A Feature-Based Model of Semantic Memory: The Importance of Being Chaotic

A. Morelli[1], R. Lauro Grotto[2], and F.T. Arecchi[3]

[1] Departement of Engineering, University of Florence, Italy
morelli@ino.it
[2] Departement of Psychology, University of Florence, Italy
[3] Departement of Physics, University of Florence, Italy

Abstract. Semantic memory representations have often be modeled in terms of a collection of semantic features. Although feature-based models show a great explanatory power with respect to cognitive and neuropsychological phenomena, they appear to be underspecified if interpreted from a neuro-computational perspective. Here we investigate the retrieval dynamics in a feature-based semantic memory model, in which the features are represented by neurons of the Hindmarsh-Rose type in the chaotic regime. We study the state of synchronization among features coding for the same or different representations and compare the correlation patterns obtained by analyzing the whole neural signal and a manipulated signal in which the sub-threshold component is ruled out. In all cases we find stronger correlations among features belonging to the same representations. We apply a formal method in order to represent the state of synchronization of features which are simultaneously coding for different representations. In this case, the synchronization and desynchronization pattern that allows for a shared feature to participate in multiple memory representations appears to be better defined when the whole signal is considered. We interpret the simulation results as suggestive of a role for chaotic dynamics in allowing for flexible composition of elementary meaningful units in memory representations.

1 Introduction

Semantic memory can be defined as our relatively permanent memory store for world knowledge: it comprises information about words meaning and allows for the recognition of meaningful perceptual stimuli. The featural description of memory representations produced accounts for a large part of the experimental phenomena described in semantic memory literature, such as basic level naming [1], typicality effects [2], context effects [3], priming [4] and category structure [5]. The existence of subgroups of shared features is at the base of the explanatory account of the models in both cases. The main problem with the feature-based account is that it appears to have a great deal of explanatory power at a general level, but it is extremely underspecified in the details. There is a remarkable lack of consensus about what could be reasonably conceived as

M. De Gregorio et al. (Eds.): BVAI 2005, LNCS 3704, pp. 328–337, 2005.

a semantic feature for different classes of stimuli [6], such as percepts and words belonging to different semantic and morpho-syntactic classes (concrete words, abstract words, verbs etc.), although recent explicit proposals in this sense have appeared in the literature [7]. Here we will approach the problem of modeling semantic memory representations with shared features starting from some assumptions on the dynamic process of memory retrieval. First of all we assume that a semantic feature is a cognitive component of the semantic representation that is encoded in the collective activity of a segregated population of neurons [8] with chaotic dynamics. In fact, although memory processes and their neural correlates have been extensively modeled in terms of Attractor Neural Networks [9] and recent approaches emphasize the role that dynamic "latching" between attractors might have in unleashing the computational capabilities of fixed point dynamics [10], simultaneous retrieval of overlapping patterns still remain very difficult to implement with more "sedate" dynamical systems. Second, recent approaches have emphasized the need to shift to dynamic paradigms in which memory representations are built 'on the fly' according to the specificity of the task demands and of the behavioral goals the subject is engaged in [11]. Chaotic dynamics might be a preferential tool in this framework due to low cost and fast transition between attractor states.

Building on some ideas that were first proposed in the case of perceptual features [12] [13] [14], we explore the possibility to resort to chaotic dynamics in order to implement a toy model of a multimodular semantic memory system in which shared features can be dynamically allocated to different semantic representations in order to allow for the co-occurring retrieval of two or more related patterns, as it is possibly needed for the memorization and retrieval of complex scenes or concepts. Taking as a starting point the multimodular structure defined in [15], we propose a richer quantitative analysis of the network behavior by applying different types of synchronicity measures. We also contrast the results obtained by the different signal manipulations in an attempt to disclose the characteristics of the neural signal that appears to be more relevant for the emergency of the hierarchical structure of memory representations.

