinguish between them.

Attractor separation

Fig. 1.3 shows the normalized histograms of all possible pairs of reciprocal distances obtained at the end of the dynamics (60 iterations) in 10 independent runs.
If $M = 1$ the majority of reciprocal distances goes to 0 indicating that some of the input patterns converge to the same basin of attraction.
When $M = 0$ (solid lines), most of the distances are large indicating that the dynamical process leads two distinct patterns almost always to different basins of attraction.
In both conditions the presence of negative weights leads to a decrease in the number of collapses and consequently to an increase in the ability of the network to distinguish among the patterns. When $M = 0$ and negative weights are allowed there is still a small number of distances close to 0, indicating the presence of a few collapses into the same basin of attraction. This imperfect pattern separation could probably be related to finite size effects, since the number of units (900) is far away from the thermodynamic limit in any practical sense. We increase the network size (from 900 to 1225 while keeping the connectivity ratios (N_{rc}/N and N_{ff}/N) and the load $\alpha = p/N_{rc}$ fixed. Fig. 1.4 shows that even if the increase in network size is not large, however there is a decrease in number of collapses, supporting then the hypothesis of finite size effects.

1.1.3 How does learning rate affect network performance?

The learning rate k has been chosen to allow the network to converge to stable points as a result of the dynamical process. We investigate whether the strength of the learning process affects the network performace, and we plot (Fig. 1.5) the percent of correct decoding as a function of learning rate for $M = 0$ and for $M = 1$. We consider only the case where negative weights are allowed.
For $M = 0$ the percent correct increases with the learning parameter k up to an asymptotic value while for $M = 1$ the percent correct is always at a

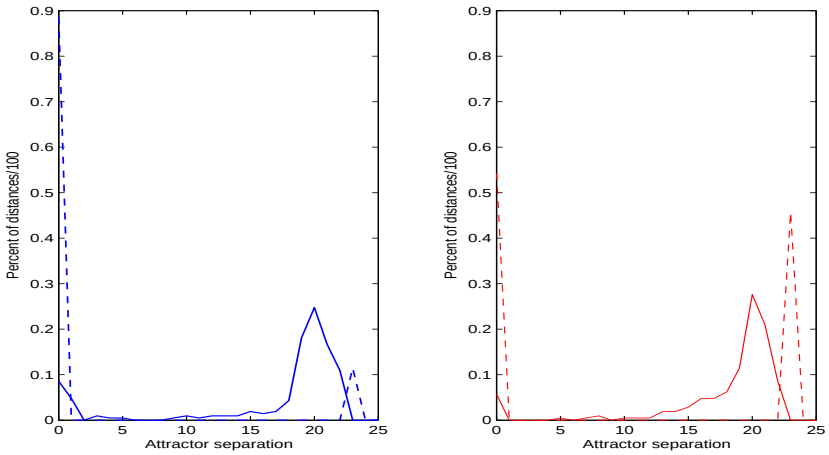

Figure 1.3: Normalized histograms of the Euclidean distances between all possible pattern pairs obtained after a 60 time-steps dynamics in 10 independent runs for $M = 0$ (solid line) and $M = 1$ (dashed line). Positive weights only (left panel), Positive and negative weights (right panel)

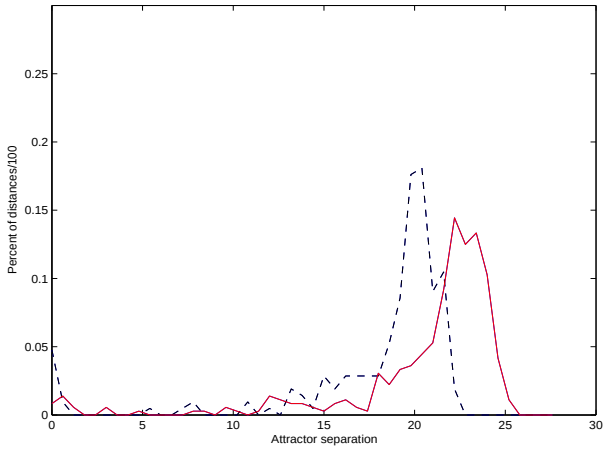

Figure 1.4: Comparison between the network with 900 units and a network with 1225 units. Normalized histograms of the Euclidean distances between all possible pattern pairs obtained after a 60 time-steps dynamics in 10 independent runs for $M = 0$ and with positive and negative weights. 900 units network (black dashed line), 1225 units network (red solid line)

poor level. There is a maximum in the percent correct for $k = 0.001$ but its value is still below 50% of correct retrieval.
The results show that over a broad range of k values the network performance is not significantly affected by the change of learning strength. So the network performance is stable to variation in the learning parameter k.

