Connected Cortical Recurrent Networks

Alfonso Renart[1], Néstor Parga[1], and Edmund T. Rolls[2]

[1] Departamento de Física Teórica, Universidad Autónoma de Madrid,
Cantoblanco, 28049 Madrid, Spain
{arenart, parga}@delta.ft.uam.es
WWW home page: http://www.ft.uam.es/~parga/index_english.html
[2] Department of Experimental Psychology, Oxford University, South Parks Road,
Oxford OX1 3UD, England
rolls@psy.ox.ac.uk

Abstract. A model of an associative memory composed of many modules working as attractor neural networks with features of biological realism is proposed and analyzed using standard statistical physics techniques. The memories of the system are stored in the synapses between neurons in the same module and the synapses between neurons in different modules provide the associations between these memories. A study of the memory storage properties as a function of the strength of the associations is performed and it is found that, if it is large, global retrieval phases can be found in which selective sustained activities induced in modules which have not been stimulated. The form of the associations is such that, in the case of a tri-modular network studied, results from a psychophysical experiment on the simultaneous processing of contradictory information [1] can be qualitatively reproduced, within the limitations imposed by the simplicity of the model.

1 Introduction

Auto-associative memory systems, implemented in recurrent neural networks, have been intensely studied both to model the associative areas of the mammalian brain and to understand their storage capacity capabilities [2]. This effort has been mostly devoted to the analysis of *isolated* networks, Although this work has been fruitful, it is an idealization of the actual situation. Neuronal structures in the brain are linked to each other: columns in a given area are connected through axonal collaterals, areas receive cortical projections from other areas, and different sensory pathways converge to multi-modal sensory areas. Despite this fact, the theoretical understanding of multimodular recurrent networks is still underdeveloped [3].

In this work we study the memory storage and information processing properties of cortical multi-modular networks with a finite number of modules. One of the most interesting results, in terms of a model with a tri-modular architecture, refers to the McGurk effect on the simultaneous processing of contradictory information. This architecture (Fig. 1) consists of two different pathways, each of which processes different features of the stimulus, which are reciprocally connected to a convergent structure in which the information is integrated. The information coming from these pathways is conveyed, in both cases, to cortical modules (**A** and **B**), to be called input modules. The module where the information is integrated will be called the convergent or bi-modal module (**C**). In our model, this convergent module does not receive information except from **A** and **B**. Although this will be our usual interpretation of the model, the analysis allows that the modules can be in pathways of either different or the same sensory modality that, at some level, interchange information through the intermodular collaterals. A bi-modular architecture will also be considered in which the convergent module does not exist and the two input modules are directly and reciprocally interconnected.

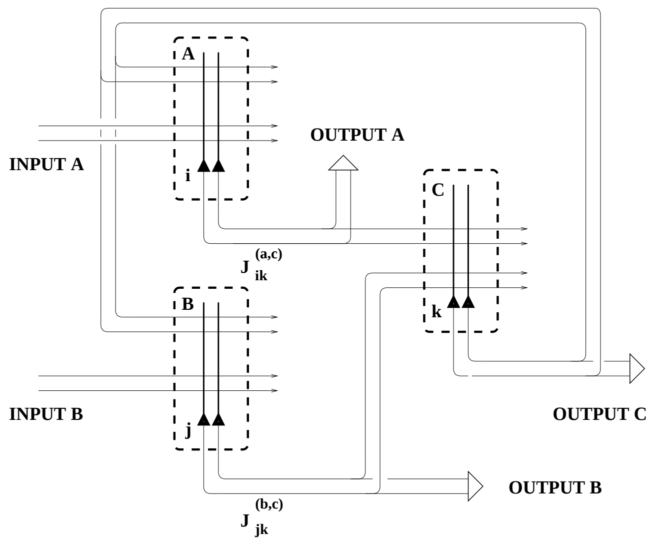

Fig. 1. An architecture with one bi-modal module. Triangles represent cell bodies, and thick lines their dendritic trees. $J_{ik}^{(c,a)}$ denotes the connection between the presynaptic unit k (in the module **C**) and the postsynaptic unit i (in the module **A**). Its relative strength is g. Similarly $J_{jk}^{(c,b)}$ is the synaptic connection between the neurons k and j located in the modules **C** and **B** respectively (with strength the same strength g). Since the synaptic matrix is assumed to be symmetric, the feedforward connections are equal to these back-projecting connections. The recurrent connections inside the modules are not shown in this diagram.