2 The Model

We study an associative neural network characterized by a *multimodular architecture*, which represents the functional segregation observed in some cortical areas (V1 and beyond [16]). The modular architecture of the network, depicted in Fig.1, is given by a set of M *feature modules*, each representing a specific dimension, or domain, in the memory pattern (e.g. color, dimension, shape, etc.). Each module includes F neurons coding for different *features* of the pattern (features are encoded by a single neuron), along the dimension specified by the module (e.g. red in color module, sphere in shape module, etc.). For the sake of computational simplicity we choose to substitute the population dynamics at the featural level with single unit dynamics. Although this is clearly a limit of the present simulations it is nevertheless known that single neuron spiking

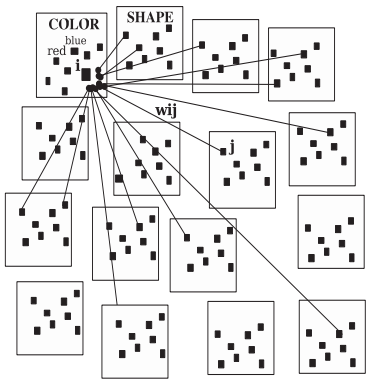

Fig. 1. Representation of the multimodular architecture. The network includes M (M=16) feature modules, represented here as boxes. Each module contains F (F=8) neurons, depicted as squares. Each single neuron i is connected to all neurons j in the other modules, through excitatory coupling (w_{ij}).

activity shares many relevant properties with network activity in terms of temporal statistics [17]. A memory pattern is defined by a vector of M features, each from a different module. In order to obtain all possible patterns, every neuron is connected via excitatory coupling to all neurons of the other modules (*cooperation*). Since neurons belonging to the same module code for mutually exclusive features (e.g. either red or yellow), we also introduce a *competitive* mechanism between them which take into account the average intra-modular activity. We use Hindmarsh-Rose model-neurons, which exhibit realistic response properties such as the presence of long interspike intervals between action potentials. Those models are characterize by a periodic or chaotic (irregular bursting) dynamic behavior, depending on a single parameter [18]. The network consists of N HR neurons ($N = 128$), belonging to M different modules ($M = 16$). In each module we have F feature neurons ($F = 8$)(i is the index for neurons in the network, j is the index identifying neurons belonging to different modules from the module of i, k identifies neurons inside the same module as i). Each HR neuron in the network is described by the first-order differential equations

$$\dot{X}_i = Y_i - aX_i^3 + bX_i^2 - Z_i + I_i + \sum_{j=1}^{F(M-1)} w_{ij} S_j(t) - \frac{1}{F-1} \sum_{k=1}^{F-1} S_k^{(i)}(t) \quad (1)$$

$$\dot{Y}_i = c - dX_i^2 - Y_i \quad (2)$$

$$\dot{Z}_i = r[s(X_i - x_0) - Z_i]. \quad (3)$$

The state of neuron i is described by three time-dependent variables, namely, the membrane potential X_i, the recovery variable Y_i, and a slow adaptation current Z_i. The external input I_i, for the standard choice of parameters ($a = 1.0$, $b = 3.0$, $c = 1.0$, $d = 5.0$, $s = 4.0$, $r = 0.006$, and $x_0 = -1.6$), is set such that the single neuron dynamics is chaotic. The synaptic input given by the firing activity of the j-th neuron on the i-th neuron is modeled in Eq.(1) by the impulse current to the i-th neuron, proportional to the synaptic strength w_{ij}, generated when the

j-th neuron is active. A neuron is here considered active whenever its membrane potential exceeds a threshold value X^* ($X^* = 0$ in our study) and its activity is coded by the variable $S_j = \Theta(X_j(t) - X^*)$, where $\Theta(x) = 1$ if $x \geq 0$ and $\Theta(x) = 0$ if $x < 0$. A local inhibition mechanism, active on the i-th neuron, is modeled in Eq.(1) by a negative impulse current to the i-th neuron, generated when the k-th neuron, belonging to the same module as i, is active. S_i here represents the activity variables of the neurons in the module (i).

Our memory patterns are defined by sets of 16 features, coded by 16 active neurons. Given two memory patterns, we distinguish between *Patterns which do not share features* (NSF) and *Patterns which share features* (SF). In the first case, vectors coding for the two patterns are orthogonal: they do not share any features. In this case, all active neurons are coding for a pattern only. In the second case, vectors are not orthogonal, so some neurons are coding for more than one pattern. We implement a *learning stage*, during which input memory patterns are stored, and a *retrieval stage*, in which the network activates some memory patterns out of the stored ones. A variable number of memory patterns is randomly generated and stored in long-term memory via updating of connection weights by a one-shot Hebbian mechanism: if two connected neurons i and j (belonging to different modules) are active at the same time, the synaptic efficacy of their connection (w_{ij}) is increased. In this work w_{ij} is defined as