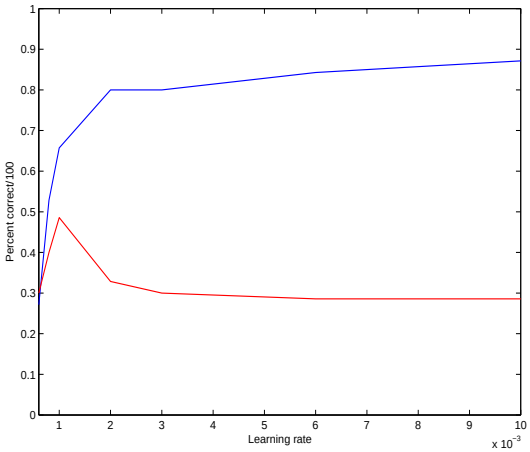

Figure 1.5: Percent correct as a function of the learning rate for $M = 0$ (blue line) and $M = 1$ (red line); Positive and negative weights are allowed.

Geometrically organized connectivity

In the model implemented as above the synaptic connections are defined randomly, without any geometrical structure. We investigate how the performance of the network and the number of collapses for $M = 0$ and $M = 1$ are modified by the introduction of a geometrically organized connectivity, where the synaptic connections are defined as a function of the reciprocal distance between the units. Fig. 1.6 shows that the network performance has the same qualitative shape as in the random connectivity case, but the percent correct is always lower. We also observe that the number of collapses is higher both for $M = 0$ and $M = 1$ even if the difference is not large.
Out model is then too rough and simplified to allow the introduction of a plausible element such as a geometrical structure that assigns higher prob-

ability to closer units.

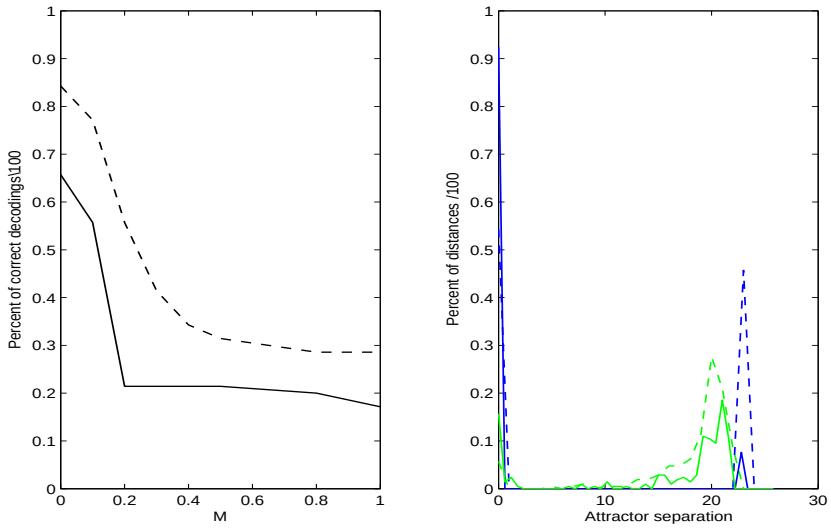

Figure 1.6: Left panel; Percent correct as a function of the parameter M for a geometrically organized connectivity ($\sigma = 6$) (solid line), and for a random connectivity ($\sigma = \text{inf}$); right panel: normalized histograms of the Euclidean distances between all possible pattern pairs for $M = 0$ (green line) and $M = 1$ (blue line) for a geometrically organized connectivity (solid lines), and for a random cpnnectivity (dashed lines); Positive and negative weights are allowed.