In both cases the multi-modular system is set up as follows: Local patterns of activity (or features) are stored in the connections between neurons belonging to the same module, whereas the connections between neurons in different modules store the associations between these features. Given the statistical properties of the sensory data, some features of the stimuli represented in one of the sensory pathways will appear temporally correlated with some of the features represented in the other module, causing these features to become associated. Both the associations between different features and the self-associations of the features with themselves are supposed to have been learned (during a learning stage not considered here) through Hebbian synaptic modifications between coactive cells.

2 The Multi-modular Network

The model is as follows:

The neurons: Neurons are described in terms of their firing rates in a way similar to the that in [4, 5]. The network dynamics is defined according to the set of equations:

$$\mathcal{T}\frac{dI_{ai}(t)}{dt} = -I_{ai}(t) + \sum_{bj} \mathcal{J}_{ij}^{(a,b)} \nu_{bj} + h_{ai}^{(ext)} \qquad a,b = 1,\ldots,M \ . \quad (1)$$

Here, I_{ai} is the afferent current into the neuron i of the module a, and ν_{bj} is the firing rate of the neuron j of the module b. The current is driven by the output spike rates of the other neurons in the network (located either in the same or in different modules), weighed with the corresponding synaptic efficacies $\mathcal{J}_{ij}^{(a,b)}$, and (in the case of the input modules) by the stimulus (or external field) $h_{ai}^{(ext)}$. The afferent current decays with a characteristic time constant $\mathcal{T}$. The conversion from currents to rates, necessary to complete the definition of the dynamics, will be indicated by $\nu = \phi(I)$, where $\phi(x)$ is the transduction function. We have chosen to study analog neurons with a hyperbolic transduction given by:

$$\phi(I) = \begin{cases} 0 & \text{if } I < \theta \\ \tanh[G\ (I-\theta)] & \text{if } I \geq \theta \end{cases} , \quad (2)$$

where G is the gain and θ is the threshold.

The stored patterns: In each module a, the number of stored patterns is denoted by P. The patterns are defined in terms of binary variables η_{ai}^{μ} ($\mu = 1,\ldots,P$; $i = 1,\ldots,N$). The η's are independent random variables which are chosen equal to one with probability f (the mean coding level of the stimuli) and equal to zero with probability $(1-f)$. Their variance is $\chi \equiv f(1-f)$.

The synaptic connections: The synaptic matrix will be denoted by $\mathcal{J}_{ij}^{(a,b)}$, where again a and b are module indices and i and j are neurons in a and b respectively. It is a multi-modular generalization of the one used in [6, 7]. The only constraint that we will impose on this matrix is symmetry under the interchange of the neuron indices. This will allow us to solve the model analytically. The intra- and inter-modular modular connections are given (respectively) by:

$$\mathcal{J}_{ij}^{(a,a)} = \frac{d_{ij}^{0a}}{\chi N_t} \sum_{\mu=1}^{P} (\eta_{ai}^{\mu} - f)\ (\eta_{aj}^{\mu} - f) \qquad i \neq j; \quad \forall a \quad (3)$$

$$\mathcal{J}_{ij}^{(a,b)} = \frac{g\ d_{ij}^{ab}}{\chi N_t} \sum_{\mu=1}^{P} (\eta_{ai}^{\mu} - f)\ (\eta_{bj}^{\mu} - f) \qquad \forall i,j; \quad a \neq b \ , \tag{4}$$