$$w_{ij} = \frac{1}{M}\frac{1}{F}(1 - exp\ (-(\frac{1}{P}\sum_{p=1}^{P} S_i(p)S_j(p)))), \tag{4}$$

where $S_l(p)$=1 if neuron l is active for pattern p, $S_l(p)$=0 otherwise; P is the number of stored patterns. The learned connection weights are kept costant during memory retrieval and successive simulations. In the later section, we report results concerning the multiple retrieval dynamics of the network. We are interested on the retrieval of patterns which share features and which do not. The numerical integration was done by using a fixed-step fourth-order Runge-Kutta method. The integration step-size was chosen equal to 0.05 ms to compare our results with experiments.

3 Results and Discussion

3.1 Retrieval Dynamics: Results and Discussion

In order to investigate the retrieval dynamics of the network, we study the temporal firing state of the neurons which are activated by input patterns (working-memory [19]). We activate those neurons coding for the 16 features of a given pattern, by setting the external input current I_i in a chaotic regime, randomly between 3.0 and 3.1 (I_i is equal to 0 for inactive neurons). We are interested on what happens when the retrieved patterns are more than one, and when they share some features (SF) or not (NSF). Simulations were run with a variable number of stored patterns, retrieved patterns and shared features. Here,

for the sake of simplicity, we report the results concerning two simulation conditions: the retrieval of two NSF patterns and that of two SF patterns with three shared features. In both conditions the number of stored patterns P is equal to 15. In order to characterize the degree of correlation within and between patterns, we analyze the normalized correlation functions with variable lag τ, between the time series $x(t), y(t)$ generated by the membrane potential X of the active neurons. In Figs.2-3 the maxima of correlation functions defined as $\mathcal{C}_{xy} = \max_{(\tau)}\{\frac{\langle x(t-\tau)y(t)\rangle_t - \langle x\rangle\langle y\rangle}{\sigma_x\sigma_y}\}$ are plotted, where $\langle . \rangle$ and σ denote time averages and standard deviations respectively. As the binarization is a standard type of manipulation of the neural signals, we use the same correlation analysis with the binarized time series of the membrane potential. We define a threshold (taken here as $thr = 0$) to encode the membrane potential $X(t)$ of the neurons as a string of 0's and 1's ($X(t) = 1$ when $X(t) > thr$ and $X(t) = 0$ otherwise). This analysis is done in order to determine if this different format encodes the same information as non binarized signals, and if this information is sufficient to describe the correlation structure of the retrieved patterns. We expect that this structure does not change dramatically for the binarized time series, due to the fact that the temporal informations about the spikes (their temporal position, length and separation from other spikes) are maintained in binarized time series.