1.2 Storage capacity

We investigate how percent correct changes as a function of the number of patterns. This measure address the question of the storage capacity, ($\alpha_c = p/N_{rc}$) the maximum number of patterns that can be stored and correctly retrieved by the network.

Fig. **??** shows that percent correct decreases almost linearly increasing the number of patterns and that there is not a sharp transition for any critical number of patterns as it happends with the Hopfield model. However the introduction of threshold linear units was shown to induce a smoother behaviour in the network performance [8].

The presentation of a larger number of patterns than the ones indicated by the storage capacity (e.g. $p = 20$, Fig. 1.8) does not prevent the convergence to stable points (figure not shown) but increases the number of collapses from $\sim 5\%$ to $\sim 18\%$, where we arbitrarily indicates with "collapses" reciprocal distances smaller the 5.

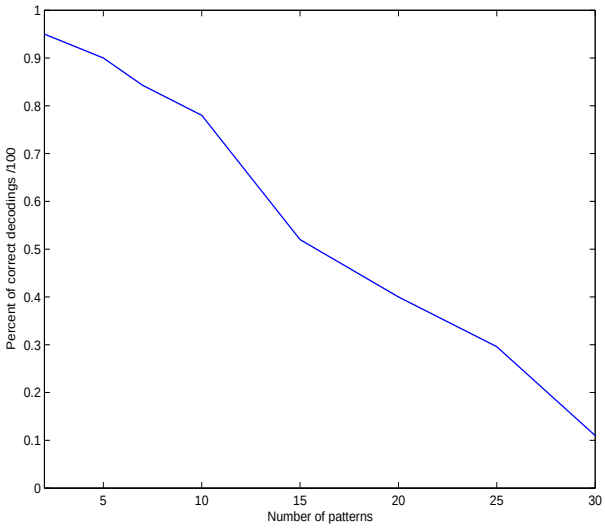

Figure 1.7: Percent correct as a function of the number of patterns.

1.3 Convergence before and after attractor formation

We investigate what is the effect of learning process in unlearnt patterns in RC not active condition and when negative weights are allowed.
We find that network performance is poor (figure not shown), but that all unleanrt patterns converge to a stable point (Fig. 1.9b), even if the convergence is slower than for learnt patterns.
The convergence to a stable point is a consequence of learning process since before weights update the state continue to chenge during dynamical process and it does not reach a stable point (Fig. 1.9a). The effect of learning process then is make the network converge to a stable point, no matter whether or not the input pattern is stored in the network.

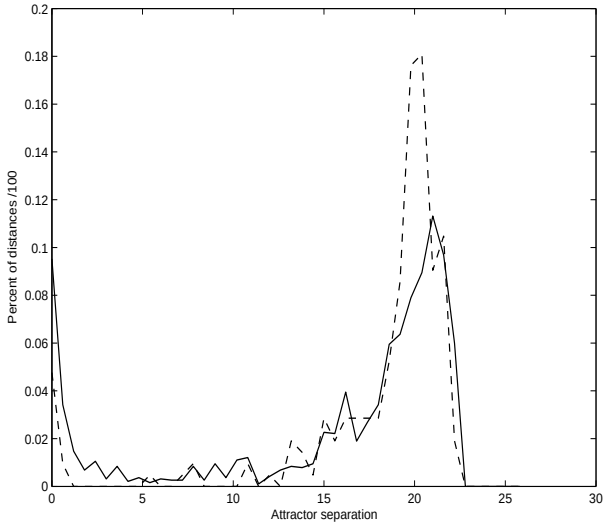

Figure 1.8: Normalized histograms of all possible reciprocal distance for 7 (dashed line) and 20 (solid line) patterns.