and $\mathcal{J}_{ii}^{(a,a)} = 0$. For the bi-modular architecture, (a, b) in eq. 4 can take the values $(\mathbf{A}, \mathbf{B})$ and $(\mathbf{B}, \mathbf{A})$, whereas in the tri-modular architecture it can be equal to $(\mathbf{A}, \mathbf{C})$ and $(\mathbf{B}, \mathbf{C})$. For this last architecture the symmetry condition has to be imposed setting $\mathcal{J}_{ij}^{(\mathbf{C},\mathbf{A})} = \mathcal{J}_{ji}^{(\mathbf{A},\mathbf{C})}$ and $\mathcal{J}_{ij}^{(\mathbf{C},\mathbf{B})} = \mathcal{J}_{ji}^{(\mathbf{B},\mathbf{C})}$. The parameter g measures the strength of the associations relative to the self-associations within the modules (which have been set equal to 1), and the variables d_{ij}^{0a} and d_{ij}^{ab} describe the connectivity within and between the modules, respectively. They are both $(0, 1)$ binary random variables which take the value one with probability d_0 and d (again respectively) independently of the pair of neurons observed and of the module they belong to. Only half of these variables are drawn randomly, the other half are set equal to their symmetric counterparts. The weight normalization is $N_t \equiv N\Lambda$ with $\Lambda = [d_0 + g\ d(M - 1)]$.

As noted before, the total number of patterns per module, P, is extensive. This means that

$$P = \alpha_a N_a \ , \tag{5}$$

where α_a is the load parameter of the module a, and N_a is the effective number of synapses afferent to a given neuron in that module. For the bi-modular architecture it is given by $N_a = N\ (d_0 + g\ d)$ for $a = \mathbf{A}, \mathbf{B}$, and it is therefore independent of a (i.e. $\alpha_a = \alpha$). For the tri-modular architecture, it depends on the module. It is $N_{\mathbf{A}} = N_{\mathbf{B}} = N\ (d_0 + g\ d)$ for the two input modules and $N_{\mathbf{C}} = N\ (d_0 + 2g\ d)$ for the bi-modal module $\mathbf{C}$.

The model was solved using statistical physics techiques [8, 5] which led to a set of self-consistency equations for the order parameters of the system. These parameters characterize the macroscopic states of the network. They allow partitioning of the space of parameters of the model into different regions, or *phases*, according to their retrieval properties. A full description of the solution of the model will be given elsewhere [9, 10] and we provide here a brief description of the results.

3 The Retrieval Diagrams

We first discuss the different retrieval regions of the bi- and tri- modular architectures, as a function of the association strength g. The retrieval diagrams can be seen in Figs. 2a-b respectively. To find out to which phase a given point in the (g,α) plane belonged to, the multimodular networks were assumed to be initially inactive, and were then persistently stimulated until a fixed point of the dynamics 1 was reached. For both diagrams in Fig. 2 only one of the input modules was stimulated in this way with a pattern equal to one of the stored features. The intensity of the stimuli was controlled by multiplying the stored pattern by a constant h.

Looking at Fig. 2a, one sees that in the large α region, the bi-modular network is in a non-retrieval or *spin glass* phase (SG) due to the large number of patterns stored. For α below a critical line, this is no longer the case, and the type of retrieval state depends also on the value of g. For g less than a critical line starting at $g \sim 0.15$ the system is in a *local retrieval* (LR) phase, in which only the stimulated module is significantly correlated with the stored pattern serving as stimulus. As g increases, the strength of the signal projected onto the other module grows, eventually causing this module to became significantly correlated with its stored pattern associated to the one serving as stimulus. When this happens we say that the network is in a *global retrieval* phase (GR). Although not shown on the diagram, if both modules are stimulated with corresponding (associated) features, the system will also reach a GR state of this type.