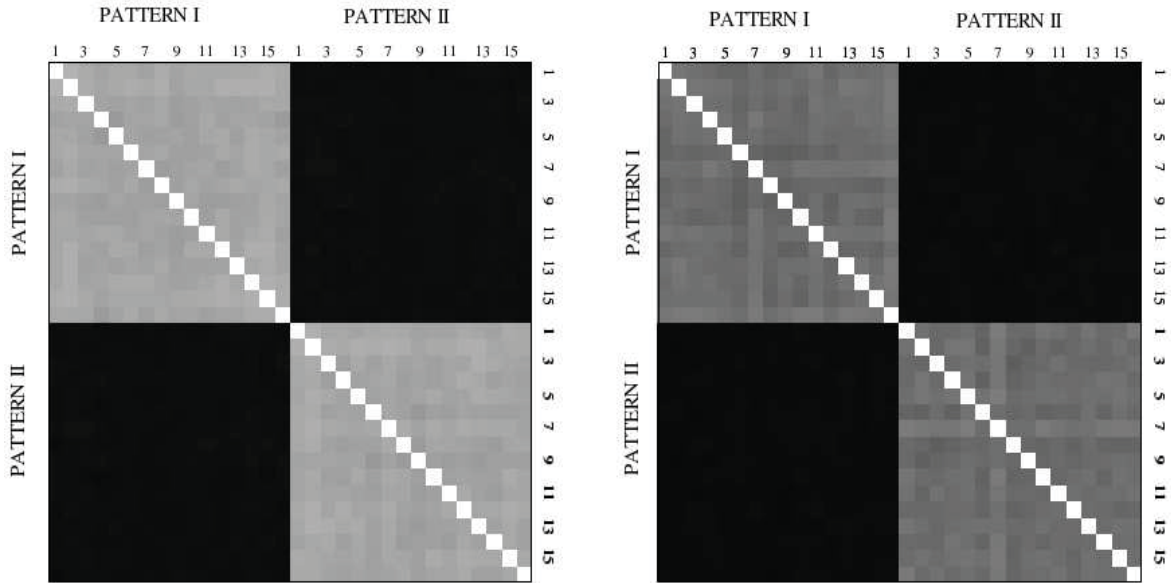

Fig. 2. Retrieval of two patterns with Not Shared Features. Maxima of correlation functions between time series (left) and binarized time series (right) of the membrane potential of active neurons (16 for Pattern I and 16 for Pattern II). Within each matrix: the maxima values for the 16 pairs of neurons belonging to Pattern I and Pattern I (top-left), Pattern I and Pattern II (top-right), Pattern II and Pattern I (bottom-left), Pattern II and Pattern II (bottom-right). Values are represented using grayscales, from 0 (black) to 1 (white).

Retrieval of Patterns with Not Shared Features (NSF). In this simulation condition, we activate two NSF patterns (Pattern I and Pattern II) out of the P stored patterns. We have 32 active neurons, each one coding for one pattern only. By analyzing the structure of correlation functions for the binarized and non binarized time series, we find stronger correlations between neurons coding for the same pattern, and weaker correlations between neurons coding for different patterns (Fig.2). The maxima of correlation functions are greater for non binarized time series compared to binarized time series, but the structure of the matrix is similar.

Retrieval of Patterns with Shared Features (SF). In this second condition, the network retrieves two SF patterns which share three features (there are three neurons which are coding for both Pattern I and Pattern II). By evaluating maxima of correlation functions for the binarized and non binarized time series (Fig.3), we observe stronger correlations between neurons coding for the same pattern and weaker correlations between neurons coding for different patterns, except for those neurons coding for shared features: they are correlated with neurons coding for both Pattern I and Pattern II. As in NSF condition, the maxima of correlation functions are greater for non binarized time series compared to binarized time series, but the structure of the matrix is similar. The neuronscoding for shared features are correlated with neurons coding for

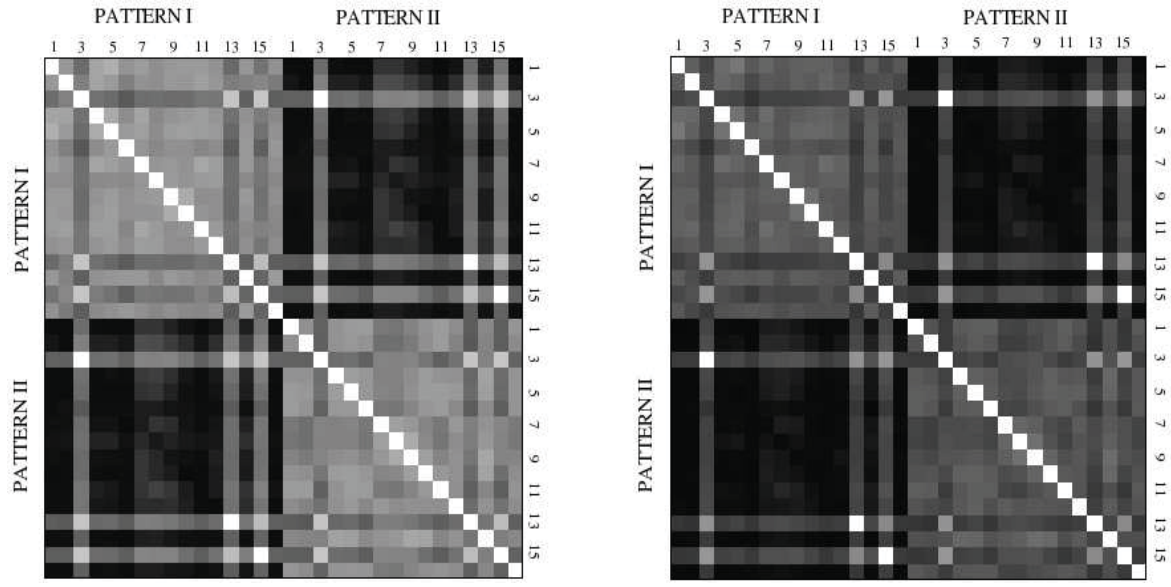

Fig. 3. Retrieval of two patterns with three Shared Features. Maxima of correlation functions between time series (left) and binarized time series (right) of the membrane potential of active neurons (16 for Pattern I and 16 for Pattern II). Within each matrix: the maxima values for the 16 pairs of neurons belonging to Pattern I and Pattern I (top-left), Pattern I and Pattern II (top-right), Pattern II and Pattern I (bottom-left), Pattern II and Pattern II (bottom-right). Values are represented using grayscales, from 0 (black) to 1 (white).