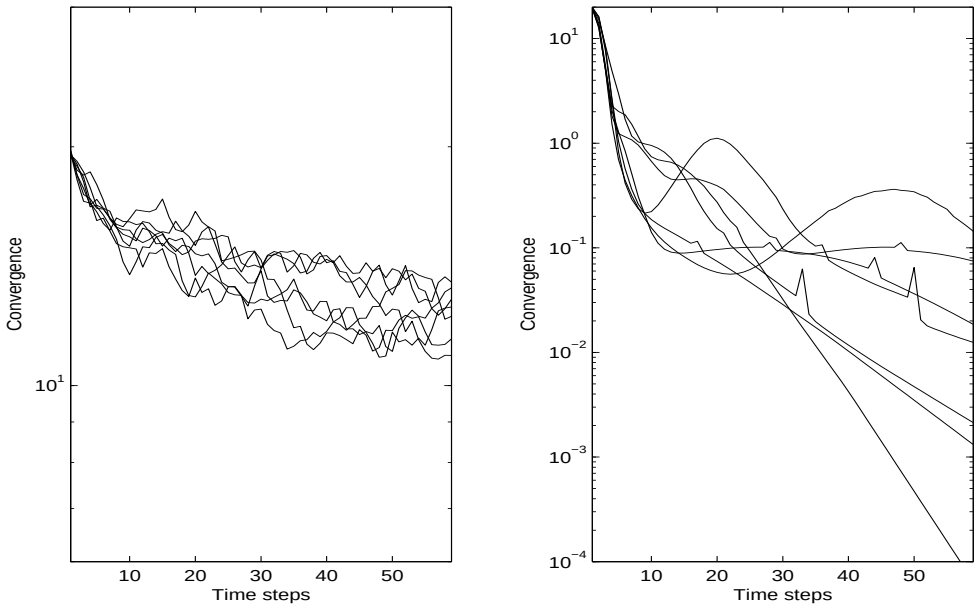

Figure 1.9: Semilogarithmic scale; Euclidean distances between two successive time steps as a function of time steps for 7 unlearnt patterns before (left panel) and after (right panel) learning process . The results shown are computed allowing negative weights, and for $M = 0$.

Chapter 2

Adaptation aftereffect

2.1 Experimental adaptation aftereffect

We now want to investigate whether the model implemented as above (in the *RC not active* condition) is able to reproduce the adaptation aftereffect. We first recapitulate the experimental study [9]. Nicola et al have demonstrated a robust adaptation effect with face stimuli, where subjects were asked to classify face images as either neutral, or emotional.

2.1.1 Experimental procedure

12 subjects classified 12 sets (28 morph strengths) of morphs. Prior to classification trials, subjects were familiarized with the set of adaptors (i.e. the primes). For a given morph set the adaptor was either the emotive extreme, or the neutral extreme of the morph set, and the direction of adaptation for each morph set was counter balanced across subjects. On each trial they saw a prime (the adaptor for 22 ms or 500 ms duration), followed by a mask of 50 ms. After 250 ms subjects were asked to classify a target face (of the same identity), as neutral or as the appropriate emotion.

2.1.2 Results

We found significant statistical aftereffects with both prime lengths (results reported fully elsewhere).

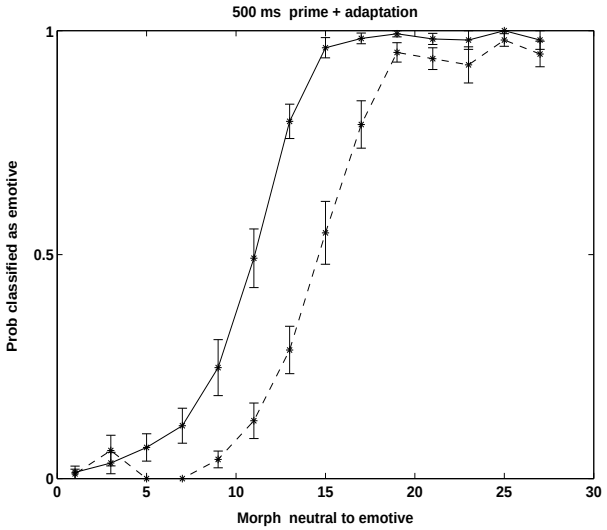

Figure 2.1: Experimental results obtained when subjects are asked to classify faces with ambiguous emotional expressions (morphs) as neutral or emotive. The probability of perceving the target stimulus as emotive is plotted against morphing scale for neutral prime (solid line) and emotional prime (dashed line). Adaptor length 500 ms.