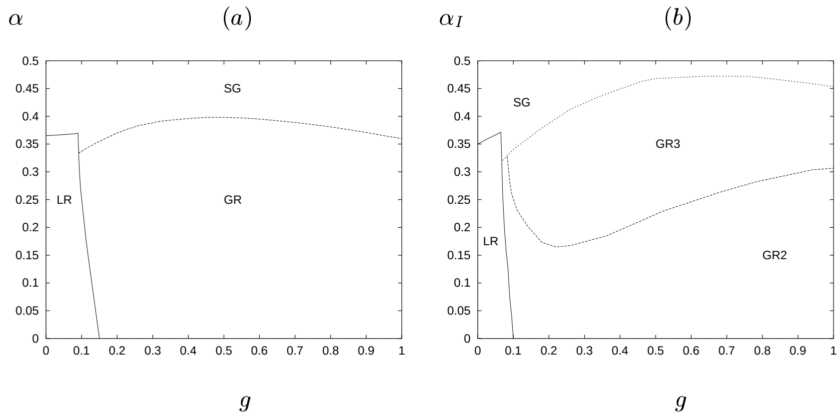

Fig. 2. Phase diagrams for the bi-modular (a) and tri-modular (b) architectures. In both cases only one of the input modules was persistently stimulated with a pattern equal to one of the stored features with intensity $h = 0.05$. Also in both cases $f = 0.22$, $d_0 = 0.1$ and $d = 0.05$, but $G = 1.3$ and $\theta = 0.02$ in (a) and $G = 1.35$ and $\theta = 0.015$ in (b). In (b) the vertical axis corresponds to the load parameter α_I of an input module (the one being stimulated). In the low g region the networks are in local phases, whereas for large g, global phases appear: sustained activity of associated features can be induced in a non-stimulated module through the reciprocal inter-modular connections.

The phase diagram for the tri-modular network (Fig. 2b) is qualitatively very similar to the previous one except for the structure of the GR phase, which is now divided in two: GR2 and GR3. In the low α portion, only the stimulated input and the convergent modules reach sustained activity states correlated with corresponding features (GR2), whereas for large α (but still below the SG phase) all three modules are able to retrieve (GR3). This is due to an increase in the firing rates of the neurons in the modules in retrieval as α increases, which in turn increases the strength of the signal projected to the non stimulated input module. In this phase (GR3), backprojections from the convergent structure are responsible for establishing selective activity in a given sensory pathway through the stimulation of a different sensory modality. This phase would again be reached if the *two* input modules were simultaneously stimulated with corresponding features.

4 The McGurk and MacDonald Effect

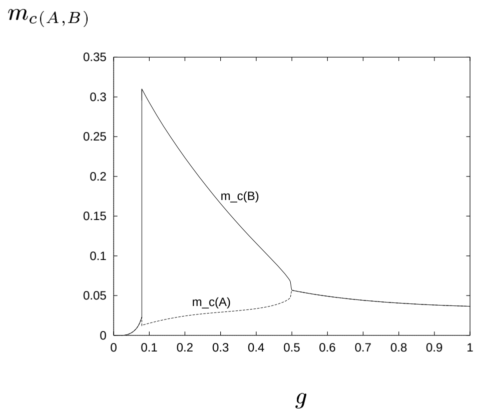

Fig. 3. Symmetry breaking and symmetry restoration as a funcion of g. With model parameters as in Fig. 2b and $\alpha = 0.15$. Both input modules were persistently stimulated with contradictory features of intensity $h = 0.05$. The overlap of the state of the convergent module with the two features associated with the stimuli are shown: the solid line corresponds to the overlap with the pattern associated with the stimulus on module B (m_c(B)) and the dotted line to the one associated with the stimulus on module A (m_c(A)). When only one line is seen is because the two overlaps are coincident. This corresponds to the symmetric region. When the symmetry is broken, the two overlaps are clearly separated.