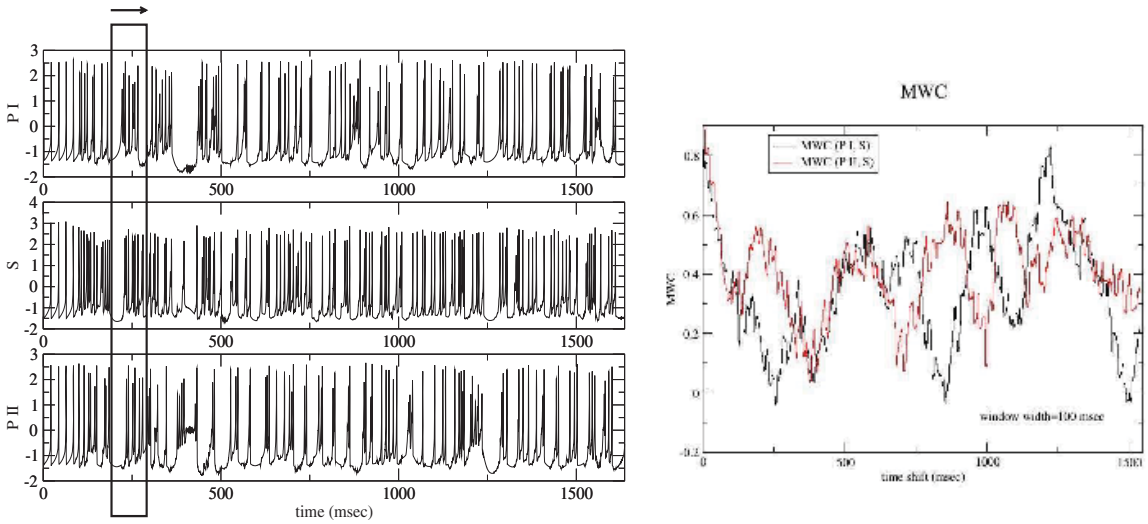

Fig. 4. Left: from top, the first time series is the membrane potential of a neuron coding for Pattern I only (PI), second one for Pattern I and Pattern II (S), and the third one for Pattern II only (PII). Over the time series the mobile window ($w = 100$ msec) for MWC is depicted. Right: the MWC between the shared neuron S and the two neurons coding for one pattern only (PI and PII), are plotted, as a function of time shift.