2.1.3 What is needed to reproduce the adaptation aftereffect?

We represented morphs by combining two of the stored patterns in different proportions, from 0 to 100. Taking all possible pairs of patterns was considered to adequately sample pattern space.

The testing phase

We have changed the testing phase to reproduce the experimental paradigm, leaving the learning phase as it is. We present a noisy version of a morph to the network that has to "decide" through a decoding algorithm if the output activity is closer to the first or the second pattern. The experimental results show a relative shift of the curves obtained with an emotive and a neutral adaptor, induced by the adaptation aftereffect.
To use the same language as the experiment, "emotive" corresponds to the first pattern in the pair of morph generators and "neutral" to the second.
In the simulations we compare the results obtained with and without an adaptor.
In the "without adaptor" condition we present a morph to the network that has to "classify" it as "neutral" or "emotive".
In the "with adaptor" condition we present a prime, which we choose as the first pattern in the pair ("emotive"); a mask,finally the morph. The mask consists in noisy activity in the output layer which is realized through the subtraction from the input activity to each cell of a term proportional to the residual activity coming from the prime. The mask has then the effect of stopping the activity coming from the prime. We found there is no shift between the two conditions, without the addition of an adaptation factor.

Adaptation in firing rates

We then introduce adaptation in firing rates as a simple model of neural fatigue. We implement it as the subtraction from the input activity of each cell of a term proportional to the difference of two exponentials with different time constants:

$$h_i^{in}(t) = h_i(t) - \alpha[r_{1i}(t) - r_{2i}(t)], \tag{2.1}$$

with

$$r_{1i}(t) = r_{1i}(t-1)\exp(-\beta_1) + r_i(t-1)$$
$$r_{2i}(t) = r_{2i}(t-1)\exp(-\beta_2) + r_i(t-1),$$

where $r_i^{in}(t)$ is the input activity to the ith cell at time t, $h_i(t)$ is the term defined in eq. (1.2) at time t, β_1 and β_2 are the time constants, $r_i^{out}(t-1)$ is the output activity of the ith cell at time $t-1$ and $r_i(0) = r_{1i}(0) = r_{2i}(0) = 0$. The input to each cell is then affected by the firing rates of all the previous time steps. The expontential decay makes the activity of the last time step more influential than the others.

The difference of the two exponential means that the effect of adaptation appears only after the second iteration, taking into account that neural fatigue does not appear instantaneously. It also makes the adaptation not too strong when t is small.

The same adaptation waveform was previously used to solve the computational conflict between pattern "completion" and "prediction" [12] in an architecture similar to ours which modeled hippocampal CA1-CA3 layers. This choice is not the only possible one, so we try, for example to model this term as the subtraction from the input activity of a square wave.

Results

The introduction of adaptation in the firing rates induces a shift (Fig. 2.2) in the curves corresponding to the two different conditions (with and without adaptor). We can then reproduce the adaptation aftereffect including neural fatigue in the model described.

An analogous result is found implementing the adaptation term as the subtraction from the input activity of a square wave. (Fig. 2.3). The size of the effect is the same for the two shapes; the only (minor) difference appears at the extremes of morphing scale. Implementing the adaptation term as a "square wave" in fact the aftereffect disappears for the extremes of morphing scale. On the contrary if we implement it as subtraction of two exponential decays the effect is still present for morphs close to 100% of correlation with the prime, while there is a small reversal effect for morphs close to 0% of correlation with the prime. This effect though is very small.

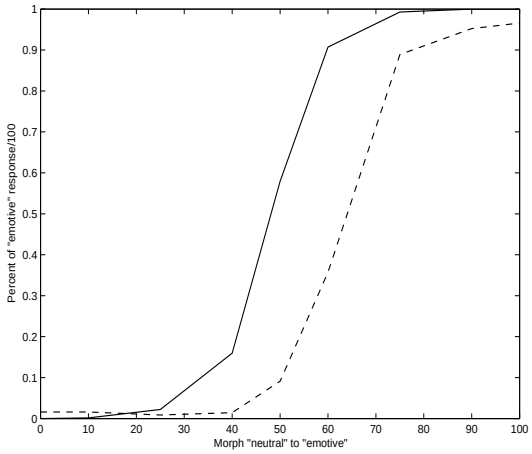

Figure 2.2: Percent of "emotive" responses as a function of the morph scale when the morph is presented without adaptor (solid line), and with adaptor (and with the introduction of adaptation in the firing rate) (dashed line).