Although the stimulation of the input modules with corresponding stimuli is certainly the natural situation under normal conditions, one may wonder what would happen if the tri-modular network was simultaneously stimulated with two contradictory (that is, non-corresponding) features (i.e. patterns that have not been associated during learning) which see each other as noise. Interference effects due to the existence of this type of noise can be helpful to understand the nature of the associations between the local features, and of the structure of the local features themselves. Psychophysical experiments have been conducted to study the simultaneous processing of contradictory information [1]. In this experiment, human subjects asked to identify a (manipulated) spoken syllable, fail to do so if the lip movements do not correspond with the syllable heard. In order to model this experience, we looked at the sustained activity states of the tri-modular network when the two input modules were simulaneously stimulated with contradictory patterns of the same intensity. We expect the different responses in the experiment to correspond to different sustained activity states of the bi-modal module. The results can be seen in Fig. 3, where the overlaps of the sustained activity state of the convergent module with the two stored patterns associated to the two patterns serving as stimuli are shown as a function of g, for a fixed value of $\alpha = 0.15$. These overlaps are a measure of the degree of similarity of the state of the module with the patterns. For g greater than $g \sim 0.5$ the two overlaps are quite small and identical. The state of the bi-modal module is symmetrical with respect to the two contradictory patterns and it does not correspond to a retrieval state for any of them. We propose that this region of g corresponds to the response category named "combination" by McGurk et al., in the sense that the convergent module has not recognized properly any of the stimuli but has an equal *preference* for the two. As g goes below this value, the association strength is smaller but the firing rates increase, effectively increasing the strength of the signal to the convergent module. When this happens, the bi-modal module is forced to *decide* between one of its alternatives. Since we have chosen the situation to be completely symmetric, which one is realized depends on the fluctuations of the dynamics on the way to the attractor. Once the final state is selected, the symmetry is spontaneously broken. This means that the final state does not exhibit the symmetry of the network. We propose that this corresponds to the visual *or* the auditory response categories in the experiment, in which the subject decides between one of the two possibilities. It should be kept in mind that these results aply only to the experimental paradigms in the experience where no *fused* responses occur.

Of course, we do not expect to find such an strictly symmetrical situation in anyone's brain, but the important point is that if the convergent module is able to reach retrieval in this extreme case, then small fluctuations in, e.g. the strength of the association to any of the input modules or the amplitude of the stimulus on each of the two sensory pathways, can lead to the different responses given by the whole subject population, with the possibility that different answers could be obtained from a given subject in *a priori* identical stimulation conditions.

5 Conclusions

The main results of the analysis performed are: For both architectures, the individual networks are able to retrieve a number of memories proportional to the number of effective connections afferent to a given neuron in the network. Also, the multimodular system as a whole is able to correctly associate the different features stored in each module, in the sense that selective sustained activity states can be reached on a given module under stimulation of a different one. For the case of a tri-modular network, when the two input modules are stimulated with contradictory stimuli, our model qualitatively reproduces the results of one of the experimental paradigms of [1]. The prediction is that the visual or auditory responses would be given by subjects in the broken symmetry phase, whereas the rest of the responses would be given by subjects in the symmetric phase, which are not able to decide between any of the two alternatives.

References

1. McGurk, H., MacDonald, J.: Hearing leaps and seeing voices. Nature **264** (1976) 746-748
2. Amit, D.: Modelling Brain Function. Cambridge University Press, Cambridge (1989)
3. O'Kane, D., Treves, A.: Short and Long Range Connections in Autoassociative Memory. Journal of Physics **A25** (1992) 5055-5069
4. Amit, D., Tsodyks, M.V.: Quantitative study of attractor neural network retrieving at low spikes rates: I. Substrate-spikes, rates and neuronal gain. Network **2** (1991) 259-273
5. Amit, D., Tsodyks, M.V.: Quantitative study of attractor neural network retrieving at low spikes rates: II. Low-rate retrieval in symmetric networks. Network **2** (1991) 275-294
6. Tsodyks, M.V., Feigel'man, M.V.: The enhanced storage capacity of neural networks with low activity level. Europhys. Lett. **6** (1988) 101-105
7. Buhmann, J., Divko, R., Schulten, K.: Associative memory with high information content. Phys. Rev. **A39** (1989) 2689-2692
8. Kuhn, R.: Statistical mechanics of neural networks near saturation. In: Garrido L. (ed.): Statistical Mechanics of Neural Networks, Springer-Verlag, Berlin Heidelberg New York (1990) 19-32
9. Renart, A., Parga, N., Rolls, E.T.: Backprojections in the cerebral cortex: implications for memory storage. Neural Computation (to appear)
10. Renart, A., Parga, N., Rolls, E.T.: Associative memory properties of multiple cortical modules. Network: Computation in Neural Systems (submitted)