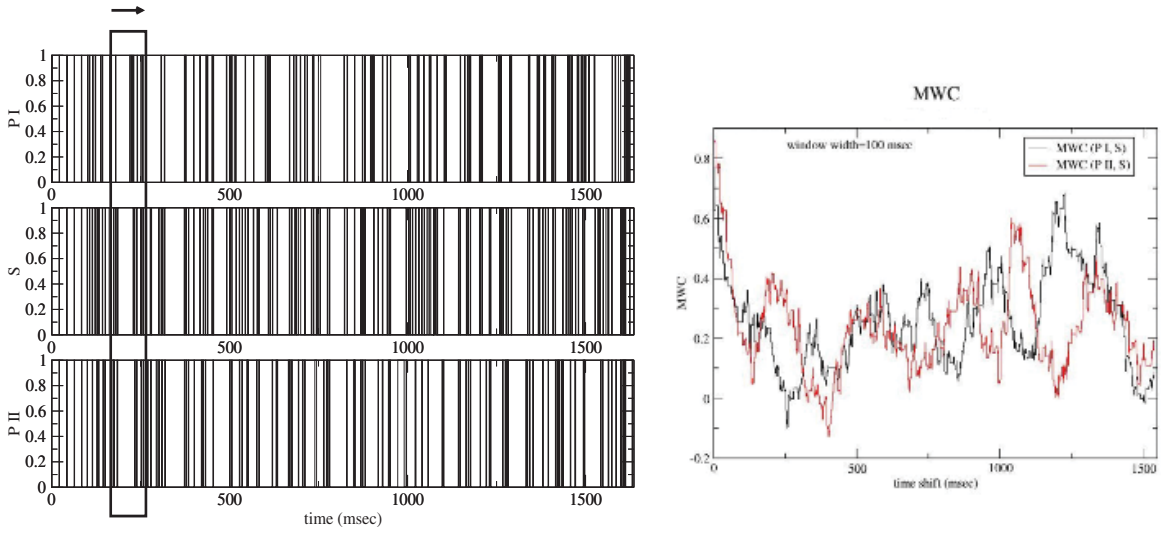

Fig. 5. Left: from top, the first time series is the binarized membrane potential of a neuron coding for Pattern I only (PI), second one for Pattern I and Pattern II (S), and the third one for Pattern II only (PII). Over the time series the mobile window ($w = 100$ msec) for MWC is depicted. Right: the MWC between the shared neuron S and the two neurons coding for one pattern only (PI and PII), are plotted, as a function of time shift.

Pattern I and Pattern II in a nonstationary way. In order to investigate this non-stationarity, we introduce the Mobile Window Correlation (MWC) (see next section).

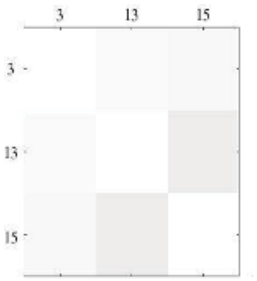

Fig. 6. The maximum of correlation functions between time series of neurons coding for the three shared features in SF simulation condition (neurons 3, 13 and 15 from Fig.3)

Mobile Window Correlation (MWC). The Mobile Window Correlation is defined as $\mathrm{MWC}(\theta) = \frac{\langle xy\rangle - \langle x\rangle\langle y\rangle}{\sigma_x \sigma_y}$, where the averages and the standard deviations are evaluated in the mobile (synchronous) window $[\theta - w/2, \theta + w/2]$. We analyze the MWC, as a function of the time shift, for the time series of a neuron coding for both patterns (indicated with 'S', in Fig.4 for non binarized time series and in Fig.5 for binarized) and two neurons, the first coding for Pattern I (MWC(PI,S)) and the second for Pattern II (MWC(PII,S)) only. As shown in Fig.4 and Fig.5, the shared neuron S is *alternatively* correlated with the neurons coding for the two patterns (PI and PII).

3.2 Hierarchical Organization in Neural Representation

Overall, the results of the correlation analysis show that units that are coding for shared features tend to be more strongly correlated among them than units that are coding for item specific features belonging to the same pattern (Fig.6). In the model it is therefore possible to disclose the emergence of pools of units coding for shared features. We suggest that these pools of units are representative of super-ordinate information with respect to what is coded at the level of Shared Features. In the semantic memory system, the stronger correlation between shared features could be at the base of the hierarchical structure of the memory representations.

4 Discussion

Previous connectionist models of semantic memory assuming feature-based representations [5] and point attractor dynamics [20] have proven to be a suitable tools to investigate and explain a great deal of cognitive and neuropshychological data. What is therefore the main advantage obtained by shifting to a chaotic dynamic regime? Our results enhance the role of chaotic dynamics in allowing much greater flexibility of semantic representations during memory retrieval. In fact, in our model the same semantic features can be dynamically allocated to different memory representations by alternating their synchronization state

with different pools of units. Furthermore the same mechanism is able to sustain the separation and concurrent retrieval of partially overlapping memory representations, and avoids spurious synchronization of unrelated semantic features. Neither attractor networks with Hopfield dynamics [20] nor models resorting to static synchronization of neural activities [21] are able to solve this computational problem. We would like to speculate that this mechanisms could play a relevant role in other cognitive domains, possibly linked to frontal lobes functions, such as conflict resolution and coherence assessment [11]. In an attempt to provide a formal description of dynamic synchronicity, we introduce the Mobile Window Correlation Analysis. In the present set of simulations dynamic synchronization shows up even when we take into account the spiking signals alone and leave out the contribution of sub-threshold neural activity (Fig.5). Nevertheless, it appears that the alternate synchronization to different pools of features is more neatly defined when the whole signal is considered (Fig.4), e.g. when the sub-threshold activity of the units is also taken into account.