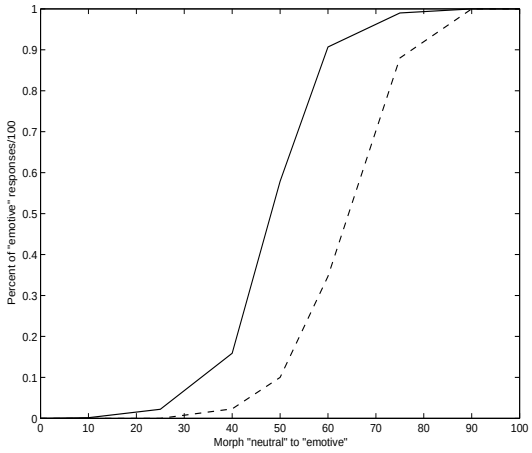

Figure 2.3: Percent of "emotive" responses as a function of the morph scale when the morph is presented without adaptor (solid line), with adaptor and with adaptation modelled as subtraction from the input activity of a square wave (dashed line).

2.2 Adaptation to repetition

We investigate more in detail what is the effect of adaptation in firing rates. We then compute the overlap between morph representation after 60 time steps and the output activity resulting from the dynamical process when a noisy version of the morph is presented to the network eq. (1.9). We plot the average overlap as a function of morphing scale (Fig. 2.4), finding that the overlaps are higher for the extremes, which are learnt patterns.
This result is compared with the case when the network is stimulated with the morph, with the mask, and with the morph again. We observe (Fig. 2.4) the the overlap is not significantly modified by the new morph presentation for the extremes, but it symmetrically decreases toward the middle of the scale. This result suggests that the adaptation in firing rate has a significant effect when the network is not in an attractor state, while it has a very small effect if the network has already reached a basin of attraction.

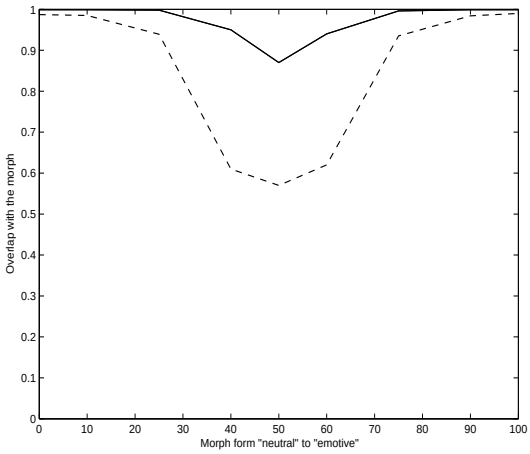

Figure 2.4: The effect of repetition on the morph; solid iine, adaptation after one repetition expressed as the overlap between the representation of the morph and the output activity of one repetition; dashed line overlap between the initial representation and the output activity resulting from the second repetition.

Do aftereffects need attractor formation?

We finally test the "with" and "without" adaptor conditions when the network was presented with morphs generated by two unlearnt patterns. We find (Fig. 2.5) that the sigmoid is not reproduced, when attractors relative to the pattern generators are not formed, but the shift between the two curves is still present, even if the size of the effect is smaller than in the previous case. Aftereffects do not strictly require, then, attractor formation: they require, of course, the presence of the adaptor. Their exact shape, though, like the shape of the nonadapted sigmoid "neurometric" curve (Fig. 2.2) does depend on the formation of attractors.

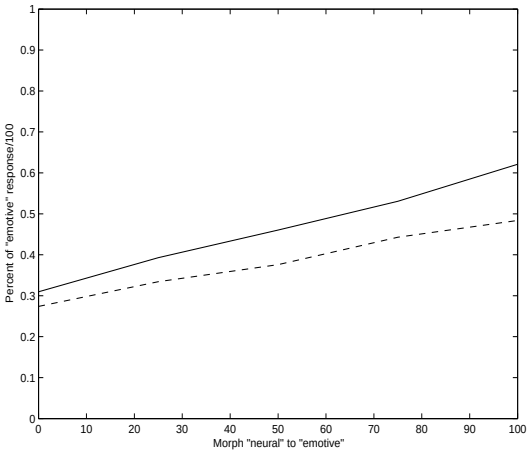

Figure 2.5: Morphs generated with unlearnt patterns. Percent of "emotive" responses as a function of the morph scale when the morph is presented without adaptor (solid line) and with adaptor (dashed line).

2.3 Discussion

We have reproduced the adaptation aftereffect with a simple associative network. In a simple but plausible model of the cortex, we found that the network is able to efficiently learn the generated patterns only when recurrent connectivity is not active in the training phase. The activation of recurrent collaterals during learning leads to a decrease in performance

which was stronger the larger was RC strength. The poor performance when RC are active during learning was related to the problem of the collapse of many pattern pairs into the same basin of attraction.
The model was able to reproduce the adaptation aftereffect only with the introduction of adaptation in the firing rates as a representation of neural fatigue. Our results suggest that the minimal model required to show this kind of effect includes an explicit model of neural fatigue, although it does not dictate its exact form.
Future work should compare the dynamics of the decay of the real aftereffect [13] with our symplified model.

Conclusion

The results obtained implementing the model suggests that the network has a much better performance when both positive and negative weights are allowed and when the activity of recurrent collaterals is switched off during learning phase. In particular the network performance monotonically decreases with the strength of recurrent collaterals during learning. This is related to the increase in the number of collapses into the same basin of attraction and to the consequent decrease in the ability to distinguish among the patterns from the network.
We observe that the number of collapses is small when recurrent collaterals are not active during learning, but it is still present. This problem could be related to finize size effects since an increase in network size reduces the number of collapses.
The learning process leads to the convergence to a stable point of the patterns presented as input, no matter whether leanrt or not. We have also found that the network performance is stable over a broad range of values of the learning paramenter k, but that the model is too simple to obtain a good performance with RC not active when a geometrically organized connectivity is introduced.

The model implemented with RC not active during learning was able to reproduce the experimental data on adaptation aftereffect only after the

introduction of an adaptation term in firing rate. The choice of adaptation term is not unique: we have in fact obtained very similar results modeling this term as a square wave and as a difference between two exponentials.
We have observed that the adaptation aftereffect does not need attractor formation, it just requires the presence of an adaptor. However we could reproduce the sigmoidal shape only in presence of attractors. Finally we have observed that the adaptation in firing rate has its major effect when the network in not in an attractor state.

Bibliography

[1] M.A. Webster, and O.H. MacLin, Psychon. Bull. Rev. **6(4)**, 647 (1999).

[2] M.A. Webster, et al, Nature **428(6982)**, 557 (2004).

[3] D. Marr, Phil. Trans. R. Soc. Lond. B **262**, 24 (1971).

[4] D.J. Amit, Modelling Brain Function. Cambridge University Press (1989), Cambridge.

[5] D.J. Amit, Behav Brain Sci. **18**, 617 (1995).

[6] K. Grill-Spector, and N. Kanwisher, Psychol Sci. **16(2)**, 152 (2005).

[7] M.E. Hasselmo, R. Schnell, J. Neurosci. **14**, 3898 (1994).

[8] E.T. Rolls, A. Treves, D. Forster, C. Perez-Vicente, ,Neural Network **10 (9)**, 1559 (1997)

[9] N van Rijsbergen, A. Jannati, A. Treves, (in preparation).

[10] I.C. Whitfield, Brain Behav. Evol. **16**, 129 (1971).

[11] A. Treves, J. Comput. Neurosci. **14**, 271 (2003).

[12] A. Treves, Hippocampus **14**, 539 (2003).

[13] D.A. Leopold, G. Rhodes, K.M. Muller, L. Jeffery Proc. Biol. Sci. **272(1566)**, 897 (2005).