5 Conclusions

In the present work we presented a toy model of the semantic memory system in which semantic features are coded by Hindmarsh-Rose neurons in the chaotic regime. We devised a formal method to quantify the level of synchronous and asynchronous activity among units coding for Shared and Not Shared Features, and we applied it to the whole signal and to different manipulated neural signals, in which the contribution of sub-threshold activity was ruled out. Although the emergence of a hierarchical structure is evident in all cases, the synchronization shifts that allow for the same feature to participate in the retrieval of multiple semantic memory representations appears to be better defined when the sub-threshold activity is also taken into account. Based on our results, we suggest that the structure of correlations typical of groups of Shared Features would be more robust with respect to damage when compared to the one of Not Shared Features. Further simulations will empirically address this issue. Overall, our results suggest that chaotic dynamics might play a relevant role in allowing for flexible composition of elementary representational states in cognition.

References

1. Rosch, E., Mervis, C., Gray, W., Johnson, D., Boyes-Braem, P.: Basic objects in natural categories. Cognitive Psychology **8** (1976) 382–439
2. Rosch, E.: On the internal structure of perceptual and semantic categories. In: Moore, T. (Ed.) Cognitive Development and the Acquisition of Language. Academic Press, New York (1973)
3. Barsalou, L.: Context-independent and context-dependent information in concepts. Memory and Cognition **10** (1982) 82–93
4. Plaut, D.: Semantic and associative priming in a distributed attractor network. In: Proceedings of the 17th Annual Conference of the Cognitive Science Society (pp. 37–42). Hillsdale, NJ: Lawrence Erlbaum Associates (1995)

5. Rumelhart, D. E.: Brain style computation: Learning and generalization. In: Zornetzer, S. F., Davis, J. L., Lau, C. (Eds), An introduction to neural and electronic networks. Academic Press, San Diego pp 405–420 (1990)
6. Malt, B.: Water is not H20. Cognitive Psychology **27** (1994) 41–70
7. Landauer, T. K., Dumais, S. T.: A solution to Plato's problem: the Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. Psychological Review **104** (1997) 211–240
8. Schyns, P. G., Goldstone, R. L., Thibaut, J.: The development of features in object concepts. Behavioral and Brain Science **21** (1998) 1–54
9. Amit, D.: Modeling brain function. Cambridge University Press, Cambridge UK (1998)
10. Treves, A.: Frontal latching networks: a possible neural basis for infinite recursion. Cognitive Neuropsychology **22** (2005) 276–291
11. Shallice, T.: Fractionation of the Supervisory System. In: Stuss, T. S., Knight, R. (Eds). Principles of Frontal Lobe Function. Oxford University Press, Oxford UK (2002)
12. Von der Malsburg, C.: The what and why of binding: The modeler's perspective. Neuron **24** (1999) 95–104
13. Engel, A. K., König, P., Kreiter, A. K., Schillen, T. B., Singer, W.: Temporal coding in the visual cortex: New vistas on integration in the nervous system. Trends Neurosci. **15** (1992) 218–226
14. Gray, C. M.: The temporal correlation hypothesis of visual feature integration: Still alive and well. Neuron **24** (1999) 31–47
15. Raffone, A. and van Leeuwen, C.: Dynamic synchronization and chaos in associative neural network with multiple active memories. Chaos **13** (2003) 1090–1104
16. Felleman, D.J., van Essen, D. C. V.: Distributed hierarchical processing in the primate visual cortex. Cereb.Cortex **1** (1991) 1–47
17. Segev, R. et al.: Long term behavior of lithographically prepared *in vitro* neuronal networks. Phys. Rev. Lett. **11** (2002) 118102–1
18. Hindmarsh, J. L., Rose, R. M.: A model of neuronal bursting using three coupled first order differential equations. In: Proc. R. Soc. London B **221** (1984) 87–102
19. Baddeley, A. D.: Working memory. Science **255** (1992) 556–559
20. LauroGrotto, R., Reich, S., Virasoro, M. A.: The computational role of conscious processing in a model of semantic memory. In: Miyashita, M., Ito, M., Rolls, E. (Eds), Cognition,Computation and Consciousness, Oxford University Press, Oxford UK pp 248–263 (1997)
21. Fujii, H., Hito, H., Aihara, K., Ichinose, N., Tsukada, M.: Dynamical cell assembly hypothesis: theoretical possibility of spatiotemporal coding in the cortex. Neural Networks **9** (1996) 1303–1350